

Computational Chemistry

Reviews of Current Trends

Volume 8

editor

Jerzy Leszczynski

World Scientific

Computational Chemistry

Reviews of Current Trends

Volume 8

Computational Chemistry: Reviews of Current Trends

Editor-in-Charge: Jerzy Leszczynski,
Dept. of Chemistry, Jackson State University, USA

Published

Vol. 1: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 2: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 3: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 4: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 5: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 6: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Vol. 7: Computational Chemistry: Reviews of Current Trends
Edited by Jerzy Leszczynski

Computational Chemistry

Reviews of Current Trends

Volume 8

editor

Jerzy Leszczynski

*Department of Chemistry
Jackson State University
USA*



World Scientific

NEW JERSEY • LONDON • SINGAPORE • SHANGHAI • HONG KONG • TAIPEI • BANGALORE

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

COMPUTATIONAL CHEMISTRY: REVIEWS OF CURRENT TRENDS

Volume 8

Copyright © 2003 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-238-702-1

PREFACE

For many years I had not doubted the most popular theory in astrophysics that our Universe is in an expanding phase. However, I have also learned how important relativity is. I have sensed that over the years my World has been shrinking. Not only am I able to move much faster from one location to another and visit many exotic places, but also telephone lines and the Internet have tightly linked many places on the Earth. This allows me to communicate quickly with colleagues in Europe, China, or Africa and to effectively collaborate with them on joint projects.

This size reduction is analogous to a continuous decrease in the dimensions of species accessible for studies and applications. Not only that “small is beautiful,” but also it enforces progress and provides a new dimension to science and technology. Chemists are finally able to work with single molecules and atoms, yet not billions of them at a time. “Nano” has become a buzzword of the new century. New “nano” journals and “nano” letters and “nano” webpages have been emerging. We are constantly reading about nano wires, sensors, transistors, and nanocomputers.

It seems that a gap between experimental objects and models for calculations in chemistry is being bridged. The size of experimental nanoobjects is decreasing, while reliable calculations are feasible for larger and larger molecular systems. The results of these calculations for isolated molecules are becoming more relevant for experiments. However, there are still significant challenges for computational methods. They include the modelling of the solid state and of solvents. Many theoretical groups are involved in the development of efficient approaches to solids and solvent effects.

This series presents reviews of current advances in computational methodologies and applications. The first chapter written by R. Cammi, B. Mennucci, and J. Tomasi provides an overview of their developments of the Polarizable Continuum Model (PCM). This approach has been particularly popularized after its implementation into the GAUSSIAN suite of programs. The authors reveal both the theoretical and numerical aspects of the PCM model. Also, promising extensions of the theory are discussed. Among the possible applications, examples concerning the evaluation of molecular-response properties and spectroscopic quantities are provided.

Significant simplifications are necessary in order to construct a solid state model. This could be achieved using the supermolecular approach. M. Guillaume, B. Champagne, F. Castet, and L. Ducasse demonstrate a multiplicative scheme that is often used for estimations of the properties of two- and three-dimensional clusters from the properties of their one-dimensional components. They apply semiempirical methods to simulate absorption spectra

and to evaluate linear and nonlinear responses. The chapter concludes with a comparison of theoretical predictions with experimental results and with a discussion of the reliability of the global approach.

Recently significant progress has been achieved in the theoretical prediction of NMR parameters. The application of *ab initio* methods allows for a reliable evaluation of the characteristics of hydrogen-bonded and van der Waals complexes. The theory and methods applied to calculations of spin-spin coupling constants are discussed by M. Pecul and J. Sadlej. The authors provide a number of applications including simple complexes, model systems for nucleic acids and proteins, and weak van der Waals systems.

Ab initio quantum chemical methods are quickly gaining popularity among researchers investigating various aspects of DNA. Reliable quantum mechanical techniques do not require empirical parameters; they yield results of experimental accuracy and allow for the simultaneous study of molecular structures. The size of DNA fragments feasible for such calculations is constantly increasing. DNA base polyads are among recently studied species. The chapter by M. Meyer and J. Suhnel reviews the properties of these complexes linked by base-base hydrogen bonds. Examples include many species of biological importance. Among them are telomeres, structures formed from triplet disease sequences, and tertiary interactions in RNA. Their occurrence in nucleic acid structures is discussed. Numerous results of computational studies are provided and used to establish a link between structural biology and computational chemistry.

DNA can be affected by chemical and environmental effects. D. M. Close presents a review of primary radiation-induced defects in nucleic acid building blocks. The discussion includes both EPR/ENDOR experiments and the application of quantum chemical calculations of spin densities and isotropic and anisotropic hyperfine couplings that could be used to assist in making free radical assignments. DFT theory has been selected for theoretical predictions. The chapter presents examples of successes and failures of the DFT calculations in the prediction of spin densities and hyperfine couplings for radiation-induced radicals observed in nucleobases. In addition, a discussion of unsolved problems and suggestions for future directions is also given.

Experimental studies of ground state properties of DNA bases have been carried out for many years. This includes studies of their molecular structures. Although one can conclude that the ground state of these important DNA constituents is well characterized, data concerning their excited state properties is scarce. The molecular geometry of such complex systems like DNA bases cannot be determined by experimental methods. It creates a need for theoretical studies that could shed light on the properties of DNA bases in the excited state. In the last chapter of this volume, M.K. Shukla and J. Leszczynski discuss available experimental data of DNA bases, base pairs, and their complexes with water. The discussion is enhanced by an overview of the results of recent

theoretical studies for these species. The main emphasis is on the electronic singlet excited state properties, phototautomerism of bases and base pairs, and their interaction with water molecules.

I would like to thank all authors for the excellent contributions and fine collaborations. The very efficient technical assistance of Dr. Manoj K. Shukla in putting together this volume is greatly appreciated. As always, your feedback is very important to me, please feel free to e-mail your suggestions to jerzy@ccmsi.us.

Jerzy Leszczynski
Jackson, MS
December 2002

CONTENTS

Preface	v
1. Computational Modelling of the Solvent Effects on Molecular Properties: An Overview of the Polarizable Continuum Model (PCM) Approach	1
<i>R. Cammi, B. Mennucci and J. Tomasi</i>	
2. Electronic and Nonlinear Optical Properties of 2-Methyl-4-Nitroaniline Clusters	81
<i>M. Guillaume, B. Champagne, F. Castet and L. Ducasse</i>	
3. <i>Ab Initio</i> Calculations of the Intermolecular Nuclear Spin–Spin Coupling Constants	131
<i>M. Pecul and J. Sadlej</i>	
4. Base Polyad Motifs in Nucleic Acids — Biological Significance, Occurrence in Three-Dimensional Experimental Structures and Computational Studies	161
<i>M. Meyer and J. Sühnel</i>	
5. Model Calculations of Radiation Induced Damage in DNA Constituents Using Density Functional Theory	209
<i>D. M. Close</i>	
6. Excited States of Nucleic Acid Bases	249
<i>M. K. Shukla and J. Leszczynski</i>	
Index	345

Chapter 1: Computational Modelling of the Solvent Effects on Molecular Properties: An Overview of the Polarizable Continuum Model (PCM) Approach

Roberto Cammi

*Dipartimento di Chimica Generale ed Inorganica, Università di Parma
Parco Area delle Scienze 1, 43100 Parma, Italy*

Benedetta Mennucci and Jacopo Tomasi

*Dipartimento di Chimica e Chimica Industriale
Università di Pisa, via Risorgimento 35, 56126 Pisa, Italy*

Abstract

An overview of the Polarizable Continuum Model (PCM) for the modelling of solvent effects on the state and the properties of quantum mechanical molecular systems is presented. The main theoretical and numerical aspects of this method are presented and discussed, together with its extension to the derivative theory. We present some selected applications concerning the evaluation of molecular response properties, and of the corresponding spectroscopic quantities, of different solvated systems.

1.1. Introduction

The long term goal of the Computational Chemistry is to calculate properties of material systems and giving an interpretation of them. Within this perspective great progresses have been obtained for the characterization of the properties of the simplest material system: a single molecule not interacting with other pieces of matter. Theoretical Chemistry has accepted

since long Quantum Mechanics (QM) as the appropriate theory and Computational Chemistry has developed appropriate and very accurate tools to determine properties of isolated molecules (of small and medium size) within the framework of the QM.

The following step, that passing from isolated molecules or atoms to more complex material systems, is still under evolution. There are many approaches in use. Some of them are specialized for specific classes of material systems where all the basic units have on equal terms role in determining the properties of the whole system: large covalently bound systems, regular crystals, pure liquids. In these cases the accurate QM description is abandoned, keeping of it some elements only.

These approaches are not the best suited ones when one is interested, as it is our case, to the accurate description of the properties of a complex material system for which there is a strong dependence on the fine characteristics of some privileged molecular components (the solute) having all the rest of the system a supporting role (the solvent). An example is given by a material in which there are chromophores immersed into a liquid solution, or more complex hosts as for example a liquid crystals or a polymeric matrix. In such a case we need an accurate QM description of the chromophore, of the same level of that used for the isolated molecules, and a faithful description of the interaction between the chromophore (or target molecular system). The perspective of treating the whole material systems at an high QM degree of accuracy is out of question, exceeding the existing computational potentialities and those available in the foreseeable future.

These considerations lead to what we call *focused models* (FM). In these models of complex material systems, the attention is focused on a small part, treated at higher level of accuracy than the remainder. The part treated at a higher level, that we call the main part **M**, should include all the molecular units of the whole systems we consider necessary to get an accurate description of the desired molecular property or process. The remainder of the system, called **S**, should have a supporting function in the determination of the property. Actually, in focused model, the definition of **M** is not univocal but must be checked on the basis of the results. To give an example, systems composed by a chromophore which exhibits strong specific interactions between with some molecular units of the host, require the inclusion into **M** of molecular entities of the hosting.

The Focused Model embraces a broad range of apparently different methods of the Computational Chemistry. We may introduce a first dis-

tinction on them.

There are FMs in which \mathbf{S} is described as a set of units conserving their molecular identities, (they will be called discrete FMs), others in which \mathbf{S} is described as a continuous medium (they will be called continuous FMs). In the case of FMs for solution also a different terminology is used: models belong to the two categories are called *explicit* and *implicit* solvation models.

There are two important advantages in using continuum models.¹ The first is a significant reduction in the number of degrees of freedom, while the second is the fact that they provide, and at very low cost, a very accurate way to treat the long-range electrostatic forces and at the same time they can easily include the polarization effects which are often neglected (due to the computational cost) in explicit methods.

The use of focused model must be accompanied by a sound strategy to test qualities, defects and potentialities. This may be performed by examining the outcomes of the model at its best, in its most complete and detailed form. The main aim of this review is to give an overview on how a specific continuum solvation approach, the Polarizable Continuum Model (PCM), has pursued this systematic research effort.

The review is organized as follow. In section II we summarize the general basis of the QM continuum solvation models. In section III we present the formal aspects of the PCM, including the theory of the analytical derivatives of the energy. In section IV we present an overview of the PCM approach to the calculation of the properties of molecular solutes.

1.2. Quantum Mechanical Theory of Solvation Models

The formulation of the QM continuum models reduces to the definition of an Effective Hamiltonian, i.e. an Hamiltonian to which solute-solvent interactions are added in terms of a solvent reaction potential. This effective Hamiltonian may be obtained from the basic energetic quantity which has the thermodynamic status of free energy for the whole solute-solvent system and for this reason is called free energy functional, $\mathcal{G}(\Psi)$. This energy functional can be expressed in the following general form:²

$$\begin{aligned}\mathcal{G}_R(\Psi) = & \langle \Psi | \hat{H}^0 | \Psi \rangle + \langle \Psi | \hat{\rho}_r | \Psi \rangle V_r^R + \frac{1}{2} \langle \Psi | \hat{\rho}_r | \Psi \rangle \mathbf{V}_{rr'}^R \langle \Psi | \hat{\rho}_{r'} | \Psi \rangle \\ & + \frac{1}{2} U_{NN} + V_{NN}\end{aligned}\quad (1.1)$$

In eq.(1.1) the Born-Oppenheimer (BO) approximation is employed. This means a standard partition of the Hamiltonian into an electronic and a nuclear part, as well as the factorization of the wave function into an electronic and a nuclear component. In this approximation eq.(1.1) refers to the electronic wavefunction with the electronic Hamiltonian dependent on the coordinates of the electrons, and, parametrically, on the coordinates of the nuclei.

In the equation above, \hat{H}^0 is the Hamiltonian describing the isolated molecule, $\hat{\rho}_r$ represents the operator of the solute electronic charge density, V_r^R is the *solvent permanent potential*, and $\mathbf{V}_{rr'}^R$ describes the *response function of the reaction potential* associated with the solvent. Here an extension of the Einstein convention on the sum has been exploited: the space variables r and r' , appearing as repeated subscripts, imply an integration in the 3-dimensional space. V_{NN} is the nuclear repulsion energy of the solute M , and U_{NN} , here introduced with factor 1/2 in order to have a free energy, represents the interaction energy between solute nuclei and the solvent reaction component generated by the nuclear part of the solute charge distribution

By applying the variational principle on this functional we can derive the time-independent nonlinear Schrödinger equation specific the system under scrutiny. If we impose that first-order variations of \mathcal{G} with respect to arbitrary variations of the solute wavefunction Ψ are zero, and that the latter is normalized, the nonlinear Schrödinger equation becomes:

$$\hat{H}_{eff} |\Psi\rangle = \left[\hat{H}^0 + \hat{\rho}_r V_r^R + \hat{\rho}_r \mathbf{V}_{rr'}^R \langle \Psi | \hat{\rho}_{r'} | \Psi \rangle \right] |\Psi\rangle = E |\Psi\rangle \quad (1.2)$$

where E is the Lagrange multiplier introduced to fulfill the normalization condition on the electronic wavefunction. Eq.(1.2) defines the specific ‘Effective Hamiltonian’, \hat{H}_{eff} , for the system considered and constitutes the key point of any QM continuum solvation model. Any stationary properties of the solute M can be reduced to the solution of the (time-independent) effective Schrödinger (1.2).

The first solvent term, V_r^R does not lead to any difficulty, neither from the theoretical point of view, neither from the practical. Many examples are known in which an external potential is introduced in the molecular calculations. On the contrary, the treatment of the reaction potential operator $\hat{\rho}_r \mathbf{V}_{rr'}^R \langle \Psi | \hat{\rho}_{r'} | \Psi \rangle$ is rather delicate, as this term induces a nonlinear character to the solute Schrödinger equation. We recall that the nonlinear eq. (1.2) is a direct consequence of the variational principle applied to \mathcal{G} . The free energy functional \mathcal{G} has a privileged role in the theory, as the

solution of the Schrödinger equation gives a minimum of this functional (that we shall indicate as G) even though it is not the eigenvalue of the nonlinear Hamiltonian, here indicated as E . We stress that in habitual linear Hamiltonians these two quantities, the Hamiltonian eigenvalue and the variational functional, coincide. The difference between E and G has, however, a clear physical meaning; it represents the polarization work made by the solute to create the charge density inside the solvent.

Within the BO approximation the free energy depends parametrically on the nuclear coordinates \mathbf{R} and acts as potential energy (hyper)surface for the nuclei. The minima of $G(\mathbf{R})$ determine the geometries of the solute and the curvature of $G(\mathbf{R})$ near the minima defines the nuclear vibrational states of the solute, within the harmonic approximation. The solution of the nuclear problem allows also to supplement G with entropic contributions due to the thermal motion connected with the nuclear degrees of freedom of the solute, so that:

$$G^S = G - RT \ln(q_{rot,s} q_{vib,s}) + RT \ln(n_{M,s} \Lambda_{M,s}^3) \quad (1.3)$$

Here $q_{rot,s}$ and $q_{vib,s}$ are the microscopic partition functions for rotation and vibration of M in the solution; while $n_{M,s}$ and $\Lambda_{M,s}$ indicate the related numeral densities and the momentum partition functions, respectively. The functional (G^S) has now the meaning of the free energy of the entire solute-solvent system, at the temperature T , with respect to a reference state given by non-interaction nuclei and electron, supplemented by the unperturbed pure liquid S , at the same temperature. Then, the fundamental energetic quantity connected with the insertion of the solute in the solvent, i.e. the free energy of solvation can be obtained as

$$\Delta G_{sol} = G^S - E^o - RT \ln(q_{rot,g} q_{vib,g}) + RT \ln(n_{M,g} \Lambda_{M,g}^3)$$

where E^o is the internal energy of the isolated solute.

A possible computational strategy to get G is to write it as a sum of various contributions of different physical origin, related to dispersion, repulsion, steric and electrostatic forces between solute and solvent molecules:

$$G = G_{dis} + G_{rep} + G_{ster} + G_{el} \quad (1.4)$$

The partition above can be further developed by introducing a parallel partition of the interaction potential \hat{V}_R :

$$\hat{V}^R = \hat{V}_{dis} + \hat{V}_{rep} + \hat{V}_{ster} + \hat{V}_{el} \quad (1.5)$$

and by explicitly writing the eventual dependence of each term from the solute charge distribution; in this way we can derive a simple classification of the various contributions in terms of their effects on the solute wave function and/or the system free energy.

The steric term, whose contribution to G is more often called ‘cavitation free energy’, G_{cav} , corresponds to the work spent in building up a cavity of appropriate shape and volume in which the solute molecule is enclosed, with all the other solute-solvent interactions switched off. Its effects on the system are only of energetic nature; in fact, as its contribution to the potential interaction, \hat{V}_{ster} , is always written as a constant operator independent from the solute electronic charge density, the corresponding wave function cannot be changed by its eventual presence and its always positive contribution to the system free energy can be evaluated independently from the quantum calculation.

For the evaluation of G_{cav} several formulas are available, based on the shape and size of the solute and on different parameters of the solvent: surface tension, isothermal compressibility, and geometrical data of the molecules.^{3,4} The first three formulas here mentioned are of empirical nature and follow almost the same philosophy of the continuum dielectric, neglecting the discrete nature of the solvent molecules but making use of experimental bulk parameters. The last formulation, on the contrary, derives from a theory based on a discrete model of fluids (the Scaled Particle Theory, SPT), even if the final expression of G_{cav} depends again on bulk solvent parameters only.

Let’s now consider in more details the terms which explicitly enter in the Effective Hamiltonian, starting from that related to electrostatic interactions.

1.2.1. Electrostatic interactions

The description of the electrostatic solute-solvent interactions has represented, historically, the keystone of continuum models. Within the continuum electrostatic frameworks the solvent is substituted with a dielectric medium and the solute occupies a cavity C within the dielectric. The main aspects to consider is thus the definition of the macroscopic characteristics of the dielectric, i.e. the form of the dielectric constant and the definition of the cavity C. Following this analysis we can define different systems: (i) homogeneous isotropic dielectrics, characterized by a constant scalar permittivity, ϵ ; (ii) homogeneous anisotropic dielectrics, characterized by a constant tensorial permittivity, ϵ ; (iii) inhomogeneous dielectrics, character-

ized by a position dependent permittivity, $\epsilon(r)$. Other specific dependencies can be found as well as complex systems formed by two or more dielectrics separated by well defined boundaries. Further interesting chemical problems directly or indirectly connected to this list are those represented by ionic solutions (i.e. liquids in which charged particles are free to move), liquids in contact with a metal, and mixtures of different isotropic liquids separated by surfaces. Clearly, this list cannot be exhaustive but a detailed analysis of all possible systems goes far beyond the scopes of the present work. What we shall do here is to focus our attention almost exclusively on the simplest, and also the most common case (a solute in an infinitely dilute solution of an homogeneous isotropic solvent).

For the selected system, i.e. a solute M in a cavity C surrounded by an infinite homogeneous isotropic dielectric with permittivity ϵ , the basic relation to be considered is the Poisson equation, with the related boundary conditions; namely:

$$\begin{cases} -\Delta V = 4\pi\rho_M & \text{in } C \\ -\epsilon\Delta V = 0 & \text{outside } C \\ [V] = 0 & \text{on } \Sigma \\ [\partial V] = 0 & \text{on } \Sigma \end{cases} \quad (1.6)$$

where V is the electrostatic potential due to the presence of the charge distribution ρ_M located inside the cavity and Σ the cavity surface. The jump condition $[V] = 0$, means that the potential V is continuous across the interface Σ , i.e. $V_e - V_i = 0$ on Σ , where the subscripts e and i indicate regions outside and inside the cavity, respectively. The equality $[\partial V] = 0$ is a formal expression of the jump condition of the gradient of the potential; in our limit isotropic system it takes the well-known form:

$$\left(\frac{\partial V}{\partial n}\right)_i - \epsilon \left(\frac{\partial V}{\partial n}\right)_e = 0 \quad (1.7)$$

where n is the outward pointing unit vector perpendicular to the cavity.

The main approaches used to solve this problem can be divided in different classes;

- (1) the apparent surface charge (ASC) methods;^{5–10}
- (2) the multipole expansion (MPE) methods;^{11–13}
- (3) the generalized Born approximation (GBA) methods.¹⁴

The following exposition will be focused only on the approach based on the apparent surface charges, and in particular in the PCM version of such

approach.

1.2.1.1. PCM methods

In the three PCM versions of the ASC method we shall consider here, the strategy consists in solving the system (1.6) and get the related potential V by exploiting an integral equation formalism.

The solution of system (1.6) can be obtained by exploiting different mathematical approaches. In particular the original version of PCM, that we shall indicate as DPCM, exploits an approach based on the direct use of Green functions, while the modern version, called Integral Equation Formalism (IEF) uses other operator functions derived from the theory of integral equations. The formal derivation of these two approaches is quite different, nevertheless it is possible, and useful for our scopes, to find a common basic description.

In all the approaches the first step is the definition of the analytical expressions for the specific operators; their knowledge enables one to transform the first two equations in (1.6) into integral equations on the surface Σ , that can be easily solved with standard numerical methods. In this framework, the solution of system (1.6) is given by the sum of two electrostatic potentials, one produced by ρ_M and the other due to a surface charge distribution σ , placed on the interface Σ , which arises from the polarization of the dielectric medium:

$$V(x) = V_M(x) + V_\sigma(x) = \int_{\mathbb{R}^3} \frac{\rho_M(y)}{|x - y|} dy + \int_{\Sigma} \frac{\sigma(s)}{|x - s|} ds \quad (1.8)$$

where the integral on the first term is taken over the entire three-dimensional space.

Once eq. (1.8) has been defined, the problem is shifted to the calculation of the screening, or apparent surface charge (ASC) density $\sigma(s)$.

We recall that the numerical approach used to treat the ASC density exploits a partition of the surface into K small portions of area a_k , called tesserae, on which a constant charge density is assumed. In this framework which can be easily linked to the analogous techniques used in physics and engineering, and known as the Boundary Element Method (BEM), the integral form of $V_\sigma(x)$ in eq.(1.8) is reduced to a finite sum running over the point charges representing the surface charge:

$$V_\sigma(x) = \sum_k^K \frac{q_k(s_k)}{|x - s_k|} \iff q_k(s_k) = \sigma(s_k)a_k \quad (1.9)$$

where vector s_k indicates the position on each tessera k at which the value of σ is computed (usually, it identifies the center of the tessera and is called the representative point).

(a). DPCM

In this method,⁵ originally developed in 1981 but then almost completely redefined in 1995, the apparent surface charge is expressed by the following classical electrostatic relation:

$$\sigma(s) = - \left(\frac{\epsilon - 1}{4\pi\epsilon} \right) E(s) \cdot n(s) \quad (1.10)$$

where E is the total electric field, i.e. the sum of the contributions, both computed on the inner side of the surface cavity, due to the solute charge distribution and to the ASC, respectively, and n is the outward normal unit vector to the cavity at position s .

Due to the dependence of $E(s)$ on $\sigma(s)$, eq.(1.10) should be solved through an iterative procedure; however, the same results can be obtained in the form of a set of linear equations. Applying the discretization technique which represents the continuous $\sigma(s)$ through point charges, the basic equations defining such charges, below collected in the column matrix \mathbf{q} , are reduced to the following:

$$\mathbf{D}\mathbf{q} = -\mathbf{E}_n^M \quad (1.11)$$

where \mathbf{E}_n^M collects the components of the solute electric field orthogonal to the cavity surface. \mathbf{D} is a square matrix, with size equal to the number K of surface tesserae, and elements determined by the value of the normal components of the field which the apparent charges induce on each other (i.e. the solvent reaction field).

(b). CPCM

CPCM⁹ is the PCM-like reformulation of the COSMO solvation method,^{7,8} the latter in turn is based on the screening conductor theory. Screening in conductors can be handled much more easily than in dielectric media; in our framework of a solute M placed in a cavity inside the medium, now represented by a conductor, the boundary condition to be fulfilled is that the total potential V defined in eq.(1.8) cancels out on the surface Σ of the cavity.

By following the strategy already used to get eq. (1.11) in the PCM method, from eq.(1.10) we can write the basic system giving the COSMO

apparent charges in the following matrix formalism⁹

$$\mathbf{BQ} = -\mathbf{V}^M \quad (1.12)$$

where the column matrix \mathbf{V}^M , of dimension equal to the number K of the tesserae, contains the solute electrostatic potential on each tessera, and the elements of the diagonal ($K \times K$) matrix \mathbf{B} are the value of the electrostatic potential the apparent charges induce on each other.

The capital letter used in eq.(1.12) for the matrix containing the apparent charges indicates that the latter are here derived for a real conductor, for which $\epsilon = \infty$; if the COSMO model is used to simulate a solvent with dielectric constant ϵ , these apparent charges have to be scaled: an effective way to do this is to multiply each charge by the factor $(\epsilon - 1)/\epsilon$.

As a final note we add that CPCM method, like the original COSMO, is correctly defined, and indeed it well reproduces solute energies and properties, only for highly polar solvents like water, for which the difference between the two values of $(1-1/\epsilon)$ obtained with $\epsilon = 78.5$ and $\epsilon = \infty$ is only 1.3%. Nevertheless, in the computational practice CPCM (and COSMO) is used also for the treatment of nonpolar solvents with $\epsilon \approx 2$, but in these cases it needs approximations which are physically by far less founded than those defined for the dielectric model, and not always easily acceptable.

(c). IEF

The main specificity of the IEF method⁶ is that, instead of starting from the boundary conditions as in the DPCM, it defines the Laplace and Poisson equations describing the specific system under scrutiny, here including also anisotropic dielectrics, ionic solutions, liquids with a flat surface boundary, quadrupolar liquids, and it introduces the relevant specifications by proper mathematical operators. The fundamental result is that the IEF formalism manages to treat structurally different systems within a common integral equation-like approach. In other words, the same considerations exploited in the isotropic DPCM model leading to the definition of a surface charge density $\sigma(s)$ which completely describes the solvent reaction response, are still valid here, also for the above mentioned extensions to non-isotropic systems.

Passing directly to the computational implementation, we recall that as for the previous ASC procedures, also in the IEF we exploit a tessellation of the cavity surface into K tesserae, and approximate the charge density σ by a piecewise constant function (i.e. a function constant on each tessera). Also in this case, the relations giving the corresponding apparent charges

q_k , can be expressed in the form of matrix equations:

$$\mathbf{C}\mathbf{q} = -\mathbf{g} \quad (1.13)$$

where \mathbf{q} is the column vector containing the apparent charges on the single tesserae, \mathbf{C} is a $K \times K$ matrix depending on the solvent dielectric constant and on the cavity parameters, and \mathbf{g} is a column vector depending on the solute electrostatic potential and the normal components of the corresponding field (both computed at the centers of the tesserae). Matrix \mathbf{C} is composed by the combination of four distinct matrices, each corresponding to different operators specific of the solvent (isotropic, anisotropic, or constituted by an ionic solution). In passing we recall that in the limit of isotropic dielectrics (i.e. in the same framework in which DPCM applies) matrix \mathbf{g} can be reduced to the potential matrix \mathbf{V}^M , with consequent reduction of computational time.

1.2.1.2. The solute cavity and its surface tessellation

In all the PCM approaches described above, one defines an empty cavity in the dielectric medium in which the solute M resides. The shape and size of the cavity are critical factors in the elaboration of a method. A cavity with a wrong shape introduces distortions in the description of the reaction field and of the related solvent effects.

The cavity shapes actually employed can be classified as follows:

- (1) regular shapes (i.e. spheres, ellipsoids, and cylinders);
- (2) molecular shapes given by the union of overlapping spheres;
- (3) molecular shapes obtained by exploiting the definition of specific molecular shape functions;
- (4) isodensity shapes.

In the following we shall limit our analysis to the molecular shapes only, being the latter the most used in the methods we have described in the previous sections. Regarding the procedures corresponding to the item 4 of our classification, namely those giving cavities determined directly from solute isodensity surfaces, we limit to recall that, in their most refined versions they may evaluate the isodensity surface in a self consistent way with the quantum mechanical calculation of the solute wave function.¹⁰

Before going inside the technical details of the procedure giving cavities of type 2), here represented by the computational technique called GEPOL,¹⁵ it is compulsory to make a further classification in terms of the

type of molecular surface we have to consider. In general we can define three main kinds of surfaces:

- (1) the proper van der Waals surface (S_w), which is the external surface resulting from a set of spheres centered on the atoms or group of atoms forming the solute;
- (2) the surface accessible to the solvent (S_a), defined as the surface generated by the center of the solvent, considered as a rigid sphere, when it rolls around the van der Waals surface;
- (3) the solvent excluding surface (S_e), which can be defined as the contact surface of a probe sphere (with radius equal to the molecular radius of the solvent molecules) rolling on the van der Waals surface.

It is easy to see that S_a can be equivalently defined as a S_w with radii enlarged by a quantity equal to the radius of the solvent, and that the last definition corresponds to the molecular cavity defined by Richards.¹⁶

GEPOL describes molecular cavities in terms of spheres centered on solute atoms, and solvent excluding surfaces (S_e) by adding some additional spheres. The latter are created with a sequential algorithm: when two sphere of the original cavity are close enough to exclude the solvent from the space between them, one or two additional spheres, not centered on atoms, are added, and the procedure is repeated to a prefixed threshold, considering each time all the possible couples of spheres, both of the first and the following generations.

The surface of each sphere is then partitioned into triangular tesserae, corresponding to the projection of the faces of a suitable polyhedron inscribed in the sphere. The latter may be defined in different ways, for example a geodesic partition based on pentakisdodecahedron giving origin to 60, 240, or 960 faces can be used. Tesserae with all the vertices inside the cavity are discarded, while those whose surface is partially covered by some other spheres are replaced by the polygonal tesserae with edges on the circles defined by each couple of intersecting spheres. In the most recent version of GEPOL the area a_k of the polygonal tessera k is calculated analytically by applying the Gauss-Bonnet theorem.

1.2.1.3. The escape charge problem

According to the Gauss theorem the sum of all the apparent charges $Q_\sigma(\text{theo})$ has to be related to the solute net charge Q_M by a simple expres-

sion:

$$Q_\sigma(\text{theo}) = -\frac{\epsilon-1}{\epsilon} Q_M \quad (1.14)$$

The difference between the theoretical (*theo*) and the computed (*comp*) value of the electronic and nuclear apparent charge

$$\Delta_Q = Q_\sigma(\text{theo}) - Q_\sigma(\text{comp}) \quad (1.15)$$

is an integral measure of the error introduced in the procedure. Since the first formulation of the DPCM method, the existence of this error was recognized and numerical corrections were introduced. The analysis of this error has been pushed more in depth later. According to this analysis there are two sources of error: the first is related to the discretization used in the BEM procedure and can be considered negligible if an accurate tessellation of the surface is exploited, the second, on the contrary, is more general as due to the presence of portions of solute electronic charge, $\rho^e(r)$, outside the cavity. In particular, this phenomenon which always happens in quantum mechanical calculations, should be taken into account by introducing an apparent charge in the volume outside the cavity:

$$\beta(r) = -\frac{\epsilon-1}{\epsilon} \rho^e(r) \quad (1.16)$$

with consequent increase of both the numerical complexity of the model and the computational costs.

In the oldest versions of DPCM the error due to the neglection of this effect was eliminated by resorting to a simple, but not accurate, renormalization of the apparent charges by multiplying each charge by a constant factor defined as follows:

$$f = \frac{Q_\sigma(\text{theo})}{Q_\sigma(\text{comp})} \quad (1.17)$$

In successive, and more accurate, versions of DPCM, a more complex procedure was introduced:¹⁷ in this procedure no numerical factors were exploited but the electronic charge lying outside the boundaries of the cavity was taken into account in an approximate way, i.e. without computing the volume charge β of eq.(1.16). The method was quite precise but also computer demanding.

The problem of the errors in the computed apparent charges is different for the IEF method. In an article appeared on the Journal of Chemical Physics,¹⁸ Chipman proposed and examined a new tractable formulation to get an accurate approximation of volume polarization effects arising from

the escaped charge without making a calculation in the outer volume. His proposal was to add to the standard ASC density σ we have defined in the previous sections, an additional apparent surface charge density representing approximately the electrostatic potential arising from the escaped charge. In this way, as most of the solute electrons are actually confined inside the cavity, the most important effects of the volume polarization are kept, and the computational effort is limited since only surface charges are used. In a successive paper¹⁹ we have shown that the method which Chipman independently formulated in 2000, is exactly equivalent to the IEF version we rewrote for isotropic solvents two years before.²⁰ Thanks to Chipman's analysis, now we have fully understood, and explained, the good results obtained by the modified IEF scheme as far as the approximation of volume polarization is concerned. Following both theoretical and numerical evidence, we can state that IEF does not need any renormalization of the surface charge.

1.2.2. Non-electrostatic contribution: repulsion-dispersion interactions

There are two different approach of modelling of dispersion and repulsion interactions in solution within the PCM framework.

The first method,^{21,22} is based on the use of pair potentials related to atoms or groups of atoms of the solvent S and the solute M , supplemented by a simplified description of the corresponding pairs distribution functions. The pairs potentials are independent on solute charge distribution; they are not involved in the QM description of the system, but they affect only the total free energy value.

The second method is based on the continuum representation of the solvent and the two contributions are treated separately. The starting point is the general expression derived from the intermolecular forces. In the previous discrete approach, however, this equation is applied to calculated or estimated potentials, V_{rep}^{AB} , available from the literature. We now substitute V_{rep}^{AB} for a suitable expression taken directly from the theory considering that, as originated from the Pauli exclusion principle, the repulsion forces between two interacting molecules increase with the overlap of the two distributions and they are strictly related to the density of electrons with the same spin.

Following this scheme and exploiting the same set of basic approximations characterizing continuum models (uniform charge distribution of the solvent, uniform response function), it is possible to derive the following

simple expression for the repulsion free energy:²³

$$G_{rep} = \alpha \int_{r \notin C} dr P_M(r) \quad (1.18)$$

where α is a suitable constant defined by some selected properties of the solvent and $P_M(r)$ is the solute electronic charge distribution. We note that in eq. (1.18) the integral is on exterior volume (i.e. excluding the cavity C) and thus G_{rep} is proportional to the fraction of solute electrons outside C ; as a consequence the minimization of the total free energy functional (1.1), when this contribution is included, provides an automatic confinement of the electronic cloud of the solute.

The continuum approach to the dispersion term has a much longer history than for the repulsion, and different procedures have been developed.

The original formal theory is expressed in terms of quantum electrodynamics with the continuum medium characterized by its spectrum of complex dielectric frequencies. A more recent formulation, derived from this theory, is based on the extension of the reaction field concept to a dipole subject to fluctuations exclusively electric in origin. Another procedure has been formulated²³ starting, as for the repulsion contribution, from the theory of intermolecular forces. Following the scheme commonly exploited to derive the electrostatic contribution to the interaction energy, the molecule B is substituted by a continuum medium, the solvent S , described by a surface charge density σ_S induced by the solute transition densities of M (the equivalent of A) and spreading on the cavity surface.

In this approximation, the dispersion free energy may be reduced to

$$G_{dis} = \frac{1}{\pi} \int_0^\infty d\omega \sum_{K \neq 0} \frac{\omega_{0K}^M}{(\omega_{0K}^M)^2 + \omega^2} \int dr_1 \times \\ \int_\Sigma \frac{dr_2}{r_{12}} P_M(0K|r_1) \sigma_S[\epsilon(i\omega), P_M(0K|r); r_2] \quad (1.19)$$

where $P_M(0K|r)$ and ω_{0K}^M are respectively transition densities and energies for the solute M (for transition to state K) and σ_S is the surface charge density induced in the solvent by the electric field of the charge distribution $P_M(0K|r)$; it depends on a solvent dielectric constant calculated at imaginary frequencies, $\epsilon(i\omega)$.

The second approach for the evaluation of the repulsion and dispersion interaction energy allows to introduce these contributions in the effective Hamiltonian through the solvnt reaction potential (1.5) and to evaluate their effects on the QM description and on the properties of the solute.

1.3. The Polarizable Continuum Model: The QM-SCRF Equations

We can now go back to the QM aspects of the PCM model by considering the methods for approximated solution of the effective non-linear Schrödinger equation for the solute. In principle, any variationally approximated solution of the effective Schrödinger equation can be obtained by imposing that first-order variation of \mathcal{G} with respect to an arbitrary variation of the solute wavefunction Ψ is zero. This corresponds to a search of the minimum of the free energy functional within the domain of the variational functional space considered. In the case of the Hartree-Fock theory, i.e. within the single determinant approximation, we finally obtain a generalized Fock operator, \hat{F}' , containing solvent terms, and the corresponding equation from which the final molecular orbitals of the solute has to be derived.

To go a step further we need to explicitly define the solvent terms. To do that, we prefer to introduce the common finite-basis approximation and a closed shell system. In this approximation we can eliminate the spin dependence (the occupied spin-orbitals occur in pairs) and expand the molecular orbitals (MOs) as a linear combination of atomic orbitals (LCAO). On performing the spin integration in the equations used so far, we find, for the electronic free energy:

$$G^e = \text{tr} \mathbf{P} \mathbf{h}^0 + \frac{1}{2} \text{tr} \mathbf{P} \mathbf{G}^0(\mathbf{P}) + \text{tr} \mathbf{P} \mathbf{h}^R + \frac{1}{2} \text{tr} \mathbf{P} \mathbf{X}^R(\mathbf{P}) \quad (1.20)$$

and, for the generalized Fock matrix:

$$\mathbf{F}' = \mathbf{h}^0 + \mathbf{G}^0(\mathbf{P}) + \mathbf{h}^R + \mathbf{X}^R(\mathbf{P}) \quad (1.21)$$

where \mathbf{P} is the one-electron density matrix and \mathbf{h}^0 and $\mathbf{G}^0(\mathbf{P})$ are the elements used in standard calculations *in vacuo* which collect one- and two-electron integrals (and in the case of DFT methods, the exchange correlation contribution), respectively.

In eqs. (1.20-1.21) the solvent contributions, previously introduced as potential functions, have been translated into a form recalling the vacuum system, just to emphasize the parallelism between the two calculations; in fact, also solvent effects can be partitioned in one- and two-electron contributions, indicated as \mathbf{h}^R and $\mathbf{X}^R(\mathbf{P})$, respectively. In general, \mathbf{h}^R and $\mathbf{X}^R(\mathbf{P})$ will contain different terms related to all possible interactions (dispersive, repulsive and electrostatic) between solute and solvent; here, we shall first focus on the electrostatic contribution when represented in

terms of the PCM apparent surface charges.

In this case, it is useful to partition the solute field (or potential) used to get the apparent charges (see eqs. 1.11-1.13) into its two sources, namely the solute nuclei and electrons. In this way we can derive a parallel partition of the apparent charges, q , into two components defined as electrons- and nuclei-induced, respectively. In this framework also the solute-solvent interaction part of the free energy may be divided into different contributions:

$$G^{el} = \int \int \frac{\rho(r)\sigma(s)}{|r-s|} dr ds = U_{ee} + U_{eN} + U_{Ne} + U_{NN} \quad (1.22)$$

where the first subscript refers to the component of the solute charge and the second to that of the apparent charges.

In the MO-LCAO approximation introduced above, the first three contributions to the interaction energy of eq.(1.22) (the last term, U_{NN} , is a constant dependent only on position and charge of solute nuclei) can be expressed in terms of matrices defined on the AOs. Actually, the two mixed $e-N$ interactions being formally equivalent can be cast in a single one-electron term, \mathbf{h}^{el} :

$$\mathbf{h}_{rs}^{el} = - \sum_k \mathbf{V}_{rs}(s_k) q^N(s_k) \quad (1.23)$$

while the $e-e$ term can be associated to a pseudo two-electron matrix \mathbf{X}^{el} defined as:

$$\mathbf{X}_{rs}^{el}(\mathbf{P}) = - \sum_k \mathbf{V}_{rs}(s_k) q^e(s_k; \mathbf{P}) \quad (1.24)$$

where we have explicitly indicated the dependence of the electron-induced charges, q^e , on the density matrix \mathbf{P} . In eqs.(1.23) \mathbf{V}_{rs} indicate AO potential integrals computed on the tesserae representative points (s_k).

Also the formalism reported above (see eq. (1.18)-(1.19)) to get a QM description of the dispersion and repulsion terms can be rewritten in a parallel way so to get the corresponding matrices; in particular, the repulsion term (eq.1.18) will give rise to a one-electron contribution only, \mathbf{h}^{rep} , while the dispersion (eq. 1.19) to both one and two-electron terms $\mathbf{h}^{dis} + \mathbf{X}^{dis}(\mathbf{P})$.²³

By combining all the electrostatic and nonelectrostatic terms we obtain the generalized Fock equation:

$$\begin{aligned} \mathbf{F}'\mathbf{T} &= [\mathbf{h}^0 + \mathbf{h}^{rep} + \mathbf{h}^{dis} + \mathbf{h}^{el} + \mathbf{G}^0(\mathbf{P}) + \mathbf{X}^{dis}(\mathbf{P}) + \mathbf{X}^{el}(\mathbf{P})] \mathbf{T} \\ &= [\mathbf{F}^0 + \mathbf{F}^R] \mathbf{T} = \mathbf{S}\mathbf{T}\varepsilon \end{aligned} \quad (1.25)$$

where \mathbf{T} is the matrix of MO coefficients, \mathbf{S} is the overlap matrix on AOs and ε contains the orbital energies. Eq.(1.25) can be solved with the same iterative procedure of the problem in vacuo; the only difference introduced by the presence of the continuum dielectric is that, at each SCF cycle, one has to simultaneously solve the standard quantum mechanical problem and the additional problem of the evaluation of the interaction matrices. In this scheme the apparent charges are obtained through a self-consistent technique which has to be nested to that determining the solute wave function; as a consequence, at the convergency, solute and solvent distribution charges are mutually equilibrated.

This brief analysis on the main quantum mechanical aspects of PCM continuum solvation models has been presented here to show the large potentialities of these models to be extended to many different QM levels including accurate correlated methods. In fact, the definition of the one- and two electrons matrix/operators \mathbf{h}^R and $\mathbf{X}^R(\mathbf{P})$, which reflect the standard \mathbf{h}^0 and $\mathbf{G}^0(\mathbf{P})$ terms, allows one to exploit theoretical and computational instruments largely used for isolated systems, without requiring any additional specificity. This is a very important point which makes the modern versions of continuum models the state-of-the-art of theoretical approaches to describe solvent effects not only on the solute energy but also on by far more complex quantities such as response properties or, more in general, molecular properties; more details on these applications will be given in following sections.

1.3.1. The analytical free energy derivatives

A further important methodological aspect of solvation methods is the computation of analytical derivatives of the solvation free energy with respect to various parameters. Here, in particular, we are focussing on PCM methods, and thus below we shall report the derivative procedures developed for such methods, however we want to stress that it is possible to obtain the same quantities also for other alternative solvation approaches.²⁴

The usual way to proceed consists in directly differentiating the expression of the solvation free energy; introducing the same approximation used in the previous section (i.e. the HF/DFT approach with MOs expanded as a linear combination of atomic orbitals (LCAO)), eq.(1.20) leads to the

following formulas for the first and second derivatives [25]:

$$\begin{aligned} G^\alpha &= \text{trPh}^\alpha + \frac{1}{2}\text{trPG}^\alpha(\mathbf{P}) + \text{trPh}_R^\alpha + \frac{1}{2}\text{trPX}_R^\alpha(\mathbf{P}) - \text{trS}^\alpha\mathbf{W} \\ &\quad + \frac{1}{2}U_{NN}^\alpha + V_{NN}^\alpha \end{aligned} \quad (1.26)$$

$$\begin{aligned} G^{\alpha\beta} &= \text{trPh}^{\alpha\beta} + \frac{1}{2}\text{trPG}^{\alpha\beta}(\mathbf{P}) + \text{trPh}_R^{\alpha\beta} + \frac{1}{2}\text{trPX}_R^{\alpha\beta}(\mathbf{P}) - \\ &\quad \text{trS}^{\alpha\beta}\mathbf{W} + \frac{1}{2}U_{NN}^{\alpha\beta} + V_{NN}^{\alpha\beta} + \text{trP}^\beta\mathbf{h}^\alpha + \text{trP}^\beta\mathbf{G}^\alpha(\mathbf{P}) + \\ &\quad \text{trP}^\beta\mathbf{h}_R^\alpha + \text{trP}^\beta\mathbf{X}_R^\alpha(\mathbf{P}) - \text{trS}^\alpha\mathbf{W}^\beta \end{aligned} \quad (1.27)$$

here we have introduced the short notation : $G^\alpha = \frac{\partial G}{\partial \alpha}$ and $G^{\alpha\beta} = \frac{\partial^2 G}{\partial \alpha \partial \beta}$ and we have omitted the subscript 0 for the *in vacuo* matrices \mathbf{h} and \mathbf{G} . Up-
perscripts of the matrices denote derivatives of the corresponding integrals over the AO basis functions. We have also introduced

$$\mathbf{W} = \mathbf{PF}'\mathbf{P} \quad (1.28)$$

and

$$\mathbf{W}^\beta = \mathbf{P}^\beta\mathbf{F}'\mathbf{P} + \mathbf{PF}'^{(\beta)}\mathbf{P} + \mathbf{PG}(\mathbf{P})\mathbf{P} + \mathbf{PX}_R(\mathbf{P})\mathbf{P} + \mathbf{PF}'\mathbf{P}^\beta \quad (1.29)$$

where

$$\mathbf{F}'^{(\beta)} = \mathbf{h}^\beta + \mathbf{h}_R^\beta + \mathbf{G}^\beta(\mathbf{P}) + \mathbf{X}_R^\beta \quad (1.30)$$

and \mathbf{P}^β denote the first derivative of the density matrix \mathbf{P} .

In analogy with the case of isolated systems, the general condition on the first derivative of \mathbf{P}

$$\text{trP}^\alpha\mathbf{F}' = -\text{trS}^\alpha\mathbf{PF}'\mathbf{P} \quad (1.31)$$

holds also for solvated systems and, as a consequence, the derivative G^α does not requires the knowledge of the density derivatives, \mathbf{P}^α .

On the contrary, the density derivatives are necessary to get higher order free energy derivatives: in particular the second free energy derivatives $G^{\alpha\beta}$ require the first derivative of the density matrix \mathbf{P}^α , which can be obtained by solving an appropriate coupled perturbed Hartree-Fock (CPHF) or Kohn-Sham (CPKS) equations.

The parameters $\alpha\beta$ introduced in eqs. (1.26)-(1.31) may be of a different nature. Here we shall consider the specific case in which they are nuclear coordinates λ ; later we shall also consider as possible derivative parameters the components of electric \mathbf{E} and magnetic \mathbf{B} static fields.

If we now restrict the analysis to the electrostatic part of the solute-solvent interactions, the derivatives of the solvent induced matrices, \mathbf{j}^λ , \mathbf{y}^λ , and $\mathbf{X}^\lambda(\mathbf{P})$ contain the derivatives of the apparent charges as obtained by differentiating the corresponding matrix equation (see eqs. 1.11, and 1.13 for the two PCM methods known as DPCM and IEF, respectively); hence, they depend on the derivatives of the geometrical quantities defining the cavity (tesserae areas, normal unit vectors, etc.). Also U_{NN}^λ , of purely nuclear origin, requires geometrical derivatives both of solute nuclear positions and of cavity surface elements. As a consequence, it follows that in evaluating electrostatic free energy derivatives $G_{el}^{\lambda\nu..}$ the important step is the analytical calculation of the derivatives of the surface geometrical elements with respect to the parameter λ ; this is in fact the time consuming step and it may also lead to uncontrolled numerical errors.

More recently, an alternative way to proceed has been proposed;²⁶ such procedure avoids computing any geometrical derivatives of the surface charges induced matrices. The new approach consists in differentiating first the electrostatic equation:

$$-\operatorname{div}(\epsilon(\lambda)\nabla V(\lambda)) = 4\pi\rho(\lambda)$$

with $\epsilon(\lambda) = 1$ inside the cavity and $\epsilon(\lambda) = \epsilon$ outside, so as to write $\partial V/\partial\lambda$ as a solution of the differentiated equation, and next inserting it in the derivative of the interaction energy formula. The only term coming from the change of the cavity with the motions of the nuclear coordinates is reduced to a simple integration on all the apparent charges lying on the moving part Γ of the cavity itself, namely:

$$\tau = \frac{\epsilon}{\epsilon - 1} \int_{\Gamma} \sigma^2(x)(U_{\Gamma}(x) \cdot n(x))dx \quad (1.32)$$

The function U_{Γ} assumes a very simple form for standard ASC cavities given as union of spheres, each of them centered on a solute nucleus, in this case we can write:

$$U_{\Gamma}(x) = \begin{cases} 0 & \text{if } x \notin \Gamma_l \\ e_{\alpha} & \text{if } x \in \Gamma_l \end{cases}$$

when λ is the α -th coordinate of the l -th nucleus. In the above expression, (e_1, e_2, e_3) is the orthonormal basis of the real space and Γ_l is the part of the cavity belonging to the sphere centered on the l -th nucleus.

Going back to the HF/DFT notation the final free energy derivative

reduces to:

$$\begin{aligned} \mathcal{G}_{el}^{\lambda} = & tr \mathbf{P} \mathbf{h}^{\lambda} + \frac{1}{2} tr \mathbf{P} \mathbf{G}^{\lambda}(\mathbf{P}) - tr \mathbf{S}^{\lambda} \mathbf{P} \mathbf{F}' \mathbf{P} + V_{NN}^{\lambda} \\ & + D(\partial \rho_M / \partial \lambda, \sigma) + \frac{1}{2} \tau \end{aligned} \quad (1.33)$$

where the first four terms are those which compose the derivative *in vacuo*. Let us notice however the Fock matrix in the third term is that given in the presence of the solvent (see eq. 1.21). On the other hand, the fifth and sixth terms are new. The former denotes the interaction energy between the variation of the solute charge distribution $\partial \rho_M / \partial \lambda$ and the apparent charge:

$$D\left(\frac{\partial \rho_M}{\partial \lambda}, \sigma\right) = \int \int_{R^3 \times R^3} \left[\frac{\partial \rho_M^N}{\partial \lambda} + \sum \mathbf{P}_{\mu\nu} \frac{\partial (\chi_{\mu} \chi_{\nu})}{\partial \lambda} \right] \frac{\sigma(y)}{|x - y|} dx dy$$

while the latter is shown in eq.(1.32).

In the MO-LCAO formalism the new $D(\partial \rho_M / \partial \lambda, \sigma)$ and τ terms can be rewritten in the following form:

$$D\left(\frac{\partial \rho_M}{\partial \lambda}, \sigma\right) = -tr [\mathbf{P}(\mathbf{D}^{Ne} + \mathbf{D}^{eN} + \mathbf{D}^{ee} + \mathbf{D}^{NN})] \quad (1.34)$$

$$\tau = -tr [\mathbf{P}(\tau^{Ne} + \tau^{ee})] + \tau^{NN} \quad (1.35)$$

where matrices \mathbf{D}^{xy} ($x, y = e, N$) describe interactions between variations of charge density x with the apparent charges produced by charge density y ; matrices τ^{xy} have a similar meaning.

Eq. (1.33) allows to obtain a computational manageable expression of the second derivatives $\mathcal{G}_{el}^{\lambda\nu}$ with respect nuclear coordinates . The standard expression:^{25,27}

$$\begin{aligned} \mathcal{G}_{el}^{\lambda\nu} = & tr \mathbf{P} [\mathbf{h}^{\lambda\nu} + \mathbf{h}_{el}^{\lambda\nu}] + \frac{1}{2} tr \mathbf{P} [\mathbf{G}^{(\lambda\nu)}(\mathbf{P}) + \mathbf{X}_{el}^{(\lambda\nu)}(\mathbf{P})] \\ & - tr \mathbf{S}^{\lambda\nu} \mathbf{W} + [\frac{1}{2} U_{NN}^{\lambda\nu} + V_{NN}^{\lambda\nu}] + tr \mathbf{P}^{\nu} \mathbf{F}'^{(\lambda)} - tr \mathbf{S}^{\lambda} \mathbf{W}^{\nu} \end{aligned} \quad (1.36)$$

can be largely simplified using eqs. (1.34)-(1.35) to get first and second derivatives of the solvent induced matrices, \mathbf{h}_{el} and \mathbf{X}_{el} , in this way in fact the order of derivation on the apparent charges reduces by one, and thus for the second derivative terms (such as $\mathbf{X}_{el}^{(\lambda\nu)}$), only first derivatives of $\{\mathbf{q}\}$ are needed.²⁸

The derivative of the density matrix \mathbf{P}^{λ} appearing in eq. (1.36) is obtained as solution of the CPHF/CPKS equation:

$$\mathbf{F}' \mathbf{P}^{\lambda} + \mathbf{F}'^{\lambda} \mathbf{P} + \mathbf{F}' \mathbf{P} \mathbf{S}^{\lambda} - \mathbf{P}^{\lambda} \mathbf{F}' - \mathbf{P} \mathbf{F}'^{\lambda} - \mathbf{S}^{\lambda} \mathbf{P} \mathbf{F}' = 0 \quad (1.37)$$

with the usual auxiliary orthonormality constraint

$$\mathbf{P}\mathbf{P}^\lambda + \mathbf{P}^\lambda\mathbf{P} + \mathbf{P}\mathbf{S}^\lambda\mathbf{P} = \mathbf{P}^\lambda$$

where both the Fock matrix and its derivatives contain explicit solvent-induced terms.

1.4. Modelling the Solute Properties

Modern quantum chemistry has been revolutionized by the ability to calculate derivatives of the wave function and associated quantities with respect to various parameters. In this section we describe how these methodologies have been modified to evaluate energy derivatives of solvated molecules with respect to nuclear positions, and external electromagnetic fields. The subject is so large and the related physical phenomena so numerous – suffice it to quote the various spectroscopies – that we shall necessarily limit our exposition to some specific aspects.

In particular, we shall focus on the PCM solvation methods only, showing how they have been linked to the most used QM formalisms exploited to evaluate geometries and molecular response properties both to electric and magnetic fields, or to the combination of them. The exposition will be thus divided accordingly to the following classification:

- (1) Energy properties
- (2) External or Response properties
- (3) Transition properties

The first category, energy properties, includes properties which can be obtained from the potential free energy surface $G(\mathbf{R})$, i.e. the free energy functional as a function of the nuclear coordinates \mathbf{R} . These include geometry and energy of critical points, in particular minima and saddle points of the first type, but also more complex structures as the intrinsic reaction paths and the corresponding energy profile. A good deal of chemical information is encoded in these properties.

The second category includes the response properties which describe the effects of any external applied field on a molecular system. This category includes the electronic and vibrational dipole (hyper)polarizabilities, both static and frequency dependent, magnetic and chiro-optic properties, etc.. Response properties are essential for a deeper understanding of molecular behaviors, and they represent the basis for an ever increasing number of technical applications.

The final category includes the transition properties between different states. The whole body of spectroscopy is based on these properties. Electronic transitions, in particular, open the way to a specific but large field of scientific enquiry and applications: the photochemistry.

The energy properties refers to stationary, time-independent, states of the solute molecules. This implies a complete equilibration between the degrees of freedom, electronic and nuclear, of the solute and of the solvent. For the last two categories, on the contrary, we have to consider not only an equilibrated solute-solvent system but also the dynamics of its response to a time dependent perturbation.

We recall that the results resumed in the following exposition have been made possible by the great advantage of the PCM approaches to reduce solvent effects to a set of operators which can be cast in a physically and formally simple form. In this framework, the inclusion of the PCM formalism into the various approaches provided by the quantum-mechanical theory becomes almost straightforward. In addition, the use of an accurate representation of the solvent field through an apparent surface charge – see previous chapter for more details – when joined to the definition of a realistic molecular cavity embedding the solute immersed in the dielectric, makes the PCM methods one of the most reliable among the various QM methods exploiting a continuum description of the solvent. Finally, their recent extension to different environments, like anisotropic solvents, ionic solutions, immiscible solvents with a contact surface, etc., allows, for the first time, the analysis of important response properties also for molecular systems immersed in non standard external matrices, such as liquid crystalline phases or symmetric crystalline frames, charged solutions like those representing the natural neighborhood of proteins and other biological molecules and membranes, just to quote few examples.

All the calculations we report in the following sections have been performed with two quantum mechanical computational codes, Gaussian²⁹ and GAMESS,³⁰ in which the PCM solvation method has been implemented.

1.4.1. Energy properties

In this section we present two studies on energy properties of solvated systems. In the first we shall analyze the solvent effects on the stability of isomers of glycine and alanine in water solution, whereas in the second we shall analyze the solvent induced changes on the structures of donor-acceptor conjugated systems. The two studies represent two examples of application of solvation continuum models in two different important fields

of modern research: biology and material science.

1.4.1.1. Isomer energy differences: glycine and alanine in aqueous solution.

An important contribution to the comprehension of a large number of biochemical problems is given by a detailed knowledge of the conformational properties of molecular species in aqueous solutions. The necessary information can be only partly inferred from experimental data and its necessary to resort to theoretical methods.

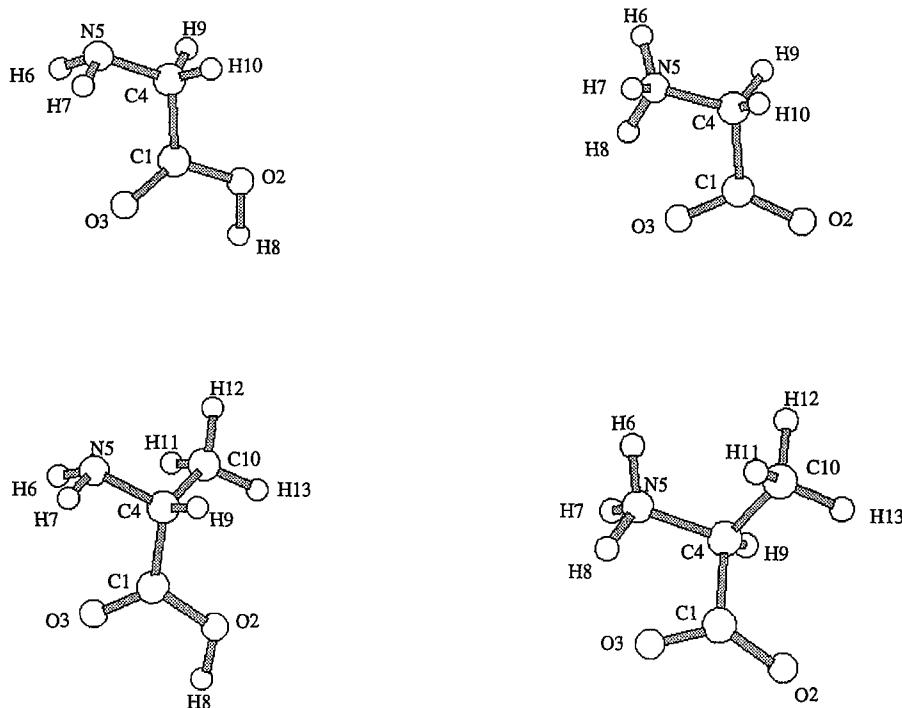
The amino acids are among the most important molecules of biochemical interest, since they constitute the building blocks of peptides and proteins. Most of their outstanding properties are due to the amphoteric nature which derives from the presence, in their structure, of two functional groups of different polarity (the amino, NH₂ and the carboxyl, COOH, groups). These group are also responsible of the existence of two tautomeric forms, neutral (NT) and zwitterionic (ZW) forms.

In the gas phase the neutral form is energetically favored, while in condensed phases, such as solutions in polar solvents and crystals, the ionic form is more stable. Glycine and alanine are zwitterionic in water solutions and their description is a rather challenging task for continuum solvation models. In the following we summarize the results of a study we have published on Theochem³¹ on structures and energies of tautomers of glycine and alanine in the framework of the IEF version of the Polarizable Continuum Model.

The geometry of zwitterionic (ZW) and neutral (NT) forms in water solution was optimized at the B3LYP/6-31G(d) level. The potential energy surface (PES) of neutral glycine and alanine has been studied in detail by several authors. In spite of their apparent simplicity, these two molecules give rise in vacuo to quite complex potential surfaces which include several minima and transition states connecting them. In both cases, in the lowest energy structure, the two hydrogen atoms of the amino group lean towards a lone pair of carboxyl oxygen; the structure for glycine has Cs symmetry. The ZW case is rather more complex. In fact, the zwitterionic form does not seem to exist in the gas phase as a stable structure; explicit water molecules, a crystal field or a continuum solvent distribution is needed to stabilize it.

Bearing in mind these apparently contradictory results, we chose in vacuo-type geometries as starting structures to perform our analysis, by applying the PCM-IEF method. The most significant optimized geometri-

cal parameters for the neutral and zwitterionic forms of the two systems in water are reported in the Figure.



Parameter (NT)	Glycine (C_2)	Alanine	Parameter (ZW)	Glycine	Alanine
d N5-H6	1.019	1.020	d N5-H6	1.030	1.030
d N5-H7	1.019	1.021	d N5-H7	1.029	1.030
d C4-N5	1.452	1.463	d C4-N5	1.503	1.515
d C1-C4	1.525	1.532	d C1-C4	1.553	1.561
d C1-O2	1.355	1.344	d C1-O2	1.247	1.247
d C1-O3	1.211	1.219	d C1-O3	1.265	1.266
d O2-H8	0.976	0.991	d O3-H8	1.834	1.800
a C1-C4-N5	112.62	112.62	a C1-C4-N5	102.59	105.60
d O3-C1-N4-N5	0.	15.78	d O3-C1-N4-N5	8.53	-9.13
d H6-N5-C4-C1	56.50	59.20	d H6-N5-C4-C1	-12.67	12.52
d H7-N5-C4-C1	-56.50	-54.31	d H7-N5-C4-C1	103.31	-103.51

^a Bond lengths in Å, angles in degrees.

Fig. 1.1 Optimized geometries of glycine and S-alanine: left, neutral forms; right, Zwitterionic forms

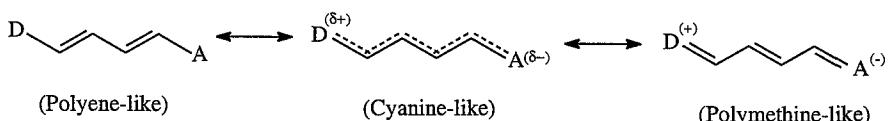
Two features worth to be noted: the introduction of a methyl group in the side chain of the NT glycine induces a deviation from planarity of about

158° in the OCCN backbone; a similar distortion can be found passing to ZW too, owing to the formation of a hydrogen bond. The latter is 1.834 Å in glycine ZW and 1.800 Å in alanine ZW. Such optimized geometries have been used for the vibrational analysis needed to calculate the thermal contributions to the free energies.

The final computed results for NT-ZW process in water are -3.42 kcal/mol for glycine and -4.48 kcal/mol for alanine. These ΔG_{ZW-NT} values are able to correctly predict ZW stability without requiring explicit water molecules and to reproduce suitably the difference between alanine and glycine. In fact, the difference between the ΔG_{ZW-NT} of the molecules is equal to 0.68 kcal/mol from the estimates and 1.06 kcal/mol from our calculation. We have to remark that reliable experimental data are not available: amino acids in fact are found to exist mainly in zwitterionic form in neutral water solutions. Unfortunately, the mole fraction of the neutral form in such media is so negligible that it cannot be revealed through experimental measurements. Therefore, the equilibrium constant of the NT-ZW process in solution is not known and thermodynamic quantities are not accessible but they can only be estimated.

1.4.1.2. Geometries of donor-acceptor solutes: push-pull molecules.

The selected systems for the analysis of solvent effects on geometrical structures belong to the category of push-pull π -conjugated molecules, i.e. systems where a conjugated linker segment is capped by a donor group on one end and an acceptor on the other. The specificity of this kind of molecules is that their ground-state (GS) structure can be represented as a combination of different resonance forms, differing in the extent of charge separation



When the donor and acceptor groups are weak, the neutral polyene-like resonance form dominates the ground state and the molecule has a structure with a distinct alternation in the bond length between neighboring carbon atoms, i.e. a high degree of bond length alternation (BLA), defined as the difference between average single and double bond distances in the conjugated pathway. When donor and acceptor substituents become

stronger, the contribution of ionic resonance forms to the ground state increases, and BLA first decreases until almost zero values (in the intermediate cyanine-like structure with partial charge separation) and then increases again towards the high negative values proper of the fully charge-separated polymethine-like (or zwitterionic) form. The relative contribution of these three resonance forms to the GS is affected by the polarity of the solvent in which the chromophore is dissolved; a more polar solvent increases the GS state polarization, which makes the partial and/or complete ionic forms more important.

In this section we shall present and discuss some numerical results regarding structures of two series of noncentrosymmetric polyenes: $\text{NH}_2(\text{CH}=\text{CH})_n\text{R}$ ($n=1,2$), with $\text{R}=\text{CHO}$ (series I) and with $\text{R}=\text{NO}_2$ (series II). These results have been extracted from a study we have published on the Journal of the American Chemical Society³² about solvent effects on electronic and vibrational components of linear and nonlinear optical properties of Donor-Acceptor polyenes.

All the calculations have been performed at SCF level with a basis set equal to the Dunning/Huzinaga valence double-zeta. All the calculations in solution have been performed using the IEF version of the PCM method and they refer to a medium having dielectric constant $\epsilon = 78.5$ corresponding to the static dielectric constants of liquid water at 298 K.

In the Tables 1.1 and 1.2 we report the bond lengths (\AA) of $\text{H}_2\text{N}(\text{CH}=\text{CH})\text{R}$ in vacuo and in water, and of $\text{H}_2\text{N}(\text{C}_4\text{H}=\text{C}_3\text{H}-\text{C}_2\text{H}=\text{C}_1\text{H})\text{R}$ in vacuo and in water, respectively.

Table 1.1 Bond lengths (\AA) of $\text{H}_2\text{N}(\text{CH}=\text{CH})\text{R}$ in vacuo and in water; if $\text{R}=\text{CHO}$ $\text{X}=\text{C}$ and $\text{Y}=\text{O}$, otherwise, when $\text{R}=\text{NO}_2$, $\text{X}=\text{N}$ and $\text{Y}=\text{O}$.

	CHO		NO_2	
	vacuum	water	vacuum	water
R_{XY}	1.2284	1.2506	1.2478	1.2696
R_{CX}	1.4543	1.4304	1.4216	1.3721
R_{CC}	1.3511	1.3696	1.3494	1.3809
R_{CN}	1.3664	1.3448	1.3552	1.3236

The two main aspects to be pointed out are the effects due to the nature of substituents and, more important, to the presence of the solvent.

From both Tables it results that more polar D-A substituents (i.e.

Table 1.2
 Bond lengths (\AA) of $\text{H}_2\text{N}(\text{C}_4\text{H}=\text{C}_3\text{H}-\text{C}_2\text{H}=\text{C}_1\text{H})\text{R}$
 in vacuo and in water; if $\text{R}=\text{CHO}$ $\text{X}=\text{C}$ and $\text{Y}=\text{O}$,
 otherwise, when $\text{R}=\text{NO}_2$, $\text{X}=\text{N}$ and $\text{Y}=\text{O}$.

	CHO		NO_2	
	vacuum	water	vacuum	water
R_{XY}	1.2274	1.2477	1.2464	1.2830
$\text{R}_{\text{C}_1\text{X}}$	1.4608	1.4385	1.4311	1.3540
$\text{R}_{\text{C}_1\text{C}_2}$	1.3483	1.3645	1.3440	1.3949
$\text{R}_{\text{C}_2\text{C}_3}$	1.4507	1.4329	1.4413	1.3896
$\text{R}_{\text{C}_3\text{C}_4}$	1.3499	1.3661	1.3550	1.3966
R_{CN}	1.3728	1.3526	1.3644	1.3229

molecules of series II) act to reduce the length of the single bonds and to increase the length of the double bonds. The same effect, even if largely amplified, is given by the presence of the solvent; this can be better quantified in terms of the bond length alternation (BLA) which decreases from 0.1067 \AA to 0.0704 \AA for $\text{NH}_2(\text{CH}=\text{CH})_2\text{CHO}$, and from 0.0917 to -0.0061 for $\text{NH}_2(\text{CH}=\text{CH})_2\text{NO}_2$, passing from gas phase to aqueous solution. This large solvent effect is in good accord with the results obtained previously by Gao and Alhambra,³³ even if in their work a completely different model is used to represent the solvent molecules (i.e. Monte Carlo simulations) and a QM/MM level of calculation is performed. In that paper, the authors find, for the first term of series I, a decrease of the single CC bond length of the order of 0.023 \AA and an increase of the double bond length of 0.021 \AA (to be compared with 0.024 \AA and 0.019 \AA of our calculations), while for the second term a decrease of 0.036 \AA in the BLA value (exactly equal to our value).

In addition, BLA data show that, as one might expect, the presence of a polar solvent like water induces a migration of electronic charge distribution along the molecular axis from D to A group; in this way the electronic structure of the molecular system, which in the gas phase is well characterized by a polyene-like structure only, in solution contains a large contribution also from the partially charge-separated cyanine-like form: in $\text{NH}_2(\text{CH}=\text{CH})_2\text{NO}_2$ in water the BLA becomes almost zero, exactly as in an ideal cyanine structure.

The same conclusions can be reached from a different way, by taking into account the changes on the net charges of the various chemical groups present in the molecule, when passing from gas phase to solution. In Table 1.3 we report the total dipole moment and the Mulliken net charges on the D-A groups as well as on the intermediate CH groups both in vacuo

and in solution:

Table 1.3 Ground state dipole moment (debye) and Mulliken net charges (a.u.) of R-(X₁=X₂-X₃=X₄)-NH₂ (with R=CHO,NO₂) both in gas phase and in aqueous solution.

μ_r	group charges (a.u.)						
	R	X ₁	X ₂	X ₃	X ₄	NH ₂	
CHO _(g)	-6.3179	-0.1555	-0.0531	+0.2699		-0.0612	
CHO _(aq)	-9.2531	-0.2602	-0.1131	+0.3475		+0.0258	
CHO _(g)	-7.7909	-0.1605	-0.0571	+0.1037	-0.0381	+0.2339	-0.0820
CHO _(aq)	-12.0932	-0.2599	-0.1253	+0.1769	-0.0935	+0.3047	-0.0029
NO _{2(g)}	-8.1474	-0.4420	+0.0705	+0.3925		-0.0211	
NO _{2(aq)}	-12.3971	-0.6449	+0.0684	+0.4582		+0.1183	
NO _{2(g)}	-10.1829	-0.4492	+0.0514	+0.2323	-0.0569	+0.2815	-0.0590
NO _{2(aq)}	-19.6363	-0.7548	+0.0296	+0.2807	-0.0656	+0.4059	+0.1041

The most evident result of the data reported in the Table is that the acceptor (A) NH₂ group which in vacuo shows significative partial negative charges in both series, in solution present a positive charge, with the only exception of the second term of series I where the net charge is almost zero. This behavior once again shows that in solution the cyanine-like structure assumes a much more important role and becomes the main one for the series II. The electron-migration towards the A group is reflected also in the net charges of the intermediate CH groups which all increase in magnitude passing from vacuum to aqueous solution

1.4.2. IR spectra: intensities and frequencies

The definition of vibrational intensities for molecules in solution requires some modifications with respect to the isolated system. The presence of the solvent in fact modifies not only the solute charge distribution but also the probing electric field acting on the molecule. As we shall see in the following sections, this is a problem of general occurrence when an external field interacts with a molecule in a condensed phase (historically it is known as 'local field effect').

Assuming the harmonic approximation as valid, vibrational frequencies for a solvated molecule can be expressed in terms of the second derivatives of the proper energy functional, the free energy \mathcal{G} , with respect to the

normal coordinates (i.e. the Hessian matrix):

$$[H]_{ij} = \left(\frac{\partial^2 \mathcal{G}(\mathbf{R})}{\partial R_i \partial R_j} \right)_{\mathbf{R}_0} \quad (1.38)$$

where the derivatives are computed at the equilibrium geometry.

In parallel, the intensities can be defined by rewriting the transition probability between two vibrational states 0 and 1 as induced by a radiation of frequency ω ; if we indicate with \mathcal{G}_0 and \mathcal{G}_1 the free energies corresponding to each state, the required quantity becomes:

$$W_{1 \leftarrow 0}^{sol} = \frac{2\pi}{\hbar} \delta(\mathcal{G}_1 - \mathcal{G}_0 - \hbar\omega) |\langle 1 | (\mu + \tilde{\mu}) \cdot \mathbf{F} | 0 \rangle|^2 \quad (1.39)$$

where \mathbf{F} is the radiation electric field in the dielectric, μ is the dipole moment of the molecule and $\tilde{\mu}$ is the dipole moment arising from the polarization induced by the molecule on the dielectric. The $\mathcal{G}_1 - \mathcal{G}_0$ “excitation energy” can be expressed, within the harmonic approximation, as a function of the eigenvalues of the Hessian (1.38). Starting from eq.(1.39) an expression for the integrated intensity can be derived:³⁴

$$A^{sol} = \frac{\pi N_A}{3n_s c^2} \left(\frac{\partial(\mu + \tilde{\mu})}{\partial Q_i} \right)^2 \quad (1.40)$$

where N_A is the Avogadro’s number, c the velocity of light in vacuo, n_s the refractive index of the solution and Q_i the mass-weighted normal coordinate associated to the vibrational mode i . The $\partial(\mu + \tilde{\mu})/\partial Q_i$ derivatives are obtained as the mixed second derivatives of the free energy with respect to Q_i and the electric field.

In the previous sections we have resumed the procedures to get the all the required free energy derivatives appearing in eqs. (1.38)-(1.40); here, we try to introduce a further aspect of the phenomenon of nuclear vibrations in solvated systems.

The basic point is how to describe the solvent response to solute vibrations, or more precisely, if, and how, to introduce solvent polarization under non-equilibrium conditions. The concept of solute-solvent non-equilibrium system involves many different aspects related to the real nature of the solvent polarization. This latter can be decomposed into different contributions related to the various degrees of freedom of the solvent molecules. The motions associated to these degrees of freedom involve different time-scales: suffice it to recall the difference (orders of magnitude) in the relaxation times typical of the translational and electronic motions. In the

common practice, such contributions are grouped in two terms only: one term accounts for all the motions which are slower than those involved in the physical phenomenon under examination, the other includes the faster contributions. The next assumption usually exploited is that only the latter are instantaneously equilibrated to the momentary molecule charge distribution whereas the former cannot readjust, giving rise to a non-equilibrium solvent-solute system. This partition and the following non-equilibrium approach has been originally formulated and commonly applied to electronic processes (for example solute electronic transitions) as well as to the evaluation of solute response to external oscillating fields (see next sections). In the case of vibrations the slow term will contain the contributions arising from the motions of the solvent molecules as a whole (translations and rotations), whereas the fast term will take into account the internal molecular motions (electronic and vibrational).³⁵ Such a division is reasonable as far as the solute vibration under study is sufficiently far in frequency from those of the solvent; when necessary this distinction may be modified.

1.4.2.1. IR spectra of ketones in solution

As an example of solvent effects on IR spectra, we summarize a study we have published on the Journal of Chemical Physics³⁵ about frequencies and IR intensities for the C=O stretching vibration of a series of dialkyl ketones: dimethyl ketone (DMK), methyl ethyl ketone (MEK), *sec*-butyl methyl ketone (SBMK), *tert*-butyl methyl ketone (TBMK). The solvents of which we have modeled the effect are: cyclohexane (cyc), carbon tetrachloride (CCl₄), benzene (C₆H₆), 1,2-dichloroethane (1,2-dce) and acetonitrile (acn).

In order to evaluate the differences in the approaches, we have performed calculations exploiting both equilibrium and non-equilibrium models within the PCM-IEF framework. All the computed frequencies and IR intensities we will report in the following are obtained in the harmonic approximation, no anharmonic effects have been considered.

The main body of the calculations have been performed using DFT with the B3LYP hybrid functional, and the 6-31G(d) basis set. The geometries of all the systems have been re-optimized in each phase. In Table 1.4 we report the optimized C=O bond distances for the four ketones in vacuo and in various solvents:

The solvent effects on geometries are reflected in the changes of frequen-

Table 1.4 B3LYP/6-31G(d) C=O bond distances in Å for the various ketones in vacuo and in various solvents.

	vac	sol				
		cyc	CCl ₄	C ₆ H ₆	1,2-dce	acn
DMK	1.216	1.218	1.218	1.218	1.221	1.222
MEK	1.216	1.218	1.219	1.219	1.222	1.222
SBMK	1.217	1.219	1.220	1.220	1.222	1.223
TBMK	1.217	1.219	1.220	1.220	1.222	1.223

cies with increasing the polarity of the solvent. Let us see what happens to the frequency values in passing from gas phase to solution, as evaluated by equilibrium and non-equilibrium models and experimentally. To make the comparison easy, we report in Table 1.5 the quantity:

$$\delta_x = \nu_{vac} - \nu_x \quad (1.41)$$

where x runs over the solvents and we have taken as reference the frequency value in gas-phase, ν_{vac}

Table 1.5 Calculated and experimental frequency shifts (cm^{-1}) with respect to gas phase for various ketones in various solvents

	DMK			MEK		
	eq	neq	exp	eq	neq	exp
cyc	12	13	18	11	12	14
CCl ₄	14	15	22	13	14	19
C ₆ H ₆	14	15	23	13	14	21
1,2-dce	29	24	27	27	22	27
acn	33	25	26	31	23	27
SBMK			TBMK			
	eq	neq	exp	eq	neq	exp
cyc	11	12	13	11	12	13
CCl ₄	15	13	17	12	13	17
C ₆ H ₆	12	13	20	12	13	19
1,2-dce	27	22	24	26	22	22
acn	31	24	24	30	23	22

The largest differences between the results obtained with the two solvent models (labelled eq and neq in the tables) are found in polar solvents; however such differences are still very small (being the largest eq-neq difference of 8 cm^{-1} for DMK in acetonitrile) as well as the computed-experimental

differences: as a consequence, it is difficult to state which model is in better agreement with experiments. Equilibrium and non-equilibrium δ values are instead very similar in the case of non-polar solvents: this is not surprising since static and optical dielectric constants are very similar in such cases.

Passing to IR intensities, we first note that both equilibrium and nonequilibrium models give the correct trend in passing from apolar to polar solvents; both the two sets of computed results and the experimental data show a net increment of the intensities from vacuo to polar solvents. For a more detailed analysis we prefer to shift from absolute values to differences as done in the previous analysis on frequencies; in Table 1.6 we report the solvent-gas phase shift on intensities for the four ketones.

Table 1.6 Calculated and experimental intensity shifts (km/mol) with respect to gas phase for various ketones in various solvents.

	DMK			MEK		
	eq	neq	exp	eq	neq	exp
cyc	33	36	25±7	39	42	29±6
CCl ₄	36	40	48±10	42	46	45±6
C ₆ H ₆	36	39	44±14	43	46	41±7
1,2-dce	92	49	46±10	99	55	55±7
acn	113	49	55±7	121	54	55±7
	SBMK			TBMK		
	eq	neq	exp	eq	neq	exp
cyc	41	45	50±7	43	47	46±6
CCl ₄	46	49	67±6	46	51	63±6
C ₆ H ₆	47	50	65±6	47	51	55±7
1,2-dce	114	64	64±7	112	61	68±6
acn	152	76	68±6	137	59	65±7

This time, when comparing the solvent shift with respect to gas-phase, we note that in polar solvents the two models show significant differences. In particular, the nonequilibrium results present a by far better agreement with experiments, leading the absolute value to be within the experimental range. On the contrary, very large discrepancies are obtained between equilibrium and experimental data. Once again, in apolar solvents the two models give very similar results and both well agree with experiments.

The results we have presented are sufficient to show that the computational procedure gives frequencies and intensities comparable with the experimental ones. The use of a non-equilibrium model for the C=O stretch

does not significantly change the frequency: on the contrary a better behavior of the non-equilibrium model turns out evident when looking at intensities in polar solvents.

1.4.3. Electric dipole polarizabilities

The dipole (hyper)polarizabilities of a molecule *in vacuo* are defined in terms of a Taylor expansion of the dipole moment when the molecule is subjected to an external homogeneous electric field; alternatively, they may be expressed in terms of the corresponding Taylor expansion of the energy of the molecule.³⁶ Analogous definitions hold for molecules in solution. In this case the external field corresponds to the perturbing field acting locally on the molecular solute and the expansion of the energy is substituted by the expansion of the free energy function G . The expansion terms are collected by order into the so-called polarizability tensors $\gamma^{(n)}$. We shall follow the usual convention which indicates as α , β , γ the tensors corresponding to the first three $\gamma^{(n)}$ sets of coefficients.

When the external field has an oscillatory behavior, all these quantities depend on the frequency of such oscillations; for a given $\gamma^{(n)}$ we have to consider the frequencies and phases of the various components of external fields that can be combined in all possible ways to give different electric molecular response. These elements constitute the essential part of the linear and nonlinear optics, a subject for which there is a remarkable interest to know the influence of solvation effects.

The molecular response to an electric field regards its whole charge distribution, electron and nuclei. We may introduce also for molecules in solution the usual partition of the theoretical chemistry into electric and nuclear parts.

1.4.3.1. Electronic contribution

The electronic contribution can be computed using two derivative schemes involving quantum mechanical calculations of the free energy or, alternatively, of the dipole moment followed by derivatives with respect to the perturbing external field, computed at zero intensity. At Hartree-Fock (HF) or Density Functional Theory level both approaches lead to the use of the coupled HF or KS theory either in its time-independent (CHF, CKS) or time-dependent (TDCHF, TDDFT) version according to the case.

The solute Hamiltonian must be now supplemented by further terms describing the interaction with the external electric field; in particular, the

corresponding one-electron contribution to the Fock matrix \mathbf{F}' becomes

$$\mathbf{h}' = \mathbf{h} + \mathbf{h}_R + \sum_a \mathbf{m}_a E_a \quad (1.42)$$

where \mathbf{m}_a is the matrix of the a th Cartesian component of the dipole moment operator, and E_a is the corresponding component of the electric field vector.

To obtain the various electric response functions we have to determine the density matrix of the unperturbed system \mathbf{P}^0 and its derivatives (\mathbf{P}^a , \mathbf{P}^{ab} , etc.) with respect to E components, corresponding to the order of the electric response functions.

To obtain time-dependent properties, we have to pass from the basic model to an extended version in which the solute is described thorough a time-dependent Schrödinger equations. In this extended version of the model we have also to introduce the time-dependence of the solvent polarization, which is expressed in terms of a Fourier expansion and requires the whole frequency spectrum of the dielectric permittivity $\epsilon(\omega)$ of the solvent.

The time-dependent Schrödinger equation may be obtained by rewriting Frenkel variation principle for the case of nonlinear Hamiltonian as follows:³⁷

$$\delta \left\langle \Psi \left| \hat{\mathcal{G}} - i \frac{\partial}{\partial t} \right| \Psi \right\rangle = 0 \quad (1.43)$$

with

$$\hat{\mathcal{G}} = \hat{H}^0 + \hat{\rho}_r V_r^R + \frac{1}{2} \hat{\rho}_r \mathbf{V}_{rr'}^R \hat{\rho}_{r'} + \hat{\rho}_r V'_r(t) \quad (1.44)$$

where the term $\hat{\rho}_r V'_r(t)$ represents the interaction of the solute with the electric field:

$$\hat{\rho}_r V'_r(t) = \hat{\rho}_r \sum_k \hat{m}_k E_k(t) \quad (1.45)$$

and \hat{m} is the electric dipole moment operator.

Applying the Frenkel variational principle, we arrive at the following time-dependent HF/KS equation

$$\left(\mathbf{F}' - \frac{\partial}{\partial t} \right) \mathbf{T} = \mathbf{T} \epsilon \quad (1.46)$$

where \mathbf{T} is the matrix containing the MO coefficients. The solutions can be obtained in the TDHF/TDDFT scheme:^{38,39} we first expand eq. (1.46) in its Fourier components and then each component is expanded in terms

of the components of the external field. The separation by orders leads to a set of coupled perturbed HF (KS) equations whose Fock matrices can be written in the following form:

$$\mathbf{F}' = \mathbf{h}' + \mathbf{G}(\mathbf{P}_0) + \mathbf{X}_0(\mathbf{P}_0) \quad (1.47)$$

$$\mathbf{F}'^a(\omega_\sigma; \omega) = \mathbf{m}_a + \mathbf{G}(\mathbf{P}^a(\omega_\sigma; \omega)) + \mathbf{X}_{\omega_\sigma}(\mathbf{P}^a(\omega_\sigma; \omega)) \quad (1.48)$$

$$\mathbf{F}'^{ab}(\omega_\sigma; \omega_1, \omega_2) = \mathbf{G}(\mathbf{P}^{ab}(\omega_\sigma; \omega_1, \omega_2)) + \mathbf{X}_{\omega_\sigma}(\mathbf{P}^{ab}(\omega_\sigma; \omega_1, \omega_2)) \quad (1.49)$$

$$\mathbf{F}'^{abc}(\omega_\sigma; \omega_1, \omega_2, \omega_3) = \mathbf{G}(\mathbf{P}^{abc}(\omega_\sigma; \omega_1, \omega_2, \omega_3)) + \mathbf{X}_{\omega_\sigma}(\mathbf{P}^{abc}(\omega_\sigma; \omega_1, \omega_2, \omega_3)) \quad (1.50)$$

where ω_σ is the frequency of the resulting wave (i.e. $\omega_\sigma = \sum_k \omega_k$, with ω_k including the sign), and matrices $\mathbf{X}_{\omega_\sigma}(\mathbf{P}^{ab..}(\omega_\sigma))$ represent the interactions with the apparent charges induced by the perturbed electron density $\mathbf{P}^{ab..}(\omega_\sigma)$ oscillating at frequency ω_σ . There is, of course a number of equations of this type, corresponding to the different combinations of frequencies, each one related to a different phenomenon of the nonlinear optics (see below).

The dynamic (hyper)polarizabilities of interest can be expressed in the following forms:

$$\alpha_a(\omega_\sigma; \omega) = -tr \mathbf{m}_a \mathbf{P}^b(\omega_\sigma; \omega) \quad (1.51)$$

$$\beta_{abc}(\omega_\sigma; \omega_1, \omega_2) = -tr \mathbf{m}_a \mathbf{P}^{bc}(\omega_\sigma; \omega_1, \omega_2) \quad (1.52)$$

$$\gamma^{abc}(\omega_\sigma; \omega_1, \omega_2, \omega_3) = -tr \mathbf{m}_a \mathbf{P}^{bcd}(\omega_\sigma; \omega_1, \omega_2, \omega_3) \quad (1.53)$$

which require the knowledge of the derivatives of the electronic densities \mathbf{P}^{abc} , etc., obtained from the solution of the TDHF/TDDFT equations.

Since the frequency-dependent response of the solvent is included in the kernel of the integrals collected into $\mathbf{X}_{\omega_\sigma}$, the resulting hyperpolarizabilities will also depend on the frequency spectrum of the dielectric function $\epsilon(\omega)$ of the solvent. When $\epsilon(\omega)$ is described by the Debye formula (i.e. in terms of a single relaxation mode), i.e.

$$\epsilon(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 - i\omega\tau_D} \quad (1.54)$$

the resulting dispersion curve depends on the frequency corresponding to the inverse of the Debye relaxation time.

We have remarked that there is a large number of coupled HF/KS equations to be solved in order to get electric response properties. However, as for the case of a molecule in vacuo, also for solutes the computational effort can be reduced by resorting the exploiting to so called $(2n+1)$ rule that permits to get the $(2n+1)$ th response using n th derivative of \mathbf{P} .³⁹

1.4.3.2. Linear response theory

The calculation of the dynamic polarizabilities by using the TDHF equations described above is equivalent to the Random Phase Approximation (RPA) within the linear response theory, LRT.^{38,40} The subject of the linear response theory is the first-order change in the expectation value, $\delta \langle A \rangle$, of a property A , as response of the system to a perturbing field $H'(t)$. For variational wavefunction $\Psi = \Psi(\{p_i\})$ the linear response can be obtained from the determination of the changes of the variational parameter $\{p_i\}$ needed to satisfy the Frenkel Variational Principle through first-order in $H'(t)$.

The basic equation of the LRT for a periodic perturbation $H'(t) = \frac{1}{2}[A_\omega e^{i\omega t} + A_{-\omega} e^{-i\omega t}]$ can be recast in the following matrix form:

$$\langle\langle B; A \rangle\rangle_\omega = \begin{pmatrix} \nabla B \\ B \nabla \end{pmatrix}^\dagger \Pi \begin{pmatrix} \nabla A \\ A \nabla \end{pmatrix} \quad (1.55)$$

where $\langle\langle B; A \rangle\rangle_\omega$ denotes the dynamic response function describing the first-order change of the property B , with associate operator \hat{B} , as response to the periodic perturbation of frequency ω , described by operator \hat{A} ; quantities ∇M and $M \nabla$ collects elements of the gradient of the expectation value $\langle M \rangle$ with respect to the variational parameters. Matrix Π is called *response matrix* and is given by

$$\Pi = [\Omega - \omega \Delta]^{-1} \quad (1.56)$$

where matrix Ω collects the full Hessian of the free energy \mathcal{G} with respect the variational parameters, and matrix Δ contains the diagonal blocks of the Hessian of the projector operator $Q = \langle \Psi | \Psi \rangle^{-1} (1 - \langle \Psi | \Psi \rangle^{-1} |\Psi \rangle \langle \Psi |)$.

The extension of the general RPA theory to solvated system has been proposed by different authors within different continuum approaches.^{41,42} Here in particular, we shall focus on the version formulated for PCM methods [42].

The dynamic dipole polarizability tensor α can be expressed through eq. (1.55) when both \hat{A} and \hat{B} operators are the electric dipole moment

operator $\hat{\mu}$:

$$\alpha_{ab} = \langle\langle \mu^a; \mu^b \rangle\rangle_\omega \quad (1.57)$$

Moreover, using the spectra representation of the *response matrix*, we can write

$$\alpha_{ab}(\omega) = \sum_q \left[\frac{\langle 0 | \mu^a | q \rangle \langle q | \mu^b | 0 \rangle}{\omega - \omega_q} - \frac{\langle 0 | \mu^b | q \rangle \langle q | \mu^a | 0 \rangle}{\omega + \omega_q} \right] \quad (1.58)$$

where the summation is over the electronic excited states obtained from the generalized eigenvalue equation

$$\Omega X_q = \omega_q \Delta X_q \quad (1.59)$$

where X_q collects the variational parameters corresponding to the excited state $|q\rangle$.

At the RPA level of theory, the molecular system is described by an HF wavefunction and for a closed shell system; if we denote with i and j orbitals which are occupied in the HF ground state and with m and n virtual orbitals, we obtain

$$\alpha_{ab}(\omega) = - \sum_{im} \sum_{jn} \langle i | \mu^a | m \rangle \Pi_{im,jm} \langle n | \mu^b | j \rangle \quad (1.60)$$

here $\Pi_{im,jm}$ are elements of the response matrix

$$\Pi = (\mathbf{M} - \omega \mathbf{Q})^{-1} \quad (1.61)$$

where matrices \mathbf{M} and \mathbf{Q} have elements

$$M_{im,jm} = F'_{mn} \delta_{ij} - F'_{ij} \delta_{mn} + 2(mi|jn) - (mn|ij) + 2B_{mi,jn}^\omega \quad (1.62)$$

$$Q_{im,jm} = 2(mi|jn) - (mj|in) + 2B_{mi,jn}^\omega \quad (1.63)$$

In eqs. (1.62)-(1.63) we have used a standard notation for the two electron integrals and $B_{mi,jn}^\omega$ denote the interaction with apparent charge distributions (it can be related to the previously defined $\mathbf{X}_{\omega_\sigma}$ matrix).

1.4.3.3. Electronic hyperpolarizabilities of solvated system: urea in water

In this section, we shall summarize a study on dynamical polarizability and hyperpolarizability tensors α , β , and γ of urea in vacuo and in water we have published on the Journal of Molecular Structure (Theochem).⁴³ Urea is a compound that has been thoroughly investigated both experimentally and

theoretically, having the advantage of corresponding to a molecule which contains classical organic atoms and both single and double bonds; an accurate evaluation of its nonlinearities can then be helpful for applications to the design of other interesting classes of organic nonlinear optical materials.

The tensors β and γ constitute the molecular origin of the second- and third-order nonlinear optical phenomena such as electro-optic Pockels effect (EOPE), optical rectification (OR), third harmonic generation (THG), electric field induced second harmonic generation (EFI-SHG), intensity dependent refractive index (IDRI), optical Kerr effect (OKE), electric field induced optical rectification (EFI-OR). To save space we do not indicate the full expressions for β and γ related to the different second and third order processes; but we introduce the notations $\beta(-\omega_\sigma; \omega_1, \omega_2)$ and $\gamma(-\omega_\sigma; \omega_1, \omega_2, \omega_3)$, where the frequency relations to be used for the various non-linear optical processes which can be obtained in the case of both static and oscillating monochromatic fields are reported in Table 1.7.

Table 1.7

ω_1	ω_2	ω_3	ω_σ	process
second order				
0	0		0	static
$+\omega$	$+\omega$		2ω	SHG
0	$+\omega$		ω	EOPE
$+\omega$	$-\omega$		0	OR
third order				
0	0	0	0	static
$+\omega$	$+\omega$	$+\omega$	3ω	THG
0	$+\omega$	$+\omega$	2ω	EFI-SHG
$+\omega$	$+\omega$	$-\omega$	ω	IDRI
0	0	$+\omega$	ω	OKE
0	$+\omega$	$-\omega$	0	EFI-OR

In the following the polarizability and the first and second hyperpolarizabilities for urea calculated at the SCF level in vacuo and in water are reported. Both static and frequency dependent nonlinear properties have been calculated, with the Coupled Perturbed Hartree-Fock (CPHF) and Time Dependent-CPHF procedures that have been described above. The solvent model is the Polarizable Continuum Model (PCM) whereas vibrational averaging of the optical properties along the C-O stretching coordinate has been obtained by the DiNa package⁴⁴ both in vacuo and in solution.

The geometry of the urea molecule has been optimized at the Restricted Hartree-Fock level using the standard polarized triple zeta basis set (TZP). The molecule is assumed to be planar as this is the conformation observed in the crystal (note that the molecule is taken to lie in the xz plane with the z -axis parallel to the C=O bond). The geometry optimization has been performed both in vacuo and in aqueous solution.

The prediction of (hyper)polarizabilities depends upon products of matrix elements of the electron position operator \hat{r} . Consequently, unlike the molecule's energy, which primarily depends upon inverse power of \hat{r} , any study of (hyper)polarizabilities must allow for an adequate description of the more diffuse regions of the molecule's wave function. Basis sets that are usually found to be adequate for a dipole moment require further extension to adequately account for the polarizability tensor, α , and even further extensions for the first hyperpolarizability, β , or for the second hyperpolarizability, γ . There are also other factors to be considered. Namely, all contributions to α have the same sign, so the larger the basis generally the better, and this is usually true for γ as well; but the relative values of the charge density must also be described in β (or $\vec{\mu}$) and that imposes additional restrictions on a balanced description. In fact it has been observed that when a larger basis set is employed the components of β are reduced, in such a way to even destroy any agreement between *ab initio* predictions and experiment. We have selected the basis set DZV+2d2p/1p (Dunning double-zeta valence set to which we add two diffuse d functions and two diffuse p functions on heavy atoms and one diffuse p function on the hydrogen atom) to be used in the following calculations.

Vibrational corrections originating from the coupling between the electronic and the nuclear motions are sometimes important in obtaining predictive accuracy for molecular (hyper)polarizabilities. Such corrections are actually of two kinds: the first one simply involves averaging of the observables over the accessible vibrational states, the second one involves contributions to the (hyper)polarizability from vibronic intermediate states. The latter is described in the next section; here we shall consider the former contribution only. In the case of urea we expect that the stretching of the C=O bond is the most relevant motion for the vibrational averaging of the non linear properties. The averaging has been obtained by a variational procedure using spline basis functions. The averaged results for the static polarizability, $\langle \alpha \rangle$ (10^{-24} esu), and the longitudinal component of the first hyperpolarizability, $\langle \beta_z \rangle$ (10^{-32} esu), both in vacuo and in solution at T=298K are shown in Table 1.8 together with the percentage correction

with respect to the not-averaged value.

Table 1.8 Vibrational averaged static polarizability (10^{-24} esu) and first hyperpolarizability (10^{-32} esu) in vacuo and in water at 298K.

	IN VACUO	IN WATER
$\langle \alpha \rangle_T$	4.70	5.64
$\Delta \alpha^a$	+1.5%	+1.7%
$\langle \beta_z \rangle_T$	26.78	-6.51
$\Delta \beta_z^b$	-8.3%	-33.7%

$$^a \Delta \alpha = [\langle \alpha \rangle_T - \langle \alpha \rangle] / \langle \alpha \rangle;$$

$$^b \Delta \beta_z = [\langle \beta_z \rangle_T - \langle \beta_z \rangle] / \langle \beta_z \rangle.$$

The vibrational averaging has a little effect on the calculated polarizability (around -2%), and a greater influence on the longitudinal component of the hyperpolarizability. It is worth noting that in aqueous solution the vibrational correction to the $\langle \beta_z \rangle$ value is markedly greater than in gas phase. Thus it is advisable to consider both the solvent and the vibrational effects, and their possible coupling, to get reliable values of molecular nonlinear properties.

Without the inclusion of the frequency dependence, the various experimentally distinct nonlinear processes would coincide in the static limit. Moreover, dispersion effects can have a relatively large effect on the observed (hyper)polarizability, and predictive calculations will frequently require more accurate dispersion values than those estimated. This would be particularly true if the exciting frequency (ω) becomes close to a resonance. A knowledge of the frequency dependence is also important in identifying possible resonance enhanced effects that might make a particular molecule suitable for a specific application. As said before, in the presence of a dielectric medium the analysis is further complicated by effects related to the solvent dynamical response to the external oscillating field. One way which allows to take into account the latter aspect is to use a dielectric permittivity varying with the exciting frequency. In the case of water as solvent the dielectric permittivity $\epsilon(\omega)$ may be described with the help of the Debye formula (1.54), with $\epsilon(0) = 78.5$, $\epsilon(\infty) = 1.7756$ and $\tau_D = 0.85 \cdot 10^{-11}$ s.

The dispersion behavior of the single electric properties will be separately analyzed in the following paragraphs.

a) *Polarizability: α .*

The static and frequency-dependent polarizabilities (10^{-24} esu) of urea in vacuo and in water are reported in Table 1.9

Table 1.9 Static and frequency-dependent polarizabilities (10^{-24} esu) in vacuo and in water.

ω (a.u.)	IN VACUO		IN WATER	
	static	$\lambda=1064\text{nm}$ (0.04282)	static	$\lambda=1064\text{nm}$ (0.04282)
α_{xx}	5.10	5.13	6.15	5.51
α_{yy}	3.51	3.53	4.12	3.69
α_{zz}	5.27	5.31	6.36	5.60
$\langle \alpha \rangle^a$	4.63	4.66	5.54	4.93

$$^a \langle \alpha \rangle = \frac{1}{3} \sum_i \alpha_{ii}.$$

These results show that the presence of the solvent leads to an increase of the $\langle \alpha \rangle$ values with respect to the vacuum, and it reverses its frequency dependence: in aqueous solution, unlike in vacuo, the molecule shows a static $\langle \alpha \rangle$ value which is larger than that at $\omega \neq 0$. This effect can be related to the shape of the $\epsilon(\omega)$ function. By applying Debye formula it is easy to see that, at the frequency considered in our calculations, the value of $\epsilon(\omega)$ is practically equal to $\epsilon(\infty)$, which is by far smaller than $\epsilon(0)$. If one considers the ‘sum-over-states’ method for the calculation of polarizabilities with these values in mind, then it is easy to give a qualitative explanation of the behavior indicated above. In fact, when the solvent response function is described by $\epsilon(\infty)$ the actual stabilization of the excited states will be less than in a situation where the same response depends on $\epsilon(0)$, hence smaller it will be the correspondent $\langle \alpha \rangle$ value.

b) *First and second hyperpolarizabilities: β and γ .*

The static and frequency-dependent first hyperpolarizabilities (10^{-32} esu) of urea in vacuo and in water are reported in Table 1.10

The most important feature of these results is given by the change of sign in $\langle \beta_z \rangle$ value going from gas phase to aqueous solution. This result seems to confirm the data obtained in dc SHG experiments, which give a positive value of the product $\mu_z \langle \beta_z \rangle$. Assuming the molecule in the same orientation with respect to z axis (C=O bond along $+z$), the experiments

Table 1.10 Static and frequency-dependent first hyperpolarizabilities (10^{-32} esu) in vacuo and in water.

	IN VACUO				IN WATER			
	static	$\lambda=1064\text{nm}$			static	$\lambda=1064\text{nm}$		
		SHG	EOPE	OR		SHG	EOPE	OR
β_{xxx}	-34.68	-37.18	-35.56	-35.56	-82.12	-62.53	-67.09	-65.75
β_{yyz}	16.42	18.81	17.14	17.14	8.79	6.94	7.88	6.64
β_{zzz}	47.48	50.57	48.47	48.47	68.46	40.35	46.75	49.93
$\langle \beta_z \rangle^a$	29.22	32.20	30.05	30.05	-4.87	-15.24	-12.46	-9.18

$$^a \langle \beta_z \rangle = \beta_{zzz} + \beta_{xxx} + \beta_{yyz}$$

give $\mu_z = -4.56\text{D}$ and $\langle \beta_z \rangle = -32 \cdot 10^{-32}\text{esu}$,⁴⁵ while from our calculation it results $\mu_z = -6.29\text{D}$ and $\langle \beta_z \rangle = -4 \cdot 10^{-32}\text{esu}$. Anyway, for completeness's sake, we have to say that a comparison of the calculated NLO properties with experimental data is difficult; this is particularly true for β , for which the calculated data are very sensitive to many parameters such as the level of quantum mechanical approach and the basis set. Analyses of second hyperpolarizabilities γ and comparison with experimental data are even more complex; here we limit to note that, in the static case, $\langle \gamma \rangle$ value increases of about 16% from vacuum to water, while in frequency dependent processes the increases are always smaller, and, in two cases (SHG and IDRI), there is even a decrease.

1.4.3.4. Nuclear contribution to electric polarizabilities

The global effect of an applied external field on a molecule involves distortions both in the electronic charge distribution and in the nuclear charge distribution; the latter leads to the so-called vibrational contribution to the (hyper)polarizabilities.³⁶ As said above, the analysis of the vibrational components reveals the presence of the distinct components,⁴⁶ the “curvature” related to the effect of the field vibrational motion and including the zero point vibrational correction (ZPV) (see above), and the “nuclear relaxation” (nr) originates from the shift of the equilibrium geometry induced by the field.

The nuclear relaxation is the dominant contribution and can be computed in two ways: by perturbation theory or by finite field approximation. We shall limit ourselves to the perturbation theory.

In the BO approximation, the perturbation theory implies the evaluation of the first terms of the expansion of the free energy function and of the properties with respect to the nuclear coordinates and to the external field. At level of double harmonicity (electric and mechanical) only linear

term are considered in the expansion of the properties and only quadratic terms are considered in the expansion of the vibrational potential. The corresponding expressions of the vibrational contribution to the electric (hyper)polarizabilities become:

$$\alpha_{\alpha\beta}^v = \sum_a^{3N-6} \left(\frac{\partial \tilde{\mu}_\alpha}{\partial Q_a} \right)_0 \left(\frac{\partial \mu_\beta}{\partial Q_a} \right)_0 \Big/ (4\pi^2 \nu_a^2) \quad (1.64)$$

$$\beta_{\alpha\beta\gamma}^v = \sum_a^{3N-6} \left[\left(\frac{\partial \tilde{\mu}_\gamma}{\partial Q_a} \right)_0 \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_a} \right)_0 + \left(\frac{\partial \tilde{\mu}_\beta}{\partial Q_a} \right)_0 \left(\frac{\partial \alpha_{\alpha\gamma}}{\partial Q_a} \right)_0 + \left(\frac{\partial \tilde{\mu}_\alpha}{\partial Q_a} \right)_0 \left(\frac{\partial \alpha_{\beta\gamma}}{\partial Q_a} \right)_0 \right] \Big/ (4\pi^2 \nu_a^2) \quad (1.65)$$

where ν_a is the harmonic frequency obtained by the eigenvalues of the nuclear Hessian computed in presence of solvent effects; its elements are the second derivatives of the free energy $G(\mathbf{R})$ with respect the mass weighted nuclear Cartesian components. Q_a denotes the normal mode associated to ν_a and each partial derivative is evaluated at the equilibrium geometry of the solvated system.³²

1.4.3.5. The vibrational hyperpolarizabilities of push-pull systems in solution

In this section we report a second extract of the study we have published on the Journal of the American Chemical Society³² about solvent effects on electronic and vibrational components of linear and nonlinear optical properties of Donor-Acceptor polyenes. In a previous section we have presented the analysis on geometries, here we report the results obtained for the electronic and vibrational (in the double harmonic approximation) static polarizability and hyperpolarizability for the two series of noncentrosymmetric polyenes: $\text{NH}_2(\text{CH}=\text{CH})_n\text{R}$ ($n=1,2$), with $\text{R}=\text{CHO}$ (series I) and with $\text{R}=\text{NO}_2$ (series II) both in vacuo and in water.

Previous calculations on other conjugated systems in vacuo⁴⁷ have given α^v/α^e ratios ranging from 0.1 to 0.4; a value of about 0.3 is computed for all the four molecules in vacuo. On the contrary, no comparisons with data from literature can be made for solvated systems, as never studied before. Anyway, our results show that the solvent effects induce either

Table 1.11 Static electronic and vibrational (hyper)polarizabilities (a.u.) of $\text{H}_2\text{N}(\text{CH}=\text{CH})_N\text{R}$ (with $\text{R}=\text{CHO}, \text{NO}_2$ and $N=1,2$) in vacuo.

N	CHO		NO_2	
	1	2	1	2
α_{rr}^e	70.57	140.51	71.32	146.73
α_{rr}^v	21.32	39.81	21.09	43.90
α_{rr}^{ev}	91.89	180.32	92.41	190.63
β_{rrr}^e	-398.7	-1437.6	-490.4	-1893.7
β_{rrr}^v	-642.0	-2489.1	-922.6	-3846.7
β_{rrr}^{ev}	-1040.7	-3926.7	-1413.0	-5740.4

Table 1.12 Static electronic and vibrational (hyper)polarizabilities (a.u.) of $\text{H}_2\text{N}(\text{CH}=\text{CH})_N\text{R}$ (with $\text{R}=\text{CHO}, \text{NO}_2$ and $N=1,2$) in water.

N	CHO		NO_2	
	1	2	1	2
α_{rr}^e	103.03	231.23	122.04	319.43
α_{rr}^v	60.67	174.81	123.87	1003.44
α_{rr}^{ev}	163.70	406.04	245.91	1322.87
β_{rrr}^e	-899.9	-4701.3	-1227.2	-1928.4
β_{rrr}^v	-2259.3	-17125.2	-5458.0	-12675.5
β_{rrr}^{ev}	-3159.2	-21826.5	-6685.2	-14603.9

a large increase in the absolute vibrational contribution and, even more evident, a net larger increase of the latter with respect to the electronic component; in the series II the ratio α^v/α^e in solution is greater than 1, and the vibrational component becomes dominant.

A similar analysis can be done on the first hyperpolarizability; also here vacuum results are in good accord with previous calculations on conjugated systems, for example for $\text{NH}_2(\text{CHCH})_2\text{NO}_2$ Kirtman and Champagne find a ratio of 2.20 with a RHF/6-31G calculation⁴⁸ to be compared with our value of 2.03. Once more, solvent effects lead to large increases in the relative importance of the vibrational contribution with respect to the electronic one multiplying the gas phase value of β^v/β^e by factors from 1.5 (in the first term of series I) to 2 (in the second term of series II).

To explain such findings it is worth going back to the equations (1.64)-(1.65) giving the formal relations between (hyper)polarizabilities and normal modes of the molecule and try to find a single mode Q^* (corresponding to frequency ω^*) which plays the most important role in the determination of α^v and β^v .

What results is that for the first molecule in vacuo two main modes (from now on indicated as mode 1 and 2) globally reproduce 40% of total α^v and 62% of β^v ; the remaining is partitioned among many other modes with contributions $\leq 5\%$. In solution, besides the two equivalents of mode 1 and 2 in vacuo which now globally give 55% of α^v and 69% of β^v , there appears another rather important mode of very low frequency (592 cm^{-1}), contributing for 8% of α^v and 13% of β^v .

The two most effective modes, either in vacuo and in solution, contain a main component related to C=C and C-C stretching of the chain, respectively. The changes induced by the solvent on these modes are evident both on frequencies and IR intensities; what we observe is a significative decrease of frequency ($\sim 117 \text{ cm}^{-1}$) and an important increase in IR intensity (by a factor of 2.5) for mode 1, and, for mode 2, a small increase of frequency ($\sim 6 \text{ cm}^{-1}$) accompanied by a very large increase in IR intensity (by a factor 8.5). Let us try to analyze these data in a more detailed way, as they can give useful informations for a more physical description of α^v and β^v . The direct relation between polarizability and spectroscopic data is explicit if we recall that IR intensity is proportional to $(\partial\mu/\partial Q)^2$; the parallel relation for β^v is a little more complex as it also involves Raman intensities, but for our scopes we can limit to consider the still present proportionality between β^v and $\partial\mu/\partial Q$, and then IR intensities.

For the series I, the observed behaviors can be related to the changes in the CC bond lengths as well as to the transfer of electronic charge from the donor to the acceptor group. As in solution we observe a lengthening of the double bond and a parallel shortening of the single CC bond, it is well acceptable that the frequencies will be smaller for mode 1 and larger for mode 2, being the two modes related to C=C and C-C stretching, respectively. On the other hand, as the additional motion associated to the main C=C stretching in mode 1 involves different groups passing from vacuum to solution, namely NH₂ and CHO, the IR intensity, as well as the contributions to α^v and β^v , in solution will be larger as larger is the dipole variation related to distortions of CHO in solution than of NH₂ in vacuo.

For mode 2, in which the groups involved remain almost equivalent in vacuo and in solution, the large amplification of IR intensity (and consequently of the contribution of this mode to (hyper)polarizabilities) passing to the solvated system, is clearly related to effects of electron-transfer towards CHO. In fact, as the main distortion characteristic of this mode, a C-C stretching, involves the C atom of CHO group, the related dipole variations and IR intensity will be largely amplified in solution.

The parallel analysis on the series II is complicated by various elements; first the number of normal modes which contribute in a significative way increases by at least one, and, secondly, the whole description looses the clear correspondence between gas phase and solution results found for the previous series. In this series in fact, the solvent effects on geometry, i.e. BLA, and electronic charge distribution are so large that the relative importance of the various modes in terms of α^v and β^v contributions completely changes passing from the gas phase to solution. From a different point of view, we can say that what observed as a secondary effect in the first series, i.e. structural differences between vacuum and solution modes which increase in the larger term of the series, is here so amplified to become the key aspect. Let us try to better understand this point by limiting our analysis to one mode (that giving the most important contribution of α^v and β^v) and to the smallest molecule only.

In this molecule, the correspondence of the modes for gas phase and solvated system is still present, and both mainly involve a C–N(O₂) stretching; the only important changes to be stressed are on the frequency which decreases of $\sim 240\text{ cm}^{-1}$, and on IR intensity which on the contrary increases of a factor of ~ 20 passing from vacuum to solution. This large solvent effect is again strongly related to changes in C–N(O₂) bond length (for which we obtain a decrease of 0.033 Å in solution) and to the charge transfer from the donor to the acceptor which leads to large dipole derivatives, and hence large IR intensities and α^v and β^v contributions, when distortion of NO₂ group are possible.

1.4.4. Magnetic response properties

Solvent effects on nuclear magnetic properties and on the related spectroscopies are well known. In general, two main different effects on NMR spectra can be distinguished: (a) shifts due to a difference in the bulk magnetic susceptibility χ of the solute and the solvent; (b) shifts arising from intermolecular interactions between solute and solvent molecules. Since the bulk susceptibility effect depends on the shape of the sample and, therefore some form of correction for it is usually applied. Of greater importance is the second component related to solute-solvent interactions; here, in particular, we shall describe some of the QM models developed so far within the PCM solvation methods to describe such effect.

In the past alternative semiclassical models have been proposed; of particular importance is the well-known analysis elaborated by Buckingham⁴⁹ to examine solvent effects on the solute nuclear shielding (see below for its

formal definition). In this scheme the solvent effect on the solute shielding for nucleus λ may be partitioned as follows

$$\Delta\sigma(\lambda) = \sigma_E(\lambda) + \sigma_w(\lambda) + \sigma_a(\lambda) + \sigma_b(\lambda) \quad (1.66)$$

where σ_E is the “polar effect caused by the charge distribution in the neighboring solvent molecules, thereby perturbing its electronic structure and hence the nuclear screening constants”, σ_w is “due to the van der Waals forces between the solute and the solvent”, σ_a “arises from anisotropy in the molecular susceptibility of the solvent molecules”, and σ_b is the “contribution proportional to the bulk magnetic susceptibility of the medium”. In the context of the subject treated in this chapter, the most important component is that due to the ‘polar effect’ σ_E , and in fact below we shall describe how this quantity can be evaluated within QM-PCM solvation methods. We also recall that such term includes also specific shorter-range interactions as Hydrogen bonding. All terms appearing in (1.66) in principle could be analytically included into solvation models; in particular for $\sigma_E(\lambda)$, and $\sigma_w(\lambda)$ continuum approaches should be sufficient while σ_a could maybe require the inclusion of discrete solvent molecules around the target solute. A more complex analysis is that required for the bulk susceptibility effect, σ_b ; a possible approach could be to consider this effect as that already mentioned for other molecular response properties computed in a condensed medium. The external field acting on the molecule is modified by the presence the solvent molecules with an extra local modification which can be related to the shape and the dimension of the volume occupied by the molecule.

1.4.4.1. Nuclear magnetic shielding

The nuclear magnetic shielding for isolated molecule can be described in terms of the influence on the total energy of the molecule of the nuclear magnetic moment and of the applied uniform magnetic field. Translating this analysis to molecular solutes in the presence of solvent interactions, leads to define the components of the shielding tensor σ , as the following second derivatives of the free energy functional:

$$\sigma_{ab}^X = \frac{\partial^2 \mathcal{G}}{\partial B_a \partial \mu_b^X} \quad (1.67)$$

where B_a and μ_b^X ($a, b = x, y, z$) are the Cartesian components of the external magnetic field B , and of the nuclear magnetic moment μ^X (X refers to a given nucleus).

It is well-known that the presence of the magnetic field introduces the problem of the definition of the origin of the corresponding vector potential. However, since σ is a molecular properties, it must be invariant with respect changes of the gauge origin. To obtain this gauge invariance in the ab-initio calculation, two ways can be adopted. One is to employ a sufficiently complete basis set so that the consequences of the choice of the gauge origin on the calculated value of σ are minimal. The second method is to introduce gauge factors into either the atomic orbitals of the basis set or the molecular orbitals of a coupled HF or KS calculation in such a manner that the results are independent on the gauge origin even though the calculation is approximate. Inclusion of gauge factors in the atomic orbitals may be accomplished by using gauge invariant atomic orbitals (GIAO).⁵⁰

$$\chi_{\nu}(B) = \chi_{\nu}(0) \exp \left[-\frac{i}{2c} (B \times R_{\nu}) \cdot r \right] \quad (1.68)$$

where R_{ν} is the position vector of the basis function, and $\chi_{\nu}(0)$ denotes the usual field-independent basis function.

The GIAO method is used in conjunction with analytical derivative theory; in this approach the magnetic field perturbation is treated in an analogous way to the perturbation produced by changes in the nuclear coordinates. In this framework, the components of the nuclear magnetic shielding tensor are obtained as:

$$\sigma_{ab} = \text{tr}[\mathbf{P} \mathbf{h}^{B_a \mu_b^X} + \mathbf{P}^{B_a} \mathbf{h}^{\mu_b^X}] \quad (1.69)$$

where \mathbf{P}^{B_a} is the derivative of the density matrix with respect to the magnetic field. Matrices $\mathbf{h}^{\mu_b^X}$ and $\mathbf{h}^{B_a \mu_b^X}$ contain the first derivative of the standard one-electron Hamiltonian with respect to the nuclear magnetic moment and the second derivative with respect the magnetic field and the nuclear magnetic moment, respectively. Both terms do no contain explicit solvent-induced contributions as the latter do not depend on the nuclear magnetic moment of the solute and thus the corresponding derivatives are zero.

On the contrary, explicit solvent effects act on the first derivative of the density matrix \mathbf{P}^{B_a} which can be obtained as solution of the corresponding first-order HF or KS equation characterized by the following derivative of the Fock matrix.⁵¹

$$\mathbf{F}'^{B_a} = \mathbf{h}^{B_a} + \mathbf{h}_R^{B_a} + \mathbf{G}^{B_a}(\mathbf{P}) + \mathbf{X}_R^{B_a}(\mathbf{P}) \quad (1.70)$$

where \mathbf{P} is the unperturbed density matrix and the solvent-induced terms,

$\mathbf{h}_R^{B_a} + \mathbf{X}^{B_a}(\mathbf{P})$, appear due to the magnetic field dependence of the atomic orbitals.

1.4.4.2. Magnetizability

The response function magnetizability χ of a molecular solute can be expressed as second derivatives of the free energy functional:

$$\chi_{ab} = -\frac{\partial^2 \mathcal{G}}{\partial B_a \partial B_b} \quad (1.71)$$

where B_a and B_b ($a, b = x, y, z$) are the Cartesian components of the external magnetic field B .

At HF level, and using a basis set depending on the magnetic field B , we have the following expression for the elements of the magnetizability:

$$\begin{aligned} \chi_{ab} = & \text{tr} \mathbf{P} \mathbf{h}^{B_a B_b} + \frac{1}{2} \text{tr} \mathbf{P} \mathbf{G}^{B_a B_b}(\mathbf{P}) + \text{tr} \mathbf{P} \mathbf{h}_R^{B_a B_b} + \frac{1}{2} \text{tr} \mathbf{P} \mathbf{X}_R^{B_a B_b}(\mathbf{P}) \\ & - \text{tr} \mathbf{S}^{B_a B_b} \mathbf{W} + V_{NN}^{B_a B_b} + \text{tr} \mathbf{P}^{B_a} \mathbf{h}^{B_b} + \text{tr} \mathbf{P}^{B_a} \mathbf{G}^{B_b}(\mathbf{P}) + \quad (1.72) \\ & + \text{tr} \mathbf{P}^{B_a} \mathbf{h}_R^{B_b} + \text{tr} \mathbf{P}^{B_a} \mathbf{X}_R^{B_b}(\mathbf{P}) - \text{tr} \mathbf{S}^{B_a} \mathbf{W}^{B_b} \end{aligned}$$

Once again, the derivative of the density matrix \mathbf{P}^{B_a} can be obtained as solution of first-order coupled-perturbed Hartree-Fock equation with derivative Fock matrix given by eq. (1.70), exactly as for the nuclear shielding.

1.4.4.3. N nuclear shielding of diazines in solution

Solute-solvent interactions may have an important effect on the nuclear shielding of nitrogen atoms, and this effect can be further amplified when nitrogen lone pair is involved in sp^2 type hybridization and is available for hydrogen bonding with protic solvents; such behavior has been also observed in diazines, also known as pyridazine (1,2 diazine), pyrimidine (1,3 diazine), and pyrazine (1,4 diazine).⁵² In particular, for diazines, the solvent induced shielding changes appear to be strongly dependent upon the relative positions of the two nitrogen atoms: an extreme situation is exhibited by 1,2 diazine where the water-to-cyclohexane shift amounts to about 41 ppm; this represents one of largest nitrogen shielding change so far found to be dependent upon a variation in solvent, with the exception of the case of a complete proton transfer occurring between solvent and solute.

Here we summarize the results of a research we have published on Journal of American Chemical Society⁵³ on the computation of solvent effects on N nuclear shielding of diazines within the PCM-IEF framework.

Geometry optimizations for all the systems both in vacuo and in the various solvents were performed at the density functional theory (DFT) using the hybrid B3LYP functional. The basis set used was the 6-31+G(d,p). Calculations of nuclear shieldings were performed at B3LYP level exploiting the gauge-including atomic orbital (GIAO) method and a 6-311+G(d,p) basis set.

In Table 1.13 we report the computed (GIAO B3LYP/6-311+G(d,p)) shift of nitrogen shielding for the three diazines in the three solutions (acetone, DMSO, and water) with respect to cyclohexane solution within the framework of the PCM-IEF continuum model. In the same Table we also reported the experimental data from Ref.[54].

Table 1.13 Calculated and experimental cyclohexane-to-solution shifts (in ppm) of diazines in vacuo and in various solvents.

	pyridazine		pyrimidine		pyrazine	
	calc	exp	calc	exp	calc	exp
Acetone	11.26	9.46	2.26	3.11	1.94	2.71
DMSO	12.29	14.38	2.61	3.60	2.26	3.17
water	12.71	41.55	3.60	16.84	2.33	16.85

If we compare the results obtained in the three solvents it is evident that for acetone and DMSO, where no hydrogens are available to interact with diazine nitrogens, the IEF continuum model gives very good results for all the three molecules. On the contrary, for water the agreement with experimental data significantly worsens; the computed shift reproduces only a part (going from 1/8 for pyrazine to 1/3 for pyridazine) of the whole observed effect. The missing term is clearly due to H-bonding which, according to the analysis reported in the previous section induces an additional contribution to the global solvent effect besides the more general polar term.

To evaluate the relative contributions of H-bonding and ‘polar’ effects, we have optimized two H-bonded clusters containing 1 and 2 water molecules, respectively, without and with an external continuum dielectric. On such optimized geometries we have computed nuclear shieldings at the same level (B3LYP/6-311+G(d,p): GIAO) used before for the single diazines. In Table 1.14 we report the calculated shieldings of the clusters without (VAC) and with the external continuum (IEF) with respect to cy-

Table 1.14 Computed shifts $\delta\sigma$ (in ppm) of water-diazine clusters in vacuo (VAC) and in the presence of an external continuum dielectric (IEF). For the 2-water clusters of pyridazine two values referring to the the two non equivalent nitrogens are reported.

	VAC		IEF		exp
	+1w	+2w	+1w	+2w	
pyridazine	25.44	35.30 (17.16)	41.92	49.43 (41.91)	41.55
pyrimidine	13.62	12.70	19.36	15.93	16.84
pyrazine	18.46	12.97	24.57	18.55	16.85

clohexane. For the clusters with just one water molecule the shielding we report refers to the H-bonded nitrogen as the other nitrogen behaves as in the monomers. For the 2-water clusters of pyridazine two values referring to the two non equivalent nitrogens are reported. Experimental data are from Ref.[52]

The results presented in Table 1.14 shows that including only specific H-bonding effects (i.e. considering gas-phase clusters) means to take into account only a part of the total solvent effect while the complete description is reproduced by adding long-range effects (here represented by the external continuum).

Before discussing in further details such finding, it is interesting to note a specificity of this property: considering only one H-bond leads to a larger effect than by assuming a completely saturated 2w-cluster. Only in pyridazine-clusters, the behavior is opposite indicating that for this molecule the effects of the two H-bonds are additive: this diversity of pyridazine with respect to the other diazines needs a further analysis. Pyridazine is characterized by the aggregation of two nitrogen atoms in adjacent ring positions: this vicinity affects the structure of the 2-water cluster which presents an asymmetry between the two H-bonds and thus makes the two nitrogens not completely equivalent. The non-equivalence of the two nitrogens is reflected in the corresponding shieldings which differ of about 13 ppm in vacuo and 6 ppm in solution (see the two values reported in the Table above for each 2w-cluster, the first one referring the nitrogen with the closest hydrogen). Such aspect has to be taken into account in the comparison with the experimental cyclohexane-to water shift: pyridazine is in fact the system for which we observe the largest computed-observed discrepancy and the only one for which the 1w-cluster gives a better result than 2w-cluster. It is worth noting that such trend is changed by consid-

ering the other nitrogen in the 2w cluster (the corresponding values are reported between parentheses in the table above): in this case the shift with respect to cyclohexane reduces to 41.91 ppm while the experimental value is 41.55 ppm.

The results reported in the tables can be collectively analyzed in terms of a final correlation plot between calculated and experimental data, where the calculated data are those obtained using a larger cavity and (in the case of water) the IEF/2w clusters. The plot is reported below together with the line resulting from a linear fitting

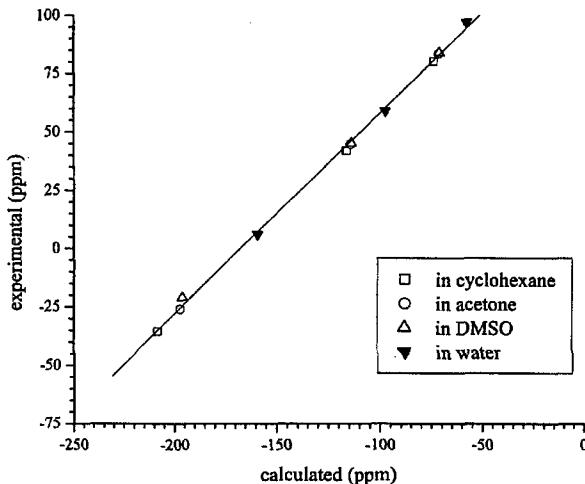


Fig. 1.2 Experimental and calculated isotropic N nuclear shielding plot of diazines in various solutions. Experimental data are reported with respect to neat liquid nitromethane.

From the analysis of the plot it is appears evident the improvement achieved introducing a combination of refinements in the continuum solvation model so to take into account both nonelectrostatic effects (through a larger cavity) and, in the case of water, more specific H-bonding interactions (through 2w/clusters). The correlation is now very good (the correlation coefficient is 0.99932) and all the major changes induced by solvent polarity and those exerted by solvent to solute hydrogen bonding are very well represented.

1.4.5. Chiroptical solvated molecules: OR and VCD

Optical activity is one of the oldest research tools that is routinely practiced by chemists. Optical rotatory dispersion (optical rotation, OR, versus wavelength) and circular dichroism (CD) are the two components of this optical activity that can be used to elucidate the absolute stereochemistry of chiral molecules. Vibrational circular dichroism (VCD), and vibrational Raman optical activity (VROA) are alternate properties that are being currently used for the same scope. The main aspect to stress here is that both sets of investigative tools are strictly related to the condensed phase, as experimental measurements are generally limited to this. Despite this, the extension of solvation models to general QM methods for predicting chiro-optical properties has been very limited. Only in this last few years some efforts (many unpublished as still in their phase of testing) have been appeared; below we shall try to give a preliminary summary of these new applications, once again with exclusive attention to PCM solvation schemes.

1.4.5.1. Optical rotation

The ab initio theoretical quantity needed to predict molecular optical rotations is the electric dipole-magnetic dipole polarizability tensor, indexelectric dipole-magnetic dipole polarizability tensor given by the expression:

$$G'_{ab} = \frac{-4\pi c}{3h} \sum_{n \neq s} \frac{\omega}{\omega_{ns}^2 - \omega^2} \text{Im} \left\{ \langle \Psi_s | \mu_a^{el} | \Psi_n \rangle \langle \Psi_n | \mu_b^{mag} | \Psi_s \rangle \right\} \quad (1.73)$$

where μ_a^{el} and μ_b^{mag} are respectively, the electric and magnetic dipole operators, and Ψ_s and Ψ_n represent the ground and excited electronic states.. The angular frequency ω_{ns} and ω are those related to the $s \rightarrow n$ transition and the exciting radiation, respectively. In addition it is usual to define the quantity

$$\beta = -\omega^{-1} (G'_{xx} + G'_{yy} + G'_{zz}) / 3 \quad (1.74)$$

to which the optical rotation, and the related specific rotation are directly (through numerical factors) proportional.

An explicit evaluation of the sum over excited states in eq.(1.73) can be avoided by noting that when $\omega_{ns} \gg \omega$ eq.(1.73) can be simplified as⁵⁵

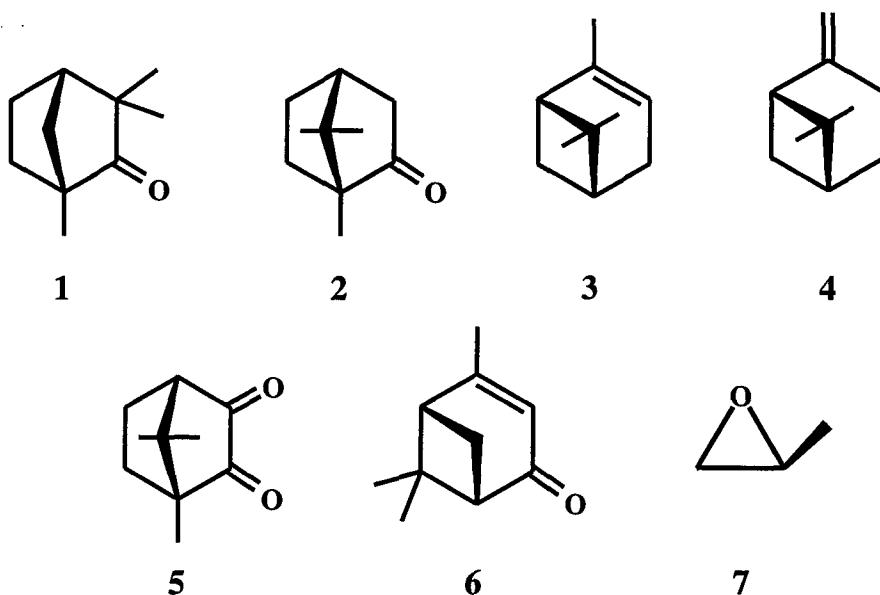
$$\omega^{-1} G'_{ab} = -\frac{hc}{3\pi} \text{Im} \left\{ \left\langle \frac{\partial \Psi_s}{\partial E_a} \middle| \frac{\partial \Psi_s}{\partial B_b} \right\rangle \right\} \quad (1.75)$$

where E_a and B_b are the Cartesian components of the applied electric and magnetic fields, respectively.

The alternative expression (1.75) permits to compute the required β by applying standard coupled HF or KS methods. In this framework it also becomes almost immediate to include solvent effects; in particular for PCM methods the corresponding CPHF or CPKS equations to solve are exactly those involved in the calculation of electric polarizabilities and nuclear magnetic shieldings (see previous sections).

1.4.5.2. OR of benchmark solvated systems

Here, we present a summary of the results of a study of solvent effects on optical rotations we have recently published on Journal of Physical Chemistry A.⁵⁶ Optical rotations are calculated using DFT, and solvent effects are incorporated using the IEF version of the PCM. The accuracy of the DFT/PCM theory is evaluated by comparison of its predictions for seven, conformationally rigid, chiral organic molecules in seven solvents (C_6H_{12} , CCl_4 , C_6H_6 , $CHCl_3$, $(CH_3)_2CO$, CH_3OH and CH_3CN) to experiment. The chiral molecules chosen are fenchone (1), camphor (2), α -pinene (3), β -pinene (4), camphorquinone (5), verbenone (6) and methyloxirane (7):



DFT calculations of sodium D line specific rotation, $[\alpha]_D$, were carried out using the B3LYP functional and the aug-cc-pVDZ basis set at B3LYP/6-31G* equilibrium geometries. The solvents were described in terms of their static, $\epsilon(0)$, and optical, ϵ_{opt} , dielectric constants: 2.028 ($\epsilon(0)=\epsilon_{opt}$) for C₆H₁₂, 2.228 and 2.129 for CCl₄, 2.247 and 2.244 for C₆H₆, 4.90 and 2.085 for CHCl₃, 20.7 and 1.841 for (CH₃)₂CO, 32.63 and 1.758 for CH₃OH, and 36.64 and 1.806 for CH₃CN. In the calculation of optical rotation, a solute-solvent nonequilibrium model is assumed (see before for details).

The most striking feature of experimental results is that the variation of $[\alpha]_D$ with solvent is highly molecule-dependent. No two molecules exhibit the same ordering of $[\alpha]_D$ values with respect to solvent variation. This is particularly surprising in the case of very similar molecules. Thus, one might expect the $[\alpha]_D$ values of fenchone, 1, and camphor, 2, two very similar molecules, to exhibit similar solvent dependence. Instead, the ordering of $[\alpha]_D$ for 2: CH₃OH < C₆H₆ < CHCl₃ < CH₃CN < CCl₄ < (CH₃)₂CO < C₆H₁₂ is, with the exception of CHCl₃, exactly opposite to that for 1: C₆H₁₂ < (CH₃)₂CO < CCl₄ < CH₃CN < C₆H₆ < CH₃OH < CHCl₃.

Passing to calculated IEF-PCM results, we note, first, that for each molecule predicted $[\alpha]_D$ values are very similar for the three non-polar, low-dielectric-constant solvents C₆H₁₂, CCl₄ and C₆H₆. Likewise, $[\alpha]_D$ values are also quite similar for the three polar, high-dielectric-constant solvents (CH₃)₂CO, CH₃OH and CH₃CN. $[\alpha]_D$ for the solvent of intermediate polarity and dielectric constant, CHCl₃, is intermediate. Thus, calculated $[\alpha]_D$ values vary essentially monotonically with solvent dielectric constant. As observed previously for PCM calculations of other properties, the variation is non-linear and exhibits "saturation" with increasing dielectric constant. At the same time, the IEF-PCM calculations predict changes in $[\alpha]_D$ with solvent which vary widely with molecule, both in magnitude and sign.

Comparison of experimental and calculated variations in $[\alpha]_D$ with solvents shows reasonably good correlation between theory and experiment for the polar solvents plus C₆H₁₂: the calculated and experimental changes in $[\alpha]_D$ from C₆H₁₂ to (CH₃)₂CO, CH₃OH and CH₃CN are compared in figure 1.3

Quantitative agreement is also quite good: the average of the absolute magnitudes of the differences between calculated and experimental changes in $[\alpha]_D$ from C₆H₁₂ to (CH₃)₂CO, CH₃OH and CH₃CN, is 3. On the contrary, for the changes from C₆H₁₂ to the other three solvents, namely CCl₄, C₆H₆ and CHCl₃, the correlation between theory and experiment is

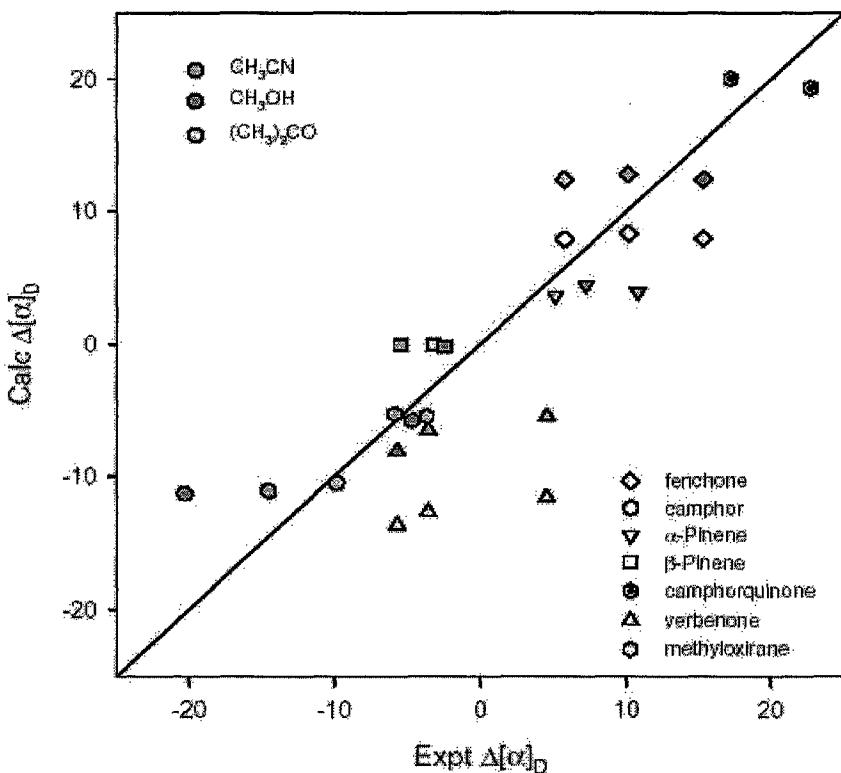


Fig. 1.3 Comparison of calculated and experimental variations in $\Delta[\alpha]_D$ values for 1-7 in $(\text{CH}_3)_2\text{CO}$, CH_3OH , and CH_3CN . $\Delta[\alpha]_D$ is the difference in $[\alpha]_D$ from that in C_6H_{12} . The line is of slope +1.

very poor. Experimental changes are on average much larger than predicted changes: the average of the absolute magnitudes of the differences between calculated and experimental changes in $[\alpha]_D$ is 8. The finding that the dynamic/non-equilibrium DFT/PCM calculations for C_6H_{12} , $(\text{CH}_3)_2\text{CO}$, CH_3OH and CH_3CN account quite well for the observed changes in $[\alpha]_D$ in these solvents suggests that the latter are primarily electrostatic in origin. Conversely, the finding that calculations for CCl_4 , C_6H_6 and CHCl_3 account poorly for the observed changes in $[\alpha]_D$ in these solvents suggests that the latter are not primarily electrostatic in origin.

To check the validity of the dynamic/non-equilibrium DFT/PCM calculations we have also carried out calculations referred to as static/equilibrium calculations, in which α is calculated in the static limit and the equilibrium solvent model is used. In molecules 1, 4 and 7, the variation in $[\alpha]_D$ with

solvent predicted by the static/equilibrium calculations is very similar to that predicted by the dynamic/non-equilibrium calculations. However, for the molecules 2, 3, 5 and 6 the results are quite different for the polar high-dielectric-constant solvents, and in much worse agreement with experiment. Overall, it is clear that the static/equilibrium calculations of solvent variations in $[\alpha]_D$ are much less accurate than the dynamic/non-equilibrium calculations.

1.4.5.3. Vibrational Circular Dichroism spectra

The intensities of vibrational circular dichroism (VCD) spectra exhibited by chiral molecules are obtained by computing the different response of the molecules to left and right circularly polarized light:

$$\Delta\epsilon(\bar{\nu}) = k\bar{\nu} \sum_i R_i f_i(\bar{\nu}, \bar{\nu}_i) \quad (1.76)$$

where k is a numerical factor and $\bar{\nu}_i$ and R_i are the excitation frequencies and the rotational strength of the fundamental transition ($0 \rightarrow 1$) of the i th normal mode; $f_i(\bar{\nu}, \bar{\nu}_i)$ is a normalized Lorentzian band shape.

The vibrational rotation strength R_i is given by

$$R_i = \text{Im} [\langle 0 | \mu_{el} | 1 \rangle_i \cdot \langle 1 | \mu_{mag} | 0 \rangle_i] \quad (1.77)$$

where $\langle 0 | \mu_{el} | 1 \rangle$ and $\langle 0 | \mu_{mag} | 1 \rangle$ are the electric and magnetic vibrational transition dipole, respectively. These two quantities are obtained within the harmonic approximation by computing, respectively, the atomic polar tensor (APTs) and the atomic axial tensors (AATs).⁵⁷

The calculation of $\Delta\epsilon$ in the presence of a solvent medium still relies on eqs.(1.76,1.77), but some refinements are needed. As already remarked in Ref.[34] with regard to infrared intensities for molecules in solution, the μ_{el} operator in eq.(1.77) has to be replaced by the sum of the dipole moment μ_{el} of the molecule and the dipole moment $\tilde{\mu}_{el}$ arising from the polarization induced by the molecule on the solvent. As already pointed out, $\tilde{\mu}_{el}$ takes into account effects due to the field generated from the solvent response to the probing field once the cavity has been created (the so-called "cavity field"). In principle also μ_{mag} should be similarly reformulated. However, by assuming the response of the solvent to magnetic perturbations to be described only in terms of its magnetic permittivity (which is usually close to unity), it is reasonable to consider that the magnetic analogous of the electric "cavity field" gives minor contributions to R_i^{sol} . For this reason we

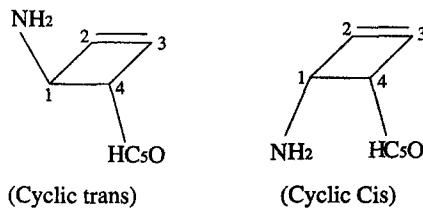
will postpone the study of such effect to future communications. Thus, the final expression for R_i^{sol} to use is:

$$R_i^{sol} = \text{Im} [\langle 0 | (\mu_{el} + \tilde{\mu}_{el}) | 1 \rangle_i \cdot \langle 0 | \mu_{mag} | 1 \rangle_i] \quad (1.78)$$

Both APTs and AATs depend on the wave function derivatives with respect to nuclear displacements; the computation of these terms for a molecule in solution has been already treated above. AATs depend also on the derivative of the wave function with respect to an external magnetic field. This derivatives is already known as it enters in the general definition of the nuclear shielding tensors discussed in the previous section. Once again, for its evaluation its possible to exploit the GIAO method.

1.4.5.4. VCD spectra of isomers in solution

We report here the analysis of vibrational circular dichroism VCD spectra for the cis and trans cyclic isomers of $\text{H}_2\text{NC}_4\text{HCHO}$ in gas-phase and in water. This report represents a part of the material elaborated for a talk at the 39th Sanibel Symposium and successively published on the International Journal of Quantum Chemistry.⁵⁸



All the calculations have been performed at the self-consistent field SCF level with a basis set equal to the Dunning-Huzinaga valence double zeta. The solvation calculations are performed within the IEF-PCM framework for a medium having dielectric constant equal to 78.5 corresponding to the static dielectric constants of liquid water at 298 K.

In the two following Figures we report the spectra computed both in vacuo and in water for the trans (Fig. 1.4) and cis (Fig. 1.5) isomers, within the frequency range of 600-1500 cm⁻¹.

The spectra are the combination of the signals of the two chiral carbons (C1 and C4) each with its specific sign.

For both systems, the spectra show significant solvent effects on the

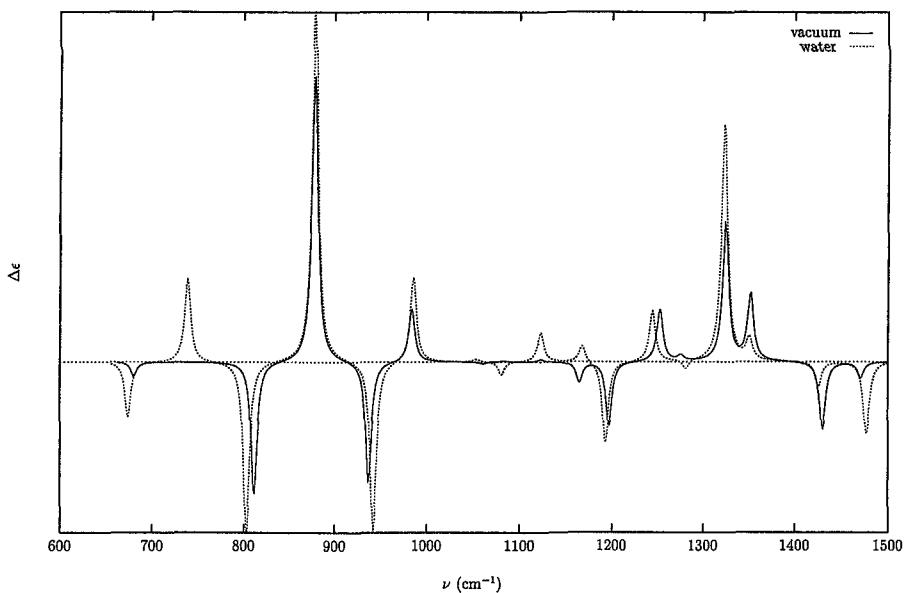


Fig. 1.4 VCD spectra of trans isomer in vacuo and in water.

positions of the peaks, i.e. the vibrational frequencies and, even more strongly, on their intensities; in addition, some additional peaks appear in solution and others change sign with respect to gas-phase. The most evident solvent effects can be observed in the spectrum of the cis isomer, where an intense peak not present in the spectrum of the isolated system appears at 650 cm^{-1} ; looking at the corresponding normal mode, we see that this peak is due to a bending of the hydrogens in the NH group.

A similar but smaller effect can be observed also in the spectrum of the solvated trans isomer in which, at 738 cm^{-1} , we found a peak that is not present in the parallel spectrum computed for the isolated molecule. Actually, a similar signal exists also in vacuo but in this case its frequency falls much below the window considered here. In particular we can observe the parallel of the cis signal at 328 cm^{-1} and that of the trans at 523 cm^{-1} ; by repeating the calculation without the solvent but still exploiting the geometries optimized in solution, these peaks are shifted again within the analyzed range of frequencies showing that this specific effect is entirely due to the modifications induced by the solvent in the nuclear geometry.

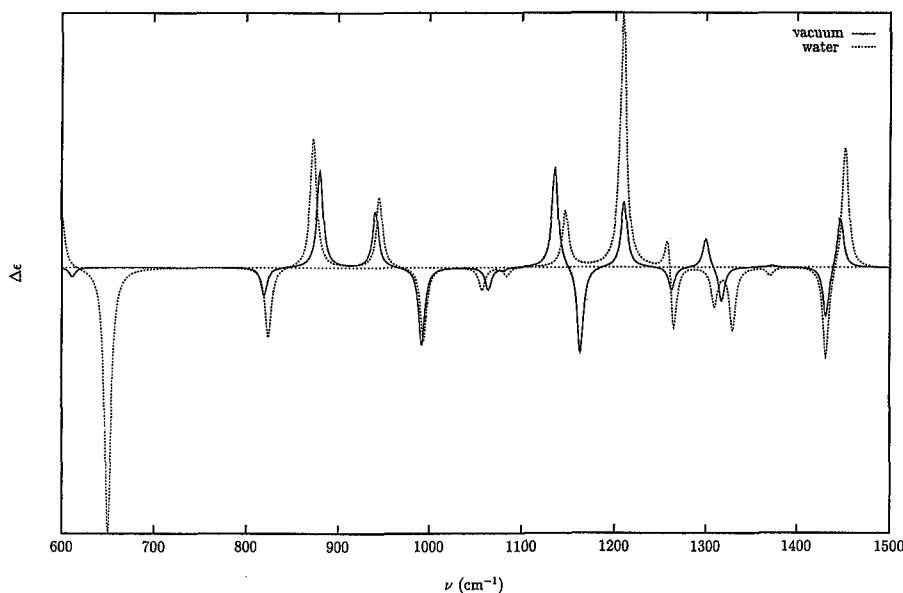


Fig. 1.5 VCD spectra of cis isomer in vacuo and in water.

Other important changes between spectra of solvated and isolated systems are found for the cis isomer in the range 1100–1300 cm^{-1} , where, for example, the large peak at 1162 cm^{-1} completely disappears in water while the next one at about 1200 cm^{-1} is largely amplified.

1.5. Excited States

The extension of the solvation models introduced above for molecules in their ground state, to electronically excited states is neither immediate nor straightforward. There are in fact many fundamental characteristics which are specific of the molecules in their excited states and which make them unique systems to be treated with completely new techniques.

This specificity of the electronically excited molecules when the molecules can interact with an external medium, introduces the concept of time progress, concept which can be safely neglected in treating solutes

in their ground states. In these cases in fact, and also when introducing reaction processes, one can always reduce the analysis to a completely equilibrated solute-solvent system. On the contrary, when the attention is shifted towards dynamical phenomena as those involved in electronically transitions (absorptions and/or emissions), or towards relaxation phenomena as those which describe the time evolution of the excited state, one has to introduce new models, in which solute and solvent have proper response times which have not to be coherent or at least not before very long times.

It is convenient to introduce a partition of the sources of the dynamical behavior of the medium into two main components.⁵⁹ One is represented by the internal molecular motions inside the solvent due to changes in the charge distribution, and in the geometry, of the solute system. As already shown, the solute when immersed in the solvent produces an electric field inside the bulk of the medium which can modify its structure, for example inducing phenomena of alignment and/or preferential orientation of the solvent molecules around the cavity embedding the solute. These molecular motions are characterized by specific time scales of the order of the rotational and translational times proper of the condensed phases (10^{-9} - 10^{-11} s). In a analogous way, we can assume that the single solvent molecules are subjected to internal geometrical variations, i.e. vibrations, due to the changes in the solute field; once again these will be described by specific shorted time scales (10^{-14} - 10^{-12} s). Both the translational, the rotational and/or the vibrational motions involve nuclear displacements and therefore, in the following, they will be collectively indicated as 'nuclear motions'. The other important component of the dynamical nature of the medium, complementary to the nuclear one, is that induced by motions of the electrons inside each solvent molecule; these motions are extremely fast (of the order of 10^{-16} - 10^{-15} s) and they represent the electronic polarization of the solvent.

These 'nuclear' and electronic components, due to their different dynamic behavior, will give rise to different effects. In particular, the electronic motions can be considered as instantaneous and thus the part of the solvent response they originate is always equilibrated to any change, even if fast, in the charge distribution of the solute. On the contrary, solvent nuclear motions, by far slower, can be delayed with respect to fast changes, and thus they can give origin to solute-solvent systems not completely equilibrated in the time interval interested in the phenomenon under study. This condition of nonequilibrium will successively evolve towards a more stable and completely equilibrated state in a time interval which will

depend on the specific system under scrutiny.

In we limit to the initial step of the whole process, i.e. the vertical electronic transition (absorption and emission), we can safely assume a Franck-Condon like response of the solvent, exactly as for the solute molecule; the nuclear motions inside and among the solvent molecules will not be able to immediately follow the fast changes in the solute electronic charge distribution and thus the corresponding part of the response (also indicated as inertial) will remain frozen in the state immediately previous to the transition.

In the PCM framework, this will lead to a partition of the apparent charges in two separate sets,^{20,60} one related to the instantaneous electronic response, σ_{fast} , and the other to the slower response connected to the nuclear motions, σ_{slow} (their sum is just the already defined total apparent charge σ). According to what said before about the origin of the apparent charges, the electronic components will depend on the instantaneous wavefunction of the solute, on the optical dielectric constant and on the complementary set of slower apparent charges. The latter, on the contrary, will still depend on the solute wavefunction of the initial state.

From a physical point of view, we can always define the equations giving the total and the electronic surface charge by considering two parallel systems of the type (1.6) in which both the operator and the related boundary conditions are defined in terms of ϵ or of ϵ_∞ , respectively. On the contrary, the orientational surface charge has to be defined as difference of the two other charges, being the orientational contribution to the solvent polarization not related to a physically stated dielectric constant.

By applying this model to the problem of a solute electronic transition from the ground state to a given excited state, we have:

$$\left\{ \begin{array}{ll} -\Delta V = 4\pi\rho_M^{fin} & \text{within the cavity} \\ -\epsilon_\infty\Delta V = 0 & \text{out of the cavity} \\ V_i - V_e = 0 & \text{on the cavity surface} \\ (\partial V/\partial n)_i - \epsilon_\infty(\partial V/\partial n)_e = 4\pi\sigma_{slow} & \text{on the cavity surface} \end{array} \right. \quad (1.79)$$

where ρ_M^{fin} is the charge distribution of the solute in its final excited state. The second jump condition on the gradient of V has here a different form with respect to that of the equilibrium situation given before (1.6), due to the presence of the constant slow charge σ_{slow} .

As shown before for ground state equilibrated solutes, the system (1.79) can be solved by defining an electronic surface charge σ_{fast} which can be discretized into point charges, q_{fast} , exploiting to the usual partition of

the surface cavity into tesserae. However, in the nonequilibrium case the definition of the electrostatic free energy has to be changed; in particular the contribution of the slow polarization has to be seen an external fixed field which does not follow the variationally derived rules; the proper equation becomes:

$$G^{neq} = G_\infty + U_{e,slow}^{neq} + U_{N,slow}^{neq} + \frac{1}{2} U_{fast,slow}^{neq} \quad (1.80)$$

where G_∞ is the equivalent of eq.(1.20) with solvent induced matrices defined in terms of the fast charges only, i.e.:

$$G_\infty = \text{tr} \mathbf{P} [\mathbf{h} + \mathbf{h}_{fast}^{el}] + \frac{1}{2} \text{tr} \mathbf{P} [\mathbf{G}(\mathbf{P}) + \mathbf{X}_{fast}^{el}(\mathbf{P})] + \frac{1}{2} U_{N,N_{fast}} + V_{NN} \quad (1.81)$$

The last three quantities in the right hand side of eq.(1.80) represent the interaction energies of the constant slow apparent charges with solute electrons, solute nuclei and fast apparent charges, respectively.

Besides vertical processes as those described above, continuum models can be generalized to treat also the more complex aspect of the relaxation of the excited state following the vertical excitation, or inversely that of the ground state after emission. As said before, this is a more general dynamical process in which solute and solvent dynamical behaviors mutually interact. In this case, it is thus necessary to know, or to model, the time dependence of the solvent permittivity so to get a set of solvent apparent charges continuously depending on time. Unfortunately, the time dependence of the permittivity is of complex nature, and specific of each solvent; in general it cannot be represented through an analytic function: what we can do is to derive semiempirical formula either by applying theoretical models based on measurements of relaxation times (as that formulated by Debye) or by knowing through experiments the behavior of the permittivity with respect to the frequency of an external applied field. These formula, in turn, will allow to obtain induced charges which take into account, through its dependence on the time-dependent dielectric constant, the dynamic behavior of the specific solvent they represent.

The extension of continuum models to relaxation phenomena, among which that related to excited states is just an example, is still in its initial phase; however some interesting proposals already appeared, and surely quantitative numerical results will not delay arriving.

1.5.1. Excited states in solution: the DMABN case

The interest in the spectroscopic properties of 4-(N,N-dimethylamino)benzonitrile (DMABN) has attracted increasing attention since the first experimental observations by Lippert et al.⁶¹ The main finding of such experiments was an abnormal emission (1L_a -type according to Platt's notation) in polar solvents, red-shifted from the normal 1L_b -type fluorescence. This phenomenon, since then known as dual fluorescence, has successively been observed in several other electron donor-acceptor compounds but DMABN still remains one of the most studied case. Several models for the explanation of the unusual behavior of DMABN, and related compounds, have been proposed. Among all the proposals that supported by the largest experimental evidence is the twisted intramolecular charge transfer (TICT) model by Grabowski and coworkers.⁶² The TICT model explains the dual fluorescence as occurring by means of a radiationless transition leading to two conformations differing in the dihedral angle between the amino group and the phenyl ring. According to the TICT model the initially promoted excited state yields another minimum on the energy surface by twisting the dialkylamino group from a planar (or nearly planar) to a perpendicular position with respect to the benzonitrile ring. The twisting is accompanied by an intramolecular charge transfer (CT) from the donor (the amine) to the acceptor moiety (the benzonitrile group) in the excited state. The latter (also called CT state), being characterized by a large intramolecular charge separation and an increased dipole, is stabilized by polar solvents thus leading to the observed 'anomalous' 1L_a fluorescence band. The 'standard' 1L_b , on the contrary, is assigned to the less polar locally excited (LE) state, where no twisting has occurred.

Numerous theoretical studies on DMABN have been carried out, and many of them confirm the greater validity of the TICT model. The main body of such calculations, however, has been limited to the isolated system, while few examples including solvent effects can be quoted.^{63–65} On the contrary, the phenomenon is strongly related to solvation and thus explicit considerations of solvent interactions are very important to get a more accurate understanding of the experimental evidence on the specific effects due to the presence of polar solvents. Here we summarize the results of the correlated study of DMABN both in vacuo and in solution we have published on the Journal of American Chemical Society.⁶⁶ In this study we have used the multireference perturbation configuration interaction (CI) method, known with the CIPSI acronym,⁶⁷ which has been coupled to the PCM-IEF solvation continuum model.⁶⁸

Table 1.15 Absorption energies (eV) for planar DMABN in vacuo and in solution ('acn' indicates acetonitrile as solvent while 'cyc' refers to cyclohexane). Experimental data refer to absorption maxima.

	calculated		exp	
	1^1B	2^1A	LE	CT
vacuum	4.29	4.78	4.25 ^a	4.56 ^a
acn	4.00	4.10	n.o.	4.2 ^b
cyc	4.00	4.32	4.1 ^b	4.4 ^b

^a Ref. [70]. ^b Ref. [71].

Ground state (GS) geometry optimizations both in vacuo and in solution have been obtained at DFT level employing the hybrid (B3LYP) functional, while gas-phase and solvated excited states have been optimized at CIS level;⁶⁹ in all cases a 6-31G(d) basis set has been used. In the CIPSI calculations we excluded the *d* basis functions for both carbons of the amino group in order to obtain a better manageable basis set (we recall that we do not freeze any virtual orbital, i.e. every molecular orbital should be correlated). We also recall that in the current PCM-IEF version the solvent perturbative operators describe both electrostatic and repulsive effects; in particular, repulsion interactions between solute and solvent we have used the model originally formulated by Amovilli and Mennucci²³ (see section [1.2.2]). In brief, we note that two different solvation models will be used, namely assuming equilibrium or nonequilibrium conditions between solute and solvent charge densities, respectively (see above).

In Table 1.15 we report absorption energies (in eV) computed for DMABN in vacuo and in the two solvents (acetonitrile, acn, and cyclohexane, cyc) through the CIPSI algorithm. The adopted geometries are those optimized for the ground states in gas phase and in each solution. As comparison we also report experimental data.

In vacuo the first excited state has B symmetry and is only a weakly allowed transition (the computed oscillatory strength is 0.082). The most intense transition is into the second excited state with A symmetry (in this case the oscillatory strength is 0.786). From experiments, it is well known that the most intense absorption band corresponding to the excitation into the 2^1A state is red-shifted in the presence of polar solvents. The charge transfer character of this state, associated to a large dipole moment, in fact allows a by far more effective solvent-induced stabilization with respect to the less

Table 1.16 Dipole (debye) of ground and excited states of planar DMABN in vacuo and in two solutions ('acn' indicates acetonitrile as solvent while 'cyc' refers to cyclohexane). Ranges of data reported for gas-phase collect many available experimental values obtained in various apolar solvents.

	calculated			exp ^a	
	GS	1 ¹ B	2 ¹ A	GS	LE
vacuum	7.728	9.791	13.610	5-7	8-11
acn	9.819	12.134	15.339		
cyc	8.589	10.821	14.233		~10

^a Ref. [72] and refs. therein.

polar ground state; as a final result the excitation energy is decreased more as more polar the solvent is. This behavior is correctly reproduced by the results reported in the Table. Both in acetonitrile and cyclohexane the excitation energies towards A are smaller than in vacuo and they decrease by increasing the polarity of the solvent. The agreement with the experimental evidence is not limited to a qualitative aspect, but for both transitions and both solvent a very good correspondence with observed results is found; for transition to 1¹B in acetonitrile no experimental data are available due to the complete overlap of the bands.

The analysis we have done on the possible interactions leading to the observed solvent effects on the absorption spectra is supported by values of dipole moments for each state in vacuo and in solution. In Table 1.16 we report the dipole moments of ground and excited states of planar DMABN both in vacuo and in the two solvents with indication of the experimental data, when available.

Upon excitation the dipole moments of all the states increase, as it can be predicted by considering the previous analysis on the nature of the orbitals involved in the two excitations. In fact the π^* orbital, which characterizes both the excited states, is more localized towards the cyano group than the π orbital (i.e. the initial orbital as concerns excitation to B), and than the mixed $\pi-n$ orbital (involved in the transition to A). Such increase in the dipole is amplified in the presence of a polar solvent which stabilizes most the states with larger dipolar character.

After the analysis on the vertical process of absorption and on the effects of apolar and polar solvents, we can now focus on the ground and low-lying singlet excited states potential curves (energy relative to the minimum of

the ground state) of DMABN in vacuo and in acetonitrile as a function of the twisting angle.

To obtain these curves the geometry has been kept frozen to that optimized for the ground state (i.e. with a wagging angle equal to zero) at any point and for any electronic state; the only changing parameter is the twisting angle θ which has been varied from 0° to 90° in steps of 15° . All the points have been calculated at the CIPSI level both in vacuo and in acetonitrile solution preserving a C_2 symmetry. In addition, for calculations in solution an equilibrium description has been used; here in fact the interest is not focused on transition phenomena, which could imply delays in the solvent response with respect to solute electronic changes, but on the analysis of sections of PES. In this case we can assume that solute and solvent can always completely equilibrate.

The results for the ground and the two lowest excited states of both isolated and solvated DMABN are presented in the graphs 1.6 and 1.7, respectively. All energies (in eV) are respect to the ground state at non-twisted geometry ($\theta = 0^\circ$).

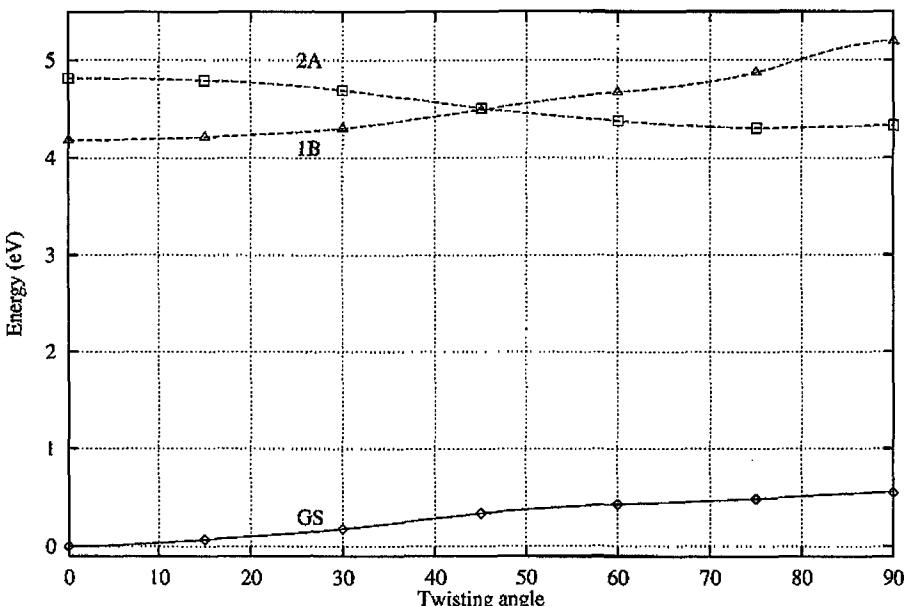


Fig. 1.6 Ground and low-lying singlet excited states potential curves (energy relative to the minimum of the ground state) of DMABN in vacuo as a function of the twisting angle. Energies are in eV and angles in degrees.

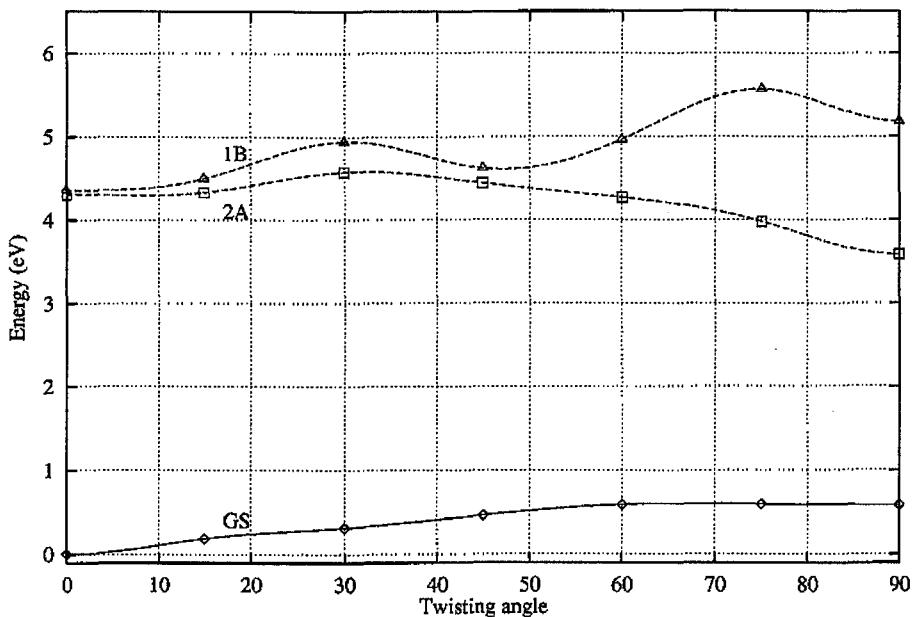


Fig. 1.7 Ground and low-lying singlet excited states potential curves (energy relative to the minimum of the ground state) of DMABN in acetonitrile as a function of the twisting angle. Energies are in eV and angles in degrees.

Inspection of the potential curves along the twisting path, as expected, shows very different features for isolated and solvated systems.

As reported in other calculations, gas phase results are characterized by the two lowest singlet excited states which cross at a twisting angle near 45° . In particular, the most stable excited state at the non-twisted geometry presents a maximum at perpendicular geometry, while the second one shows a completely reversed behavior thus inducing the crossing of the corresponding curves. The two states for which 1^1B and 2^1A labels are used according to the C_2 symmetry, can be identified with the L_b -type, or LE, state and the L_a -type, or CT, state, respectively. The observed crossing is here allowed due to the differences in the intrinsic symmetry character of the two states. An analysis of the CIPSI wavefunction shows that the first excited state (1^1B) at non-twisted conformation is dominated by the $\text{HOMO} \rightarrow \text{LUMO}+1$ and $\text{HOMO}-1 \rightarrow \text{LUMO}$ single excitations, where the involved orbitals are benzene ring π (for the two occupied) and π^* (for

the two unoccupied) orbitals. Also the other state (2^1A) is dominated by a single excited configuration. At the non-twisted geometry the wave function is mainly composed of the HOMO \rightarrow LUMO excitation; it should be noted that the HOMO orbital includes a strong mixing with the N amino lone pair orbital (lpN). It is the excitation from this orbital (i.e. the transition lpN \rightarrow LUMO or better $n \rightarrow \pi^*$) that becomes the most important at the twisted conformation. At this conformation the lone pair orbital is no more mixed with the π system and, as a consequence, it is destabilized up to the HOMO level. The partial lpN character of the HOMO orbital observed in the non-twisted wavefunction of the 2^1A state shows that some of the charge transfer inferred by the $n \rightarrow \pi^*$ transition is already included. As we shall show, this mixing acquires an important role in presence of a polar solvent.

In the presence of a polar solvent the same curves become significantly different as the highly polar 2^1A state can be stabilized much more effectively than the less polar 1^1B state thus inverting their relative energies even at non-twisted geometry. Some further comments for such behavior can be derived from the orbital analysis. The 2^1A state is mainly dominated by the singly excited configuration HOMO \rightarrow LUMO exactly as in vacuo; here, however, the partial mixing among benzene π orbitals and N lone pair we have already indicated for the HOMO orbital in the isolated system is largely increased, as well as the percent of charge transfer character of the state. Qualitatively this means that the solvent can interact more effectively with this state lowering its energy below the other one involving only $\pi \rightarrow \pi^*$ excitations.

The second important characteristic of the solution curves is the important stabilization of the 2^1A state at twisted geometry which represents the global minimum in the curve; the twisted conformation lies ~ 19.5 kcal/mol below the planar one, and the thermodynamic equilibrium of the charge transfer reaction is therefore almost totally shifted to the twisted conformation. This result gives an indirect support to the interpretation of the dual fluorescence in terms of the TICT model. It seems that this large stabilization and not the relative values of the barrier between non-twisted and twisted geometries is the most important point. In fact these curves do not represent the lowest energy path on the multidimensional energy surface and thus the barrier we can estimate from them is a very rough approximation. Anyway, assuming valid the definition of such barrier as the energy difference between the minimum of the B state and the energy maximum in the path going to the A minimum, the computed results show

Table 1.17 Fluorescence energies (eV) at planar and twisted conformations for DMABN in vacuo and in two solvents ('acn' indicates acetonitrile and 'cyc' cyclohexane). Experimental data refer to fluorescence maxima (the parentheses indicate that such value is very uncertain).

	calculated		exp	
	planar	twisted	LE	CT
vac	4.24	3.53		
acn	3.87	2.50	(3.4) ^a	2.6 ^a

^a Ref. [71] and refs. therein.

a significant decrement of the barrier height (from 7 to 5 kcal/mol from vacuo to solution) in qualitative agreement with experiments (see ref.[65] and references therein).

For the emission spectra, the experimental finding is that two distinct bands are found in acetonitrile (centered at 3.4 and 2.6 eV, respectively), thus leading to the phenomenon of dual fluorescence, and that in the non-polar cyclohexane a very weak CT band at 3.2 eV is found besides the strong normal LE fluorescence at 3.6 eV; no clear data are available for gas phase. The geometries used for the excited states are those obtained with CIS optimizations both in vacuo and in solution. The main aspect to observe for all the excited structures when compared to the corresponding ground state geometry is the strong increase in the phenyl bond lengths at the substituent-bearing carbons. On the contrary, the bond lengths of the carbons in between (here represented by R_{C₂C₃}) become shorter. In both the twisted structures a significant decrease with respect to the corresponding ground state value of the amino-ring bond is observed; this can be tentatively explained in terms of the charge transfer from the donor to acceptor part of the molecule which leads to a positively charged amino group.

The fluorescence energies (eV) at planar and twisted conformations for DMABN in vacuo and in two solvents along with experimental results are summarized in Table 1.17. The parentheses indicate that such value is very uncertain as the band is apparent as a broadening of the emission spectrum without a distinct maximum. Experimental data refer to fluorescence maxima.

More than gas-phase data for which no experimental confirmation can

be achieved, it is interesting to analyze results obtained in acetonitrile. The results reported in Table 1.17 show a very good agreement for the second, or anomalous, fluorescence (the discrepancy is 0.1 eV), while a less good behavior is found for the first, or normal, fluorescence. However, we have to recall that in polar solvents the CT (2^1A) band dominates and the LE (1^1B) band is observed as a shoulder on the very broad CT band. Furthermore, the lifetime of LE state is very short, in the picosecond region (it is in fact a largely permitted state), while the lifetime of CT state is in the nanosecond region, all these aspects should lead to a large uncertainty in the experimental datum for LE.

1.6. Conclusions

The overview of applications of the description of solvent effects on molecular properties we have presented in the previous sections is limited to the use of PCM model developed in our laboratory. However, it is important to stress that we have not exposed all the results we have thus far collected, that other groups have used PCM for similar purposes, and, more important, that there are other continuum solvation methods that allow the calculation of the properties we have considered.

The wealth of information about the performances of continuum solvation methods in computing molecular properties is nowadays considerable and it is steadily growing. In general, a sizeable number of positive hits are the necessary prerequisite to pass to a second step in the elaboration of the model: the examination of "difficult" cases and the introduction of refinements with respect to the basic model. This second step, which can be (and actually is in the present case) accompanied by what we call here "extensions", has little sense when the model gives wrong answers in a number of cases comparable to the correct ones. We may say that this fundamental preliminary test has been successfully overcome by PCM, and, by other continuum codes. We may thus try to resume, from the results we have displayed and from other data coming from our laboratory or from other sources, some points regarding the second step of the investigation.

1) Accuracy of the version of the model employed in a specific application. It can be noticed that several results presenting low agreement with experiments regard incomplete versions of the model. This happened, in particular, for the determination of the optimized geometries and the corresponding energies. The effects of conformational equilibria and/on reaction energy barriers are cases in which a relatively poor description of the ge-

ometry may have important consequences. The case of glycine/alanine we have presented is just an example: previous studies on the same systems but with less accurate solvation models, in fact, arrived at the conclusion that continuum models must be accompanied by specific solvent molecules to properly describe equilibria in the α -amino acids.

2) The need of introducing explicit solvent molecules. This is indeed an important issue to consider in all continuum solvation models, however, only an accurate comparison of the descriptions obtained with different approaches (from the basic one, in which the solvent is entirely described by a continuum dielectric, to those containing also discrete solvent molecules) can lead to a reliable and correct analysis of the problem. Only such comparison in fact allows to define the nature, and the relative importance, of the interactions to be included in the model. In the previous sections, we have given an example of this issue when applied to the calculation of N nuclear shielding of diazines in water solution, but we have many other different cases in our files or in other published papers. What appears evident is that this problem is very complex and that, even for a given property, as the nuclear shielding for example, there is not a unique answer but different descriptions have to be considered according to the specific solute-solvent system under scrutiny. In particular, an important question should always be made as a preliminary step: do the solute-solvent interactions, which we are trying to model by adding some solvent molecules, belong to the category of specific and permanent interactions? Distinctions should, in fact, be done among interactions that are specific, permanent, or both as they require different approaches. It should be evident to all, that this is a very complex problem and that it cannot be discussed in details here, however, we have remarked it as it represents one of the most challenging problem for any theoretical model.

3) The introduction in the model of time-dependent aspects. This issue actually belongs to the category of extensions rather than to that of refinements. The changes in the model that such aspect requires, have been sufficiently explained in the previous sections; here, we remark that the perspective of using time-dependent, or nonequilibrium, formulations is by far larger than that we have described. Among the properties we have not considered in the previous sections as we have just started to work on, we quote the solvent relaxation phenomena produced by changes in the solute for which we have to use the whole complex permittivity $\epsilon(\omega, k)$ spectrum. In some cases, time-dependent effects belong to the category of refinements; a typical one is that of molecular vibrations. There is a small refinement in

frequency due to consideration of the inner dynamics of the vibration, and a larger effect in the IR and Raman intensities.

4) The refinement of the model to reproduce the conditions exploited in the experiments and thus to get a more correct comparison between computed and observed data. This issue mostly applies to response properties of molecular systems to external electromagnetic fields. In these cases in fact, we had to reconsider the well-known problem of the "local field effect". Local field effects are of general occurrence when an external field interacts with a molecule in a condensed phase; they can be qualitatively defined as the effects arising from the fact that the field locally experienced by the molecule is different from the macroscopic (Maxwell) field in the medium. In the last few years we have given new numerical recipes to compute the local correction to the applied field; these have shown to be more general, and more accurate, than that given by old, and celebrated, models. This approach applies, with due changes, to all the properties we have considered in this paper, hyperpolarizabilities, IR, Raman and VCD intensities, OR, as well as to others we have not considered, as Cotton-Mouton and Kerr effects.

The set of checks, refinements, extensions, we have shown here is just a small part of what is required for a method that aims at broad field of applications as continuum solvation models are. Much more has still to be done, but the results we have tried to summarize in the previous sections seem to show that some of the fundamental issues have been already identified and realized.

Bibliography

1. (a) J. Tomasi and M. Persico (1994) *Chem. Rev.* **94**, 2027; (b) C.J. Cramer and D.G. Truhlar (1999) *Chem. Rev.* **99**, 2161.
2. C. Amovilli, V. Barone, R. Cammi, E. Cancès, M. Cossi, B. Mennucci, C.S. Pomelli and J. Tomasi (1998) *Adv. Quant. Chem.* **32**, 227.
3. H. Reiss, H.L. Frisch, J.L. Lebowitz (1959) *J. Chem. Phys.* **31**, 369.
4. R. A. Pierotti (1976) *Chem. Rev.* **76**, 717.
5. (a) S. Miertuš, E. Scrocco and J. Tomasi (1981) *Chem. Phys.* **55**, 117; (b) R. Cammi and J. Tomasi (1995) *J. Comp. Chem.* **16**, 1449.
6. (a) E. Cancès, B. Mennucci and J. Tomasi (1997) *J. Chem. Phys.* **107**, 3032; (b) B. Mennucci and E. Cancès, J. Tomasi (1997) *J. Phys. Chem.*

- B **101**, 10506; (c) E. Cancès and B. Mennucci (1998) *J. Math. Chem.* **23**, 309.
- 7. (a) A. Klamt and G. Schuurmann (1993) *J. Chem. Soc., Perkin Trans. 2*, 799; (b) J. Andzelm, C. Kolmel and A. Klamt (1995) *J. Chem. Phys.* **103**, 9312.
 - 8. (a) T.N. Truong, E.V. Stefanovich (1995) *J. Chem. Phys.* **103**, 3709; (b) E.V. Stefanovich, T.N. Truong (1996) *J. Chem. Phys.* **105**, 2961.
 - 9. V. Barone and M. Cossi (1998) *J. Phys. Chem. A* **102**, 1995.
 - 10. J.B. Foresman, T.A. Keith, K.B. Wiberg, J. Snoonian and M.J. Frisch (1996) *J. Phys. Chem.* **100**, 16098.
 - 11. J.-L. Rivail, D. Rinaldi, in *Computational Chemistry Review of Current Trends*, J. Leszczynski (ed.), World Scientific, New York, 1995; and reference therein.
 - 12. (a) K.V. Mikkelsen, E. Dalgaard, P. Swanstrøm (1987) *J. Phys. Chem.* **91**, 3081; (b) K.V. Mikkelsen, P. Jørgensen, H.J.Aa. Jensen (1994) *J. Chem. Phys.* **100**, 6597.
 - 13. M.M. Karelson, M. Zerner (1992) *J. Phys. Chem.* **96**, 6949.
 - 14. C.J. Cramer, D.G. Truhlar in *Reviews on Computational Chemistry*, D.B. Boyd and K.B. Lipkowitz (eds.), Vol.6, VCH, New York, 1995; and references therein.
 - 15. J.L. Pascual-Ahuir, E. Silla, J. Tomasi and R. Bonaccorsi (1987) *J. Comp. Chem.* **8**, 778.
 - 16. B. Lee and F.M. Richards (1971) *J. Mol. Biol.* **55**, 379.
 - 17. B. Mennucci and J. Tomasi (1997) *J. Chem. Phys.* **106**, 5151.
 - 18. D. M. Chipman (2000) *J. Chem. Phys.* **112**, 5558.
 - 19. E. Cancès and B. Mennucci (2001) *J. Chem. Phys.* **114**, 4744.
 - 20. B. Mennucci, R. Cammi and J. Tomasi (1998) *J. Chem. Phys.* **109**, 2798.
 - 21. J. Langlet, P. Claverie, J. Caillet and A. Pullman (1988) *J. Phys. Chem.* **92**, 1617.
 - 22. F.M. Floris and J. Tomasi (1989) *J. Comput. Chem.* **10**, 616; F.M. Floris, J. Tomasi and J.L. Pascual-Ahuir (1991) *J. Comput. Chem.* **12**, 784.
 - 23. C. Amovilli and B. Mennucci (1997) *J. Phys. Chem. B* **101**, 1051.
 - 24. (a) K.V. Mikkelsen, H. Ågren, H.J.Aa. Jensen and T. Helgaker (1988) *J. Chem. Phys.* **89**, 3086; (b) V. Dillet, D. Rinaldi, J. Bertran, J.L. Rinaldi (1996) *J. Chem. Phys.*, **104**, 9437; (c) T. Zhu, J. Li, D.A. Liotard, C.J. Cramer, D.G. Truhlar (1999) *J. Chem. Phys.* **110**, 5503.
 - 25. (a) R. Cammi and J. Tomasi (1994) *J. Chem. Phys.* **100**, 7495; (b) R.

- Cammi and J. Tomasi (1994) *J. Chem. Phys.* **101**, 3888.
26. (a) E. Cancès and B. Mennucci (1998) *J. Chem. Phys.* **109**, 249; (b) E. Cancès, B. Mennucci and J. Tomasi (1998) *J. Chem. Phys.* **109**, 260.
 27. M. Cossi and V. Barone (1998) *J. Chem. Phys.* **109**, 6246.
 28. B. Mennucci, R. Cammi and J. Tomasi (1999) *J. Chem. Phys.* **110**, 6858.
 29. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, C. S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, A. G. Baboul, J. Cioslowski, B. B. Stefanov, G. Liu, C. A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, C. M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople (Gaussian, Inc., Pittsburgh, PA, 1998); *Gaussian 99, Development Version*.
 30. M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, J.A. Montgomery (1993) *J. Comp. Chem.* **14**, 1347.
 31. L. Gontrani, B. Mennucci and J. Tomasi (2000) *J. Mol. Struct. (Theochem)* **500**, 113.
 32. R. Cammi, B. Mennucci and J. Tomasi (2000) *J. Phys. Chem. A* **104**, 4690.
 33. J. Gao and C. Alhambra (1997) *J. Am. Chem. Soc.* **119**, 2962.
 34. R. Cammi, C. Cappelli, S. Corni and J. Tomasi (2000) *J. Phys. Chem. A* **104**, 9874.
 35. C. Cappelli, S. Corni, R. Cammi, B. Mennucci, J. Tomasi (2000) *J. Chem. Phys.* **113**, 11270.
 36. (a) D.M. Bishop (1990) *Rev. Mod. Phys.* **62**, 343; (b) D.M. Bishop (1998) *Adv. Chem. Phys.* **104**, 1.
 37. (a) R. Cammi and J. Tomasi, *Int. J. Quantum Chem.* **60**, 1165 (1996); (b) R. Cammi, M. Cossi, B. Mennucci and J. Tomasi, *J. Chem. Phys.* **105**, 10556 (1996).
 38. R. McWeeny, *Methods of Molecular Quantum Mechanics*, (Academic

- Press, London, 1992); second edition.
- 39. (a) H. Sekino, R.J. Bartlett (1986) *J. Chem. Phys.* **85**, 976; (b) S.P. Karna, M. Dupuis (1991) *J. Comp. Chem.* **12**, 487.
 - 40. J. Olsen, P. Jorgensen, *Time-dependent Response theory with application to SCF and MCSCF wavefunctions*, in *Modern Electronic Structure Theory*, D.R. Yarkony (ed.) (World Scientific, 1995).
 - 41. (a) D. Jonsson, P. Norman, H. Ågren, Y. Luo, K.O. Sylvester-Hvid, K.V. Mikkelsen (1998) *J. Chem. Phys.* **109**, 6351; (b) K.V. Mikkelsen, Y. Luo, H. Ågren, P.J. Jorgensen (1994) *J. Chem. Phys.* **100**, 8240.
 - 42. R. Cammi, B. Mennucci (1999) *J. Chem. Phys.* **110**, 9877.
 - 43. B. Mennucci, R. Cammi, M. Cossi, J. Tomasi (1998) *Theochem*, **426**, 191.
 - 44. (a) C. Minichino, and V. Barone (1994) *J. Chem. Phys.* **100** 3717; (b) V. Barone, and C. Minichino (1995) *Theochem* **330** 365.
 - 45. I. Ledoux, and J. Zyss (1982) *Chem. Phys.* **73** 203.
 - 46. (a) D.M. Bishop, B. Kirtman (1991) *J. Chem. Phys.* **95**, 2646; (b) D.M. Bishop, B. Kirtman (1992) *J. Chem. Phys.* **97**, 5255.
 - 47. (a) B. Champagne, E.A. Perpète, J.M. André, B. Kirtman (1995) *J. Chem. Soc. Faraday Trans.* **91**, 1641; (b) B. Champagne, H. Vanderhoeven, E.A. Perpète, J.M. André (1996) *Chem. Phys. Lett.* **248**, 301.
 - 48. B. Kirtman, B. Champagne (1997) *Int. Rev. Phys. Chem.* **16**, 389.
 - 49. (a) A.D. Buckingham (1960) *Can. J. Chem.* **38**, 300; (b) A.D. Buckingham, T. Schaefer, W.G. Schneider (1960) *J. Chem. Phys.* **32**, 1227.
 - 50. (a) R. Ditchfield (1974) *Mol. Phys.* **27**, 789; (b) K. Wolinski, J.F. Hinton, P. Pulay (1990) *J. Am. Chem. Soc.* **112**, 8251.
 - 51. (a) R. Cammi (1998) *J. Chem. Phys.* **109**, 3185; (b) R. Cammi, B. Mennucci and J. Tomasi (1999) *J. Chem. Phys.* **110**, 7627.
 - 52. M. Witanowski, L. Stefaniak, G.A. Webb, in *Annual Reports on NMR Spectroscopy*; G.A. Webb (ed.), Vol.25, Academic Press: London, 1993.
 - 53. B. Mennucci (2002) *J. Am. Chem. Soc.* **124**, 1506.
 - 54. M. Witanowski, W. Sicinska, S. Biernat and G.A. Webb (1991) *J. Magn. Reson.* **91**, 289.
 - 55. R.D. Amos (1982) *Chem. Phys. Lett.* **87**, 23.
 - 56. B. Mennucci, J. Tomasi, R. Cammi, J.R. Cheeseman, M.J. Frisch, F.J. Devlin, S. Gabriel and P. Stephens (2002) *J. Phys. Chem. A*, ASAP.
 - 57. J.R. Cheeseman, M.J. Frisch, F.J. Devlin and P.J. Stephens (1996) *Chem. Phys. Lett.* **252**, 211.
 - 58. (a) J. Tomasi, R. Cammi and B. Mennucci (1999) *Int. J. Quantum Chem.* **75**, 783; (b) B. Mennucci, R. Cammi and J. Tomasi (1999) *Int.*

- J. Quantum Chem.* **75**, 767.
59. (a) Y. Ooshika (1954) *J. Phys. Soc. Jpn.* **9**, 594; (b) R. Marcus, *J. Chem. Phys.* **24**, 966 (1956); (c) E.G. McRae, *J. Phys. Chem.* **61**, 562 (1957); (d) S. Basu (1964) *Adv. Quantum. Chem.* **1**, 145; (e) V.G. Levich (1966) *Adv. Electrochem. Eng.* **4**, 249; (f) H.J. Kim, and J.T. Hynes (1990) *J. Chem. Phys.* **93**, 5194 and 5211; (g) A.M. Berezhkosvskii (1992) *Chem. Phys.* **164**, 331; (h) M.V. Basilevski, and G.E. Chudinov (1990) *Chem. Phys.* **144**, 155; (i) K.V. Mikkelsen, A. Cesar, H. Ågren, and H. J. Aa. Jensen (1995) *J. Chem. Phys.* **103**, 9010.
 60. (a) M.A. Aguilar, F.J. Olivares del Valle, and J. Tomasi (1993) *J. Chem. Phys.* **98**, 7375; (b) R. Cammi and J. Tomasi (1995) *Int. J. Quantum Chem: Quantum Chem. Symp.* **29**, 465; (c) M.L. Sánchez, M.A. Aguilar and F.J. Olivares del Valle (1995) *J. Phys. Chem.* **99**, 15758.
 61. E. Lippert, W. Lüder and H. Boos, in *Advanced molecular spectroscopy*; A. Mangini (ed.), Vol. 1, Pergamon Press, New York, 1962.
 62. (a) K. Rotkiewicz, K.H. Grellmann, Z.R. Grabowski (1973) *Chem. Phys. Lett.* **19**, 315; (b) Z.R. Grabowski, K. Rotkiewicz, A. Siemarczuk, D.J. Cowley and W. Baumann (1979) *Nouv. J. Chim.* **3**, 443.
 63. A. Broo and M.C. Zerner (1994) *Chem. Phys. Lett.* **227**, 551.
 64. (a) G.D. Scholes, D. Phillips and I.R. Gould (1997) *Chem. Phys. Lett.* **266**, 521; (b) G.D. Scholes, I.R. Gould, A.W. Parker, D. Phillips (1998) *Chem. Phys.* **234**, 21.
 65. P. Gedeck and S. Schneider (1997) *J. Photochem. Photobiol. A: Chem.* (a) **105**, 165; (b) (1999) *ibidem* **121**, 7.
 66. B. Mennucci, A. Toniolo and J. Tomasi (2000) *J. Am. Chem. Soc.* **122**, 10621.
 67. (a) B. Huron, J.P. Malrieu and P. Rancurel (1973) *J. Chem. Phys.* **58**, 5745; (b) S. Evangelisti, J. Daudey and J.P. Malrieu (1983) *Chem. Phys.* **75**, 91; (c) F. Spiegelmann and J.P. Malrieu (1984) *J. Phys. B* **17**, 1235; (d) R. Cimiraglia, *J. Chem. Phys.* **83**, 1746 (1985); (e) R. Cimiraglia and M. Persico (1987) *J. Comp. Chem.* **8**, 39.
 68. B. Mennucci, A. Toniolo and C. Cappelli (1999) *J. Chem. Phys.* **110**, 6858.
 69. R. Cammi, B. Mennucci and J. Tomasi (2000) *J. Phys. Chem. A*, **104**, 5631.
 70. C. Bulliard, M. Allan, G. Wirtz, E. Haselbach, K.A. Zachariasse, N. Detzer and S. Grimme (1999) *J. Phys. Chem. A* **103**, 7766.

71. J. Lipinski, H. Chojnacki, Z.R. Grabowski and K. Rotkiewicz (1980) *Chem. Phys. Lett.* **70**, 449.
72. A.B.J. Parusel, K. Kohler and S. Grimme (1998) *J. Phys. Chem. A* **102**, 6297.
73. (a) R. Cammi, B. Mennucci and J. Tomasi (1998) *J. Phys. Chem. A* **102**, 870; (b) (2000) *ibidem* **104**, 4690.

Chapter 2: Electronic and Nonlinear Optical Properties of 2-Methyl-4-Nitroaniline Clusters

Maxime Guillaume^a, Benoît Champagne^a, Frédéric Castet^b and
Laurent Ducasse^b

^a*Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires
Notre-Dame de la Paix, 61 rue de Bruxelles, B-5000 Namur, Belgium*

^b*Laboratoire de Physico-Chimie Moléculaire, UMR 5803 CNRS
Université Bordeaux I, 33405 Talence Cedex, France*

Abstract

The supermolecule approach is used to study the linear and second-order nonlinear susceptibilities of the 2-methyl-4-nitroaniline crystal. The packing effects on these properties, evaluated at the time-dependent Hartree-Fock level with the AM1 Hamiltonian, are assessed as a function of the size and shape of the clusters. A simple multiplicative scheme is demonstrated to be often reliable for estimating the properties of two- and three-dimensional clusters from the properties of their constitutive one-dimensional arrays. The electronic absorption spectra are simulated at the ZINDO level and used to rationalize the linear and nonlinear responses of the 2-methyl-4-nitroaniline clusters. Comparisons with experiment are also provided as well as a discussion about the reliability of the global approach.

1. Introduction

Materials presenting remarkable electronic, optical and magnetic properties are the ferment of the emerging technologies. Among these, nonlinear optical (NLO) materials can find a niche for applications ranging from frequency doublers, all-optical and electro-optical switches. To reach this goal, many quantum mechanical approaches have been elaborated and widely employed to understand the underlying phenomena of interactions between light and matter, to deduce structure-property relationships, and to design new materials. However, these efforts have, to a large extend, been restricted to the molecular hyperpolarizabilities. Indeed, at the microscopic level, several hierarchies of quantum chemistry methods have been developed to determine the first (β) and second (γ) hyperpolarizabilities. For small molecules, high level *ab initio* methods including electron correlation, vibrational contributions and frequency dispersion enable to attain an accuracy surpassing that of experiment. On the other hand, for larger systems, the most interesting in view of applications, the predictions have been more qualitative and related to the deduction of structure-property relationships. These aspects of hyperpolarizability calculations have recently been reviewed for small and medium-size molecules [1] as well as for large molecules and polymers [2].

The problem becomes more complex when studying solid phases because the microscopic NLO responses do not provide the full information about their macroscopic counterparts, the second- and third-order nonlinear susceptibilities, $\chi^{(2)}$ and $\chi^{(3)}$. To make the transition between the **microscopic** and **macroscopic**, it is necessary to know the structure of the condensed phases as well as the nature and the effects of the intermolecular interactions in the bulk of the material. In both the Physics and Chemistry arena, several schemes have been proposed to characterize the NLO responses of solid phases. One of the authors has recently contributed to review these approaches [3] of which one of the extremes is occupied by the oriented gas approximation that consists in performing a tensor sum of the microscopic NLO properties to obtain the macroscopic responses of the crystal. The other extreme consists in performing a complete treatment of the solid by using the supermolecule method or by taking advantage of the spatial periodicity in crystal orbital calculations. In between these techniques, one finds the interaction schemes and the semi-empirical approaches.

For molecular crystals, the interaction schemes and the supermolecule approaches have focused much attention. However, as mentioned in Ref. 3, the later has mostly been applied to small ensembles of organic molecules [4-14], typically dimers and trimers. Three years ago, in Namur and in Bordeaux, we began to extend these investigations to larger molecular ensembles. The molecular crystal of 2-methyl-4-nitroaniline (MNA) which presents one of the largest $\chi^{(2)}$ values was chosen. In this paper, we review these investigations on MNA clusters by summarizing the most important results of Refs. 15 and 16 as well as by providing new data and a more complete analysis. The aspects that are tackled here (Section IV) are i) the rationalization of the effects of the surroundings with respect to the size and the shape of the clusters on the first hyperpolarizabilities and excitation energies, ii) the extrapolation of the NLO cluster properties to determine the properties of the crystal, iii) the assessment of the differences between static and dynamic phenomena and their connection with the UV/Visible absorption spectroscopy. In addition, Section II summarizes the experimental data on the MNA crystal while Section III describes the adopted methodological approaches within the range of methods for computing molecular (hyper)polarizabilities and excitation energies.

2. The 2-methyl-4-nitroaniline crystal

When starting our investigations, two X-ray diffraction studies describing the structure of the crystal of MNA were available [17-18]. The crystal of MNA is monoclinic (space group Cc) and contains four MNA molecules per unit cell (Fig. 1). In Ref. 17, the unit cell parameters are $a = 11.17 \text{ \AA}$, $b = 11.60 \text{ \AA}$, $c = 7.90 \text{ \AA}$, $\beta = 137^\circ$, the density amounts to 1.48 g/cm^3 for a unit cell volume of 698.1 \AA^3 whereas $a = 11.57 \pm 0.01 \text{ \AA}$, $b = 11.62 \pm 0.01 \text{ \AA}$, $c = 8.22 \pm 0.01 \text{ \AA}$, $\beta = 139.18 \pm 0.02^\circ$, $\rho = 1.477 \text{ g/cm}^3$, and $V = 722.4 \text{ \AA}^3$ in Ref. 18. The MNA crystals, which are bright yellow, adopt the form of thin plates of approximately $50 \mu\text{m}$ thick, up to 0.5 cm wide and $1\text{-}2 \text{ cm}$ long [17]. Contrary to the parent *p*-nitroaniline compound where the crystal packing is centrosymmetric and $\chi^{(2)} = 0$, the methyl groups of MNA molecules prevent the molecules to stack head to tail. Furthermore, the non-centrosymmetric packing of the MNA is stabilized by the presence of weak hydrogen bonds between nitro and amino groups ($d_{\text{H...O}} = 2.33 \text{ \AA}$). The polar axes of the individual molecules are nearly aligned with each other along the crystallographic a -axis. In a dimer, the angle between the two *quasi-C₂*

molecular axis is about 16° . This means that the first hyperpolarizabilities of the individual molecules almost add together to produce a dominant β tensor component parallel to the a crystallographic axis (see also Section IV) and that the (1,4) MNA dimer can be considered as the $\chi^{(2)}$ -building block. In other words, the crystal structure preserves the microscopic properties, *i.e.* the large second-order NLO response which is attributed to the presence of the intramolecular charge-transfer between the amino and nitro groups.

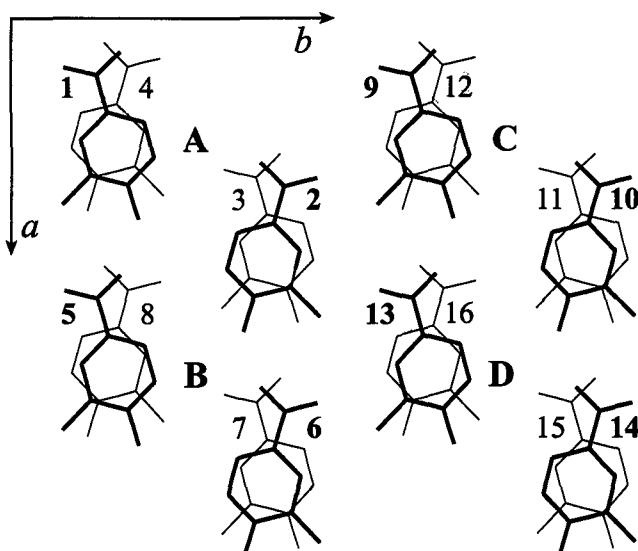


Fig. 1. MNA crystal structure. Following Figs. 3 and 4 of Ref. 17, a , b , and c refer to the crystallographic axes for a nonstandard setting (Ia) of the space group Cc ($a = 7.58 \pm 0.01$, \AA , $\beta = 94.08^\circ$). A-D correspond to unit cell labels. The unit cell A contains four molecules (1 to 4). Reprinted with permission from F. Castet and B. Champagne, J. Phys. Chem. A **105**, 1366 (2001). Copyright 2001 American Chemical Society.

In addition, MNA crystals present an almost optimal structure with respect to phase matching conditions [17]. This is important for effects such as frequency-doubling, although not for electro-optical effect, and explains why the MNA crystal presents one of the largest effective d_{12} coefficients among the class of push-pull aromatic compounds [19]. The d_{11} coefficient has also a remarkable figure of merit which is 2000 times larger than in LiNbO_3 , one of

the best known inorganic NLO crystals. d_{11} can be made phase-matchable in a waveguide configuration.

The $\chi_{111}^{(2)}(-2\omega; \omega, \omega) = 2d_{11}(-2\omega; \omega, \omega) \approx \chi_{aaa}^{(2)}(-2\omega; \omega, \omega)$ coefficient, measured with a laser source at 1064 nm, amounts to $300 \pm 75 \text{ pm/V}$ [17] after accounting for the new quartz reference [20]. The dc-Pockels electro-optic r_{11} coefficient has been measured using a wavelength equal to 632.8 nm and amounts to $67 \pm 25 \text{ pm/V}$ [18]. Since $n_1(632.8 \text{ nm}) = 2.0$ [21], this corresponds to $\chi_{111}^{(2)}(-\omega; \omega, 0) = 2d_{11}^{\text{EO}}(-\omega; \omega, 0) = [n_1^4(\omega)/2] \times r_{11}(-\omega; \omega, 0) = 536 \pm 200 \text{ pm/V}$. Using the two-state approximation, the electronic component of the r_{11} coefficient estimated by Bosshard *et al.* [19] from the SHG d_{11} value amounts to $70 \pm 14 \text{ pm/V}$ which enables to conclude that the phonon (vibrational) contribution to the electro-optic effect is negligible. In view of applications in waveguide optics, the refractive indices of the MNA crystal were measured by Morita *et al.* [21] at different wavelengths and fitted using the Sellmeier expression

$$n^2(\lambda) = A + \frac{B}{1 - \frac{C}{\lambda^2(\mu\text{m})}} \quad (1)$$

The reported parameters A, B and C for the different crystallographic axes are given in Table I. The structure of the MNA molecule is presented in Fig. 2. The geometrical parameters taken from the work of Lipscomb *et al.* [18] show the distortion of the benzene ring by the donor-acceptor (D/A) pair.

Table I. Sellmeier coefficients for the refractive indices of the MNA crystal.

	A	B	C
$n_x \approx n_a$	1.27	1.56	0.172
$n_y = n_b$	1.37	0.78	0.174
n_z	1.53	0.26	0.204

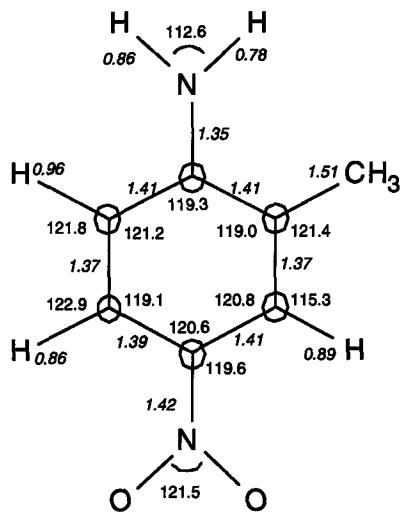


Fig. 2. Bond lengths (\AA) and angles ($^\circ$) of the MNA molecule according to Ref 17. The hydrogen linked to the benzene ring are located in the plane of the ring. The dihedral angle between the benzene ring and the oxygen (hydrogen) atoms of the nitro (amino) group is 1.1° (7.2°). Reprinted with permission from F. Castet and B. Champagne, J. Phys. Chem. A **105**, 1366 (2001). Copyright 2001 American Chemical Society.

3. Methodological aspects

The macroscopic optical responses of a medium are given by its linear and nonlinear susceptibilities that are the expansion coefficients of the material polarization, P , in terms of the Maxwell fields, E . When neglecting magnetic and electric field gradient effects, the defining equation reads [22]:

$$P = P_0 + \chi^{(1)} E + \chi^{(2)} E^2 + \chi^{(3)} E^3 + \dots \quad (2)$$

where P_0 is the permanent polarization, $\chi^{(1)}$ the linear susceptibility, $\chi^{(2)}$ and $\chi^{(3)}$ the second- and third-order nonlinear susceptibilities. At the microscopic level the n^{th} Cartesian component of the electric dipole moment

oscillating at a pulsation ω_σ is expressed in a Taylor series expansion of the external static and/or dynamic electric fields :

$$\begin{aligned} \mu_\eta(\omega_\sigma) = & \mu_\eta^0 \delta_{\omega_\sigma, 0} + \sum_\zeta \alpha_{\eta\zeta}(-\omega_\sigma; \omega_1) E_\zeta(\omega_1) \\ & + \frac{1}{2} K^{(2)} \sum_{\zeta\chi} \beta_{\eta\zeta\chi}(-\omega_\sigma; \omega_1, \omega_2) E_\zeta(\omega_1) E_\chi(\omega_2) \\ & + \frac{1}{6} K^{(3)} \sum_{\zeta\chi\xi} \gamma_{\eta\zeta\chi\xi}(-\omega_\sigma; \omega_1, \omega_2, \omega_3) E_\zeta(\omega_1) E_\chi(\omega_2) E_\xi(\omega_3) + \dots \end{aligned} \quad (3)$$

where the frequency-dependence and tensor nature of the (hyper)polarizabilities have been exemplified. In the right hand side, we may identify successively, the permanent dipole moment, the first-order response as the polarizability (α), and the second- and third-order responses as the first and second hyperpolarizabilities. These responses depend on the (Cartesian) electric field directions, indicated by the Greek subscripts, ζ , χ , and ξ as well as on the (circular) optical frequencies ω_1 , ω_2 , and ω_3 . In Eq.(3), $\omega_\sigma = \sum_i \omega_i$ while $K^{(2)}$ and $K^{(3)}$ are such that the β and γ associated with different NLO processes converge towards the same static value. For instance, at the second order, $\beta(0; 0, 0)$, $\beta(-\omega; \omega, 0)$, $\beta(0; \omega, -\omega)$, and $\beta(-2\omega; \omega, \omega)$ define the static, dc-Pockels, optical rectification, and second harmonic generation (SHG) responses, respectively. Sometimes the Taylor series expansion in Eq. (3) is replaced by a power series (like in Eq. (2) for the macroscopic analog). At the exception of crystal orbital approaches, molecular properties constitute necessary and non-avoidable quantities for estimating the macroscopic responses [3]. Accurate molecular properties constitute therefore a first condition to suitably describe the macroscopic properties. In addition, the effects of the surroundings on these molecules have to be accounted for. They are generally divided into two contributions. First, the surrounding modifies the molecular wavefunctions and therefore the molecular properties through the van der Waals forces, hydrogen bondings, donor/acceptor interactions, charge transfers, ... From a more physical angle, these "chemical" effects can be classified as electrostatic interactions, polarization, dispersion, and exchange-overlap [23]. The properties of these molecular entities embedded into a chemical environment are generally called **dressed** (hyper)polarizabilities. The second or **local field** effect, originates from the differences between the local electric fields that act on the individual molecules and the macroscopic or Maxwell fields of Eq. (2). These two contributions to the

effects of the surroundings on the NLO properties are, in principle, accounted for when applying the supermolecule approach. However, an accurate description of all these effects often requires the use of very elaborated techniques. Nevertheless, effects like London dispersion forces which are much sensitive to electron correlation description have a rather limited impact on the dressing effects. Moreover, since most of the effects of the surroundings have an electrostatic or polarization origin, an accurate evaluation of the molecular properties like the dipole moment, polarizability, and hyperpolarizabilities turns out to be, a second time, the key element for predicting the macroscopic NLO responses of materials.

They are different ways to classify the approaches for computing the (hyper)polarizabilities. On the one hand, the distinction can be made between the perturbation approaches where the (hyper)polarizabilities are given by summation-over-states (SOS) expressions and the differentiation approaches where (electric field) perturbed wavefunctions and/or electron densities are explicitly calculated. The theoretical level of treatment can also be used to define this classification. It ranges from empirical and semi-empirical schemes to high-level *ab initio* methods that incorporate systematically electron correlation effects and to the density functional approaches that account for electron correlation effects through approximations to the exact exchange-correlation functionals. Another classification refers to the physical phenomena that are included in the linear and nonlinear responses. Indeed, the application of external electric fields modifies in a global way the molecular characteristics, *i.e.* the electronic and nuclear distributions. In this global approach, the molecule is considered as a whole and the (hyper)polarizabilities can be given under the form of summations over rovibronic states [24]. However, this approach (as well as its vibronic analog) is computationally too demanding for a systematic screening of the molecular NLO properties in view of designing new molecules and materials. It is usual to replace it by a two-step procedure which follows closely the Born-Oppenheimer methodology and which treats sequentially, rather than simultaneously, the effects of the applied electric fields upon the motions of the electrons and nuclei. This is known as the clamped-nucleus approximation [25]. First, it consists in determining the effects on the electronic distribution and therefore in evaluating the **electronic** contribution to the (hyper)polarizabilities. Then, the nuclear motion on the potential energy surface until a new, field-dependent, nuclear configuration is attained gives rise to the **nuclear relaxation** contribution to the (hyper)polarizabilities whereas the **curvature** term originates from the field-dependence of the zero-point vibrational energy.

The nuclear relaxation and curvature terms form the **vibrational** contribution to the (hyper)polarizabilities that can also be decomposed into a pure vibrational contribution and a zero-point vibrationally-averaged (ZPVA) correction [1-2, 26]. Several schemes have been developed and employed to characterize small, medium, and large molecular systems [27]. Numerous applications have been carried out and have shown that the pure vibrational component can be very large and even dominate the total response for large organic molecules [2, 26, 28-30]. However, following Ref. 19 which reports that the pure vibrational contribution to dc-Pockels second-order nonlinear susceptibility is small in MNA crystals, these contributions are not treated in this work and the term “(hyper)polarizabilities” will implicitly refer to the “electronic (hyper)polarizabilities”. Less has been achieved for the ZPVA correction and its importance for large conjugated systems remains nowadays a matter of investigation [31].

Coming back to the distinction between the SOS and the differentiation approaches, using perturbation theory, the polarizability and first hyperpolarizability read [32]:

$$\alpha_{\eta\zeta}(-\omega_\sigma; \omega_1) = \sum P_{-\sigma,1} \sum_{n \neq 0} \frac{\langle 0|\mu_\eta|n\rangle \langle n|\mu_\zeta|0\rangle}{\Delta E_{0n} - \omega_\sigma} \quad (4)$$

$$\beta_{\eta\zeta\chi}(-\omega_\sigma; \omega_1, \omega_2) = \sum P_{-\sigma,1,2} \sum_{n \neq 0} \sum_{m \neq 0} \frac{\langle 0|\mu_\eta|n\rangle \overline{\langle n|\mu_\zeta|m\rangle} \langle m|\mu_\chi|0\rangle}{(\Delta E_{0n} - \omega_\sigma)(\Delta E_{0m} - \omega_2)} \quad (5)$$

where the sums run over all excited states $|n\rangle$ and $|m\rangle$ of energy E_n and E_m . $|0\rangle$ is the ground state wavefunction of energy E_0 . $\langle n|\mu_\zeta|m\rangle = \langle n|\mu_\zeta|m\rangle - \langle 0|\mu_\zeta|0\rangle \delta_{nm}$ and $\Delta E_{0n} = \hbar\omega_{0n} = \omega_{0n} = E_n - E_0$. $\sum P_{-\sigma,1}$ and $\sum P_{-\sigma,1,2}$ are sums over the two or six permutations of the pairs $(-\omega_\sigma, \eta)$, (ω_1, ζ) and (ω_2, χ) . In the SOS scheme, the accuracy on the (hyper)polarizabilities is dictated by the method used for determining the ground and excited state wavefunctions as well as the corresponding excitation energies and dipole transition moments. The most-widely employed method in combination with the SOS scheme is certainly the Configuration Interaction Singles (CIS) method with a semi-empirical Hamiltonian, typically INDO/S [33] or CNDOVSB [34]. This method has been widely popularized by Brédas [35], Morley [36] and their co-workers for deducing structure-property relationships in a large range of conjugated molecules. As discussed later, in

addition to its applicability to large molecules, one of the advantages of the SOS schemes is the interpretation of the NLO responses in terms of spectroscopic quantities. On the other hand, these approaches suffer from size inconsistency due to the truncation of the CI expansion. This truncation is particularly damaging when investigating the effects of the size of the system upon the NLO responses. In the polarization propagator, response, and equation-of-motion approaches, the central quantities are also SOS expressions but they are recast under the form of superoperator resolvents which are then evaluated by following a different formalism [37]. These later approaches have mostly been applied *ab initio* within electron correlated schemes going from Møller-Plesser perturbation theory method limited to second order (MP2) to the successive coupled-cluster approximations.

Aside the SOS approaches, the (hyper)polarizabilities can be evaluated by differentiating the energy (E_0) or the dipole moment with respect to the external electric fields. Since, in the electric dipole approximation, the field-dependent dipole moment is related to the energy through the following expression:

$$\mu(E) = - \left(\frac{\partial E_0(E')}{\partial E'} \right)_{E=E}, \quad (6)$$

a variety of schemes for computing α , β , ... can be derived from combining Eqs. (3) and (6):

$$\alpha(E) = - \left(\frac{\partial^2 E_0(E')}{\partial E'^2} \right)_{E'=E} = \left(\frac{\partial \mu(E')}{\partial E'} \right)_{E=E} \quad (7)$$

$$\beta(E) = - \left(\frac{\partial^3 E_0(E')}{\partial E'^3} \right)_{E'=E} = \left(\frac{\partial^2 \mu(E')}{\partial E'^2} \right)_{E'=E} = \left(\frac{\partial \alpha(E')}{\partial E'} \right)_{E=E} \quad (8)$$

These relations are also true for approximate wavefunctions provided the Hellmann-Feynman theorem is satisfied. This differentiation procedure is either numerical in the so-called finite-field (FF) method, analytical, or a combination of both. The FF method, first proposed by Cohen and Roothaan [38] for atomic properties, can only be employed for derivatives with respect to static electric fields. It consists in adding the perturbation potential, $-\vec{\mu} \cdot \vec{E}$, to the Hamiltonian and to determine the wavefunctions and energies as in the field-free case. The elements of the (hyper)polarizability tensors are then evaluated from numerical differentiation of the energy or dipole moment

obtained for electric fields of different magnitudes and directions. This approach can be applied to any level of theory for which the energy or dipole moment can be evaluated but it may suffer from numerical accuracy problems as well as from the unbounded character of the dipole moment operator. On the other hand, analytical schemes are required to get frequency-dependent (hyper)polarizabilities. At the Hartree-Fock level, it is the time-dependent Hartree-Fock (TDHF) scheme which is one of the most widely-employed techniques to evaluate the NLO responses. Its static equivalent, which provides the same (hyper)polarizability values as the HF-FF method, is known as the coupled-perturbed Hartree-Fock (CPHF) method. The TDHF approach is size consistent. It consists in expanding the matrices of the TDHF equation in Taylor series of the perturbation (static and dynamic electric fields) and in solving it order by order [39]. The so-obtained successive field-derivatives of the density matrix are then inserted into the (hyper)polarizability expressions. In the case of β ,

$$\beta_{\eta\zeta\chi}(-\omega_c; \omega_1, \omega_2) = -\text{Tr}[\mathbf{m}_\eta \mathbf{D}^{\zeta\chi}(\omega_1, \omega_2)] \quad (9)$$

where $\mathbf{D}^{\zeta\chi}(\omega_1, \omega_2)$ is the second-order derivative of the density matrix with respect to electric fields of pulsation ω_1 and ω_2 oriented in the ζ and χ directions, respectively. Different schemes have also been proposed to account for electron correlation effects. They range from the time-dependent MP2 approach of Hättig and Hess [40], to the *pseudo*-energy derivative method of Rice and Handy [41], and to the *quasi*-energy derivative scheme of Aiga and co-workers [42]. These approaches are also closely related to the polarization propagator, response or equation-of-motion schemes mentioned in the previous paragraph. As shown in Ref. 1, these correlated derivative and/or response approaches have mostly been applied *ab initio* to accurately characterize small systems in comparison with experimental data. On the other hand, the CPHF and TDHF schemes have also been adopted to study large conjugated compounds and to determine which donor/acceptor pair and which conjugated segment will confer optimal NLO responses to chromophores [2]. At a lower extend, the TDHF approach has also been used in combination with semi-empirical Hamiltonians in order to study larger systems [43]. As described and substantiated below, this is the type of approach that was employed for studying clusters containing from 2 to 32 MNA molecules.

In addition to these traditional wavefunction approaches methods based on density functional theory (DFT) have achieved a significant breakthrough during the last decade and they are now used worldwide to determine the structure and properties of small, medium as well as large molecules. In

particular, several schemes have been elaborated to calculate the molecular (hyper)polarizabilities [44] and improved exchange-correlation functionals (non-local, hybrid, asymptotically correct) have been proposed [45]. For small molecules, these methods provide polarizability values which are in good agreement with *post* Hartree-Fock methods whereas the hyperpolarizability values are usually overestimated and may not be better than the uncorrelated Hartree-Fock results. For spatially extended systems, the DFT overestimates of α , β , and γ as well as of μ become quite dramatic. This has been related to the ultra-locality of the conventional exchange-correlation functionals which incorrectly describe the ultra non-local polarization effects due to donor/acceptor substitutions or external electric fields [46-47]. To cure this failure, exact exchange [48] and current-density-functional [49] approaches have been designed and have shown that a better agreement with *post* Hartree-Fock methods can also be achieved for large molecules. However, to our knowledge, these methods have not yet been employed to study the linear and nonlinear properties of donor-acceptor conjugated systems like MNA.

This description of quantum mechanical methods for computing (hyper)polarizabilities demonstrates why, nowadays, the determination of hyperpolarizabilities of systems containing hundreds of atoms can, at best, be achieved by adopting, for obvious computational reasons, semi-empirical schemes. In this study, the evaluation of the static and dynamic polarizabilities and first hyperpolarizabilities was carried out at the Time-Dependent Hartree-Fock (TDHF) [39] level with the AM1 [50] Hamiltonian. The dipole moments were also evaluated using the AM1 scheme. The reliability of the semi-empirical AM1 calculations was addressed in two ways. For small and medium-size push-pull polyenes, the TDHF/AM1 approach was compared to Hartree-Fock and *post* Hartree-Fock [51] calculations of the static and dynamic longitudinal first hyperpolarizability. Except near resonance, the TDHF/AM1 scheme was shown to perform appreciably better than the *ab initio* TDHF scheme. Then, the static electronic first hyperpolarizabilities of the MNA molecule and dimer have been calculated [15] with various *ab initio* schemes and compared to the AM1 results. In particular, the inclusion of electron correlation at the MP2 level leads to an increase of β_{aaa} by about 50% with respect to the CPHF approach, similar to the effect calculated by Sim *et al.* [52] for the longitudinal β tensor component of *p*-nitroaniline. The use of AM1 Hamiltonian predicts a β_{aaa} value that is smaller than the correlated MP2/6-31G result but larger than any of the CPHF ones, which results from the implicit treatment of correlation effects, characteristic of the semi-empirical methods. This comparison confirms that a part of the electron

correlation necessary to the description of nonlinear polarization effects is present in the AM1 parameterization. Moreover, and more important, the AM1 Hamiltonian reproduces to within 3% the crystal packing ratio - estimated by dividing the first hyperpolarizability of the dimer by twice the monomer value - determined at the MP2 level. On the basis of this dimer investigation it appeared that the AM1 parameterization is a method of choice for determining the crystal packing effects on the first hyperpolarizability of MNA crystals because it provides the quality of *ab initio* correlated methods at a reduced computational price. In addition, improved crystal nonlinear susceptibility can be obtained by combining the AM1 crystal packing ratios with the *ab initio* correlated dimer first hyperpolarizability. The TDHF/AM1 calculations were performed using the MOPAC2000 program [53] while static values can also be obtained using the FF procedure implemented into the AMPAC 6.0 code [54] according to Kurtz *et al.* [55].

The determination of excitation energies and excited state properties of large molecules or clusters of molecules is also a difficult task where electron correlation plays a crucial role. On the one hand, *ab initio* wavefunction approaches are still limited to rather small systems. Indeed, the excitation energies of the first $\pi-\pi^*$ transitions in benzene are typically overestimated by 0.3-0.4 eV at the CCSD level with respect to both experiment and the CCSD(T) method [56]. This already corresponds to an underestimation of the wavelength by about 10-20 nm while most of the investigations on molecular clusters cannot be achieved at such highly-correlated level. On the other hand, the methods of density functional theory still suffer from the ultra locality problem of the conventional exchange-correlation functionals when describing long-range phenomena like charge-transfer transitions [57]. In order to combine computational feasibility with quality of the results, the excitation energies were calculated at the ZINDO level, *i.e.* using the CIS scheme with Zerner's semi-empirical INDO/S Hamiltonian [33]. Indeed, the ZINDO approach is known to yield accurate predictions for the electronic excitation energies because, on the one hand, most of the low-lying excitations are dominated by single excitations and, on the other hand, the INDO/S parameters were fitted to reproduce spectroscopic data [33]. The CIS scheme also explicitly accounts for intra- and intermolecular charge-transfer excitations, including their coupling, and as it has already been demonstrated by Canuto *et al.* [58], it includes up to the second-order in the perturbation theory the dispersion interactions between the subsystems in a supermolecule approach. In ZINDO calculations on molecular clusters, the size of the excitation manifold is also a crucial issue. Indeed, the ZINDO approach has

been parameterized for an active space made of the 10 highest-energy occupied and 10 lowest-energy unoccupied molecular orbitals, *i.e.* for 100 single excitations. On the other hand, since only the full CIS approach is size-consistent, keeping the same excitation manifold while increasing the system size can lead to errors. This is indirectly illustrated in Fig. 3 when varying the CIS space for the MNA unit cell [16].

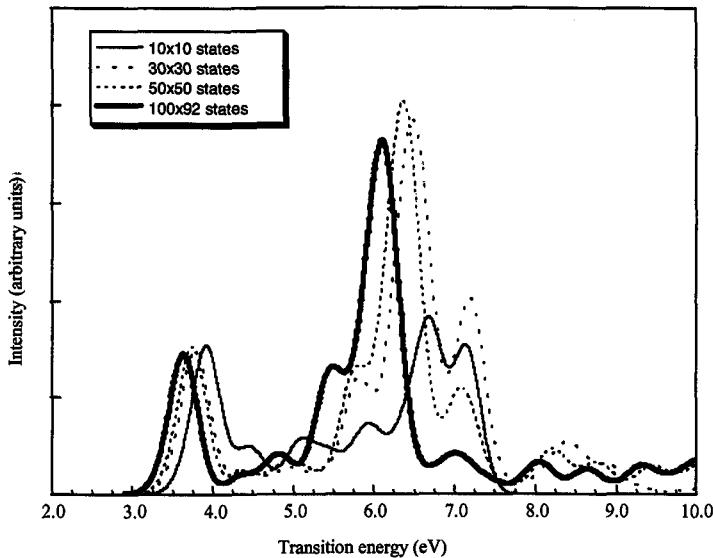


Fig. 3. Simulated ZINDO electronic absorption spectrum of the MNA unit cell as a function of the size of the configuration space. The total number of occupied and unoccupied molecular orbitals for the MNA unit cell is 116 and 92, respectively. FWHM = 0.4 eV. Reprinted from M. Guillaume *et al.*, Int. J. Quantum Chem. **90**, 1378 (2002) by permission of John Wiley & Sons, Inc.

Moreover, Table II shows that, for the dominant low-energy intramolecular charge-transfer excitation of clusters containing N MNA molecules, reducing the excitation manifold to $(10N) \times (10N)$ leads to a systematic hypsochromatic shift of *circa* 0.1 eV. Consequently, in order to reduce most of this inconsistency, comparisons between clusters containing one, two and three unit cells have been based on calculations using 30×30 , 60×60 and 90×90 excited states, respectively. The ZINDO calculations were carried out using the MOS-F program [59]. In order to assess the effects of including higher excitations in the CI space on the excitation energies, Table II compares the energy of the first intense intramolecular charge-transfer

excitation obtained at the ZINDO level with those calculated with CIS and truncated CISD schemes in combination with the AM1 hamiltonian. The AM1 calculations have been performed using the AMPAC program [54]. Like the CIS/INDO/S scheme, the CIS/AM1 approach provides small variations of the excitation energy with the size of the cluster whereas the CISD/AM1 method gives a much larger hypsochromatic shift when going from the dimer to the unit cell. In the case of the monomer, the inclusion of double excitations improves the agreement with the experimental value of 4.31 eV [19] whereas for the dimer (and the unit cell), there is no available experimental data. In such a case, it would be interesting to check these results with respect to highly-correlated methods such as the coupled clusters schemes [56, 60] but this is computationally impossible for us at this moment. Since the AM1 scheme has not been parameterized to reproduce excitation energies, it appears wise just to conclude that both the CIS and CISD schemes predict an increase of the excitation energy when going from the monomer to the dimer.

Table II. Excitation energy (in eV) of the dominant low-energy intramolecular charge-transfer excitation obtained from CIS and CISD schemes.

	ZINDO ^a	ZINDO ^b	CIS/AM1 ^c	CISD/AM1 ^d
Monomer	3.48	3.58	3.66	4.11
Dimer	3.60	3.72	3.74	4.27
Unit cell	3.65	3.77	3.88	5.27

^a full CIS space with a maximum of 100 occupied or virtuals.

^b 10N × 10N CIS space

^c 4N × 4N CIS space. A maximum of 3600 Slater determinants were kept in the CI matrix.

^d 4N × 4N CISD space. A maximum of 5600 Slater determinants were kept in the CI matrix.

4. Results

In order to address the crystal packing effects and to determine the properties of the MNA crystal, molecular clusters of increasing size were studied and their properties were extrapolated to the infinite crystal size. This is similar to the oligomeric approach used to study polymers although for crystals the system grows in the three space directions. As mentioned in Section II, the dimer appears as the building block. Table III lists the components of the

Table III. Tensor components of the AM1 dipole moment, static and dynamic polarizabilities and first hyperpolarizabilities for the monomer (1), the dimer (1, 4) and the unit cell (1, 2, 3, 4) in the MNA crystal frame. The a , b , c indices refer to the crystallographic axes. The dipole moments are given in D (for the conversion to a.u., 1 a.u. of dipole moment = 8.478358×10^{-30} Cm = 2.5415 D) whereas the other values in a.u. (1 a.u. of polarizability = 1.6488×10^{-41} C 2 m 2 J $^{-1}$ = 0.14818×10^{-24} esu; 1 a.u. of first hyperpolarizability = 3.2063×10^{-53} C 3 m 3 J $^{-2}$ = 8.641×10^{-33} esu). $\beta_{\mu} = 3/5 (\beta_{aa} + \beta_{bb} + \beta_{cc})/\|\boldsymbol{\beta}\|$ with $\beta_i = 1/3[(\beta_{iaa} + \beta_{iba} + \beta_{aai}) + (\beta_{iab} + \beta_{bib} + \beta_{bbi}) + (\beta_{iac} + \beta_{cic} + \beta_{cci})]$ and $\|\boldsymbol{\beta}\|$ the norm of the dipole moment.

A. Dipole moment

$\underline{\mu}$	μ_a	μ_b	μ_c
monomer	7.7	1.7	0.2
dimer	14.3	0.0	0.4
unit cell	28.0	1.8	0.8

B. Polarizability

$\alpha(0;0)$						$\alpha(-\omega;\omega)$ ($\hbar\omega = 1.16$ eV)					
	α_{aa}	α_{ab}	α_{ba}	α_{ac}	α_{bc}	α_{aa}	α_{ab}	α_{ba}	α_{ac}	α_{bc}	α_{cc}
monomer	131	8	90	-2	-1	19	139	9	93	-2	-1
dimer	230	0	163	-3	-1	40	241	0	168	-4	-1
unit cell	457	28	356	7	0	79	472	30	367	7	0

		C. First hyperpolarizability									
$\beta(0;0,0)$		β_{aaa}	β_{abb}	β_{acc}	β_{bbh}	β_{baa}	β_{bcc}	β_{caa}	β_{ccb}	β_{abc}	$\beta_{a\bar{a}}$
monomer		1346	-43	-14	-184	504	4	2	-62	-6	-18
dimer		1733	-62	-31	-2	-3	0	5	-91	-9	-7
unit cell		3227	-118	-60	-73	171	3	8	-179	-21	-16
$\beta(\omega;\omega,0) (\hbar\omega = 1.16 \text{ eV})$										$\beta(-2\omega;\omega,\omega) (\hbar\omega = 1.16 \text{ eV})$	
monomer		1599	-50	-14	-202	591	5	2	-70	-6	-20
dimer		2026	-68	-31	-2	3	0	5	-102	-10	-9
unit cell		3760	-127	-59	-72	210	4	9	-199	-23	-18
monomer		2484	-143	-10	-256	882	7	2	-82	-5	-32
dimer		3008	-172	-25	-2	-3	0	5	-120	-9	-13
unit cell		5534	-337	-47	-62	378	5	9	-235	-21	-29

dipole moment, the static and dynamic polarizabilities as well as first hyperpolarizabilities for the monomer, dimer and unit cell evaluated at the HF/AM1 or TDHF/AM1 level of approximation. Except the case of the monomer, the component along the a crystallographic axis of μ and β is, at least, one order of magnitude larger than the others. Furthermore, the formation of the dimer induces strong changes for the a -axis component of all the properties whereas the building of the unit cell modifies less these properties (with respect to the dimer). In most cases, the response of the unit cell tends to be two times the response of the dimer. This building block can therefore be compared to a one-dimensional chromophore.

4.1 Effect of the Crystal Packing on the Linear and Nonlinear Optical Properties

We have calculated at the TDHF/AM1 level of approximation several properties of different clusters of MNA, *i.e.* the static and dynamic polarizabilities and the static, dc-Pockels, and second harmonic generation first hyperpolarizabilities, as well as the dipole moment. Only the non negligible components were analyzed, *i.e.* the component of the dipole moment along the a crystallographic axis, the three diagonal components of the polarizability tensor and β_{aaa} . In order to represent the effects of the intermolecular interactions on the cluster properties, a packing ratio has been defined. For instance, in the case of the first hyperpolarizability of a cluster containing N MNA dimers the packing ratio reads as :

$$R = \frac{\beta^{\text{cluster}}}{N\beta^{\text{dimer}}} \quad (10)$$

Since $R = 1$ in the ideal additive approximation, R informs about the deviation with respect to it. If the intermolecular interactions enhance the total response of the cluster, the ratio will be larger than unity whereas if the response diminishes, the ratio will be smaller than unity.

For the growing clusters, the (1,4) dimer is always considered as the principal building block when determining packing ratios. Three types of clusters were considered: i) clusters built from unit cells, such as the one containing the unit cells A and B (Fig. 1). This type of cluster will be further referred to as two-dimensional (2D) clusters; ii) mono-dimensional (1D) clusters built from dimers along each crystallographic axes. For such clusters extending along the a - and c -axis, odd or even numbers of dimers were studied whereas along the b -axis, only even numbers of stacked dimers were

considered. Otherwise, in the later case, the configuration of the unit cells would lead to large off-diagonal β tensor components; and iii) one three-dimensional cluster of 8 unit cells, *i.e.* containing two unit cells along each crystallographic axis. The 1D clusters were used to get extrapolated values for the packing ratios of infinite 1D systems whereas the 2D and the 3D clusters, to assess the validity of a simple multiplicative scheme [15]. In this work, 2D clusters containing up to six unit cells, three along one crystallographic axis and two along another, 1D clusters containing up to ten dimers along one axis were studied.

4.1.1 Dipole Moment

For the different types of clusters the ground state dipole moments were evaluated. They are reported together with the packing ratios in Tables IV and V for 1D and 2D clusters, respectively. For 1D clusters, Table IV shows the evolution of these properties with the number N of dimers along each of the three crystallographic axes. For 2D arrays, we present first the clusters containing two MNA unit cells, then the three, four and six unit cell clusters, and finally the eight unit cell 3D cluster. The denomination of the clusters is based on the axis along which they are built. For instance, a cluster containing two unit cells along the a -axis is denoted $2a$. Similarly, the notation $3a-2b$ for a cluster built along two axes means three unit cells along a and two unit cells along b . The eight unit cell clusters will be denoted $2a-2b-2c$. For clusters built along two crystallographic axes, the estimated packing ratios were determined by multiplying the packing ratios associated with each of the axes along which the clusters extend. For instance, $R_{2a-2c}^{\text{est}} = R_{2a} \times R_{2c}$. Similarly, for the 3D cluster, $R_{2a-2b-2c}^{\text{est}} = R_{2a} \times R_{2b} \times R_{2c}$.

A first conclusion is that the packing effect differs as a function of the growing axes. For the " a " clusters, *i.e.* $2a$, $3a$, $4a$, and the 1D arrays of dimers along a , the ratio is always larger than unity, which means that the packing enhances the dipole moment of the dimer in the cluster compared to the isolated dimer (1,4). This effect was expected because the dipole moment of the dimer presents its dominant component along this axis (Table III). Indeed, by relying on simple electrostatic interactions which constitute the major packing effects, the electric field induced by the dipole moments of the (1,4) dimer acts "constructively" on the dipole moment of the (5,8) dimer (Fig. 4). On the other hand, when considering the two other cases (with stacking along the b and c axes), the effect is reverse (the induced field acts "destructively") and the ratios are smaller than unity. Moreover, as a result of the distances

between the dimers in the crystal and their relative position, the decrease is more important for clusters built along the *c* axis then along the *b* axis. These modifications of the ground state properties when going from isolated to condensed phases are often referred to as the **dressing effect**. It is also interesting to note that the packing effect is enhanced when the cluster grows because of cooperative dressing effects of all the building blocks and because of the long-range character of the electrostatic interactions. This clearly appears when looking at the evolution in 1D clusters of the packing ratios with the array size (Table IV).

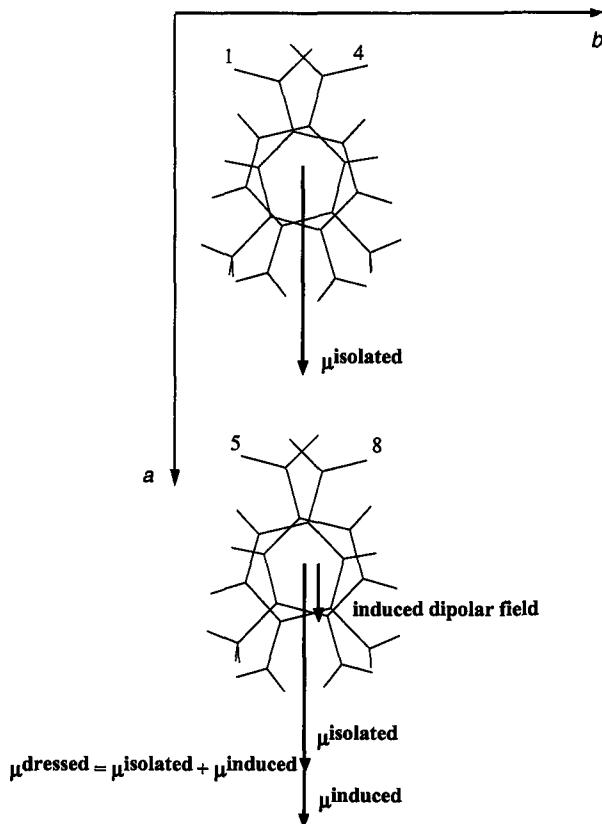


Fig. 4. Schematic representation of the dressing effect from the (1,4) dimer upon the (5,8) dimer.

Concerning the clusters built along two or three axes, the effect is intermediate in comparison with the cases described above for each axis separately. This justifies, although only in a qualitative way, the assessment of the multiplicative scheme where two-dimensional packing ratios are estimated from the product of the corresponding one-dimensional packing ratios. For instance, the packing effect for the $2a\text{-}2b$ cluster containing four unit cells is a combination of the packing effect for the $2a$ cluster and the packing effect for the $2b$ cluster. Effectively, $R(2a\text{-}2b) = 1.068$ is situated between $R(2a) = 1.119$ and $R(2b) = 0.950$. The estimated ratio by the multiplicative scheme which amounts to $R(2a) \times R(2b) = 1.061$ represents a credible characteristic of the packing effects in the $(2a\text{-}2b)$ cluster. Since the error is always smaller than 2% for the two-dimensional or three-dimensional clusters, the multiplicative scheme appears to be an efficient tool. Therefore, with this "multiplicative scheme", one can expect to predict the dipole moment of a cluster $nx\text{-}my$ or $nx\text{-}my\text{-}lz$ from the response of the clusters nx and my (and lz), and reduce considerably the amount of calculations.

Table IV. AM1 dipole moments and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The values of the dipole moment are given in D. The infinite extrapolated ratio is reported at the bottom of the table.

μ	a-axis		b-axis		c-axis	
	N	μ_a	R	μ_a	R	μ_a
2	33.1	1.159	28.1	0.983	27.0	0.946
3	53.4	1.246	/	/	39.6	0.924
4	74.2	1.299	54.2	0.950	52.1	0.912
5	95.2	1.334	/	/	64.6	0.905
6	116.4	1.359	80.2	0.936	77.1	0.900
7	137.6	1.377	/	/	89.5	0.896
8	158.9	1.391	106.0	0.929	102.0	0.893
9	180.2	1.402	/	/	114.5	0.891
10	*	*	131.9	0.924	126.9	0.889
extrapolated value : 1.42 ± 0.04			0.91 ± 0.01		0.88 ± 0.01	

*: SCF cycle not converged

Table V. AM1 dipole moments and the corresponding calculated and estimated packing ratios for MNA clusters containing from two to eight unit cells. The values of the dipole moment are given in D. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	μ_a	R	R^{est}	error (%)
2 unit cells				
2a	63.9	1.119	/	/
2b	54.2	0.950	/	/
2c	52.5	0.920	/	/
3 unit cells				
3a	103.2	1.205	/	/
3b	80.2	0.936	/	/
3c	76.5	0.893	/	/
4 unit cells				
4a	144.3	1.263	/	/
4b	106.0	0.929	/	/
4c	100.3	0.879	/	/
2a-2b	122.0	1.068	1.062	-0.6
2a-2c	118.7	1.040	1.029	-1.0
2b-2c	99.8	0.875	0.874	-0.1
6 unit cells				
3a-2b	197.5	1.153	1.145	-0.7
3a-2c	193.1	1.127	1.109	-1.6
3b-2a	179.1	0.856	0.861	0.6
3b-2c	146.6	1.046	1.047	0.1
3c-2a	171.6	1.002	0.999	-0.3
3c-2b	144.2	0.842	0.848	0.7
8 unit cells				
2a-2b-2c	221.7	0.971	0.977	0.6

4.1.2 Static and Dynamic Polarizabilities

In the same way, we have calculated the static and dynamic polarizabilities for the MNA clusters described in the previous section. For the dynamic part, $\hbar\omega = 1.16 \text{ eV}$ which corresponds to a wavelength of 1064 nm. The results for the “longitudinal” component of the polarizability tensor (α_{aa}) are reported in Tables VI and VII for 1D arrays, and in Tables VIII and IX for 2D and 3D clusters. Tables X, XI, XII, and XIII list the results for the other diagonal tensor components in 2D and 3D clusters.

With respect to the dipole moment, the R values for α_{aa} are more different from unity showing larger environmental effects upon α_{aa} than upon μ_a . In this case, two types of effects are present : 1) the **dressing** effect, already described in the previous section, which modifies the electronic structure of neighboring molecules, and 2) the **local field** effect which appears when an external electric field is applied on a condensed phase. Each molecule responds to the applied field through a charge reorganization. The dominant effect is generally characterized by an induced dipole moment. To each induced dipole moment, in a similar way as for the dressing effect, corresponds an induced electric field which interferes "constructively" or "destructively" with the external electric field depending on the packing direction (Fig. 5). This is a screening effect which modifies the electric field sampled by each MNA molecule. When considering field effects parallel to the permanent dipole moment (along the *a*-axis), the two effects (dressing and screening) are additive. Another point to notice is the difference between the static and dynamic cases: the packing effects are stronger for the dynamic polarizability than for the static response. For instance, the packing ratios for clusters extending along the *a*-axis are larger than for the static polarizability. Moreover, the multiplicative scheme remains suitable for estimating the packing ratios of 2D and 3D clusters.

For the two other diagonal tensor components of the static and dynamic polarizabilities, packing along a given axis leads to an increase of the diagonal tensor component associated with this axis whereas for the other tensor components, R is either smaller or larger than unity. This is due to the fact that the dressing and local field effects can compensate each other : the directionality of the dressing effect is, to a good approximation, driven by the dipole axis (*a*-axis) whereas the local field effects depend upon the orientation of the external field. In addition, the dressing effects on α_{bb} and, more particularly, on α_{cc} are small because μ_b and μ_c as well as α_{ab} and α_{ac} are small. Moreover, the local field effects are smaller on α_{cc} than α_{bb} because $\alpha_{cc} < \alpha_{bb}$. Another important – and related - difference is the poor accuracy of the estimated packing ratios for the α_{bb} component.

Table VI. TDHF/AM1 static polarizabilities (in a.u.) and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The infinite extrapolated ratios are reported at the bottom of the table.

static <i>N</i>	<i>a</i> -axis		<i>b</i> -axis		<i>c</i> -axis	
	α_{aa}	R	α_{bb}	R	α_{cc}	R
2	564	1.228	451	0.981	419	0.913
3	936	1.359	/	/	607	0.881
4	1325	1.442	870	0.947	793	0.863
5	1722	1.500	/	/	979	0.853
6	2123	1.541	1284	0.932	1165	0.846
7	2526	1.572	/	/	1351	0.841
8	2931	1.596	1698	0.924	1537	0.837
9	3337	1.615	/	/	1723	0.834
10	*	*	2111	0.919	1908	0.831
extrapolated value :		1.66 ± 0.04		0.91 ± 0.01		0.82 ± 0.01

*: SCF cycle not converged

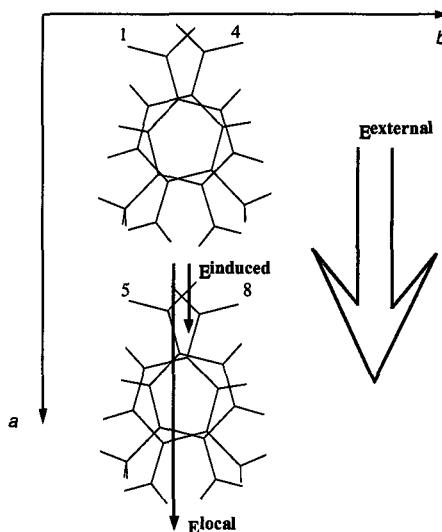


Fig. 5. Schematic representation of the local field effect associated with the induced dipole moment on the (1,4) dimer upon the (5,8) dimer.

Table VII. TDHF/AM1 dynamic polarizabilities (in a.u.) ($\hbar\omega = 1.16 \text{ eV}$) and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The infinite extrapolated ratios are reported at the bottom of the table.

dynamic	<i>a</i> -axis		<i>b</i> -axis		<i>c</i> -axis	
<i>N</i>	α_{aa}	R	α_{aa}	R	α_{aa}	R
2	599	1.244	472	0.980	438	0.909
3	1003	1.388	/	/	632	0.875
4	1427	1.482	909	0.944	826	0.857
5	1862	1.546	/	/	1019	0.846
6	2302	1.593	1342	0.928	1212	0.838
7	2745	1.628	/	/	1405	0.833
8	3190	1.656	1773	0.920	1598	0.829
9	3636	1.677	/	/	1790	0.826
10	*	*	2204	0.915	1983	0.823
extrapolated value :		1.73± 0.06		0.91± 0.01		0.82± 0.01

*: SCF cycle not converged

Table VIII. AM1 static polarizabilities (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{aa} (static)	R(static)	R ^{est} (static)	error (%)
2 unit cells				
2a	1084	1.180	/	/
2b	870	0.947	/	/
2c	814	0.886	/	/
3 unit cells				
3a	1801	1.307	/	/
3b	1284	0.932	/	/
3c	1170	0.849	/	/
4 unit cells				
4a	2565	1.396	/	/
4b	1698	0.924	/	/
4c	1523	0.829	/	/
2a-2b	2058	1.120	1.117	-0.3
2a-2c	1947	1.060	1.045	-1.4
2b-2c	1545	0.841	0.839	-0.3
6 unit cells				
3a-2b	3415	1.239	1.237	-0.1
3a-2c	3252	1.180	1.158	-1.9
3b-2a	3014	1.094	1.100	0.6

3b-2c	2265	0.822	0.826	0.5
3c-2a	2781	1.009	1.002	-0.7
3c-2b	2201	0.799	0.804	0.6
8 unit cells				
2a-2b-2c	3618	0.985	0.990	0.5

Table IX. TDHF/AM1 dynamic ($\hbar\omega = 1.16 \text{ eV}$) polarizabilities (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{aa} (dyn.)	R (dyn.)	$R^{\text{est.}}$ (dyn.)	error (%)
2 unit cells				
2a	1148	1.191	/	/
2b	909.2	0.944	/	/
2c	848.4	0.881	/	/
3 unit cells				
3a	1922	1.330	/	/
3b	1342	0.928	/	/
3c	1217	0.842	/	/
4 unit cells				
4a	2752	1.428	/	/
4b	1773	0.920	/	/
4c	1583	0.822	/	/
2a-2b	2172	1.127	1.124	-0.3
2a-2c	2049	1.063	1.049	-1.3
2b-2c	1606	0.834	0.831	-0.3
6 unit cells				
3a-2b	3628	1.255	1.255	-0.1
3a-2c	3444	1.192	1.171	-1.7
3b-2a	3176	1.099	1.106	0.7
3b-2c	2353	0.814	0.817	0.4
3c-2a	2918	1.010	1.003	-0.4
3c-2b	2283	0.790	0.795	0.7
8 unit cells				
2a-2b-2c	3793	0.984	0.990	0.6

Table X. AM1 static polarizability components along the *b*-axis (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{bb} (static)	R(static)	R ^{est} (static)	error (%)
2 unit cells				
2a	693	1.060	/	/
2b	768	1.175	/	/
2c	665	1.017	/	/
3 unit cells				
3a	1022	1.042	/	/
3b	1190	1.214	/	/
3c	968	0.987	/	/
4 unit cells				
4a	1349	1.032	/	/
4b	1615	1.236	/	/
4c	1269	0.971	/	/
2a-2b	1541	1.179	1.245	5.6
2a-2c	1268	0.970	1.078	11.1
2b-2c	1451	1.110	1.195	7.6
6 unit cells				
3a-2b	2291	1.169	1.224	4.8
3a-2c	1851	0.944	1.060	12.3
3b-2a	2268	1.157	1.235	6.7
3b-2c	2428	1.238	1.286	3.9
3c-2a	1827	0.932	1.046	12.3
3c-2b	2118	1.080	1.160	7.3
8 unit cells				
2a-2b-2c	2861	1.095	1.266	15.7

Table XI. TDHF/AM1 dynamic ($\hbar\omega = 1.16 \text{ eV}$) polarizability components along the *b*-axis (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{bb} (dyn.)	R(dyn.)	R ^{est} (dyn.)	error (%)
2 unit cells				
2a	714	1.060	/	/
2b	794	1.178	/	/
2c	684	1.016	/	/
3 unit cells				

3a	1053	1.043	/	/
3b	1231	1.219	/	/
3c	995	0.985	/	/
4 unit cells				
4a	1390	1.032	/	/
4b	1673	1.242	/	/
4c	1304	0.968	/	/
2a-2b	1593	1.183	1.250	5.7
2a-2c	1304	0.968	1.077	11.3
2b-2c	1496	1.111	1.197	7.8
6 unit cells				
3a-2b	2368	1.172	1.229	4.8
3a-2c	1902	0.942	1.059	12.5
3b-2a	2343	1.160	1.238	6.8
3b-2c	2514	1.244	1.293	3.9
3c-2a	1876	0.929	1.045	12.5
3c-2b	2183	1.081	1.161	7.5
8 unit cells				
2a-2b-2c	2949	1.095	1.269	15.9

Table XII. AM1 static polarizability components along the *c*-axis (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{cc} (static)	R(static)	R ^{est} (static)	error (%)
2 unit cells				
2a	155	0.972	/	/
2b	157	0.981	/	/
2c	163	1.021	/	/
3 unit cells				
3a	231	0.964	/	/
3b	234	0.977	/	/
3c	248	1.035	/	/
4 unit cells				
4a	307	0.960	/	/
4b	312	0.976	/	/
4c	334	1.044	/	/
2a-2b	306	0.958	0.954	-0.5
2a-2c	321	1.004	0.992	-1.2
2b-2c	323	1.012	1.001	-1.1
6 unit cells				

3a-2b	454	0.948	0.946	-0.2
3a-2c	478	0.996	0.984	-1.2
3b-2a	483	1.008	0.997	-1.1
3b-2c	457	0.953	0.950	-0.3
3c-2a	490	1.022	1.006	-1.5
3c-2b	493	1.028	1.016	-1.2
8 unit cells				
2a-2b-2c	634	0.991	0.973	-1.8

Table XIII. TDHF/AM1 dynamic ($\hbar\omega = 1.16 \text{ eV}$) polarizability components along the *c*-axis (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

	α_{cc} (dyn.)	R(dyn.)	$R^{est.}$ (dyn.)	error (%)
2 unit cells				
2a	158	0.972	/	/
2b	159	0.981	/	/
2c	165	1.021	/	/
3 unit cells				
3a	234	0.964	/	/
3b	238	0.977	/	/
3c	252	1.036	/	/
4 unit cells				
4a	311	0.960	/	/
4b	316	0.975	/	/
4c	339	1.044	/	/
2a-2b	311	0.958	0.953	-0.5
2a-2c	326	1.005	0.992	-1.2
2b-2c	328	1.012	1.001	-1.1
6 unit cells				
3a-2b	461	0.947	0.945	-0.2
3a-2c	484	0.996	0.984	-1.2
3b-2a	490	1.008	0.998	-1.1
3b-2c	463	0.952	0.950	-0.3
3c-2a	497	1.022	1.007	-1.5
3c-2b	500	1.029	1.016	-1.3

8 unit cells				
2a-2b-2c	643	0.991	0.973	-1.8

4.1.3 Static, dc-Pockels and Second Harmonic Generation First

Hyperpolarizabilities

In the same way as in the two previous sections, the first hyperpolarizabilities associated to three different nonlinear effects were calculated: the static response and two dynamic effects, the dc-Pockels response characterized by $\beta(-\omega; \omega, 0)$ and the second harmonic generation response, $\beta(-2\omega; \omega, \omega)$. Again, in the dynamic case $\hbar\omega = 1.16$ eV. The results are listed in Tables XIV, XV, and XVI for the 1D clusters, and in Tables XVII, XVIII, and XIX for the 2D and 3D clusters. Similarly to the dipole moment case, only the major tensor component, β_{aaa} , is considered.

The first observation concerns the amplitude of the packing ratios which are much different from those of μ_a and α_{aa} . Indeed, in the static limit the packing ratio for β_{aaa} ranges for the 1D arrays from 0.52 (10 dimers along c) to 4.73 (9 dimers along a) and for 2D clusters from 0.45 ($3c-2b$) to 2.82 ($4a$) whereas for dc-Pockels they go, for the same clusters, from 0.51 to 5.14 for 1D systems and from 0.44 to 2.98 for 2D clusters. They further increase or decrease in the case of the second harmonic generation : for 1D systems, 0.49 for 10 dimers along c and 6.81 for 9 dimers along a , while for 2D systems they go from 0.34 for ($3c-2a$) to 3.50 for ($4a$). This can be explained. On the one hand, Ref. 61 shows that the dressing effects become larger and larger for higher-order properties whereas, on the other hand, the local field effects are simply squared when going from linear to quadratic responses.

The second observation deals with the packing ratios and their estimation using the multiplicative scheme. In many cases, the error remains reasonable (smaller than 5 %) but gets larger in several instances: for dynamic processes with several optical frequencies as well as, in the static limit, for the ($3b-2a$), ($3c-2a$) and ($2a-2b-2c$) clusters. In these cases, the multiplicative scheme overestimates the packing ratios. In addition one has encountered convergence problems in the TDHF procedure for the ($3a-2b$), ($3a-2c$) and ($2a-2b-2c$) clusters for SHG process. Several reasons can be given to explain the limitations of the multiplicative scheme and the TDHF convergence problems. The TDHF scheme is known to work well provided the frequency of the generated light is below the resonance associated with electronic excitations. As it will be shown later, packing along the a -axis creates low-energy excited states which can make difficult or impossible the TDHF calculations when

$\omega_1 + \omega_2$ is not small enough. Moreover, in larger clusters, other physical phenomena like inter-molecular charge-transfer excited states, can contribute to the first hyperpolarizability. Beyond these problems and limitations, the multiplicative scheme has also proved to be efficient in another work [51] treating the sum-frequency generation first hyperpolarizabilities of 2D MNA layers. In that work, the largest errors on the estimated packing ratios are smaller than 4%.

Table XIV. TDHF/AM1 static first hyperpolarizability (in a.u.) and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The infinite extrapolated ratios are reported at the bottom of the Table.

<i>static</i>	<i>a</i> -axis		<i>b</i> -axis		<i>c</i> -axis	
<i>N</i>	β_{aaa}	R	β_{aaa}	R	β_{aaa}	R
2	6318	1.823	3227	0.931	2549	0.735
3	13573	2.611	/	/	3367	0.648
4	22370	3.227	5643	0.814	4179	0.603
5	32010	3.694	/	/	4988	0.576
6	42117	4.051	7997	0.769	5794	0.557
7	52499	4.328	/	/	6600	0.544
8	63042	4.547	10332	0.745	7404	0.534
9	73688	4.725	/	/	8208	0.526
10	*	*	12658	0.730	9011	0.520
extrapolated value :		5.1 ± 0.2	0.72 ± 0.01		0.51 ± 0.01	

*: SCF cycle not converged

Table XV. TDHF/AM1 dc-Pockels first hyperpolarizability (in a.u.) ($\hbar\omega = 1.16$ eV) and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The infinite extrapolated ratios are reported at the bottom of the Table.

<i>dc-Pockels</i>	<i>a</i> -axis		<i>b</i> -axis		<i>c</i> -axis	
<i>N</i>	β_{aaa}	R	β_{aaa}	R	β_{aaa}	R
2	7596	1.875	3760	0.928	2954	0.729
3	16645	2.739	/	/	3888	0.640
4	27778	3.428	6546	0.808	4815	0.594
5	40084	3.957	/	/	5739	0.567
6	53051	4.364	9259	0.762	6661	0.548
7	66406	4.682	/	/	7581	0.535
8	79993	4.935	11950	0.737	8500	0.524
9	93729	5.140	/	/	9418	0.517

10	*	*	14630	0.722	10335	0.510
extrapolated value :	5.7 ± 0.3		0.71 ± 0.01		0.50 ± 0.01	

*: SCF cycle not converged

Table XVI. TDHF/AM1 SHG first hyperpolarizability (in a.u.) ($\hbar\omega = 1.16 \text{ eV}$) and the corresponding packing ratios for MNA 1D clusters containing from two to ten dimers along the three crystallographic axes. The infinite extrapolated ratios are reported at the bottom of the Table.

<i>SHG</i>	<i>a</i> -axis		<i>b</i> -axis		<i>c</i> -axis	
<i>N</i>	β_{aaa}	R	β_{aaa}	R	β_{aaa}	R
2	12275	2.040	5533	0.920	4287	0.713
3	28689	3.179	/	/	5588	0.619
4	50001	4.156	9505	0.790	6882	0.572
5	74350	4.944	/	/	8171	0.543
6	100527	5.570	13376	0.741	9457	0.524
7	127818	6.070	/	/	10741	0.510
8	155801	6.474	17216	0.715	12023	0.500
9	184229	6.805	/	/	13304	0.491
10	*	*	21040	0.700	14585	0.485
extrapolated value :	7.9 ± 0.4		0.68 ± 0.01		0.47 ± 0.01	

*: SCF cycle not converged

Table XVII. TDHF/AM1 static first hyperpolarizabilities (in a. u.) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

static	β_{aaa}	R	$R^{\text{estimated}}$	error (%)
2 unit cells				
2a	10716	1.546	/	/
2b	5643	0.814	/	/
2c	4550	0.656	/	/
3 unit cells				
3a	23123	2.224	/	/
3b	7997	0.769	/	/
3c	5872	0.565	/	/
4 unit cells				
4a	39156	2.824	/	/
4b	10332	0.745	/	/
4c	7181	0.518	/	/
2a-2b	17242	1.244	1.258	1.2

2a-2c	13915	1.004	1.015	1.1
2b-2c	7491	0.540	0.534	-1.1
6 unit cells				
3a-2b	37129	1.785	1.810	1.4
3a-2c	30957	1.489	1.460	-1.9
3b-2a	23324	1.122	1.189	6.0
3b-2c	10337	0.497	0.505	1.6
3c-2a	16477	0.792	0.873	10.2
3c-2b	9355	0.450	0.460	2.2
8 unit cells				
2a-2b-2c	18993	0.685	0.826	20.6

Table XVIII. TDHF/AM1 dc-Pockels first hyperpolarizabilities (in a. u.) ($\hbar\omega = 1.16 \text{ eV}$) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

dc-Pockels	β_{aaa}	R	$R^{\text{estimated}}$	error (%)
2 unit cells				
2a	12777	1.577	/	/
2b	6546	0.808	/	/
2c	5253	0.648	/	/
3 unit cells				
3a	28099	2.312	/	/
3b	9259	0.762	/	/
3c	6752	0.555	/	/
4 unit cells				
4a	48212	2.975	/	/
4b	11950	0.737	/	/
4c	8238	0.508	/	/
2a-2b	20303	1.253	1.274	1.7
2a-2c	16082	0.992	1.022	3.0
2b-2c	8599	0.531	0.524	-1.3
6 unit cells				
3a-2b	44606	1.835	1.867	1.8
3a-2c	36947	1.520	1.498	-1.4
3b-2a	27259	1.121	1.201	7.1
3b-2c	11841	0.487	0.494	1.4
3c-2a	18722	0.770	0.876	13.7
3c-2b	10693	0.440	0.449	2.0
8 unit cells				
2a-2b-2c	20596	0.635	0.826	30.0

Table XIX. TDHF/AM1 SHG first hyperpolarizabilities (in a. u.) ($\hbar\omega = 1.16$ eV) and the corresponding calculated and estimated packing ratios for MNA clusters. The error made on the estimated ratios by using the multiplicative scheme is given in the last column.

SHG	β_{aaa}	R	R ^{estimated}	error (%)
2 unit cells				
2a	20081	1.669	/	/
2b	9505	0.790	/	/
2c	7541	0.627	/	/
3 unit cells				
3a	46984	2.603	/	/
3b	13376	0.741	/	/
3c	9591	0.531	/	/
4 unit cells				
4a	84231	3.500	/	/
4b	17216	0.715	/	/
4c	11631	0.483	/	/
2a-2b	30310	1.260	1.318	4.7
2a-2c	22425	0.932	1.046	12.3
2b-2c	12156	0.505	0.495	-2.0
6 unit cells				
3a-2b	*	*	2.056	*
3a-2c	*	*	1.632	*
3b-2a	38813	1.075	1.237	15.0
3b-2c	16643	0.461	0.465	0.7
3c-2a	12151	0.337	0.887	163.5
3c-2b	14950	0.414	0.420	1.4
8 unit cells				
2a-2b-2c	*	*	0.827	*

*: TDHF cycle not converged

4.2 Effect of the Crystal Packing on the Electronic Excitations

The excited states were studied at the ZINDO level of approximation. The MNA being a donor/acceptor system, the main excited states are charge-transfer states where electronic charges move from the amino to the nitro groups. For the ground as well as the excited states the charge distributions are analyzed by using the Mulliken population scheme. One of our objectives was to determine what occurs to these charge-transfer excited states upon stacking

MNA molecules. Another objective was to exemplify the link between the excited state properties (excitation energies, dipole transition strengths or oscillator strengths, and excited state dipole moments) and the polarizabilities and first hyperpolarizabilities (Eqs. (4) and (5)). In particular, for the diagonal components and $\omega_1 = \omega_2 = \omega$, using the two-state approximation ($m = n = e$) [62], the relationships to spectroscopic quantities become more obvious and simpler:

$$\alpha_{aa}(-\omega; \omega) = \frac{2\Delta E_{0e} |\langle 0|\mu_a|e \rangle|^2}{\Delta E_{0e}^2 - \omega^2} \quad (11)$$

$$\beta_{aaa}(-2\omega; \omega, \omega) = \frac{6|\langle 0|\mu_a|e \rangle|^2 \Delta \mu_{e0}^a \Delta E_{e0}^2}{(\Delta E_{0e}^2 - \omega^2)(\Delta E_{0e}^2 - 4\omega^2)} \quad (12)$$

Where $\langle 0|\mu_a|e \rangle$ is the a^{th} cartesian component of the transition dipole moment between states 0 and e and $\Delta \mu_{e0}^a = \mu_e^a - \mu_0^a$ is the change in the dipole moment (a -component) upon excitation. The oscillator strength for excited state e, given by

$$f_{0e} = \frac{2}{3} \mu_{0e}^2 \Delta E_{0e}, \quad (13)$$

is proportional to the transition probability and was used to simulate the absorption spectra with a FWHM of 0.4 eV. Thus, large polarizabilities are associated with large μ_{0e} and small ΔE_{0e} values while small $\Delta \mu_{0e}$ leads to small first hyperpolarizabilities. The analysis concentrated on these spectroscopic quantities which define the two-state contributions to the diagonal α and β tensor components because they are dominant.

4.2.1 Excited States of MNA Molecule, Dimer and Unit Cell

The simulated absorption spectra for the monomer, dimer and unit cell are displayed in Fig. 6. All single excitations (up to a maximum of 100×100) were included in the ZINDO treatment. In the monomer, the typical intramolecular charge-transfer excited state corresponds to an excitation energy of 3.48 eV and a transfer of 0.6 electron from the amino group and the phenyl ring to the nitro moiety. It is associated with a large change in dipole moment and it dominates the low-energy region of the absorption spectrum.

Therefore, this low-energy intramolecular charge-transfer excited state provides the dominant contribution to β . Another broad absorption band is located around $\Delta E = 6.0$ eV and is related to excited states characterized by small transfers of charge or by transfers going in the reverse direction

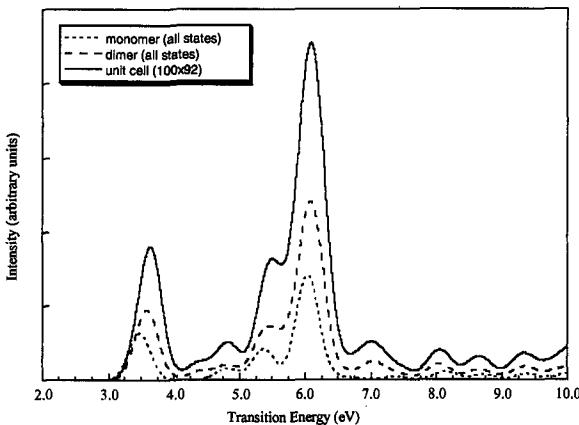


Fig. 6. Simulated ZINDO electronic absorption spectra of the monomer, dimer and unit cell.

The intermolecular interactions lead to splittings of the excited state energy levels: one transition in the monomer gives rise to two transitions in the dimer, or four transitions in the unit cell. For the dominant low-energy intramolecular charge-transfer excited state of MNA, the packing is associated with a hypsochromatic shift. Indeed, in the dimer, the lower state ($\Delta E = 3.40$ eV) is characterized by a small dipole moment transition while the upper state ($\Delta E = 3.60$ eV) presents a strong absorption. As discussed in Ref. 16, the excited state wavefunctions of the dimer are developed over excitations involving MOs delocalized on the two monomers, and thus these two states contain both intra- and inter-molecular contributions. Moreover, for these two transitions ($\Delta E = 3.40$ eV, 3.60 eV), the amount of intermolecular charge-transfer is small. The calculated spectra also evidence several transitions associated with large intermolecular charge-transfer. However, the corresponding μ_{0e} are very small so that these states are not visible in Fig. 6, and their contributions to β are negligible. This is the case of the excitation at $\Delta E = 2.60$ eV associated with a transfer of 0.26 electron between the two

molecules whereas for higher-energy excitations ($\Delta E \geq 6.0$ eV) the intermolecular charge-transfer can attain almost one electron.

In the unit cell, there are four low-energy intramolecular charge-transfer excitations ($\Delta E = 3.41, 3.45, 3.59, 3.65$ eV), the highest-energy state presents also the largest μ_{0e} value and is mostly responsible for the first absorption band and the first hyperpolarizability. Like in the dimer, the first excited state ($\Delta E = 2.05$ eV) presents a substantial character of intermolecular charge-transfer and is associated with a negligible μ_{0e} . Moreover, at $\Delta E = 3.89$ eV and at higher excitation energies, there are other transitions having important intermolecular charge-transfer character and small oscillator strength. The $\Delta E = 3.89$ eV transition corresponds to the transfer of almost one electron from the (1,4) to the (2,3) stacked dimer and is characterized by a dipole moment variation almost antiparallel to the unit cell dipole moment.

4.2.2 Excited States of Clusters

The next step was the study of the evolution of the electronic absorption spectra as a function of the number of unit cells in one-dimensional arrays. When stacking unit cells along the *c* axis, the first absorption band ($\Delta E \sim 4$ eV) presents a hypsochromatic shift (Fig. 7). This band contains several intramolecular charge-transfer transitions of which the most intense has the highest ΔE value. Moreover, its relative intensity decreases with respect to the single unit cell case (Table XX). On the other hand, there is almost no shift of the three-peak absorption band at higher energy of which the intensity grows almost linearly with the number of unit cells. The first transition ($\Delta E = 2.12$ eV) keeps its important intermolecular charge-transfer character and a very weak transition dipole moment. The excited states associated with the transfer of charge from the (1,4) to the (2,3) stacked dimers (as well as their analogs in the other unit cells) remain around 3.6 – 3.9 eV while the corresponding μ_{0e} becomes larger than in the case of a single unit cell. However, μ_{0e} is small with respect to the intramolecular charge-transfer band.

When packing along the *b* crystallographic axis the ΔE shift is small and the relative intensities of the absorption peaks remain almost unchanged (Fig. 7, Table XX). In addition to the excitation-induced electron transfer from the (1,4) to (2,3) stacked dimers belonging to the same unit cell ($\Delta E = 3.85$ eV for 2 unit cells along *b*), there also exist transfers between dimers belonging to adjacent unit cells. Again, the corresponding transition dipole moments are very small and most of the first hyperpolarizability can be associated to the intramolecular charge-transfer excitations.

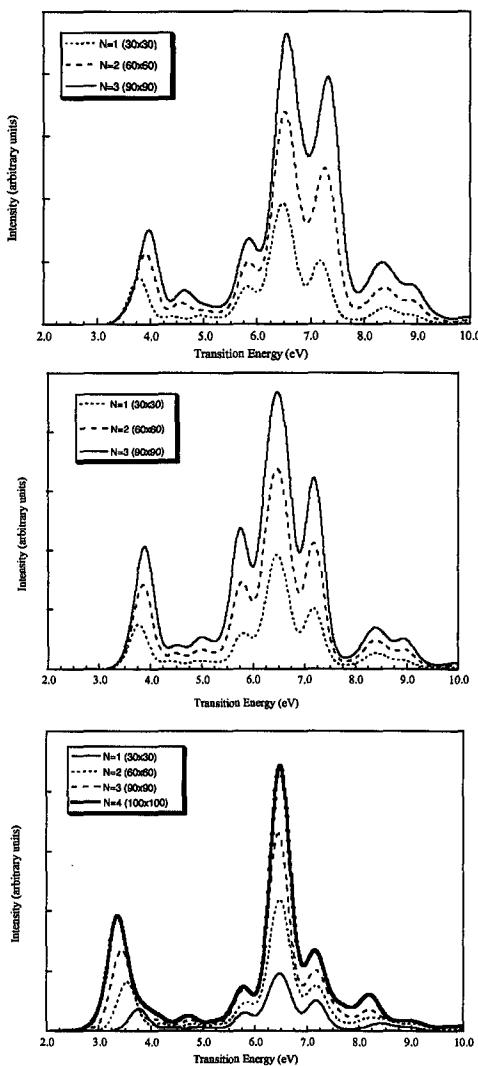


Fig. 7. Simulated ZINDO electronic absorption spectrum of increasingly large arrays of MNA unit cells extending along the *c* (top), *b* (middle) and *a* (bottom) crystallographic axes. *N* is the number of unit cells. Reprinted from M. Guillaume *et al.*, Int. J. Quantum Chem. **90**, 1378 (2002) by permission of John Wiley & Sons, Inc.

Table XX : ZINDO characteristics of the dominant low-energy intramolecular charge-transfer excited state for different one-dimensional arrays of MNA. Relative Intensity (RI) = $f^{\text{cluster}} / (N \times f^{\text{unit cell}})$ with N the number of unit cells in the cluster and f the corresponding oscillator strength. Reprinted from M. Guillaume *et al.*, Int. J. Quantum Chem. **90**, 1378 (2002) by permission of John Wiley & Sons, Inc.

	ΔE_{0e} (eV)	μ_{0e} (D)	$\Delta\mu_{0e}$ (D)	f_{0e}	RI
Unit cell	3.65	3.74	9.33	0.194	1.00
2 unit cells along <i>a</i> (2<i>a</i>)	3.50	5.54	10.29	0.407	1.05
2 unit cells along <i>b</i> (2<i>b</i>)	3.87	4.98	10.32	0.364	0.94
2 unit cells along <i>c</i> (2<i>c</i>)	3.93	4.50	10.08	0.302	0.78

The situation becomes different when stacking unit cells along the *a* crystallographic axis. Indeed, it is associated with a bathochromatic shift of the first absorption band and its relative intensity increases with the number of unit cells. In these aggregates, the intermolecular charge-transfer can take place over much longer distances and the corresponding $\Delta\mu_{0e}$ can be very large. Although their transition dipole moments are still small, these states are therefore expected to play some role in the NLO responses when approaching resonance. In addition to these states, there are a number of intermolecular charge-transfer excited states below the dominant intramolecular band which present non negligible oscillator strengths and energies in the 2.0 – 2.5 eV region and therefore, which contribute slightly to β .

The differences in behavior of ΔE and μ_{0e} for the three directions of packing are consistent with the differences in the evolution of the effective linear and nonlinear optical properties when building one-dimensional arrays along the *a*, *b*, or *c* crystallographic axes. In addition, the position of the first intense absorption band between 3.5 and 4.0 eV is consistent with the yellow color of the MNA crystals [17].

Finally, in two-dimensional structures for which the multiplicative scheme leads (in some cases) to important errors in estimating β , the effect of packing is intermediate between the effects of the constitutive one-dimensional clusters. Indeed, in the case of the 2*a*-2*b* cluster, the first dominant peak ($\Delta E = 3.75$ eV) is at a slightly higher energy than in the (A + B) cluster ($\Delta E = 3.50$ eV) but at a slightly lower energy than in the (A + C) cluster ($\Delta E = 3.87$ eV) (Fig. 8). The most striking feature is, however, the appearance of small peaks at ΔE between 1.0 and 2.5 eV which correspond to stabilized intermolecular charge-transfer states presenting less negligible transition dipole moments. It is therefore not excluded to believe that these transitions are at the origin of the

low-energy absorption features detected for a 50- μm plate of MNA [17]. Moreover, the cluster (2a-3c) presenting an intermolecular charge-transfer peak at the lowest excitation energy in Fig. 8 is also associated with the largest error of the multiplicative scheme for predicting β (Tables XVII, XVIII, and XIX). From these calculations and their analysis it was concluded [16] that, due to packing effects, new intermolecular charge-transfer excited states appear in large clusters and can contribute to the first hyperpolarizability and they affect the nice predictive character of the multiplicative scheme. In addition, due to their small excitation energy, they can prevent the TDHF iterative procedure of converging for frequency combinations approaching the resonance region.

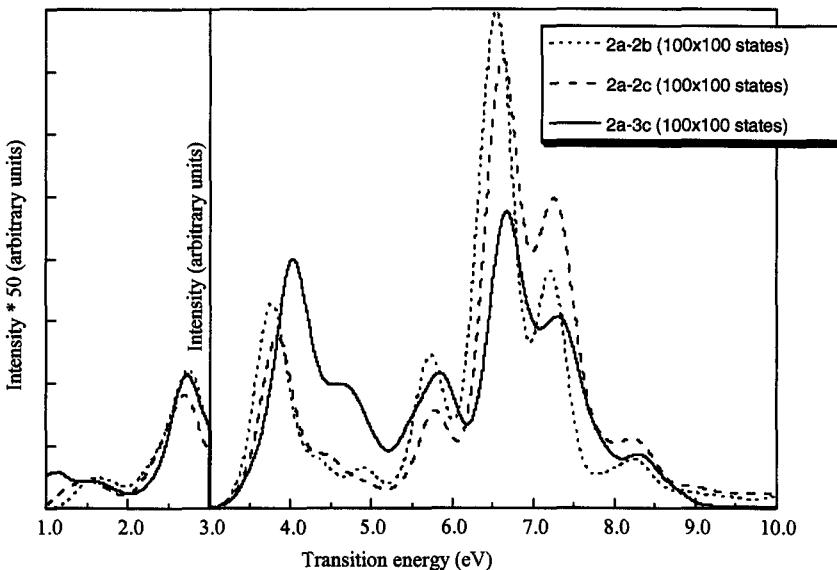


Fig. 8. Simulated absorption spectra for the 2a-2b, 2a-2c, and 2a-3c clusters. The intensity has been multiplied by a factor of fifty for the 0-3 eV window.

Nevertheless, a word of caution is necessary : i) different levels of theory were used for computing the excitation energies (ZINDO) and (hyper)polarizabilities (TDHF/AM1) so that the excitation energies do not exactly match the poles of the latter; ii) in these polar systems where excited states exhibiting multiple charge-transfer can be stabilized by electrostatic interactions with the surroundings, it appears of interest to consider explicitly the effects of including double, triple, ... excitations on the optical and

electronic properties of MNA clusters. As already discussed in Section III, this will require to go beyond the ZINDO scheme. One possibility would consist in performing high-level *ab initio* calculations but, due to the enormous computational costs, alternative solutions like the effective fragment approach proposed in Ref. 63 extended to include local configuration interactions will have to be considered to study the collective electronic and optical properties of the crystal of MNA.

4.3 Extrapolation to Infinite Clusters and Discussion

Following the procedure designed in Ref. 64, the packing ratios for different properties of one-dimensional arrays were extrapolated to the infinite array size. They are listed at the bottom of the Tables of Section IV-1 [Table IV for μ_a , Table VI for $\alpha_{aa}(0;0)$, Table VII for $\alpha_a(-\omega;\omega)$, Table XIV for $\beta_{aaa}(0;0,0)$, Table XV for $\beta_{aaa}(-\omega;\omega,0)$, and Table XVI for $\beta_{aaa}(-2\omega;\omega,\omega)$]. For instance, in the case of the *a*-axis, this leads to a ratio of 7.9 for the SHG response, which almost represents a factor of eight between the response of an infinite number of isolated dimers (non-interacting) and the same number of dimers interacting in a one-dimensional crystal. For the arrays extending along the *b*-axis, the ratio amounts to 0.68 for SHG whereas for the third direction, it attains 0.47. The ratios for the other properties are closer to unity, although still quite different. In addition, it can be noticed that the extrapolated ratios for α_{aa} and β_{aaa} are almost independent of the frequency for packings along the *b* and *c* axis, the maximum dispersion being achieved for $\beta(\text{SHG}) : R_b^\infty(\text{SHG}) = 0.68$ while $R_b^\infty(\text{static}) = 0.72$ and $R_c^\infty(\text{SHG}) = 0.47$ while $R_c^\infty(\text{static}) = 0.51$. On the other hand, when packing along the *a* axis, the effects due to frequency dispersion are substantial. These facts further evidence the strong anisotropy of the MNA crystal.

Though the multiplicative scheme has been criticized due to its limitations in several cases, it was used to estimate the packing ratios of the full three-dimensional, crystal, lattice:

$$R^{\text{crystal}} = R_a^\infty \times R_b^\infty \times R_c^\infty \quad (14)$$

Using these crystal packing ratios, effective properties of the MNA dimer in the MNA crystal can be computed (Table XXI). For example, for the first hyperpolarizability, it reads as:

$$\beta_{\text{dimer}}^{\text{eff.}} = \beta_{\text{dimer}}^{\text{isolated}} \times R^{\text{crystal}} \quad (15)$$

Table XXI : Properties (P) of the isolated dimer, extrapolated crystal packing ratio, $R_{\text{crystal}}^{\text{crystal}}$, for the different properties and NLO processes as well as effective properties of the dimer in the crystal obtained from AM1 and TDHF/AM1 calculations. The dipole moments are given in D whereas the other quantities in a.u.

	$P_{\text{dimer}}^{\text{isolated}}$	$R_{\text{crystal}}^{\text{crystal}} = R_a^{\infty} \times R_b^{\infty} \times R_c^{\infty}$	$P_{\text{dimer}}^{\text{effective}}$
μ_a	14.3	1.14 ± 0.06	16.3 ± 0.9
$\alpha_{aa}(0;0)$	230	1.24 ± 0.06	285 ± 14
$\alpha_{aa}(-\omega;\omega)$	241	1.29 ± 0.07	311 ± 17
$\beta_{aaa}(0;0,0)$	1733	1.87 ± 0.14	3241 ± 243
$\beta_{aaa}(-\omega;\omega,0)$	2026	2.02 ± 0.18	4093 ± 365
$\beta_{aaa}(-2\omega;\omega,\omega)$	3008	2.52 ± 0.22	7580 ± 657

The macroscopic values can therefore be evaluated after accounting for the crystal density. Within one of the conventions of the MKS system [65], these relations read:

$$P_a^{(0)} = \frac{2 \times \mu_a^{\text{eff}}}{(\epsilon_0 V^{\text{unit cell}})} \quad (16)$$

$$\chi_{aa}^{(1)} = \frac{2 \times \alpha_{aa}^{\text{eff}}}{(\epsilon_0 V^{\text{unit cell}})} \quad (17)$$

$$\chi_{aaa}^{(2)} = \frac{2 \times \beta_{aaa}^{\text{eff}}}{(\epsilon_0 V^{\text{unit cell}})} \quad (18)$$

Following Ref. 18, the unit cell volume is chosen to be 722.4 \AA^3 whereas, similarly to Ref. 15, no $1/2$ factor is used to account for the *possible* difference of convention between the expressions of the field-dependent dipole moment and polarization. The macroscopic quantities are given in the 2nd column of Table XXII. In the MKS system [65], the refractive indices are expressed from the first-order susceptibilities as :

$$n^2 = 1 + \chi^{(1)}, \quad (19)$$

For $\lambda = \infty$ and $\lambda = 1064 \text{ nm}$, n_a amounts to 1.56 ± 0.03 and 1.62 ± 0.03 , respectively. This can be compared to the values calculated from the Sellmeier expression: 1.68 and 1.76. Therefore, theory slightly underestimates the refractive index whereas for SHG second-order susceptibilities, theory is off

by a factor of 4 [From Ref. 18, $\chi_{aaa}^{(2)}(-2\omega, \omega, \omega) \approx \chi_{111}^{(2)}(-2\omega, \omega, \omega) = 300 \pm 75 \text{ pm/V}$ for $\hbar\omega = 1.16 \text{ eV}$]. Differences between theory and experiment can have several origins. As described in Section III, comparisons with high-level *ab initio* calculations demonstrate that the TDHF/AM1 scheme underestimates the longitudinal first hyperpolarizability of MNA monomer and dimer. A correction factor of 1.3, corresponding to the ratio between the MP2 and CPHF/AM1 β_{aaa} values, enables to improve, although weakly, the agreement between theory and experiment (third column of Table XXII). For α_{aa} , this ratio amounts to 0.99 and has almost no effect on the predicted linear susceptibilities or refractive indices. Another correction arises from difference between the external field, E_0 , and the Maxwell or macroscopic field, E_{Maxwell} . Indeed, the calculations on the clusters correspond to taking the field-derivatives with respect to the external field whereas the macroscopic linear and nonlinear susceptibilities are taken with respect to the Maxwell field [65]. Although efficient procedures have been proposed within the interaction scheme approach [66], the E_0/E_{Maxwell} ratio is difficult to evaluate because it is sensitive to the shape of the cluster [for a parallelepiped-shape cluster inside the plates of a condenser, it is equal to the relative refractive index of the material, ϵ_r , whereas for a spherical cluster, it is given by $(2 + \epsilon_r)/3$]. Its $\chi^{(1)}$ -dependence (through ϵ_r) makes this ratio even more difficult to estimate. However, since the estimates of the refractive indices are in good agreement with experiment, this correction was assumed to be negligible for our clusters.

In summary, despite the approximations made to relate the microscopic properties of the MNA molecule to the macroscopic value of the crystal, a qualitative accuracy has been reached since the refractive index is estimated with an error of 8 % while for $\chi_{aaa}^{(2)}(-2\omega; \omega, \omega)$ the order of magnitude is reproduced.

These "multiplicative scheme" estimates of the macroscopic properties were then compared to those obtained by using the oriented gas approximation. For reasons explained in Section II, the dimer was considered as the building unit. In such a case, the SHG second-order nonlinear susceptibility reads:

$$\chi_{aaa}^{(2)}(-2\omega, \omega, \omega) = N f_a^{2\omega} f_a^\omega f_a^\omega \beta_{aaa}^{\text{dimer}}(-2\omega, \omega, \omega) \quad (20)$$

Where N is the dimer density ($1/V^{\text{dimer}}$), $\beta_{aaa}^{\text{dimer}}$ is the first hyperpolarizability of the isolated dimer calculated at the TDHF/AM1 level (Table III) and f_a is the local field factor along the a -axis. Analogous expressions are used for the static and electro-optical responses. Two versions

Table XXII : Crystal susceptibilities calculated within different schemes. MS = supermolecule approach with the multiplicative scheme approximation. OG = oriented gas approximation.

	MS/AM1	MS/AM1 + MP2	OG/f = 1	OG/f = Lorentz
P_a	$1.70 \pm 0.09 \text{ } 10^{10} \text{ V/m}$		$1.57 \pm 0.08 \text{ } 10^{10} \text{ V/m}$	
$\chi_{aa}^{(1)}(0;0)$	1.45 ± 0.09		1.44 ± 0.09	
$\chi_{aa}^{(1)}(-\omega; \omega)$	1.62 ± 0.10		1.60 ± 0.10	
$\chi_{aaa}^{(2)}(0;0,0)$	$32.6 \pm 0.2 \text{ pm/V}$		$42.4 \pm 0.3 \text{ pm/V}$	17.5 pm/V
$\chi_{aaa}^{(2)}(-\omega; \omega, 0)$	$41.1 \pm 0.3 \text{ pm/V}$		$53.4 \pm 0.4 \text{ pm/V}$	20.4 pm/V
$\chi_{aaa}^{(2)}(-2\omega; \omega, \omega)$	$76.1 \pm 0.3 \text{ pm/V}$		$98.9 \pm 0.4 \text{ pm/V}$	30.3 pm/V
				212 pm/V

of the oriented gas approximation were used. In the first case, the local field factors were assumed to be unity. This corresponds to an additive scheme based on the dimer properties without taking into account the dressing and local field effects. In the second case, the anisotropic Lorentz spherical cavity expression,

$$f_a^{2\omega,\omega,0} = \frac{(n_a^{2\omega,\omega,0})^2 + 2}{3}, \quad (21)$$

was used with the refractive indices evaluated by using the Sellmeier coefficients (Table I). For $\lambda = 1064$ nm and $\lambda = 532$ nm, $n_a = 1.763$ and $n_a = 2.291$, respectively. The results are listed in Table XXII. The factor of four to seven between these two sets of $\chi^{(2)}$ results emphasizes the importance of taking into account, even in a simple way, the influence of environment on the macroscopic susceptibilities. In this case, the estimated value for $\chi_{aaa}^{(2)}(-2\omega;\omega,\omega)$ is now in rather good agreement with experiment. The agreement gets even better if using the $\beta_{aaa}(-2\omega;\omega,\omega)$ values corrected for electron correlation effects : $\chi_{aaa}^{(2)}(-2\omega;\omega,\omega) = 1.3 \times 212 \text{ pm/V} = 276 \text{ pm/V}$. These values can also be compared to the oriented gas estimate of Ref. 19 which amounts to $558 \pm 120 \text{ pm/V}$. The latter was evaluated using the EFISH-extracted molecular first hyperpolarizabilities.

5. Conclusions and Outlook

The supermolecule approach has been employed to characterize the linear and nonlinear optical properties of the 2-methyl-4-nitroaniline (MNA) crystal. These properties have been evaluated at the TDHF/AM1 level to assess the crystal packing effects as a function of the size and shape of the MNA clusters. When building one-dimensional arrays of MNA unit cells along the a crystallographic axis (axis parallel to the permanent crystal polarization), the effective properties or the crystal packing ratios strongly increase whereas stacking unit cells along the b and c axes leads to a reduction of these quantities. These effects increase with the order of the electric response (small for the dipole moment, larger for the polarizability and still larger for the first hyperpolarizability) and are magnified in the case of frequency-dependent phenomena. These "direction" differences have been explained by using simple electrostatic models. A simple multiplicative scheme has been shown to be reliable for many properties as well as for many clusters. It enables to

determine the crystal packing ratios for two- and three-dimensional arrays from the crystal packing ratios of their constitutive one-dimensional arrays. However, this scheme fails when the frequencies get closer to the resonant regime.

In order to further rationalize these effects, the excitation energies of the MNA clusters have been determined at the ZINDO level of approximation and the corresponding UV/visible absorption spectra have been simulated. For all clusters, the spectra display an intense low-energy absorption band in the 3.0-4.0 eV region that is associated with a strong intramolecular charge transfer and therefore that gives rise to important contributions to the first hyperpolarizability. When stacking unit cells along the *b* and *c* crystallographic axes, this absorption band exhibits a hypsochromatic shift whereas for the third crystal direction, the shift is bathochromic. Moreover, the transition dipole moment is larger for stacking along the *a* axis than for stacking along its perpendicular directions. These behaviors are consistent with the differences in the evolution of the effective linear and nonlinear optical properties when building one-dimensional arrays along the *a*, *b*, or *c* crystallographic axes. In addition to these intramolecular charge-transfer excited states which dominate the UV/visible absorption spectra, a number of intermolecular charge-transfer states have been detected. In the small MNA aggregates as well as for clusters extending along the *b* and *c* crystallographic axes, these excited states possess very small transition dipole moments although the corresponding energies of excitation can be as small as 2 eV. On the other hand, when stacking MNA unit cells along the *a* axis (or simultaneously along two axes), these intermolecular charge-transfer states acquire non negligible transition dipole moments and their excitation energies decrease. As a consequence, it can be understood that i) although small, their contributions to β increase, especially for dynamic processes and that ii) in the near resonant region, they will lead to convergence problems of the TDHF scheme which lacks damping effects. Such intermolecular charge-transfer excited states are stabilized by electrostatic interactions because the change in dipole moment is almost antiparallel to the dipole moment of the cluster.

Using the extrapolated crystal packing ratios for one-dimensional arrays of MNA, the global packing ratios were calculated within the multiplicative scheme and then used to evaluate the effective dipole moment, polarizability and first hyperpolarizability of the MNA dimer in the crystal. Finally, the corresponding macroscopic quantities were evaluated. Although electron correlation effects have been indirectly accounted for, our estimates still underestimate the available $\chi^{(2)}$ experimental values. On the other hand, the

accuracy on the refractive index is acceptable. The simple oriented gas approximation was also shown to predict accurate crystal second-order susceptibilities provided gas phase properties are used in combination with the anisotropic Lorentz spherical cavity expression of the local field factors. In addition to the low-accuracy of the experimental measurements, the difference between theory and experiment can be traced back to the limitations of the semi-empirical Hamiltonians, of the supermolecule approach, and of the multiplicative scheme. Indeed, the domain of validity of the latter still needs to be defined by considering other molecular crystals and/or by deriving the conditions where Eq. (14) is satisfied. As pointed out in several places, other questions concern the validity of the semi-empirical schemes (TDHF/AM1 and ZINDO) for describing the effects of the surroundings upon the intra- and intermolecular charge-transfers. Conceptual questions arise with respect to the large dipole moment of the clusters and the presence of surface charges to compensate for it as well as with respect to the determination of the ratio between the external and the Maxwell electric fields.

At the end of most of the theoretical investigations reported here, another X-ray characterization of the MNA crystal has been published [67]. Although the structure is more precise, the cell parameters (nonstandard setting *Ia* of the space group *Cc*) as well as the hydrogen-bonding geometry are fully consistent with those used in this study.

Acknowledgements

M.G. thanks the "Fonds pour la Formation et la Recherche dans l'Industrie et dans l'Agriculture" (FRIA) for financial support as well as the European community for his doctoral Marie Curie Fellowship. B.C. thanks the Belgian National Fund for Scientific Research for his Senior Research Associate position. This study has benefited from a Tournesol scientific cooperation established and supported by the Centre National de la Recherche Scientifique (CNRS), the Belgian National Fund for Scientific Research (FNRS), and the Commissariat Général aux Relations Internationales (CGRI) de la Communauté Wallonie-Bruxelles. The authors thank Profs. J.M. André, A. Fritsch, B. Kirtman, D. Liotard as well as Dr. E. Botek for fruitful discussions. The calculations were performed thanks to computing time made available by the SiMoA (Simulation et Modélisation en Aquitaine, France) as well as on the Pentium III cluster of the CTA lab for which the authors acknowledge the financial support of the FNRS.

References

- [1] D.M. Bishop and P. Norman, in *Handbook of Advanced Electronic and Photonic Materials and Devices*, edited by H.S. Nalwa (Academic Press, New York, 2001), Vol 9, Chap.1, p. 1.
- [2] B. Champagne and B. Kirtman, in *Handbook of Advanced Electronic and Photonic Materials and Devices*, edited by H.S. Nalwa (Academic Press, New York, 2001), Vol. 9, Chap. 2, p. 63.
- [3] B. Champagne and D.M. Bishop, *Adv. Chem. Phys.*, **126**, 41 (2003).
- [4] C.W. Dirk, R.J. Twieg, and G. Wagnière, *J. Am. Chem. Soc.* **108**, 5387 (1986).
- [5] T. Yasukawa, T. Kimura, and M. Uda, *Chem. Phys. Lett.* **169**, 159 (1990).
- [6] J. Perez and M. Dupuis, *J. Phys. Chem.* **95**, 6525 (1991).
- [7] S. Di Bella, M.A. Ratner, and T.J. Marks, *J. Am. Chem. Soc.* **114**, 5842 (1992).
- [8] J.D. Auspurger and C.E. Dykstra, *Int. J. Quantum Chem.* **43**, 135 (1992).
- [9] T. Hamada, *J. Phys. Chem.* **100**, 8777 (1996).
- [10] S.Y. Chen and H.A. Kurtz, *J. Molec. Struct. (THEOCHEM)* **388**, 79 (1996).
- [11] V. Moliner, P. Escribano, and E. Peris, *New J. Chem.* **22**, 387 (1998).
- [12] O. Xie and C.W. Dirk, *J. Phys. Chem. B* **102**, 9378 (1998).
- [13] B. Champagne and B. Kirtman, *J. Chem. Phys.* **109**, 6450 (1998).
- [14] Y. Okuno, S. Yokoyama, and S. Mashiko, *J. Phys. Chem. B* **105**, 2163 (2001).
- [15] F. Castet and B. Champagne, *J. Phys. Chem. A* **105**, 1366 (2001).
- [16] M. Guillaume, E. Botek, B. Champagne, F. Castet, and L. Ducasse, *Int. J. Quantum Chem.* **90**, 1378 (2002).
- [17] B.F. Levine, C.G. Bethea, C.D. Thurmond, R.T. Lynch, and J.L. Bernstein, *J. Appl. Phys.* **50**, 2523 (1979).
- [18] G.F. Lipscomb, A.F. Garito, and N.S. Narang, *J. Chem. Phys.* **75**, 1509 (1981).
- [19] Ch. Bosshard, K. Sutter, Ph. Prêtre, J. Hulliger, M. Flosheimer, P. Kaatz, and P. Günter, *Organic Nonlinear Optical Materials*, (Gordon and Breach, Basel, 1995).
- [20] D.A. Roberts, *IEEE J. Quantum Electr. QE* **28**, 2507 (1992); K. Hagimoto, C. Takahashi, *Nonlinear Opt.* **13**, 3 (1995).
- [21] R. Morita, N. Ogasawara, S. Umagaki, and R. Ito, *Jpn. J. Appl. Phys.* **26**, 1711 (1987).
- [22] P.N. Butcher and D. Cotter, *The Elements of Nonlinear Optics* (Cambridge University Press, Cambridge, 1990).
- [23] A.D. Buckingham, *Adv. Chem. Phys.* **12**, 107 (1967).
- [24] D.M. Bishop, *Rev. Mod. Phys.* **62**, 343 (1990).
- [25] D.M. Bishop, B. Kirtman, and B. Champagne, *J. Chem. Phys.* **107**, 5780 (1997).
- [26] D.M. Bishop, *Adv. Chem. Phys.* **104**, 1 (1998).
- [27] in addition to Refs. 1, 2, 26, and 27, see in particular, J.M. Luis, M. Duran, B. Champagne, and B. Kirtman, *J. Chem. Phys.* **113**, 5203 (2000); M. Torrent-Sucarrat, M. Sola, M. Duran, J.M. Luis, and B. Kirtman, *J. Chem. Phys.* **116**, 5363 (2002); O. Quinet, B. Kirtman, and B. Champagne, *J. Chem. Phys.* **118**, 505 (2003).
- [28] B. Kirtman and B. Champagne, *Int. Rev. Phys. Chem.* **16**, 389 (1997).

- [29]B. Champagne, « *Élaboration de Méthodes de Chimie Quantique pour l'Évaluation des Hyperpolarisabilités Vibrационnelles – Conséquences pour l'Optique non Linéaire* », (PUN, Namur, 2001).
- [30]B. Champagne, M. Spassova, J.B. Jadin, and B. Kirtman, *J. Chem. Phys.* **116**, 3935 (2002).
- [31]J.M. Luis, B. Champagne, and B. Kirtman, *Int. J. Quantum Chem.* **80**, 471 (2000); O. Quinet, B. Champagne, and B. Kirtman, *J. Molec. Struct. (THEOCHEM)*, submitted.
- [32]B.J. Orr and J.F. Ward, *Mol. Phys.* **20**, 513 (1971); D.M. Bishop, *J. Chem. Phys.* **100**, 6535 (1994).
- [33]J.E. Ridley and M.C. Zerner, *Theoret. Chim. Acta (Berl.)* **32**, 111 (1973); J.E. Ridley and M.C. Zerner, *Theoret. Chim. Acta (Berl.)* **42**, 223 (1976); A.D. Bacon and M.C. Zerner, *Theoret. Chim. Acta (Berl.)* **53**, 21 (1979); M.C. Zerner, G.H. Loew, R.F. Kirchner, and U.T. Mueller-Westerhoff, *J. Am. Chem. Soc.* **102**, 589 (1980); M.C. Zerner, in *Reviews of Computational Chemistry*, edited by K.B. Lipkowitz and D.B. Boyd (VCH, New York, 1991), Vol. 1, 313; C.H. Martin and M.C. Zerner, in *Inorganic Electronic Structure and Spectroscopy*, edited by E.I. Solomon and A.B.P. Lever, (Wiley, New York, 1999), Vol. 1, 555.
- [34]V.J. Docherty, D. Pugh, and J.O. Morley, *J. Chem. Soc., Faraday Trans. 2* **81**, 1179 (1985).
- [35]see for example, in the case of second-order NLO responses, F. Meyers, J.L. Brédas, and J. Zyss, *J. Am. Chem. Soc.* **114**, 2914 (1992); C. Dehu, F. Meyers, E. Hendrickx, K. Clays, A. Persoons, S.R. Marder, and J.L. Brédas, *J. Am. Chem. Soc.* **117**, 10127 (1995).
- [36]see for example, J.O. Morley, *J. Chem. Soc. Faraday Trans.* **87**, 3009 (1991); J.O. Morley, *J. Phys. Chem.* **99**, 10166 (1995).
- [37]Y. Luo, H. Ågren, P. Jørgensen, and K.V. Mikkelsen, *Adv. Quantum Chem.* **26**, 165 (1995); P.B. Rozyczko and R.J. Bartlett, *J. Chem. Phys.* **107**, 10823 (1997); O. Christiansen, P. Jørgensen, and C. Hättig, *Int. J. Quantum Chem.* **68**, 1 (1998).
- [38]H.D. Cohen and C.C.J. Roothaan, *J. Chem. Phys.* **43**, S34 (1965).
- [39]H. Sekino and R.J. Bartlett, *J. Chem. Phys.* **85**, 976 (1986); S.P. Karna and M. Dupuis, *J. Comp. Chem.* **12**, 487 (1991).
- [40]C. Hattig and B.A. Hess, *J. Chem. Phys.* **105**, 9948 (1996).
- [41]F. Aiga and R. Itoh, *Chem. Phys. Lett.* **251**, 372 (1996).
- [42]J.E. Rice and N.C. Handy, *Int. J. Quantum Chem.* **43**, 91 (1992).
- [43]C.P. de Melo and R.J. Silbey, *J. Chem. Phys. Lett.* **140**, 537 (1987); M. Nakano, I. Shigemoto, S. Yamada, and K. Yamaguchi, *J. Chem. Phys.* **103**, 4175 (1995).
- [44]J. Guan, P. Duffy, J.T. Carter, D.P. Chong, K.C. Casida, M.E. Casida, and M. Wrinn, *J. Chem. Phys.* **98**, 4753 (1993); S.J.A. van Gisbergen, J.G. Snijders, and E.J. Baerends, *J. Chem. Phys.* **109**, 10657 (1998); H.H. Heinze, F. Della Sala, and A. Görling, *J. Chem. Phys.* **116**, 9624 (2002).
- [45]R. van Leeuwen and E.J. Baerends, *Phys. Rev. A* **49**, 2421 (1994); D.J. Tozer and N.C. Handy, *J. Chem. Phys.* **108**, 10180 (1998); P.R.T. Schipper, O.V. Gritsenko, S.J.A. van Gisbergen, E.J. Baerends, *J. Chem. Phys.*, **112**, 1344 (2000); H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, *J. Chem. Phys.* **115**, 3540 (2001).

- [46]B. Champagne, E.A. Perpète, S.J.A; van Gisbergen, J.G. Snijders, E.J. Baerends, C. Soubra-Ghaoui, K.A. Robins, and B. Kirtman, *J. Chem. Phys.* **109**, 10489 (1998); *erratum* **110**, 11664 (1999).
- [47]B. Champagne, E.A. Perpète, D. Jacquemin, S.J.A; van Gisbergen, E.J. Baerends, C. Soubra-Ghaoui, K.A. Robins, and B. Kirtman, *J. Phys. Chem. A* **104**, 4755 (2000).
- [48]S.J.A. van Gisbergen, P.R.T. Schipper, O.V. Gritsenko, E.J. Baerends, J.G. Snijders, B. Champagne, and B. Kirtman, *Phys. Rev. Lett.* **83**, 694 (1999); M. Grüning, O.V. Gritsenko, and E.J. Baerends, *J. Chem. Phys.* **116**, 6435 (2002).
- [49]M. van Faassen, P.L. de Boeij, R. van Leeuwen, J.A. Berger, and J.G. Snijders, *J. Chem. Phys.* **118**, 1044 (2003).
- [50]M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *J. Am. Chem. Soc.* **107**, 3902 (1985).
- [51]E. Botek and B. Champagne, *Appl. Phys. B* **74**, 627 (2002).
- [52]F. Sim, S. Chin, M. Dupuis, and J. E. Rice, *J. Phys. Chem.* **97**, 1158 (1993).
- [53]MOPAC2000, © Fujitsu Limited, 1999; J.J.P. Stewart, Quantum Chemistry Program Exchange, n°. 455.
- [54]AMPAC 6.0, Semichem, 7204 Mullen, Shawnee, KS 66216, 1997.
- [55]H.A. Kurtz, J.J.P. Stewart, and K.M. Dieter, *J. Comput. Chem.* **11**, 82 (1990).
- [56]J.E. Del Bene, J.D. Watts, and R.J. Bartlett, *J. Chem. Phys.* **106**, 6051 (1997).
- [57]D.J. Tozer, R.D. Amos, N.C. Handy, B.J. Roos, and L. Serrano-Andrés, *Mol. Phys.* **97**, 859 (1999).
- [58]S. Canuto, K. Coutinho, and M.C. Zerner, *J. Chem. Phys.* **112**, 7293 (2000).
- [59]MOS-F (Semi-empirical Molecular Orbital Package for Spectroscopy) V4, © Fujitsu Limited, 1998.
- [60]C. Hättig and A. Köhn, *J. Chem. Phys.* **117**, 6939 (2002).
- [61]B. Kirtman, B. Champagne, and D.M. Bishop, *J. Am. Chem. Soc.* **122**, 8007 (2000).
- [62]J.L. Oudar and D.S. Chemla, *J. Chem. Phys.* **66**, 2664 (1977).
- [63]G. Brunaud, F. Castet, A. Fritsch, M. Kreissler, and L. Ducasse, *J. Phys. Chem. B* **105**, 12665 (2001).
- [64]B. Champagne, D. Jacquemin, J. M. André, and B. Kirtman, *J. Chem. Phys. A* **101**, 3158 (1997).
- [65]R.W. Boyd, *Nonlinear Optics*, (Academic Press, San Diego, 1992), Appendix A.
- [66]R.W. Munn, *Mol. Phys.* **64**, 1 (1988).
- [67]G. Ferguson, C. Glidewell, J.N. Low, J.M.S. Skakle, J.L. Wardell, *Acta Cryst. C* **57**, 315 (2001).

Chapter 3: Ab Initio Calculations of the Intermolecular Nuclear Spin-Spin Coupling Constants

Magdalena Pecul and Joanna Sadlej

Department of Chemistry, University of Warsaw

Pasteura 1, 02-093 Warsaw

Poland

The ab initio calculations of the hydrogen-bond- and van der Waals interactions-transmitted indirect spin-spin coupling constants are discussed, starting with a short introduction to the theory of the spin-spin coupling constants and presentation of the techniques used in their calculations. The available numerical results for simple complexes with traditional hydrogen bond, charged or acid-base systems with strong hydrogen bond, models of the biochemical systems: nucleic acids and proteins and, finally, weak van der Waals complexes are reviewed. The most important practical and theoretical consequences of these results are discussed.

1. Introduction

The characterization of the intermolecular interactions is one of the basic goals of theoretical and experimental chemistry. The hydrogen bond, one of the strongest intermolecular interactions, is particularly intensively investigated on account of its importance in biochemistry^{1,2}. Lately these efforts have intensified because of the possible role of strong hydrogen bonds in enzymatic catalysis^{3,4,5,6,7}. NMR spectroscopy has been extensively used for hydrogen bond detection and characterization, especially for biomolecules, providing such parameters as isotropic and anisotropic chemical shifts^{1,8,9,10}, isotope shifts⁸, quadrupole coupling constants, and hydrogen-exchange rates (see Refs.^{10,11,12,13} for references). Considerable progress has been also made in employing indirect intramolecular nuclear spin-spin coupling constants, such as $^1J(\text{NH})$, as parameters describing hydrogen bonds^{11,12,13,14,15,16,17,18,19}.

All the spectroscopic methods mentioned above rely on detection and characterization of hydrogen bonds or other van der Waals interactions through changes in some properties of monomers (eg. IR frequencies or NMR shieldings). In recent years, new spectroscopic parameters — intermolecular indirect spin-spin coupling constants have been measured — providing an unique direct experimental evidence for hydrogen bond formation^{11,12,13,15,16,17,18,19,20,21,22,23,24}.

The phenomenon of spacial proximity increasing the value of the indirect coupling between magnetic nuclei formally separated by several bonds has been known for a long time^{25,26}. It has been suggested as early as 1961 that in some cases $^{19}\text{F}-^{19}\text{F}$ ²⁶ and $^1\text{H}-^{19}\text{F}$ ²⁵ coupling constants are transmitted by a "through-space" mechanism (not to be confused with the direct coupling) rather than by a "through-bond" mechanism. The empirical correlation between fluorine-fluorine couplings and internuclear distance in difluorocyclophanes has been given in Ref.²⁷. The fluorine coupling constants in fluorine-substituted naphthalenes, exhibiting the effect of enhanced values due to the spacial proximity, have been recently calculated by means of the density functional theory²⁸. The "through-space" coupling constants between nuclei belonging to the same molecule (but separated by several bonds) have been detected also for the other nuclei, for example in ^{113}Cd resonanse¹⁵, where the indirect $^1\text{H}-^{113}\text{Cd}$ couplings mediated through $\text{NH}\cdots\text{S}$ hydrogen bonds have been measured.

The interest in the intermolecular indirect coupling constants has grown dramatically in 1998, when the intermolecular indirect spin-spin coupling constants were measured between nuclei belonging to complementary pairs

of nucleic bases in RNA¹³ and DNA²⁰. What in particular attracted the interest were surprisingly large values of those couplings: ¹⁵N-¹⁵N hydrogen-bond transmitted couplings fall within 6—7 Hz range, while many ¹⁵N-¹⁵N indirect coupling constants between covalently bound nuclei are smaller than that. The analogous hydrogen-bond-transmitted ¹⁵N-¹⁵N couplings measured between histidine side chains in a protein²¹ have even larger values. Also the ¹H-¹⁵N hydrogen-bond transmitted couplings have substantial values: 2–4 Hz²⁰. Soon after that other measurements of the indirect intermolecular coupling in nucleic acids^{13,17}, proteins^{11,12,15,16,21,22,23}, and fluorine-containing clusters^{18,19} followed.

We will employ, after Ref.¹¹, the notation ⁿ*h* *J*(XY) for the indirect intermolecular coupling constant between nuclei X and Y, where the number *n* in superscript denotes the numbers of bounds through which the intermolecular coupling is formally transmitted (including the hydrogen bond). In the cases when the path of the intermolecular coupling between nuclei X and Y is difficult to establish (as in weak van der Waals complexes), we will use the notation ^h *J*(XY) instead of ⁿ*h* *J*(XY).

The NH· · · OC hydrogen-bond-transmitted ³*h* *J*(NC') couplings between backbone chains in proteins range from 0.1 Hz to 0.9 Hz^{11,16} and the sign has been assumed to be negative¹⁶. The other experimentally detected intermolecular coupling constants in biomolecules are ²*h* *J*(HC') couplings in proteins^{22,23}, of the magnitude between 0.4 Hz and 0.6 Hz (sign not determined experimentally). The long-range ⁴*h* *J*(NN) coupling in nucleic acid (G-G-G-G tetrad) has also been measured, even though its absolute value is only 0.14 Hz²⁴. Other types of experimentally detected hydrogen-bond-transmitted coupling constants are ²*h* *J*(PH) (0.5–1.6 Hz) and ³*h* *J*(PN) (approx. 1.7 Hz) couplings in flavoprotein²⁹. The experimental studies of the hydrogen bond-transmitted coupling constants are reviewed in Ref.³⁰.

Much of the interest of the theoretical chemists in the intermolecular coupling constants is connected to the debate on the alleged covalent character of the hydrogen bond^{12,31,32}. The experimental discovery of the hydrogen-bond transmitted intermolecular coupling constants has been treated by some authors as an evidence of the partial covalent character of the hydrogen bond^{12,32}. However, nothing in the nature of indirect spin-spin coupling constants requires this to be the case. In particular, prediction of the non-negligible values of the coupling constants transmitted through weak van der Waals interactions^{31,33,34,35} refutes the hypothesis of the transmission of the indirect coupling being the evidence of covalent character of the interaction. The analysis of the density distribution in the hy-

drogen bond-transmitted spin-spin coupling constants by means of Bader's theory of Atoms in Molecules³⁶ also provides evidence for the lack of covalent character in hydrogen bonds, with an exception of the low-barrier proton-shared ones³¹.

The other source of interest in the intermolecular coupling constants lies in their potential for structure determination. The ${}^3hJ(\text{CN})$ coupling in proteins have been found to correlate well with the chemical shift of the proton forming the hydrogen bond¹¹ and therefore with the strength of the interaction³⁷. The experimental correlation of ${}^3hJ(\text{CN})$ with the hydrogen bond length has also been established¹⁶. Consequently, most of the theoretical modelling of the hydrogen-bond-transmitted couplings in biomolecules is focused on correlating their magnitudes with the structural parameters, in particular with the hydrogen bond length.

The indirect nuclear spin-spin coupling constants are still not calculated routinely, mainly because of the extremely large electron correlation effects involved³⁸. The quantum mechanical calculations of the intermolecular spin-spin coupling constants have started only recently. However, the number of papers reporting these type of calculations is growing rapidly. The reason for this lies both in the potentially wide applicability of these parameters for detection and characterization of hydrogen bonds (and, possibly, other van der Waals interactions) and in the development of computational methods and computer hardware, rendering the spin-spin coupling constants in the sizable systems accessible by means of ab initio and DFT methods. In this situation, a review of the work conducted in this field is due.

The presentation of the numerical results available in the literature is preceded in this review by a short discussion of a theory of the spin-spin coupling and of available computational methods. After that, the calculations of the intermolecular spin-spin couplings reported in the literature are discussed in four subsections. The calculations on the small, model systems, such as water or ammonia dimers, with medium strength hydrogen bond, are considered first. In the next subsection the calculations of the hydrogen-bond-transmitted couplings in the systems with strong hydrogen bond are discussed. The following subsection reviews the works in which hydrogen bonds in the systems of biological importance: nucleic acids and proteins, are modelled. Finally, in the last subsection the calculations of the intermolecular spin-spin coupling constants in the van der Waals complexes bound by intermolecular forces weaker than hydrogen bonds are discussed.

2. Methods of calculation of the spin-spin coupling constants

2.1. Quantum mechanical theory of the spin-spin coupling constants

Before discussing the methods used for ab initio calculations of the spin-spin coupling constants we will shortly present the basic theory of these NMR parameters. The quantum mechanical expressions for the NMR shielding constant and the spin-spin coupling constant were first derived by Ramsey^{39,40}. In this chapter a notation of Ref. ³⁸ is employed.

Effective NMR spin Hamiltonian NMR spectra arise from the absorption of electromagnetic radiation by nuclei with non-zero magnetic moment \mathbf{M}_K , i.e. from radiative transitions between nuclear spin energy levels, which are split in the presence of an external magnetic field \mathbf{B} ^{38,39,40}. The empirical interpretation of NMR spectra consists in finding two types of static parameters: shielding constants σ_K (or chemical shifts δ_K) and spin-spin coupling constants J_{KL} to fit the observed spectrum, using the effective spin Hamiltonian.

The spectral parameters appearing in the effective NMR spin Hamiltonian depend on the electronic structure around the nuclei. The NMR effective spin Hamiltonian has the following form:

$$\hat{H} = - \sum_K \mathbf{B}^T (\mathbf{1} - \sigma_K) \mathbf{M}_K + \frac{1}{2} \sum_{k \neq l, l} \mathbf{M}_K^T (\mathbf{D}_{KL} + \mathbf{K}_{KL}) \mathbf{M}_L \quad (1)$$

\mathbf{M}_K denotes the nuclear magnetic dipole moment operator, obtained by multiplication of the nuclear spin operator \mathbf{I}_K by the magnetogyric factor γ_K .

$$\mathbf{M}_K = \gamma_K \hbar \mathbf{I}_K \quad (2)$$

\mathbf{B} denotes the external magnetic field.

In the hamiltonian 1 the electric quadrupole moments of the nuclei are neglected.

The tensors appearing in Eq. 1 has the following interpretation:

- The NMR shielding constant σ_K describes a modification of the external magnetic field at the nucleus by the presence of electrons.
- The direct (dipolar) nuclear spin-spin coupling constant \mathbf{D}_{KL} represents the classical through-space interaction of the magnetic mo-

ments of nuclei K and L.

$$\mathbf{D}_{KL} = \frac{\mu_0}{4\pi} \frac{3\mathbf{R}_{KL}\mathbf{R}_{KL} - 1\mathbf{R}_{KL}^2}{R_{KL}^5} \quad (3)$$

- The reduced indirect (also sometimes labeled, somewhat misleadingly, "scalar") nuclear spin-spin coupling constant \mathbf{K}_{KL} describes the interaction of the magnetic nuclei, transmitted through the surrounding electrons. \mathbf{K}_{KL} is four orders of magnitude smaller than \mathbf{D}_{KL} .

\mathbf{D}_{KL} , \mathbf{K}_{KL} and $\boldsymbol{\sigma}_K$ are second rank tensors. \mathbf{D}_{KL} , in contrast to \mathbf{K}_{KL} and $\boldsymbol{\sigma}_K$, is a traceless tensor. Hence, the averaged effective spin Hamiltonian of a rapidly tumbling molecule (as in a liquid or in a gas phase) in a uniform magnetic field B_0 parallel to the z axis has the form:

$$\hat{H}_{iso} = - \sum_K (1 - \sigma_K^{iso}) B_0 M_{Kz} + \frac{1}{2} \sum_{L, K \neq L} K_{KL}^{iso} \mathbf{M}_K \cdot \mathbf{M}_L \quad (4)$$

where

$$\sigma_K^{iso} = \frac{1}{3} \text{Tr } \boldsymbol{\sigma}_K = \frac{1}{3} ((\sigma_K)_{xx} + (\sigma_K)_{yy} + (\sigma_K)_{zz}) \quad (5)$$

$$K_{KL}^{iso} = \frac{1}{3} \text{Tr } \mathbf{K}_{KL} = \frac{1}{3} ((K_{KL})_{xx} + (K_{KL})_{yy} + (K_{KL})_{zz}). \quad (6)$$

In experimental practice the spin-spin coupling constant J_{KL} is used instead of the reduced spin-spin coupling constant K_{KL} .

$$J_{KL} = h \frac{\gamma_K \gamma_L}{2\pi 2\pi} K_{KL} \quad (7)$$

NMR parameters as energy derivatives Now we define NMR parameters as energy derivatives using the expression for the molecular electronic Hamiltonian in the presence of the magnetic field.

The energy of the closed-shell molecule perturbed by the external magnetic field and the magnetic field of the nuclei can be expanded in the magnetic induction \mathbf{B} and in the nuclear moments \mathbf{M}_K around zero field and zero magnetic moments.

$$\begin{aligned} E(\mathbf{B}, \mathbf{M}) = E_0 &+ \frac{1}{2} \mathbf{B}^T \mathbf{E}^{(20)} \mathbf{B} + \sum_K \mathbf{B}^T \mathbf{E}_K^{(11)} \mathbf{M}_K \\ &+ \frac{1}{2} \sum_{L, K \neq L} \mathbf{M}_K^T \mathbf{E}_{KL}^{(02)} \mathbf{M}_L \end{aligned} \quad (8)$$

The following notation is employed in Eq. 9:

$$\mathbf{E}_K^{(11)} = \frac{d^2 E(\mathbf{B}, \mathbf{M})}{d\mathbf{B} d\mathbf{M}_K} \Big|_{\mathbf{B}=0, \mathbf{M}=0} \quad (9)$$

$$\mathbf{E}_{KL}^{(02)} = \frac{d^2 E(\mathbf{B}, \mathbf{M})}{d\mathbf{M}_K d\mathbf{M}_L} \Big|_{\mathbf{B}=0, \mathbf{M}=0} \quad (10)$$

$$\mathbf{E}_{KL}^{(20)} = \frac{d^2 E(\mathbf{B}, \mathbf{M})}{d\mathbf{B} d\mathbf{B}} \Big|_{\mathbf{B}=0, \mathbf{M}=0} \quad (11)$$

In the expansion of $E(\mathbf{B}, \mathbf{M})$ in Eq. 9 all terms higher than second-order are omitted which is motivated by small strength of the perturbation. The first-order terms vanish for the closed-shell systems.

Comparison of the expansion Eq. 9 with the effective spin Hamiltonian of Eq. 1 leads to identification of $\mathbf{E}_K^{(11)}$ with the NMR shielding constant

$$\mathbf{E}_K^{(11)} = \sigma_K - 1 \quad (12)$$

and of $\mathbf{E}_{KL}^{(02)}$ with the nuclear spin-spin coupling constant

$$\mathbf{E}_{KL}^{(02)} = \mathbf{K}_{KL} + \mathbf{D}_{KL}. \quad (13)$$

Sum-over-states expression The time-independent second-order properties, such as the nuclear spin-spin coupling constant (see Eq. 13), can be expressed by means of the perturbation theory as

$$\frac{d^2 E(\mathbf{x})}{dx_i dx_j} = \left\langle 0 \left| \frac{d^2 H}{dx_i dx_j} \right| 0 \right\rangle - \sum_{n \neq 0} \frac{\left\langle 0 \left| \frac{dH}{dx_i} \right| n \right\rangle \left\langle n \left| \frac{dH}{dx_j} \right| 0 \right\rangle}{E_n - E_0} \quad (14)$$

where the derivatives are taken at zero perturbation \mathbf{x} (here magnetic moments of the nuclei \mathbf{M}_K). In the case of magnetic properties the first contribution (the expectation value) is called the diamagnetic part and the second sum-over-states contribution is known as the paramagnetic part. The phenomenon of spin-spin coupling was analyzed theoretically in terms of perturbation theory by Ramsey in 1953 ⁴⁰.

Linear response theory expression Alternatively, the spin-spin coupling constant can be expressed using the linear response theory formalism. Let us write the electronic energy of the system perturbed by the nuclear magnetic dipole moments \mathbf{M}_K in the form $E = E(\mathbf{M}_K, \lambda)$, where λ are the variational parameters of the wave function. λ may represent orbital rotation parameters for the SCF wave function, or orbital rotation parameters and coefficients of the configuration interaction expansion for the MCSCF

wave function. For the non-variational wave function ansatz (eg. coupled clusters) the response theory can be formulated analogously, invoking Lagrange's method of undetermined multipliers.

The reduced spin-spin coupling constant may now be calculated as

$$\mathbf{K}_{KL} = \frac{\partial^2 E}{\partial \mathbf{M}_K \partial \mathbf{M}_L} + \frac{\partial^2 E}{\partial \mathbf{M}_K \partial \lambda} \frac{\partial \lambda}{\partial \mathbf{M}_L} - \mathbf{D}_{KL} \quad (15)$$

The derivatives of the variational parameter λ with respect to the perturbations \mathbf{M}_K can be obtained by solving (iteratively) the set of linear equations

$$\frac{\partial^2 E}{\partial \lambda^2} \frac{\partial \lambda}{\partial \mathbf{M}_K} = - \frac{\partial^2 E}{\partial \lambda \partial \mathbf{M}_K} \quad (16)$$

In the calculations of spin-spin coupling constants the number of linear response equations to be solved depends on the number of magnetic nuclei in the system. In computational practice, for N magnetic nuclei N linear response equations should be solved for the Fermi contact terms, $3N$ for the paramagnetic spin-orbit terms and $9N$ for the spin-dipole terms (or $6N$, if a full symmetry of the tensor was used, see Eq. 166 and following in Ref. ³⁸, or Ref. ⁴¹). The diamagnetic spin-orbit terms are obtained as expectation values. Therefore even when most time-consuming SD term is omitted, $4N$ sets of the linear response equations remain to be solved for N magnetic nuclei.

Electronic Hamiltonian of a molecules perturbed by magnetic field The nonrelativistic molecular electronic Hamiltonian of a system with magnetic nuclei in the static magnetic field \mathbf{B} can be written as

$$\begin{aligned} H(\mathbf{B}, \mathbf{M}) = & \frac{1}{2} \sum_i (\nabla_i + \mathbf{A}^{tot}(\mathbf{r}_i))^2 - \sum_i \mathbf{m}_i \cdot \mathbf{B}^{tot}(\mathbf{r}_i) - \sum_{iK} \frac{Z_K}{r_{iK}} \\ & + \frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}} + \frac{1}{2} \sum_{K \neq L} \frac{Z_K Z_L}{R_{KL}} - \sum_K \mathbf{M}_K \cdot \mathbf{B}^{tot}(\mathbf{R}_K) \\ & + \alpha^2 \sum_{K \neq L} \frac{R_{KL}^2 (\mathbf{M}_K \cdot \mathbf{M}_L) - 3 (\mathbf{M}_K \cdot \mathbf{R}_{KL}) (\mathbf{R}_{KL} \cdot \mathbf{M}_L)}{R_{KL}^5} \end{aligned} \quad (17)$$

In Eq. 17 i and j indices are used for electrons, K and L indices are used for nuclei, α is the fine-structure constant $\alpha \approx 1/137$, and the meaning of $\mathbf{A}^{tot}(\mathbf{r}_i)$ and $\mathbf{B}^{tot}(\mathbf{r}_i)$ is explained below. \mathbf{m}_i is the permanent magnetic moment of electron. The other symbols in Eq. 17 have the usual meaning.

$\mathbf{A}^{tot}(\mathbf{r}_i)$ is the vector potential of the total magnetic field, measured at the position of electron i . The curl of $\mathbf{A}^{tot}(\mathbf{r}_i)$ reproduces the magnetic

induction $\mathbf{B}^{tot}(\mathbf{r}_i)$ arising from the external magnetic field and the NMR active nuclei.

$$\mathbf{B}^{tot}(\mathbf{r}_i) = \nabla_i \times \mathbf{A}^{tot}(\mathbf{r}_i) \quad (18)$$

The total magnetic field has two sources: the uniform external magnetic field and magnetic nuclei.

$$\mathbf{A}^{tot}(\mathbf{r}_i) = \frac{\mathbf{B} \times \mathbf{r}_{iO}}{2} + \alpha^2 \sum_K \frac{\mathbf{M}_K \times \mathbf{r}_{iK}}{r_{iK}^3} \quad (19)$$

$$\begin{aligned} \mathbf{B}^{tot}(\mathbf{r}_i) = \mathbf{B} + \alpha^2 \sum_K & \frac{3(\mathbf{M}_K \cdot \mathbf{r}_{iK}) \mathbf{r}_{iK} - r_{iK}^2 \mathbf{M}_K}{r_{iK}^5} \\ & + \frac{8\pi}{3} \alpha^2 \sum_K \delta(\mathbf{r}_{iK}) \mathbf{M}_K \end{aligned} \quad (20)$$

The second term in Eq 21 is the dipole-dipole interaction between the electrons and the nucleus. The third term gives rise to the Fermi contact interaction of the electrons with the nuclei. The vector potential \mathbf{A}^{tot} depends on the choice of the gauge origin \mathbf{O} , although the magnetic induction \mathbf{B}^{tot} does not. This results, for approximate wave functions, in the dependence of the NMR shielding constant (but not of the spin-spin coupling constant) on the gauge origin for \mathbf{A}^{tot} .

The paramagnetic contributions The operators associated with the paramagnetic contributions to the spin-spin coupling are obtained as derivatives of the Hamiltonian 17 with respect to the magnetic moments of the nuclei \mathbf{M} . It leads to the relation:

$$\frac{dH}{d\mathbf{M}_K} = \mathbf{h}_K^{pso} + \mathbf{h}_K^{sd} + \mathbf{h}_K^{fc} \quad (21)$$

The singlet paramagnetic spin-orbit (PSO) operator

$$\mathbf{h}_K^{pso} = \alpha^2 \sum_i \frac{\mathbf{l}_{iK}}{r_{iK}^3}, \quad (22)$$

the triplet Fermi contact (FC) operator

$$\mathbf{h}_K^{fc} = -\frac{8\pi\alpha^2}{3} \sum_i \delta(\mathbf{r}_{iK}) \mathbf{m}_i, \quad (23)$$

and the triplet spin-dipole (SD) operator

$$\mathbf{h}_K^{sd} = \alpha^2 \sum_i \frac{r_{iK}^2 \mathbf{m}_i - 3(\mathbf{m}_i \cdot \mathbf{r}_{iK}) \mathbf{r}_{iK}}{r_{iK}^5} \quad (24)$$

contribute to the spin-spin coupling constant. As a rule, the dominant contribution to the isotropic coupling originates from the FC term. The PSO term is usually second in importance. The SD contribution to the isotropic coupling is generally not very important, but the mixed SD-FC term (vanishing for the isotropic coupling) strongly contributes to the anisotropy of the coupling constant^{42,43,44}.

The diamagnetic contributions The diamagnetic contributions to the spin-spin coupling originate from the operators obtained by double differentiation of the Hamiltonian 17 with respect to the nuclear magnetic moments

$$\frac{d^2 H}{d\mathbf{M}_K d\mathbf{M}_L} = \mathbf{D}_{KL} + \mathbf{h}_{KL}^{dso} \quad (25)$$

The diamagnetic operator \mathbf{h}_{KL}^{dso} has the form

$$\mathbf{h}_{KL}^{dso} = \frac{\alpha^4}{2} \sum_i \frac{(\mathbf{r}_{iK} \cdot \mathbf{r}_{iL}) \mathbf{1} - \mathbf{r}_{iK} \mathbf{r}_{iL}^T}{r_{iK}^3 r_{iL}^3} \quad (26)$$

The diamagnetic spin-orbit (DSO) contribution to the spin-spin coupling constants is usually small, but nonnegligible, especially for the proton-proton coupling constants.

To summarise, there are several physical mechanisms contributing to spin-spin coupling. They can be divided into those originating from the coupling of the nuclear spins through the spins of electrons, i.e. the SD (spin-dipolar) and FC (Fermi-contact) contributions, and from the coupling of the nuclear spins through the orbital motion of electrons, i.e. the DSO and PSO (dia- and paramagnetic spin-orbit) contributions.

2.2. Methods of calculation of the spin-spin coupling constants

In this subsection we will shortly discuss the computational methods used for calculation of the spin-spin coupling constants. Two main approaches available are: ab initio theory from Hartree-Fock (or self-consistent field SCF) technique to its correlated extensions, and density function theory (DFT), where the electron density, instead of the wave function, is the fundamental quantity. The discussion here is limited to the methods actually used for calculation of the intermolecular spin-spin coupling constants, i. e. multiconfigurational self consistent field (MCSCF) theory, coupled cluster (CC) theory and density functional theory (DFT). For example, the second order polarization propagator method (SOPPA) approach^{45,46} is not

considered here since it has not been used, to our knowledge, for the calculations of the intermolecular coupling constants. The basis set requirements for the calculations of the spin-spin coupling constants are briefly discussed. For a more complete survey of the subject the reader is referred to the recent reviews by Helgaker, Jaszuński and Ruud ³⁸ and Fukui ⁴⁷.

Multiconfigurational Self Consistent Field Theory The MCSCF approximation provides a framework for treating the systems for which the SCF method fails, i.e. for those which are not properly described by a single Slater determinant. MCSCF does not treat efficiently the dynamic correlation effects, but it handles well the static correlation (or near-degeneracy) effects. It also performs well in the calculations of the spin-spin coupling constants, where the single reference SCF model breaks down because of its instability to the triplet SD and FC perturbations. The spin-spin coupling constants calculations (all four terms) at the linear response MCSCF level are implemented in the DALTON program ⁴⁸.

There are two basic variations of the MCSCF method: CAS (Complete Active Space) and RAS (Restricted Active Space) techniques. The determinants to be included in the wave function expansion are constructed by exciting electrons from the molecular orbitals that are occupied in the HF wave function to the virtual MOs. In the CAS technique the molecular orbitals are divided into inactive space, consisting of doubly occupied orbitals, active space, where all occupations are allowed and external space, containing unoccupied orbitals. The number of determinants in the CAS wave function grows factorially with the size of the active space. In the RAS technique the orbitals are divided into inactive space (in analogy to CAS), RAS1 space from which only the given maximum number of electrons can be excited (i.e., with a limited number of holes), RAS2 space, analogous to CAS active space, RAS3 space to which only a limited number of electrons can be excited (from RAS1 or RAS2) and the external space containing orbitals which remain unoccupied. The computational experience suggests that for the calculations of the spin-spin coupling constants it is advantageous to use extended RAS3 space, but to limit RAS1 and RAS2 spaces to the orbitals occupied in the HF wave function. In this review we will denote the MCSCF active space by the symbol (number of inactive MOs/number of MOs in RAS1/number of MOs in RAS2/number of MOs in RAS3 number of electrons excited to RAS3).

MCSCF method has an advantage of being easily extendable, at least in principle, to the full configuration iteration limit. Moreover, it can handle smoothly, in contrast to the other methods, open shell or nearly degener-

ate systems. However, the choice of the active space is often problematic, the size-consistency problems arise, and in practice the spin-spin coupling constants calculations cannot be easily performed for molecules containing more than 7-10 second row atoms. In spite of these complications, MCSCF linear response method has been successfully applied in the calculations of hydrogen-bond- and van der Waals-interactions-transmitted spin-spin coupling constants^{33,34,49,50,51}.

A single determinant SCF method can in principle also be applied to calculate the spin-spin coupling constants. However, it gives notoriously very poor results, in some cases rendering even the order of magnitude wrong (see for example Ref.⁴¹). This is connected to the triplet instabilities problem in a single determinant wave function ansatz and makes this basic model of ab initio theory practically useless for calculations of the spin-spin coupling constants.

Coupled cluster theory Coupled cluster theory has been employed for calculations of the spin-spin coupling constants by Perera, Sekino and Bartlett^{52,53}. They have derived the indirect spin-spin coupling constant as a derivative of the coupled cluster singles and doubles (CCSD) energy⁵² (see Eqs 10 and 13). An alternative formulation employs sum-over-states formalism^{52,53} (see Eq. 14). This method is referred to as equation-of-motion coupled cluster singles and doubles (EOM-CCSD) method and is implemented in the development version of AcesII⁵⁴ program. All four terms contributing to the spin-spin coupling constants can be evaluated in the EOM-CCSD approach. Numerous calculations of the hydrogen-bond-transmitted spin-spin coupling constants, especially in the charged systems with strong hydrogen bond, have been carried out by means of this method^{55,56,57,58,59,60,61,62}.

Gauss and Auer⁶³ have recently proposed another scheme of calculating the spin-spin coupling constants, also based on the coupled cluster ground state wave function, but employing the linear response formalism. In their computational method there is a possibility of improving the quality of the calculations by including triple excitations (approximate CC3, CCSD(T) or full CCSDT approach)⁶³. This approach (at the CCSD level) has been utilized for calculation of the spin-spin coupling constants in He₂³³ and in FHF⁻³⁴.

Density Functional Theory The most popular technique for calculations of the spin-spin coupling constants employing the density functional theory is implemented in deMon program by Malkina, Salahub and Malkin^{64,65}. The FC term is calculated there by means of a finite field

method, using a spin-unrestricted approach. The PSO term is computed using the sum-over-states density perturbation theory and spin-restricted approach. The DSO term, depending only on the unperturbed ground-state density, is computed by means of a straightforward numerical integration. The calculation of the SD term is not implemented in deMon, but the FC-SD cross term (contributing to the coupling anisotropy) can be derived using the finite perturbation theory approach. This "mixed" finite-field and sum-over-states implementation of DFT has been used in several calculations of the hydrogen-bond-transmitted coupling constants^{18,66,67}.

Another approach to the calculations of the spin-spin coupling constants by means of DFT has been recently proposed by Helgaker, Watson and Handy⁴¹. All four contributions to the spin-spin coupling constants are derived analytically there, using the linear response formalism. In contrast to the approach of Malkina, Salahub and Malkin^{64,65} and to that of Dickson and Ziegler⁶⁸, where only the local-density and generalized-gradient approximations can be used, the implementation of Helgaker, Watson and Handy⁴¹ allows for the use of Hartree-Fock exchange. This is in favour of the latter approach, since it has been shown⁴¹ that the use of hybrid functionals such as B3LYP improves considerably the results of the spin-spin coupling constants calculations. The approach of Helgaker, Watson and Handy is implemented in the development version of the DALTON program⁴⁸. It has been employed in the calculations of the hydrogen-bond-transmitted spin-spin coupling constants in $(\text{NH}_3)_2$ and N_2H_7^+ ⁶⁹.

The Fermi-contact contribution, usually the dominant one, can be relatively easily calculated by means of the finite field method. This approach, which does not require extensive programming, has been used by several groups for calculating the spin-spin coupling constants at the DFT level, also for the hydrogen-bond-transmitted couplings^{17,70,71}.

The use of DFT in the calculations of the coupling constants is already quite widespread, particularly for large systems, where the use of MCSCF or CC is not possible^{66,72,73}. On the other hand, DFT possesses certain limitations. It usually performs satisfactorily (and in some cases surprisingly well⁶⁵) for the couplings involving protons and carbons. However, the performance of DFT rapidly deteriorates with the number of lone pairs associated with the coupled nuclei, so in particular for the fluorine couplings it is not reliable^{18,41,65}.

The basis set requirements The problem of the basis set requirements for the spin-spin coupling constants calculations is cumbersome, and only a few suggestions can be provided⁷⁴. As for the energy, much larger basis sets

are required in the calculations aiming at a representation of the Coulomb hole (eg. CC) than in SCF or small MCSCF calculations. There are also some basis set requirements specific for the magnetic properties which make the calculations more difficult than those of the energy alone. It follows from the computational experience collected up to now³⁸ that the standard basis sets are not adequate, mostly because the dominant Fermi-contact contribution requires a more accurate description of the nuclear region than the usual basis sets offer. The correlation-consistent basis sets cc-pVXZ^{75,76,77} perform poorly, as the MCSCF study of Helgaker *et al.*⁷⁴ indicates. This can be improved by the use of (considerably larger) core-valence basis sets cc-pCVXZ⁷⁸ or by the decontraction of the s orbitals and the addition of n tight s orbitals, which leads to basis sets cc-pVXZ-sun, still tractable computationally. Satisfactory results can also be obtained with IGLO HIII and HIV basis sets^{79,80} and the basis sets of Ahlrichs *et al.*⁸¹. It should be kept in mind, however, that the reliable description of the systems interacting by van der Waals forces, like those discussed in this review, may require an addition of diffuse functions.

3. Review of the numerical results

3.1. Small systems with medium strength hydrogen bond

Let us begin with the calculations for the water dimer, for which one of the first ab initio calculations of the hydrogen-bond transmitted coupling constants have been carried out (by means of the MCSCF method) by Pecul and Sadlej⁴⁹. The ${}^1hJ(\text{OH})$ value is 4.3 Hz at RAS (2/0/8/14 2e)/HIII basis set level and is large enough to be interesting from the experimental point of view, even considering difficulties connected with the NMR measurements for quadrupolar nuclei such as ${}^{17}\text{O}$. The intermolecular ${}^{2h}J(\text{OO})$ coupling constants is ca 1 Hz at the same level of theory. Its value is very sensitive to the electron correlation effect. The coupling constants in the $(\text{H}_2\text{O})_2$ dimer have been calculated also by means of the EOM-CCSD method, yielding 1.29 Hz for the ${}^{2h}J(\text{OO})$ coupling⁵⁸.

The $(\text{HF})_2$ and $(\text{NH}_3)_2$ dimers have been the next small dimers under study^{34,82}. Table 1, reproduced from Ref.⁸², collects the reduced X-H \cdots Y hydrogen-bond-transmitted couplings, the interaction-induced changes of the intramolecular ${}^1J(\text{XH})$ coupling in the proton donor and ${}^1J(\text{YH})$ in the proton acceptor (averaged value in the case of the $(\text{NH}_3)_2$ dimer) as well as the interaction energies calculated at a uniform level of theory for $(\text{HF})_2$, $(\text{H}_2\text{O})_2$ and $(\text{NH}_3)_2$ ⁸³. The coupling constants have been obtained at the MCSCF level^{34,82} and the SAFT (Symmetry Adapted Perturbation

Theory) interaction energies are taken from Ref. ⁸³

Table 1. The comparison of the interaction energies ΔE^a , the reduced intermolecular coupling constants ${}^{2h}K(XY)$ and ${}^{1h}K(HY)$, and the changes in the reduced coupling constants in the proton donor $\Delta^1K(XH)$ and proton acceptor $\Delta^1K(YH)^b$ in the X-H...Y hydrogen-bonded complexes.

	ΔE (kJ/mol)	${}^{2h}K(XY)$ ($10^{-19}T^2/J$)	${}^{1h}K(HY)$ ($10^{-19}T^2/J$)	$\Delta^1K(XH)$ ($10^{-19}T^2/J$)	$\Delta^1K(YH)$ ($10^{-19}T^2/J$)
$(NH_3)_2$	-10.46	16.76 (17.04)	-1.31 (-1.40)	2.53 (2.90)	0.13 (0.15)
$(HF)_2$	-15.69	-0.58 (1.63)	-2.44 (-2.65)	1.65 (3.55)	2.94 (3.84)
$(H_2O)_2$	-17.66	5.41 (6.09)	-2.27 (-2.46)	2.75 (3.92)	1.38 (1.61)

^a SAPT results from Ref. ⁸³.

^b Fermi contact term in parenthesis, the coupling constants calculated at RAS (2/0/8/8 4e)/aug-cc-pVDZ-su1 level.

The first observation from Table 1 is that the reduced ${}^{2h}K(XY)$ coupling constant in the sequence of dimers under study has little in common with the interaction energy. Even the sign of ${}^{2h}K(XY)$ varies, from positive in $(H_2O)_2$ and $(NH_3)_2$ to negative in $(HF)_2$ (dominated by the paramagnetic spin-orbit term ³⁴). The sign remains constant when only the Fermi contact term is analysed, but the value of ${}^{2h}K(XY)$ still does not correlate with the interaction energy. This is further confirmed by the fact that the reduced ${}^{2h}K(NO)$ coupling in $H_2O \cdots NH_3$ dimer calculated by Janowski and Jaszuński ⁸⁴ is $18.73 \cdot 10^{19} T^2 J^{-1}$, a value similar to the one obtained for $(NH_3)_2$, although the interaction energy in $H_2O \cdots NH_3$ is approximately 2.5 times larger than that in $(NH_3)_2$. Thus it can be concluded that the reduced intermolecular coupling constant ${}^{2h}K(XY)$ is not a measure of the hydrogen bond strength when the coupling constants of different nuclei are compared.

The ${}^{1h}K(HY)$ coupling seems to be a better general parameter of the strength of the hydrogen bond. Its value in $(HF)_2$ is larger than in $(H_2O)_2$, which is opposite to what one will expect from the comparison of the interaction energies, but the discrepancy is not substantial. The reduced ${}^{1h}K(NH)$ coupling in the *trans* form of the $H_2O \cdots NH_3$ dimer is $-2.89 \cdot 10^{19} T^2 J^{-1}$ ⁸⁴, while the interaction energy is approximately -26.8 kJ/mol. It fits therefore in the tendency described here.

Much better correlation with the interaction energy is observed for the X-H...Y hydrogen-bond-induced changes in the Fermi contact terms of the intramolecular one-bond coupling ${}^1K(XH)$ (of the hydrogen-bond-forming

proton) than for the intermolecular coupling constants. The changes in the total $^1K(XH)$ couplings do not correspond so well to the interaction energy since in $(HF)_2$, unlike in $(H_2O)_2$ and $(NH_3)_2$, dimerization changes to a large extent the PSO term. In the $H_2O \cdots NH_3$ dimer the corresponding reduced coupling $^1K(OH)$ increases by approximately $5.4 \cdot 10^{19} T^2 J^{-1}$, which correlates well with the interaction energy -26.8 kJ/mol⁸⁴. Predictably, the hydrogen-bond-induced change in the reduced coupling constant between non-hydrogen-bonded proton and proton acceptor $^1K(XH)$ exhibits no correlation with the hydrogen bond strength.

The MCSCF studies on $(HF)_2$ and $(NH_3)_2$ have led to two more observations of interest. Firstly, the $^{2h}J(FF)$ coupling constant in $(HF)_2$ is completely dominated by the PSO term³⁴, while in all other short-range hydrogen-bond-transmitted coupling constants the FC term prevails. It means that the calculations of the fluorine intermolecular coupling constants may lead to qualitatively wrong results if only the FC term is included. Secondly, the calculations for three structures of $(NH_3)_2$ (assymetric global minimum, cyclic local minimum and linear saddle point) indicate that the intermolecular coupling constants are determined primarily by the internuclear distances, with little angular dependence.

Table 2. The comparison of the interaction energies ΔE^a , the reduced intermolecular coupling constants $^{2h}K(XY)$ and $^{1h}K(HY)$, and the changes in the reduced coupling constants in the proton donor $\Delta^1K(XH)$ and proton acceptor $\Delta^1K(YH)^b$ in the X-H...Y hydrogen-bonded complexes.

	ΔE (kJ/mol)	$^{2h}K(XY)$ ($10^{-19} T^2 / J$)	$^{1h}K(HY)$ ($10^{-19} T^2 / J$)	$\Delta^1K(XH)$ ($10^{-19} T^2 / J$)	$\Delta^1K(YH)$ ($10^{-19} T^2 / J$)
$(NH_3)_2$	-10.46	16.76 (17.04)	-1.31 (-1.40)	2.53 (2.90)	0.13 (0.15)
$(HF)_2$	-15.69	-0.58 (1.63)	-2.44 (-2.65)	1.65 (3.55)	2.94 (3.84)
$(H_2O)_2$	-17.66	5.41 (6.09)	-2.27 (-2.46)	2.75 (3.92)	1.38 (1.61)

^a SAPT results from Ref.⁸³.

^b Fermi contact term in parenthesis, the coupling constants calculated at RAS (2/0/8/8 4e)/aug-cc-pVDZ-su1 level.

The next group of hydrogen-bonded molecules investigated by the MCSCF method are the complexes of simple organic molecules: $CH_3OH \cdots H_2O$, $H_2O \cdots H_2CO$, $(HCOOH)_2$ and $HCCH \cdots H_2O$, two systems in which water molecule is a donor of electrons and two systems in which the carbonyl group is a donor of electrons⁵⁰ (in the notation above, the acceptor of the electron pair is listed first). These molecules have been chosen as models of hydrogen bonds in alcohols, aldehydes and carboxylic

acids. The spin-spin coupling constants calculated for these systems and the corresponding interaction energies⁵⁰ are collected in Table 2. Among the intermolecular coupling constants between nuclei directly engaged in the XH...Y hydrogen bond, the ${}^1hJ(OH)$ intermolecular coupling constants correlates best with the interaction energy in two subgroups and is promising as a parameter of hydrogen bond. The intermolecular oxygen-oxygen coupling constant ${}^2hJ(OO)$ approximately correlates with the H-bond energy, but the carbon-oxygen coupling constant in acetylene-water does not fit into this pattern (in terms of the reduced coupling constant).

Another finding of Ref.⁵⁰ is that the long-range intermolecular proton-proton coupling constants may have significant values, ranging from -0.5 Hz to +0.5 Hz. The $(HCOOH)_2$ is exceptional in this respect since the coupling constants between protons forming adjacent hydrogen bond is 4.4 Hz, and some of the long-distance couplings are more than 1 Hz (with a negative sign). In contrast to the short-distance hydrogen-bond-transmitted couplings (except ${}^2hJ(FF)$ in $(HF)_2$ ³⁴) for which Fermi contact terms dominate, these intermolecular long-distance proton-proton coupling constants are determined primarily by the spin-orbit terms.

The nitrogen-nitrogen coupling constants in neutral complexes of HCN and several proton acceptors forming N—H...N hydrogen bond have been calculated by means of the EOM-CCSD method⁸⁵. It has been concluded that the type of the bonding of the proton acceptor does not influence significantly the calculated ${}^2hJ(NN)$ coupling constant which is determined primarily by the internuclear distance: the dependence of ${}^2hJ(NN)$ on the N-N distance for five different complexes can be fitted by the same curve. The coupling constant for N-N distance corresponding to the A-U and G-C nucleic base pairs has been interpolated and the values are in agreement with experiment.

The vibrationally averaged EOM-CCSD hydrogen-bond-transmitted ${}^2hJ(NN)$ coupling constant in the CNH...NCH dimer has been also calculated⁸⁶. The value calculated for the ground vibrational state is 7.05 Hz, larger than the equilibrium value of 6.37 Hz. Predictably, ${}^2hJ(NN)$ changes significantly with the vibrational state.

3.2. Complexes with strong hydrogen bond

The spin-spin coupling constants transmitted through strong low-barrier hydrogen bonds have been particularly well investigated in the literature on account of the possible role of strong hydrogen bonds in enzymatic catalysis^{3,4,5,6,7}. Most of these calculations have been carried out at the EOM-

CCSD level^{56,57,58,59,60,61,87} although MCSCF^{18,34} and DFT¹⁸ methods have also been used.

Table 3. The fluorine coupling constants in $[F(HF)_n]^-$ complexes, $n = 1\dots 4$ (in Hz): the comparison of DFT, MCSCF and EOM-CCSD performance.

	DFT ^a Ref. 18	MCSCF ^b Ref. 34	EOM-CCSD ^c Ref. 56	exp Ref. 18
FHF^-				
$^1J(FH)$	6	122.0	100	124 ± 3
$^2J(FF)$	-133	327.1	225 ^d	$\approx 220^e$
$[F(HF)_2]^-$				
$^{1h}J(FH)$	-71	-20.8	-45	24 ± 3
$^1J(FH)$	229	359.8	388	354 ± 3
$^2J(FF)$	-73	194.3	179	146 ± 4
$[F(HF)_3]^-$				
$^{1h}J(FH)$	-72	-30.0	-57	41 ± 4
$^1J(FH)$	300	417.6	426	430 ± 4
$^2J(FF)$	-100	147.5	101	92 ± 5
$[F(HF)_4]^-$				
$^{1h}J(FH)$	-63.4	-31.7	-57	<10
$^1J(FH)$	336	474.1	492	480 ± 5
$^2J(FF)$	-144	110.5	11	<15

^a SD term not included

^b RASSCF calculations with large (27 orbitals) active spaces, all terms included. For clusters with $n > 1$ the calculations carried out on FHF^- with the geometry distorted to emulated larger cluster.

^c The qz2p basis set of Ahlrichs⁸¹

^d CCSD calculations in Ref.⁵⁷ resulted in $^2J(FF)$ equal to 254.4 Hz for an equilibrium bond length and 212.7 Hz for a ground vibrational state.

^e Extrapolation from the $^2J(FF)$ coupling constants of the larger clusters.

One of the first systems in which the intermolecular indirect spin-spin coupling constants have been measured are $[F(HF)_n]^-$ complexes, $n = 1\dots 4$ ¹⁸. In the same work the theoretical calculations of these coupling have been reported. The $^{19}F-^{19}F$ and $^1H-^{19}F$ coupling constants in FHF^- have been calculated by means of the linear response MCSCF method⁸⁸, while the couplings in the larger complexes have been obtained by means of the DFT^{64,65} method. Later on, the spin-spin coupling constants in these complexes have been studied by means of the EOM-CCSD method by Perera and Bartlett^{56,57}, and by means of the linear response MCSCF method by Pecul, Sadlej and Leszczynski³⁴. The vibrational corrections to the $^{2h}J(FF)$ couplings in FHF^- and FDF^- have been calculated at the

EOM-CCSD level⁵⁷. The results of these calculations are summarized in Table 3.

DFT leads to the results which are clearly very discrepant with experiment, for both proton-fluorine and fluorine-fluorine couplings. This is not surprising, since the inadequacy of DFT in the case of fluorine spin-spin coupling constants has been already known before⁶⁵. MCSCF is definitely superior in this respect, although the convergence of the calculated fluorine-fluorine coupling in FHF⁻ when the active space is extended is slow³⁴. CCSD yields the fluorine-fluorine coupling which is consistent with experiment, especially when the vibrational correction are included⁵⁷. On the other hand, proton-fluorine couplings are in a better agreement with experiment when the MCSCF method is employed. Even the couplings evaluated for the larger complexes using a model of FHF⁻ with geometry corresponding to that of these complexes are in agreement with experimental data. It means that geometric parameters, rather than influence of the other bound atoms, are the most important factors governing the value of the intermolecular proton-fluorine coupling. This coincides with the findings of Refs.^{85,89} for nitrogen couplings. It is worth mentioning that the vibrational averaging has an opposite effect on the $^{2h}J(FF)$ coupling in FHF⁻⁵⁷ than it has on the $^{2h}J(NN)$ coupling constant in the CNH \cdots NCH dimer⁸⁶.

The oxygen coupling constants, although difficult to measure experimentally, have also been calculated in charged and strongly bonded complexes at the EOM-CCSD level^{58,59}. The EOM-CCSD study of the spin-spin coupling constants in (H₂O)₂, H₃O₂⁻, and H₅O₂⁺, including the analysis of the distance dependence, has been conducted by Del Bene, Perera and Bartlett⁵⁸. Protonation and decrease of the O-O distance when going from (H₂O)₂ to H₅O₂⁺ have been shown to increase the $^{2h}J(OO)$ coupling from 1.29 Hz to 39.54 Hz, while in the anionic H₃O₂⁻ complex this coupling constant is 16.28 Hz. In all cases the $^{2h}J(OO)$ coupling constant is dominated by the FC term and decreases in an exponential manner with the O-O distance. In the same work⁵⁸ the coupling constants in neutral, protonated and deprotonated complexes of HCN with H₂O and HCN have been calculated. In all cases the coupling constant transmitted though a proton-shared hydrogen bond is predicted to be several times larger than its traditional hydrogen bond analogue.

Del Bene, Perera and Bartlett have also carried out the study of the spin-spin coupling constants in a number of positively charged systems with proton-shared hydrogen bond: protonated dimers of H₂O, H₂CO, CO, HNO and HPO with OH⁺ \cdots O hydrogen bonds and protonated dimers

of NH_3 , N_2 , HCN, HNO and HPNH with $\text{NH}^+ \cdots \text{N}$ hydrogen bonds⁵⁹. For some of these systems the distance dependence of the hydrogen-bond-transmitted couplings has been analysed. It has been found that the $^{2h}J(\text{XY})$ coupling constant transmitted through $\text{X-H} \cdots \text{Y}$ hydrogen bond is largest for a symmetric hydrogen bond. The hybridization of proton-donor or proton-acceptor atoms plays little role in determining the magnitude of hydrogen-bond transmitted coupling. All nitrogen and oxygen hydrogen-bond-transmitted coupling constants are determined by the FC term and are strongly distance dependent. No correlation of $^{2h}J(\text{XY})$ with the interaction energy has been found.

Another type of hydrogen-bond transmitted coupling which has been studied theoretically (only the FC term) by means of the EOM-CCSD method are $^{4h}J(\text{PP})$ coupling constants transmitted through $\text{NH}^+ \cdots \text{N}$ hydrogen bond⁶². A small model system has been chosen to emulate the internal hydrogen bond in large organic cations⁹⁰. The coupling constant calculated for the model compound is significantly overestimated in comparison with experiment, which is attributed to the inadequacies of the model.

A short review of the EOM-CCSD spin-spin coupling constants transmitted through $\text{NH} \cdots \text{N}$ and $\text{NH}^+ \cdots \text{N}$ hydrogen bonds is provided in Ref.⁸⁹, together with some previously unpublished results of the EOM-CCSD calculations of the intermolecular coupling constants in the neutral complexes of HCN and HNC. An analysis of the $^{2h}J(\text{NN})$ coupling constant in a wide range of complexes confirmed its good correlation with the N-N distance and small dependence on the bonding in the proton acceptor and donor. According to the authors, the proposed fit expressing the distance dependence of $^{2h}J(\text{NN})$ can be used to calculate the N-N distance from an experimental NMR spectrum with accuracy of about 0.05 Å.

Strong hydrogen bonds can be formed in principle not only in ionic complexes but also in acid-base pairs, such as HCl-pyridine complex. The hydrogen bonds in these complexes span a wide range of types, from weak traditional hydrogen bonds to the interactions involving proton transfer. The hydrogen-bond-transmitted coupling constants have been intensively studied in the complexes of HCl with nitrogen-containing bases^{60,61} at the EOM-CCSD level, in order to correlate the spin-spin couplings with the hydrogen bond strength and geometry. The calculations in Ref.⁶⁰ have been carried out for the $\text{ClH} \cdots \text{NH}_3$ complex for various strength of the electric field imposed along the hydrogen bond Cl—H \cdots N. The nature of the complex changes then from a traditional hydrogen bond at zero field

to a proton-shared hydrogen bond at intermediate fields and to an ion-pair hydrogen bond at higher fields. These transitions are mirrored by the $^{2h}J(\text{ClN})$ spin-spin coupling constant (calculated in an absence of the field but for the structure optimized for a given field strength), which assumes the maximum value (in terms of an absolute value) of -11.8 Hz for an intermediate field strength when the hydrogen bond is nearly symmetric. The $^{2h}J(\text{ClN})$ coupling constant in $\text{ClH}\cdots\text{NH}_3$ is dominated by the FC term, therefore in the next study of HCl-pyridine complex⁶¹ only this term has been evaluated.

The $^{2h}J(\text{ClN})$ coupling constant in the $\text{HCl}\cdots$ pyridine complex has been studied in Ref.⁶¹ by means of a methodology similar to the one employed for the $\text{ClH}\cdots\text{NH}_3$ complex⁶⁰. The conclusions are basically the same as for the $\text{ClH}\cdots\text{NH}_3$ complex, with an additional advantage of being partly supported by the experimental NMR measurements for this complex⁹¹. A summary of the CCSD calculations of the spin-spin coupling constants in acid-base pairs is presented in Ref.⁹².

The last paper published in that series⁸⁷ reports on the EOM-CCSD calculations of the spin-spin coupling constants carried out for a series of substituted pyrroles as proton donors and seven nitrogen bases as proton acceptors. Additionally, the spin-spin coupling constants have been recalculated in the $\text{CNH}\cdots\text{NCH}$ complex with geometry optimized in the presence of electric field, which favours a formation of a proton-shared or even ion-pair complex. It has been found that increasing the electric field leads to an increase in hydrogen-bond-transmitted coupling, and so does increasing the base strength. However, in uncharged complexes in the absence of external electric field no proton-shared or ion-pair complexes have been observed.

3.3. Models of biological systems

Most of the interest in the hydrogen-bond-transmitted spin-spin coupling constants lies in their potential application for determination of the higher order structures of biomolecules: proteins and nucleic acids. Consequently, the theoretical papers reporting calculations of the hydrogen-bond-transmitted spin-spin coupling constants in models of biological systems are abundant.

The MCSCF method has been used to calculate the intermolecular spin-spin coupling constants in formamide-formamide and formamide-formamidine dimers, modelling the hydrogen bond in peptides and adenine-tymine (or adenine-uracil) pair in a nucleic acid, respectively⁵¹. The obtained hydrogen-bond-transmitted coupling constants are visualized in Fig-

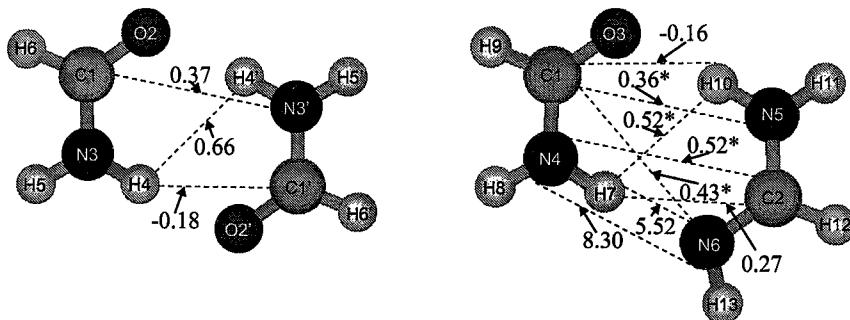


Fig. 1. The hydrogen-bond-transmitted coupling constants calculated at the MCSCF level for formamide-formamide and formamide-formamidine dimers⁵¹. Asterix denotes the numbers for which SD terms are omitted.

ure 1. ${}^3hJ(NC')$ couplings have been found to be positive, which contradicts the tentative assignment of sign in the experimental spectra¹⁶. Both ${}^1hJ(NH)$ and ${}^2hJ(NN)$ coupling constants in the formamide-formamidine dimer have been found to be positive, and their absolute values are in a satisfactory agreement with the experimental results for nucleic acids (6—7 Hz for ${}^2hJ(NN)$, 2—4 Hz for ${}^1hJ(NH)$ ²⁰). In this case there is no controversy concerning the sign⁷⁰. A slight overestimation of the theoretical results probably originates from a too short length of the hydrogen bond in the employed model. The investigation of the dependence of the intermolecular coupling constants on the hydrogen bond length has led to the conclusion that the short-distance hydrogen bond-transmitted couplings (e.g. ${}^1hJ(NH)$ and ${}^2hJ(NN)$) decrease fast (in an exponential manner) with increase in the hydrogen bond length, in contrast to the long-distance proton-proton intermolecular couplings. The reason for this behaviour lies in the dominance of the FC terms in the short-distance hydrogen bond-transmitted couplings, while the long-distance ones are determined primarily by less distance-dependent spin-orbit terms (see also Ref.⁵⁰). Another interesting finding of Ref.⁵¹ is the relatively significant values of the coupling constants between protons engaged in adjacent hydrogen bonds.

Similar calculations, also by means of the MCSCF method, have been undertaken for the methyleneimine dimer⁹³. The distance and angular dependence of the ${}^2hJ(NN)$ have been studied. The exponential decay of the coupling with the internuclear distance has been observed and it has been concluded that the hydrogen bond length is the most important fac-

tor determining the value of the coupling but that for accurate structure determination also the angular dependence should be considered.

Realistic models of hydrogen bonds in peptides and nucleic acids are necessarily systems of significant size. Therefore most of the calculations of the spin-spin coupling constants in such systems have been carried out using various implementations of the DFT method, which has the most favourable scaling with the size of the system.

Soon after the experimental measurement of the hydrogen-bond-transmitted spin-spin coupling constants in proteins and nucleic acids they have been modelled by the sum-over-states density functional perturbation theory as implemented in the deMon program⁶⁵. As a model system for peptides *N*-methylacetamide has been chosen⁶⁶. Additionally, the intermolecular $^{2h}J(\text{NN})$ couplings between uracil-adenine, cytosine-guanine and adenine-thymine pairs have been calculated⁶⁶. The results have been found to agree reasonably well with the experiment. For proteins, good correlation between $^{3h}J(\text{NC}')$ and chemical shift of hydrogen-bond forming proton, previously found experimentally¹¹, has been confirmed. This means that $^{3h}J(\text{NC}')$ correlates also with the hydrogen bond length. It should be mentioned at this point that the $^{3h}J(\text{NC}')$ coupling constants calculated at the DFT level^{66,71} are negative, in contrast to the MCSCF results⁵¹. This is a controversy which calls for more extended theoretical studies.

DFT calculations for the 16-atom model of a Watson-Crick guanine-cytosine pair have been reported in Ref.¹⁷. They also suggest positive sign of $^{1h}J(\text{NH})$ and $^{2h}J(\text{NN})$ couplings, although the method of calculation is less sophisticated than in Ref.^{51,66} since only FC terms are calculated.

A cyclic formamidine dimer has been used as a model of interaction between nucleic bases in the DFT calculations⁹⁴ (deMon implementation⁶⁵). The positive signs of the $^{1h}J(\text{NH})$ and $^{2h}J(\text{NN})$ coupling constant has been confirmed, and the calculated values have been found to be in agreement with experiment. The changes of the coupling constants during the proton transfer in the dimer have been evaluated, and the values have been fitted into simple equations correlating the proton-nitrogen and nitrogen-nitrogen coupling constants with the degree of assymetry of the $\text{NH}\cdots\text{N}$ bond.

Another DFT calculations of the interresidual coupling constants in DNA have been reported in the paper by Barfield et al⁷⁰. Here the model has been extended to the DNA triplets T·A-T and C⁺·G-C, but the calculations have been restricted to the FC terms obtained by means of the finite field method. The DFT calculations for several different base pairs separated by various interresidual distances have led to a good correlation between $^{2h}J(\text{NN}')$ coupling and $\delta(^1\text{H})$, confirmed by the experimental data

reported in the same study. An interesting finding is a prediction that the $^{1h}J(\text{NH})$ coupling becomes negative for charged C⁺.G pair, which is characterized by the shortest intermolecular distance. Although the method employed in Ref. ⁷⁰ cannot be expected to yield quantitative results, this last finding seems probable, considering similar effects reported for fluorine couplings ^{18,34,56} and the fact that one-bond proton-nitrogen coupling transmitted through a covalent bond is negative. A simple equation based on a three orbital model has been proposed to predict the couplings in the NH...N fragment ⁷⁰. However, it has been found to yield the values of $^{2h}J(\text{NN}')$ and $^{1h}J(\text{NH})$ coupling overestimated by a factor of 5. Moreover, it predicted $^{1h}J(\text{NH})$ of a wrong sign, at least for longer hydrogen bond lengths.

The coupling constants in a linear formamide dimer, treated as a model of hydrogen bonds in peptides, have been calculated by means of the finite field DFT method by Bagno ⁷¹. In the same work, the couplings between selected residues of ubiquitin are calculated, with all aminoacids except the ones of interest removed or replaced (adjacent ones) by -COCH₃ groups.

The geometric dependence of the $^{3h}J(\text{PN})$ and $^{2h}J(\text{PH})$ coupling constants in the complexes between Mg[PO₄(CH₃)] or [PO₄(CH₃)₂]⁻ (modelling a nucleotide) and several aminoacids have been evaluated by means of the DFT method ⁹⁵ (deMon implementation ⁶⁵). The results show that the sign of these couplings varies in different complexes. The $^{3h}J(\text{PN})$ and $^{2h}J(\text{PH})$ coupling constants range from negative values of several herz to positive values not exceeding 0.5 Hz, exhibiting a strong hydrogen bond length and angular dependence.

To sum up, numerous equations correlating the coupling constants in models of biological with the geometric parameters have been proposed on the basis of ab initio calculations, and have been found helpful for structural studies. In most cases the intermolecular coupling constants depend primarily on the internuclear distance, as discussed above ^{82,85,89}. Another purpose of the theoretical investigations of the spin-spin coupling constants in biological systems is to establish their signs since most of the NMR experiments yield only the absolute values. The importance of this is obvious, considering that the sign of the coupling constants may vary with the intermolecular distance, which in turn can affect structural assignments.

3.4. Weak van der Waals complexes

The calculations of the spin-spin coupling constants transmitted through van der Waals interactions weaker than hydrogen bonds have twofold mo-

tivation: first, to predict whether these couplings, not detected experimentally so far, can have measurable values; second, to get some insight into an alleged correlation between a magnitude of the coupling constant and covalent character of the interaction.

One of the simplest van der Waals complexes is the helium dimer. The small size of the system has made it possible to evaluate the ${}^3\text{He}-{}^3\text{He}$ spin-spin coupling constant in an accurate manner, at the full configuration interaction level ³³. The Fermi-contact term has been found to have non-negligible value of 1.3 Hz at $R=5.6$ a.u. (close to the energy minimum), while the other contributions are practically zero. The coupling decreases very fast, in an exponential manner, with the internuclear distance R . For R equal to 4 a.u. it is over 22 Hz, while for R over 7 a.u. it falls below 0.1 Hz. The non-contact terms decay with R even faster than the FC term. It is concluded therefore that the distance dependence of the FC term is similar in the hydrogen-bonded and van der Waals complexes, which suggests a common mechanism of the coupling. The exponential character of this dependence suggests a connection with the exchange energy. A non-negligible value of the ${}^3\text{He}-{}^3\text{He}$ spin-spin coupling constant for the energy minimum of in He_2 has been confirmed by CCSD calculations with large basis sets ³³. The CCSD implementation used has been that of Gauss and Auer ⁶³.

Another noble gas dimer in which the coupling constant (the FC term only) has been calculated is the Xe dimer (DFT calculations with a local functional) ³⁵. The result suggests that the reduced indirect spin-spin coupling in Xe_2 , close to the minimum of energy, is one order of magnitude larger than in He_2 ³³. However, rather unexpectedly, the sign is opposite. Obviously, the computational approach is widely different in Refs. ^{33,35} and a direct comparison is probably not possible.

In the series of fluorine-containing systems for which the MCSCF calculations have been carried out ³⁴ there is also a van der Waals complex, $\text{CH}_4 \cdots \text{HF}$. It has been found that the long-distance proton-fluorine coupling is substantial in this molecule, even larger than the ${}^3hJ(\text{HF})$ coupling in $(\text{HF})_2$ ³⁴. Also the distance-dependence of the intermolecular proton-fluorine coupling in $\text{CH}_4 \cdots \text{HF}$ is analogous to the trends in the hydrogen-bonded complexes.

The ${}^{19}\text{F}-{}^{19}\text{F}$ coupling constant in difluoromethane dimer, obtained at the DFT level, is also substantial ³¹. It should be remembered, however, that the DFT method, used for its evaluation, is not reliable for the calculations of the fluorine couplings ^{18,65}.

Another work in which DFT (deMon-NMR ⁶⁵ implementation) has been

used to calculate the couplings transmitted through weak van der Waals interaction is by Bagno *et al.*⁶⁷ on $\text{CH}_4 \cdots \text{C}_6\text{H}_6$ and $\text{C}_6\text{H}_6 \cdots \text{C}_6\text{H}_6$ dimers. Their findings confirm the conclusions of Pecul, Sadlej and Leszczynski^{33,34} that such couplings have measurable values. Another conclusion of Ref. ⁶⁷ is that the intermolecular ^1H - ^1H and ^{13}C - ^1H couplings are not always dominated by the FC terms and that spin-orbit terms also contribute considerably. This is in agreement with the observations of Refs ^{50,51}, too.

4. Summary

Numerous efforts have been undertaken in a relatively short time span to correlate the hydrogen-bond-transmitted spin-spin coupling constants with other energetical and structural properties of the hydrogen bonds, employing ab initio techniques to calculate spin-spin couplings. Let us summarize the most important findings. It has been shown on many examples that the $^{2h}J(\text{XY})$ coupling constant is an excellent indicator of the intermolecular distance in the $\text{X-H} \cdots \text{Y}$ hydrogen bond, although it does not correlate with the interaction energy. Better, although not general, correlations with the interaction energy have been obtained for the $^{1h}J(\text{HY})$ coupling constant. The ab initio calculations of the hydrogen-bonds-transmitted coupling constants in models of proteins and nucleic acids have led to the conclusion that a wide variety of the intermolecular coupling constants can be employed in order to establish the structures of biomolecules. The ab initio calculations allowed also to identify signs of the hydrogen-bond-transmitted coupling constants. This issue is of particular importance since the sign of the intermolecular coupling may vary even if the formal type of the coupling does not.

The calculations of spin-spin coupling constants in charged and strongly hydrogen-bonded complexes demonstrate that the $^{2h}J(\text{XY})$ hydrogen-bond-transmitted spin-spin coupling constant may serve as a fingerprint of the hydrogen bond type, since it assumes maximum values for the symmetric proton-shared hydrogen bonds. The results for the van der Waals complexes bound by dispersion forces, such as He_2 and Xe_2 , indicate that transmission of the indirect spin-spin coupling by an interaction is not an evidence of the covalent character of this interaction. The coupling constants between nuclei belonging to different molecules bound by weak van der Waals forces have not yet been detected experimentally, but the theoretical results suggest that such measurements are feasible.

Most of the hydrogen-bond- and van der Waals interaction-transmitted coupling constants have been found to be dominated by the Fermi contact

terms. However, similarly as in the case of covalent-bond-transmitted couplings, there are several exceptions: some of the fluorine-fluorine, fluorine-proton and proton-proton intermolecular coupling constants are determined primarily by the spin-orbit terms.

Acknowledgments

We are indebted to M. Jaszuński for helpful discussions. This work has been supported by the 7 TO9A 11121 KBN Grant.

References

1. Z. R. Wu, S. Ebrahimian, M. E. Zawrotny, L. D. Thornburg, G. C. Perez-Alvarado, P. Brothers, R. M. Pollack, and M. F. Summers. *Science*, **276**, 415, (1997).
2. C. L. Perrin. *Science*, **266**, 1665, (1994).
3. W. W. Cleland, P. A. Frey, and J. A. Gertl. *J. Biol. Chem.*, **273**, 25529, (1998).
4. S. Scheiner and T. Kar. *J. Am. Chem. Soc.*, **117**, 6970, (1995).
5. J. P. Guthrie and R. Kluger. *J. Am. Chem. Soc.*, **115**, 11569, (1993).
6. A. Waesel, A. Papazyan, and P. A. Kollman. *Science*, **269**, 103, (1995).
7. J. A. Gertl and P. G. Gassman. *J. Am. Chem. Soc.*, **115**, 11552, (1994).
8. P. A. Frey, S. A. Whitt, and J. B. Tobin. *Science*, **264**, 1927, (1994).
9. G. A. Kumar and M. A. McAllister. *J. Org. Chem.*, **63**, 6968, (1998).
10. J. E. Del Bene, S. A. Perera, and R. J. Bartlett. *J. Phys. Chem. A*, **103**, 8121, (1999).
11. F. Cordier and S. Grzesiek. *J. Am. Chem. Soc.*, **121**, 1601, (1999).
12. G. Cornilescu, J.-S. Hu, and A. Bax. *J. Am. Chem. Soc.*, **121**, 2949, (1999).
13. A. J. Dingley and S. Grzesiek. *J. Am. Chem. Soc.*, **120**, 8293, (1998).
14. N. Juranic, P. K. Ilich, and S. Macura. *J. Am. Chem. Soc.*, **117**, 405, (1995).
15. P. R. Blake, J. B. Park, M. W. W. Adams, and M. F. Summers. *J. Am. Chem. Soc.*, **114**, 4931, (1992).
16. G. Cornilescu, B. E. Ramirez, M. K. Frank, G. Marius Clore, A. M. Gronenborn, and A. Bax. *J. Am. Chem. Soc.*, **121**, 6275, (1999).
17. A. J. Dingley, J. E. Massee, R. D. Peterson, M. Barfield, J. Feigon, and S. Grzesiek. *J. Am. Chem. Soc.*, **121**, 6019, (1999).
18. I. G. Shenderovich, S. H. Smirnov, G. S. Denisov, V. A. Gindin, N. S. Golubev, A. Dunger, R. Reibke, S. Kirpekar, O. L. Malkina, and H.-H. Limbach. *Ber. Bunsenges. Phys. Chem.*, **102**, 422, (1998).
19. N. S. Golubev, I. G. Shenderovich, S. N. Smirnov, G. S. Denisov, and H.-H. Limbach. *Chem. Eur. J.*, **5**, 492, (1999).
20. K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, and K. Wüthrich. *Proc. Natl. Acad. Sci. USA*, **95**, 14147, (1998).
21. M. Henning and B. H. Geierstanger. *J. Am. Chem. Soc.*, **121**, 5123, (1999).
22. F. Cordier, M. Rogowski, S. Grzesiek, and A. Bax. *J. Magn. Reson.*, **140**, 510, (1999).

23. A. Meissner and O. W. Sorensen. *J. Magn. Reson.*, **143**, 387, (2000).
24. A. Liu, A. Majumdar, W. Hu, A. Kettani, E. Skripkin, and D. J. Patel. *J. Am. Chem. Soc.*, **122**, 3206, (2000).
25. D. R. Davis, R. P. Lutz, and J. D. Roberts. *J. Am. Chem. Soc.*, **83**, 246, (1961).
26. L. Petrakis and C. H. Sederholm. *J. Chem. Phys.*, **35**, 1243, (1961).
27. L. Ernst and K. Ibrom. *Angew. Chem.*, **107**, 2010, (1995).
28. J. E. Peralta, V. Barone, and R. H. Contreras. *J. Am. Chem. Soc.*, **123**, 9162, (2001).
29. F. Löhr, S. G. Mayhew, and H. Rüterjans. *J. Am. Chem. Soc.*, **122**, 9289, (2000).
30. A. J. Dingley, F. Cordier, and S. Grzesiek. *Concepts in Magnetic Resonance*, **13**, 103, (2001).
31. W. D. Arnold and E. Oldfield. *J. Am. Chem. Soc.*, **122**, 12835, (2001).
32. S. Borman. *Chem. Eng. News*, **77**, 36, (1999).
33. M. Pecul. *J. Chem. Phys.*, **113**, 10835, (2000).
34. M. Pecul, J. Sadlej, and J. Leszczynski. *J. Chem. Phys.*, **115**, 5498, (2001).
35. F. R. Salsbury Jr and R. A. Harris. *Mol. Phys.*, **94**, 307, (1998).
36. R. F. W. Bader. *Atoms in Molecules - A Quantum Theory*. Oxford University Press, Oxford, (1990).
37. J. F. Hinton and K. Wolinski. In D. Hadzi, editor, *Theoretical Treatments of Hydrogen Bonding*, chapter Ab initio GIAO Magnetic Shielding Tensor for Hydrogen bonded Systems. J. Wiley & Sons Ltd., (1997).
38. T. Helgaker, M. Jaszunski, and K. Ruud. *Chem. Rev.*, **99**, 293, (1999).
39. N. F. Ramsey. *Phys. Rev.*, **78**, 699, (1950).
40. N. F. Ramsey. *Phys. Rev.*, **91**, 303, (1953).
41. T. Helgaker, M. Watson, and N. C. Handy. *J. Chem. Phys.*, **113**, 9402, (2000).
42. J. Kaski, P. Lantto, J. Vaara, and J. Jokisaari. *J. Am. Chem. Soc.*, **120**, 3993, (1998).
43. J. Kaski, J. Vaara, and J. Jokisaari. *J. Am. Chem. Soc.*, **118**, 8879, (1996).
44. J. Vaara. *Computational Studies on the NMR parameters of molecular probes in liquids and solids*. PhD thesis, Department of Physical Sciences, University of Oulu, (1997).
45. J. Geertsen and J. Oddershede. *Chem. Phys.*, **90**, 301, (1984).
46. J. Geertsen, J. Oddershede, and G. E. Scuseria. *Int. J. Quantum Chem.: Quantum Chem. Symp.*, **21**, 475, (1987).
47. H. Fukui. *Prog. NMR Spectrosc.*, **35**, 267, (1999).
48. T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, J. Olsen, K. Ruud, H. Ågren, T. Andersen, K. L. Bak, V. Bakken, O. Christiansen, P. Dahle, E. K. Dalskov, T. Enevoldsen, B. Fernandez, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, R. Kobayashi, H. Koch, K. V. Mikkelsen, P. Norman, M. J. Packer, T. Saue, P. R. Taylor, and O. Vahtras. *Dalton, an ab initio electronic structure program, Release 1.0* (1997). See <http://www.kjem.uio.no/software/dalton/dalton.html>.
49. M. Pecul and J. Sadlej. *Chem. Phys. Lett.*, **308**, 486, (1999).
50. M. Pecul, J. Leszczynski, and J. Sadlej. *J. Chem. Phys.*, **112**, 7930, (2000).
51. M. Pecul, J. Leszczynski, and J. Sadlej. *J. Phys. Chem. A*, **104**, 8105, (2000).

52. S. A. Perera, M. Nooijen, and R. J. Bartlett. *J. Chem. Phys.*, **104**, 3290, (1996).
53. S. A. Perera, H. Sekino, and R. J. Bartlett. *J. Chem. Phys.*, **101**, 2186, (1994).
54. J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale, and R. J. Bartlett. Aces ii, an *ab initio* program system.
55. J. E. Del Bene and R. J. Bartlett. *J. Am. Chem. Soc.*, **122**, 10480, (2000).
56. S. A. Perera and R. J. Bartlett. *J. Am. Chem. Soc.*, **122**, 1231, (2000).
57. J. E. Del Bene, M. J. T. Jordan, S. A. Perera, and R. J. Bartlett. *J. Phys. Chem. A*, **105**, 8399, (2001).
58. J. E. Del Bene, S. A. Perera, and R. J. Bartlett. *J. Am. Chem. Soc.*, **122**, 3560, (2000).
59. J. E. Del Bene, S. A. Perera, and R. J. Bartlett. *J. Phys. Chem. A*, **105**, 930, (2001).
60. J. E. Del Bene and M. J. T. Jordan. *J. Am. Chem. Soc.*, **122**, 4794, (2000).
61. K. Chapman, D. Crittenden, J. Bevitt, M. J. T. Jordan, and J. E. Del Bene. *J. Phys. Chem. A*, **105**, 5442, (2001).
62. J. E. Del Bene, S. A. Perera, R. J. Bartlett, I. Alkorta, and J. Elguero. *J. Phys. Chem. A*, **104**, 7165, (2000).
63. A. A. Auer and J. Gauss. *J. Chem. Phys.*, **115**, 1619, (2001).
64. V. G. Malkin, O. L. Malkina, and D. R. Salahub. *Chem. Phys. Lett.*, **221**, 91, (1994).
65. O. L. Malkina, D. R. Salahub, and V. G. Malkin. *J. Chem. Phys.*, **105**, 8793, (1996).
66. C. Scheurer and R. Brüschiweiler. *J. Am. Chem. Soc.*, **121**, 8661, (1999).
67. A. Bagno, G. Saielli, and G. Scorrano. *Angew. Chem. Int. Ed.*, **40**, 2532, (2001).
68. R. M. Dickinson and T. Ziegler. *J. Phys. Chem.*, **100**, 5286, (1996).
69. M. Pecul and J. Sadlej. work in progress.
70. M. Barfield, A. J. Dingley, J. Feigon, and S. Grzesiek. *J. Am. Chem. Soc.*, **123**, 4014, (2001).
71. A. Bagno. *Chem. Eur. J.*, **6**, 2925, (2001).
72. M. Hricovini, O. L. Malkina, F. Bzik, L. T. Nagy, and V. G. Malkin. *J. Phys. Chem. A*, **101**, 9756, (1997).
73. T. Onak, J. Jaballas, and M. Barfield. *J. Am. Chem. Soc.*, **121**, 2850, (1999).
74. T. Helgaker, M. Jaszuński, K. Ruud, and A. Górska. *Theor. Chim. Acc.*, **99**, 175, (1998).
75. T. H. Dunning Jr. *J. Chem. Phys.*, **90**, 1007, (1989).
76. D. E. Woon and T. H. Dunning Jr. *J. Chem. Phys.*, **98**, 1358, (1993).
77. R. A. Kendall, T. H. Dunning Jr., and R. J. Harrison. *J. Chem. Phys.*, **96**, 6796, (1992).
78. D. E. Woon and T. H. Dunning Jr. . *J. Chem. Phys.*, **103**, 4572, (1995).
79. M. Schindler and W. Kutzelnigg. *J. Chem. Phys.*, **76**, 1919, (1982).
80. Ch. van Wüllen. PhD thesis, Ruhr-Universität Bochum, Bochum, (1992).
81. A. Schäfer, H. Horn, and R. Ahlrichs. *J. Chem. Phys.*, **97**, 2571, (1992).
82. M. Pecul and J. Sadlej. *Chem. Phys. Lett.*, **360**, 272, (2002).
83. J. Langlet, J. Caillet, and M. Caffarel. *J. Chem. Phys.*, **103**, 8043, (1995).
84. T. Janowski and M. Jaszuński. *Int. J. Quantum Chem.* accepted.

85. J. E. Del Bene and R. J. Bartlett. *J. Am. Chem. Soc.*, **122**, 10480, (2000).
86. M. J. T. Jordan, J. S.-S. Toh, and J. E. Del Bene. *Chem. Phys. Lett.*, **346**, 288, (2001).
87. J. S. S. Toh, M. J. T. Jordan, B. C. Husowitz, and J. E. Del Bene. **573**, 11, (2001).
88. O. Vahtras, H. Ågren, P. Jørgensen, H. J. Aa. Jensen, S. B. Padkjær, and T. Helgaker. *J. Chem. Phys.*, **96**, 6120, (1992).
89. J. E. Del Bene, S. A. Perera, and R. J. Bartlett. **39**, 109, (2001).
90. J. Laynez, M Menéndez, J.L.S. Velasco, A. L. Llamaz-Saiz, C. Foce-Foces, J. Elguero, P. Molina, and M. Alajarín. *J. Chem. Soc., Perkin Trans. 2*, page 709, (1993).
91. N. S. Golubev, G. S. Denisov, S. N. Smirnov, D. N. Shchepkin, and H.-H. Limbach. **196**, 73, (1996).
92. J. E. Del Bene and M. J. T. Jordan. *J. Mol. Struct. (Theochem)*, **573**, 11, (2001).
93. D. L. Bryce and R. E. Wasylishen. *J. Biomolec. NMR*, **19**, 371, (2001).
94. H. Benedict, I. G. Shenderovich, O. L. Malkina, V. G. Malkin, G. S. Denisov, N. S. Golubev, and H.-H. Limbach. *J. Am. Chem. Soc.*, **112**, 1979, (2000).
95. J. Czernek and R. Brüschweiler. *J. Am. Chem. Soc.*, **123**, 11079, (2001).

Chapter 4: Base Polyad Motifs in Nucleic Acids — Biological Significance, Occurrence in Three-Dimensional Experimental Structures and Computational Studies

Michael Meyer^a and Jürgen Sühnel^b

^a*Revotar Biopharmaceuticals AG, Neuendorfstr. 24b,
D-16761 Hennigsdorf / Germany*

^b*Biocomputing Group, Institute of Molecular Biotechnology, Jena
Centre for Bioinformatics, Beutenbergstr. 11, D-07745 Jena / Germany*

Abstract

Base polyads are base complexes linked by base-base hydrogen bonds. They occur in nucleic acid structures involved in many important biological functions. Examples include telomeres, structures formed from triplet disease sequences, oncogenes, the centromere, triplexes involved in gene targeting and also tertiary interactions in RNA. Due to this biological importance an increasing number of three-dimensional structures with base polyad motifs has been determined and the first computational studies have been reported. Here a comprehensive overview on the occurrence of base polyads in nucleic acid structures is provided and computational studies on these nucleic acid building blocks are reviewed, thereby attempting to establish a link between structural biology and computational chemistry.

1. Introduction

Since the 1953 paper by Watson and Crick on a structure for desoxyribose nucleic acid, the structural and functional importance of base pairs linked by hydrogen bonds (H-bonds) is one of the central themes of molecular and

structural biology.¹ The textbook picture of DNA usually shows the beautiful and regular straight antiparallel double strand or duplex held together by H-bonds between the bases adenine (A) and thymine (T), or uracil (U) in RNA, as well as between guanine (G) and cytosine (C) thereby forming the Watson-Crick base pairs AT or AU and GC. Subsequent research has confirmed the Watson-Crick structural model of the DNA B-conformation originally derived from fiber diffraction data directly by X-ray crystallography.² However, it became clear very early³ and has later been shown, in particular by the rapidly increasing number of nucleic acid structures known at atomic resolution, that the structure of DNA (and RNA) is not that simple and regular.

The primary information resources for three-dimensional nucleic acid structural data are the Protein Data Bank (PDB)^{4,5} and the Nucleic Acid Database (NDB).⁶ The NDB includes almost all X-ray structure entries from the PDB and only a few additional X-ray structures but does not take into account theoretical models and structures determined by nuclear magnetic resonance (NMR) spectroscopy. Therefore, the overwhelming majority of nucleic acid structures is available from the PDB. Currently (February 9, 2002) the PDB holds 1776 nucleic-acid containing structures. Among them are 52 theoretical models, 501 structures determined by nuclear magnetic resonance (NMR) spectroscopy and 1223 structures obtained by diffraction or other methods. The nucleic acid structure dataset includes currently 725 protein-nucleic acid complexes and 512 RNA-containing structures. Finally, it should be noted that in addition to the PDB and NDB other resources including OCA,⁷ RNAbase,⁸ PDBsum⁹ and the IMB Jena Image Library of Biological Macromolecules¹⁰ offer useful structural information on nucleic acids.

The structural diversity of nucleic acids is, among other factors, due to variations in strand number, helical axis geometry, groove width and fine structure of base pair and base pair step geometries. In addition, tertiary interactions between secondary structural elements are often crucial for the overall fold. Base pair interactions, usually via H-bonds, play a central role both for nucleic acid structure and recognition processes. One should realize, however, that H-bonding in base pairs is only one force relevant to nucleic acid structures. This is clearly shown in Fig. 1, where in addition to a complete DNA structure the sugar-phosphate backbone and base pair parts are shown separately. Other important interactions are base-base stacking, base-backbone and backbone-backbone interactions as well as effects exerted by the 'environment', such as the interaction with solvents, ions, small-molecule ligands and proteins.

Particularly from the increasing number of RNA structures we have learned that in addition to the Watson-Crick base pairs AT (or AU) and GC a

plethora of other so-called non-canonical pairs with two or three direct H-bonds can occur and that tertiary interactions between secondary structural elements are often crucial for the overall fold.¹¹⁻¹³ Bases may also be linked via standard H-bonds and additional water-mediated interactions¹⁴⁻¹⁷ or C-H...O/N contacts.¹⁸⁻²⁰ From the fact that difluorotoluene, a non-polar isostere for T, codes efficiently and specifically for A in DNA replication it has even been concluded that not H-bonding but shape complementarity is the crucial aspect in nucleic acid structure and recognition.^{21,22} Eventually, it is important to realize that a more general concept of H-bonding begins to emerge. This concept involves not only N-H and O-H as donor groups, but also C-H and not only O and N as acceptors, but also π -systems.^{23,24} Nevertheless, standard H-bond interactions between bases are of utmost importance. Therefore, the following analysis is confined to this type of interactions.

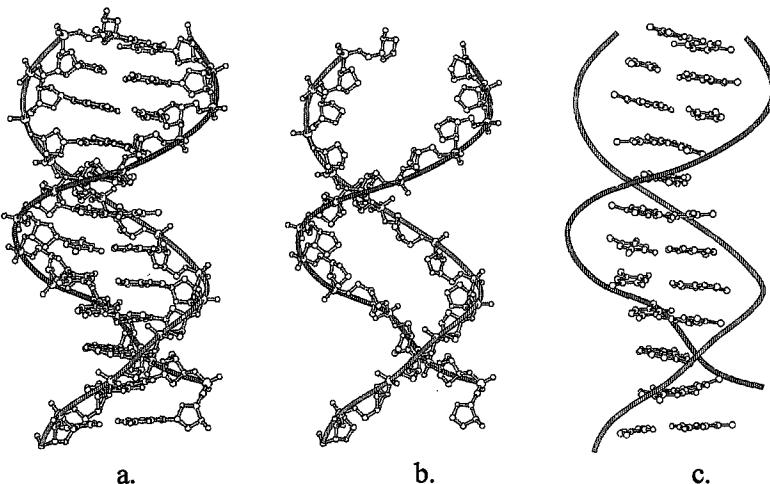


Fig. 1. X-ray structure of a B-DNA dodecamer (PDB code: 1bna).² a. complete structure. b. sugar-phosphate backbone only. c. base pairs only. Note, that as for almost all x-ray structures the hydrogen atom positions were not determined. The ribbons are drawn through the P atom of the phosphate group. In this and the following Figures the nucleic acid structures and polyads have been drawn with the programs PREPI (S. A. Islam, M. J. E. Sternberg, Imperial Cancer Research Fund, London, U.K.) and InsightII from Accelrys, Inc.

This review is focused on higher-order base motifs where more than two bases are linked by H-bonds. We call these nucleic acid building blocks base triads, tetrads, pentads, ... and in general base polyads. Alternative names for these base motifs are triple, quartet, quintet, Multistranded structure that

contain usually a few successive polyads are called triplexes, tetraplexes, pentaplexes. Alternative names are trimers and quadruplexes. Note, however, that also single base polyads can occur in nucleic acids.

2. Biological significance of nucleic acid structures with base polyads

It has been known for a long time that nucleic acids can form structures consisting of more than two strands.²⁵ It is very likely that such structures will contain higher-order base motifs. Originally, no biological function could be assigned to multi-stranded nucleic acids. The last decade, however, has seen an increased interest in structures containing base polyads. The most recent exciting observation in this field is related to the fragile X syndrome, a disease considered to be the most common cause of inherited mental retardation. It is well-known that the disease is due to an expansion of the CGG repeat in the 5'-untranslated region of the *FMR1* gene. This genetic disorder prevents the formation of the FMRP protein and thus leads to the disabling cognitive effects. It has now been shown that FRMP binds with high affinity to mRNA containing a G-tetraplex with G-tetrads as a structural element.²⁶ Tetraplex forming sequence motifs also occur in telomeres at the ends of linear chromosomes. Telomeres consist of tandem repeats of G-rich sequences associated with various proteins including telomerase, an RNA-dependent DNA polymerase. Proposed telomere functions are maintenance of the structural integrity of the genome and ensurance of complete replication at the chromosome termini. It has been shown that the G-rich sequences form G-tetrads and that G-tetraplex structures inhibit telomerase activity.²⁷⁻²⁹ Similar sequence motifs do also occur in regulatory regions of oncogenes. Recently, it has been proposed that targeted control elements of the *c-myc* oncogene adopt an intrastrand fold-back G-tetraplex.³⁰ A tetraplex structure has also been found for a centromere sequence.³¹ Further potential tetraplex forming sequence motifs occur in the promoter regions of many genes, at recombination hotspots and in triple repeat DNA that is responsible for various human genetic disorders by expansion of the number of repeats of a triplet sequence.³² Finally, various proteins are known to recognize DNA tetraplex structures.³³ Triplet forming oligonucleotides play a basic role in gene targeting.³⁴⁻³⁶ From these biological functions it is obvious that tetraplex DNA structures are a potential target for drug design.³⁷⁻⁴² Both the determination of three-dimensional nucleic acid structures with base polyad motifs and

computational studies can greatly improve the basic understanding of the biological functions.

3. Base pair classification and polyad topologies

With the increasing number and importance of non-canonical base pairs there have been attempts to compile all possible base combinations with at least two H-bonds.⁴³⁻⁴⁵ In view of the large number of possible non-canonical base pairs there clearly is a need for classification. Recently, such a classification has been proposed for RNA base pairs.⁴⁶ It is based on the fact that (approximately) planar base pairs with two or three H-bonds involve one of three edges: the Watson-Crick edge, the Hoogsteen edge and the sugar edge (which includes the 2'-OH group), Fig. 2. The second geometrical parameter necessary for the classification is the relative orientation of the glycosidic bonds of the interacting bases (*cis* or *trans*). This gives rise to 12 base pair types with at least two base-base H-bonds. Further relevant parameters are the relative local strand direction and the *syn* or *anti* orientation of individual nucleotides.

Base polyads are understood as base complexes linked by H-bonds. Even though many of these base motifs are approximately planar the notion of base polyads as used here does also cover strongly non-planar structures. The selection criterion is the base-base H-bond link and not the overall planarity of a certain base polyad. We adopt the following geometrical criteria to classify an interaction of the type D-H...A (D – H-bond donor, A - H-bond acceptor) as an H-bond: distance H...A < 2.5 Å, angle D-H...A > 90°. Interactions with smaller angles or larger distances are not considered as H-bonds.

In contrast to base pairs comprised of two bases only, the formation of base polyads requires that one or more bases interact with at least two neighboring bases. The simplest case of a base polyad is a triad where the central base is linked to two base neighbors, which in turn are not linked by H-bonds with each other. This leads to a non-cyclic triad topology. Alternatively, all three bases can interact with two base neighbors and this results in a cyclic triad. These are the only two possible triad topologies. The number of polyad topologies for a given number of bases depends on the number of neighbors that can interact with a base via H-bonds. If the maximum number of neighbors were two, the resulting topologies would be rather restricted. Independently of the number of bases involved, only unbranched cyclic or non-cyclic chains could occur. If, however, interactions of one base with three or more other bases are taken into account, the number of possible topologies becomes larger and larger with an increasing number of bases forming the

polyad motif. From the currently known experimental structures we know already that the maximum number of base neighbors may be three. So the topology space is not restricted to the simple unbranched cases. Therefore, we have recently proposed a topological classification of base polyads.⁴⁷

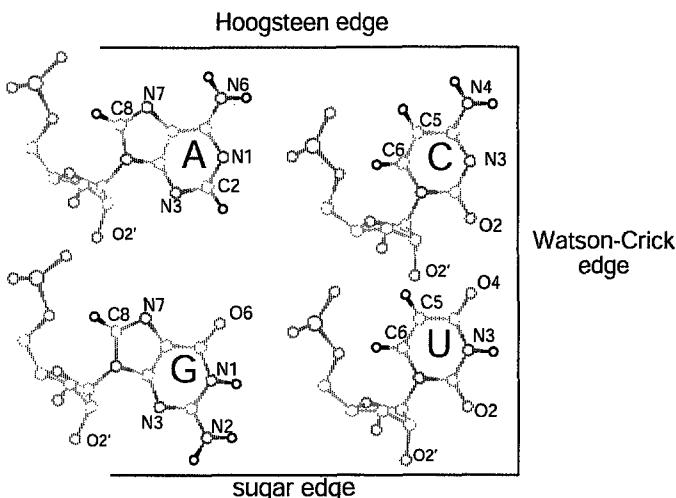


Fig. 2. Watson-Crick, Hoogsteen and sugar edges of RNA bases according to the classification of Leontis and Westhof.⁴⁶ In the sugar part of the nucleotides the hydrogen atoms have been removed. Potential H-bond donor/acceptor atoms in the edges are:

- Watson-Crick edge [G: N2, N1, O6; A: (C2), N1, N6; T, U: O2, N3, O4; C: O2, N3, N4]
 - Hoogsteen edge [G: O6, N7, (C8); A: N6, N7, (C8); T, U: (C5), (C6); C: N4, (C5), (C6)]
 - sugar edge [G: N2, N3, (O2'), A: (C2), N3, (O2'); T, U: O2, (O2'); C: O2, (O2')]
- C and O2' atoms are given in parentheses. With slight modifications this classification can also be applied to DNA bases.

Table. 1. Topologies of base polyads occurring in currently known experimental structures (see Table 2). The number of different topologies for a given number of bases N depends on the number of neighboring bases of each individual base constituent of the polyad. Here, it is assumed that bases in polyads can have 1, 2 or 3 neighbors. The maximum number is indicated by n_{\max} . A shorthand notation describing the topologies is $3_{n_3}, 2_{n_2}, 1_{n_1}$, where n_3 , n_2 and n_1 is the number of bases with 3, 2 and 1 neighbors, respectively. The sum of n_1 , n_2 and n_3 is equal to N. In some cases different topologies are described by an identical notation. Therefore, the polyad topologies are also described by a topology class index. The first part of the topology classification is given as a Roman numeral and indicates the number of

Polyad bases. The second part of the classification is simply a counting index i. The general description of a base polyad is thus given by the notation $N.i(3_{n_3}, 2_{n_2}, 1_{n_1})$. The unbranched non-cyclic polyads have the notation $N.1(2_{N-2}1_2)$. The notation for the unbranched cyclic structures is $N.2(2_N)$.

Polyad	n_{\max}	Topology classification	Topology	Examples (PDB code)
Triad	2	III.1 (2_11_2)		DNA triplex (1d3x) ⁴⁸
Triad	2	III.2 (3_2)		RNA/protein complex (1qa6) ⁴⁹
Tetrad	2	IV.1 (2_22_1)		RNA ribozyme (1gid) ⁵⁰
Tetrad	2	IV.2 (2_4)		DNA tetraplex (139d) ⁵¹
Tetrad	3	IV.3 (3_11_3)		RNA aptamer (1ddy) ⁵²
Tetrad	3	IV.4 ($3_12_21_1$)		RNA aptamer (1flt) ⁵³
Pentad	3	V.3 ($3_12_31_1$)		DNA tetraplex (1jjp) ⁵⁴
Hexad	3	VI.8 ($3_22_21_2$)		DNA tetraplex (1eeg) ⁵⁵
Hexad	3	VI.12 ($3_22_21_2$)		RNA aptamer (1ddy) ⁵²
Heptad	3	VII.3 ($3_32_11_3$)		DNA tetraplex (not yet deposited) ⁵⁶
Heptad	3	VII.1 (2_21_5)		RNA pseudoknot (437d) ⁵⁷
Octad	3	VIII.3 (3_4)		RNA tetraplex (1j8g) ⁵⁸

Topologies occurring in experimental structures currently known are compiled in Table 1. To explore polyad properties, questions at different levels have to be addressed. From a topological point of view it is interesting to ask: Which of the theoretically possible topologies do actually occur in experimentally determined nucleic acid structures? What is the maximum number of bases that form an uninterrupted network linked by base-base H-bonds? What is the maximum number of bases a single base can interact with? The next step would be from topology to geometry. Here, the focus of interest may be on different types of H-bond patterns, planar versus non-planar geometries and on the location of metal ion binding sites. It is also important to know whether the polyad geometries are primarily governed by intrinsic properties of the polyads alone or if they are primarily enforced by the nucleic acid environment the polyads are embedded in. Finally, energetic properties are relevant. They include stability, interaction energy and possibly cooperativity. All the above aspects can be studied by a combination of experimental and computational approaches.

4. Occurrence of base polyads in experimentally determined nucleic acid structures

Due to the biological significance of polyad containing nucleic acids an increasing number of three-dimensional structures of nucleic acid triplexes and tetraplexes has been determined within the last decade. This has substantially improved our understanding of the biological functions. The structures have been described in detail in a number of excellent reviews.⁵⁹⁻⁶³ A further source of base polyads are RNA structures. The structural diversity of RNA has first been recognized in tRNA structures solved in the late 1970s.¹³ Within the last few years very interesting structures of RNA pseudoknots, aptamers and ribozymes have been reported. The most recent highlight in RNA structural biology⁶⁴ is represented by the high-resolution structures of the small 30S ribosomal subunit of *Thermus Thermophilus* at 3.0 and 3.05 Å resolution (PDB codes: 1fjf, 1fjg),⁶⁵⁻⁶⁷ of the large 50S subunit from *Haloarcula marismortui* at 2.40 Å (PDB code: 1ffk)⁶⁸ and of the complete 70S ribosome of *Thermus Thermophilus* at 5.5 Å (PDB code: 1gix, 1giy).⁶⁹ Finally, base polyads play a role in supra-molecular chemistry. Lipophilic nucleobases have been shown to extract metal ions from water into organic solvents. X-ray structures of complexes formed from a guanosine analogue display stacked G-tetrads with Pb²⁺ or K⁺ located between the tetrad planes.^{70,71} In addition, interesting polyad topologies have been designed using metal-modified

bases.^{72,73} Although in the latter case the base-base H-bonds are replaced in part by metal links, it would certainly be worthwhile to compare these structures with other base polyads linked by H-bonds only. Of particular interest is the recent observation that in the presence of alkali metal ions guanosine nucleoside forms relatively regular tubular polymeric structures.⁷⁴ Finally, it should be noted that in addition to structural information thermodynamic and kinetic studies of the tetraplex formation are of utmost importance.⁷⁵

In the following an overview on experimentally determined base polyad containing nucleic acid structures deposited at the PDB is given, see Appendix (Table 2). In many cases the triplex or tetraplex fold formed from on two, three or four strands is the major determinant of a structure. The compilation in Table 2 is intended to be comprehensive for this structure type. On the other hand, especially in RNA structures only a few or even only one polyad(s) may occur. Comprehensiveness for these structures would require a systematic analysis of all currently known RNA structures. Recently, a database of non-canonical RNA base pairs has been set up.⁷⁶ In addition to base pairs it also lists 331 base triads and 25 base tetrads (March 7, 2002). We have done a few random checks of the database and noticed that it represents a rather good starting point for further analysis but is not comprehensive. Table 2 includes only a few representative cases of polyads in RNA structures.

4.1 Triads

Base triads do, of course, occur in nucleic acid triplexes.^{59,60,62} However, tetraplex structures may also contain triads and in RNA structures triads often play a crucial role. Triplexes are formed by the interaction of a third strand in the major groove of a double helix. The duplex has to be composed of a homopurine-homopyrimidine sequence (purine – R, pyrimidine – Y). The third strand can bind in a parallel or antiparallel orientation to one of the duplex strands. In parallel orientation, a homopyrimidine third strand binds to the homopurine strand of the duplex (RYR). This leads to the two canonical triads TAT and C+GC. Protonation of C (C⁺) at N3 is required for the formation of two H-bonds between C and G. Therefore, parallel triplexes are pH dependent. These structures have two canonical base triads TAT and C+GC. For an antiparallel orientation of the third strand relative to the binding duplex strand, a homopurine sequence is required that binds to the homopurine strand of the duplex (RRY). This results in the canonical triads GGC, AAT and TAT, where however the TAT triad is different to the corresponding triad in parallel triplexes. In addition to these standard triads, triplexes can also accommodate non-canonical base triads. Fig. 3 shows the two canonical triads C+GC and TAT in an intra-molecular triplex consisting of a DNA duplex and

one RNA strand. Only recently, it has been shown that base triads may also occur in DNA tetraplexes.⁷⁷⁻⁸⁰ Finally, it should be noted that triplexes with modified bases are known, as are triplexes where two strands have been replaced by peptide nucleic acids (PDB code: 1pnn).⁸¹

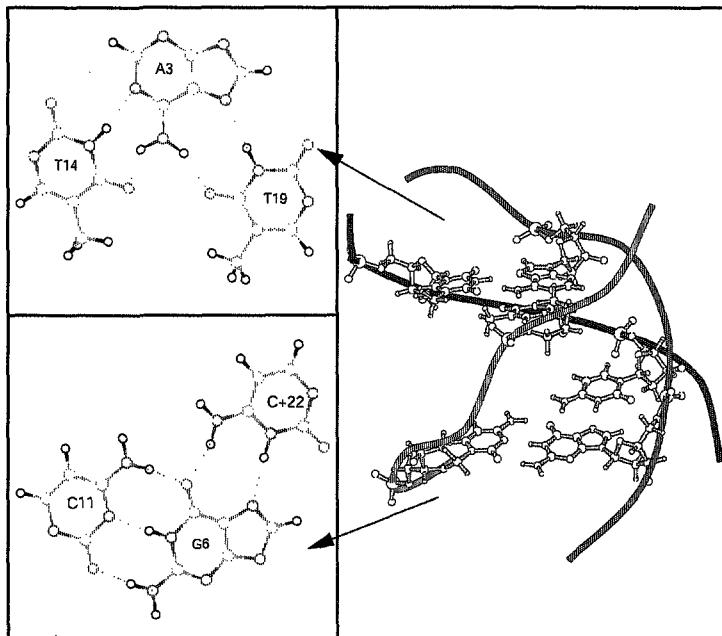


Fig. 3. Canonical base triads TAT and C+GC in an intramolecular DNA triplex structure solved by NMR spectroscopy (PDB code: 1d3x).⁴⁸ The triplex is linked by hexakis ethylene glycol units (EG) and has the sequence d(AGAGAGAA-(EG)₆-TTCTCTCT-(EG)₆-TCTCTCTT). The triads have a non-cyclic topology III.1(2₁₁). In this and the following Figures the dotted lines indicate H-bonds.

In addition to the relatively regular triplexes, base triads occur in many RNA structures. Within the last decade, RNA structural biology has come of age and now provides a rapidly increasing number of new structures with rather different structural motifs.¹¹⁻¹³ Therefore, RNA structures represent a treasury of higher-order base complexes. It is very likely that the first base triads ever seen at atomic detail are from yeast phenylalanine tRNA.⁸² Fig. 4 shows three base triads located in the central part of this structure. The ribosomal subunits (PDB codes: 1ffk, 1fjf, 1fg) contain of course a rather large number of base triads.⁶⁵⁻⁶⁸ However, small structures of pseudoknots, aptamers and ribozymes may also include more than one triad. For example, the

pseudoknot structure (PDB code: 437d) with just 28 nucleotides has four triads, where however two of them are part of a more complex motif.⁵⁷ This will be discussed in more detail below. Other triads occurring in RNA structures are listed in Table 1.

Among the two possible triad topologies the non-cyclic variant III.1(2₁1₂) is the preferred one. However, cyclic triads with the topology III.2(2₃) do occur in a ribosomal protein/RNA complex (PDB code: 1qa6), for example, Fig. 5.

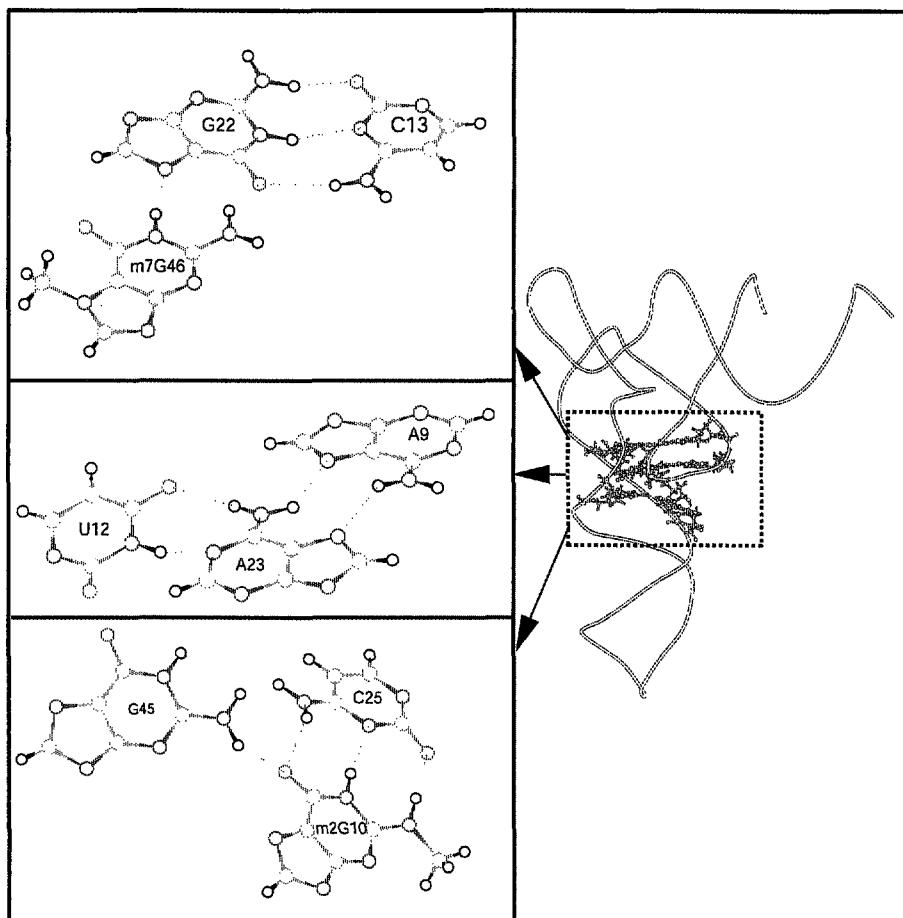


Fig. 4. Base triads in the crystal structure of yeast phenylalanine transfer RNA (PDB code: 4tna).⁸² All triads have the non-cyclic topology III.1(2₁1₂).

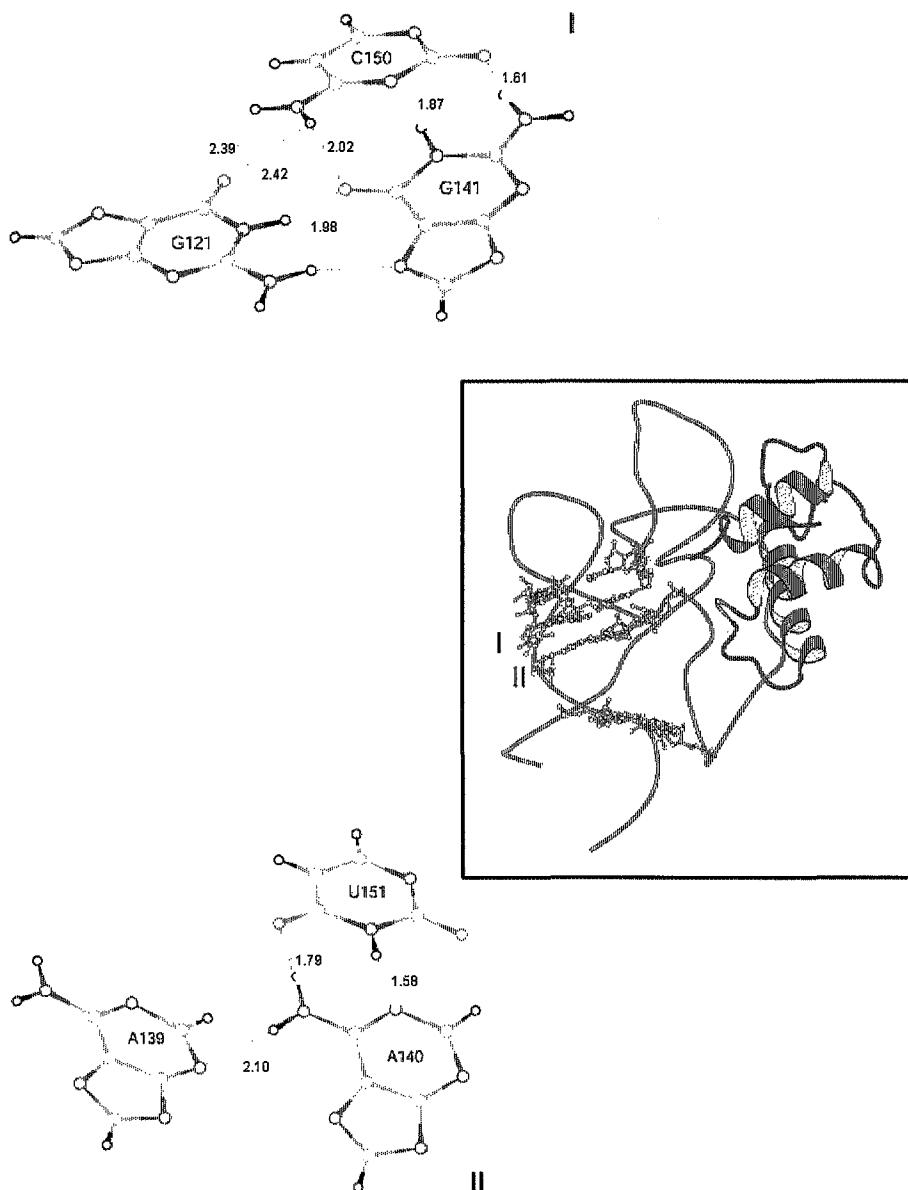


Fig. 5. Two cyclic and two non-cyclic triads in an RNA-protein complex (PDB code: 1qa6).⁴⁹ Example I: G121, G141, C150; topology III.2(2₃), example II: A139, A140, U151; topology III.1(2₁2). Only chains A (protein) and C (RNA) of the nucleic acid-protein complex are shown.

4.2. Tetrads

Base tetrads have primarily been found in DNA tetraplex structures.^{59,61} Most of the structures studied contain G-rich sequences and the G-tetrad is therefore the basic building block. The first three-dimensional G-tetraplex structures have been reported in 1992.⁸³⁻⁸⁵ It is now known that different folding different topologies and strand orientations are possible. In addition, the nucleotides may adopt a *syn* or *anti* orientation with respect to the glycosidic bond linking the base with the sugar moiety. Metal or NH₄⁺ ions play an essential role in tetraplex formation. The G-tetrad is approximately planar and the guanines interact via their Watson-Crick and Hoogsteen edges forming two H-bonds. A parallel-stranded DNA tetraplex with stacked G-tetrads is shown in Fig. 6. The four guanine oxygen atoms of the G-tetrad with a negative partial charge point towards the tetrad center. Therefore, the metal ion binding site is very likely located in the tetrad interior. Most of the known tetrad structures unfortunately provide no direct information on the exact location of these binding sites. There is, however, a crystal structure of a parallel-stranded G-tetrad at a resolution of 0.95 Å that clearly shows well-resolved Na⁺ ions both in the tetrad centers and between tetrads.⁸⁶ It should be noted that it is a notoriously difficult task to identify metal ion binding sites in biopolymer structures.⁸⁷ At this point theoretical studies including quantum-chemical, molecular dynamics (MD) and other approaches, such as Brownian-dynamics simulations, can play a very useful role.⁸⁸

Much less is known about non-G-tetrads. Even though the telomeric sequences are G-rich, they also include other bases. Therefore, the existence of non-G-tetrads cannot be excluded. Indeed, A-, T- and C-tetrads have recently been found in DNA tetraplex structures.^{89,91} Moreover, an interesting U-tetrad linked by C-H...O contacts only occurs in an exceptionally stable RNA tetraplex with G- and U-tetrads.⁹² Very recently a high-resolution x-ray structure for the same sequence has become available, that allows for a comparison between the x-ray and NMR results.⁵⁸

In addition to tetrads comprised of just one base type, mixed tetrads have been observed. Approximately planar GCGC tetrads occur in a tetraplex structure with the fragile X syndrome repeat sequence (PDB code: 1a8w)⁹³ and a tetraplex structure from adeno-associated viral DNA (PDB code: 1a8n).⁹⁴ On the other hand, significantly non-planar GCGC tetrad are also known, see Fig 7. A particularly interesting observation has been made for GCGC tetrads. In Na⁺ solution the GCGC tetrads exist. In K⁺ solution, however, the distance between the two GC pairs increases thereby disrupting the inter-base H-bonds.

Tetraplexes can also be formed from C-rich sequences at acidic pH. This tetraplex architecture is called the i-motif.^{59,61,95} It basically consists of two head-to-tail, intercalated, parallel-stranded duplexes held together by

hemiprotonated CC+ pairs. Therefore, usually no base tetrads are formed in these structures. However, TGGT tetrads have been found in the interacting loop regions.³¹ This tetrad is formed by dimerization of two GT pairs via GG H-bonds and is significantly non-planar. Twisted GCGC tetrads have also been found in another tetraplex motif called bi-loop.⁹⁶

Thus far only very few base tetrads have been identified in RNA structures. The database of non-canonical base pairs lists 25 examples in four different structures: two ribosomal subunits (PDB code: 1ffk;⁶⁶ PDB code: 1fjg;⁶⁷), a ribosomal frameshifting pseudoknot (PDB code: 437d)⁵⁷ and the malachite green aptamer (PDB code: 1flt).⁵³ We have repeated the analysis for the pseudoknot and the aptamer structures. The tetrad G24.G29.(C7).A31 in the aptamer could be confirmed. C7 is directly H-bonded to the three surrounding bases G24, G29 and C7 and this leads to a tetrad structure with topology IV.3(3₁3). For the pseudoknot it turned out that the tetrad motif A25.C8.G12.C26 listed in the database is more complex, see Fig. 11. This structure basically contains four base triads. However, the two triads C26.G12.C8 and C14.G7.A24 are linked by A25 and this finally leads to the heptad motif C26.G12.C8.A25.C14.G7.A24. Further tetrads occur in the P4-P6 group I intron ribozyme domain (PDB code: 1gid),⁵⁰ in an RNA with the sequence from an HIV-1 dimerization initiation site (PDB code: 462d),⁹⁷ in the malachite green aptamer (PDB code: 1flt)⁵³ in the vitamin B12 aptamer (PDB code: 1ddy).⁵² The tetrad structure from the P4-P6 domain is shown in Fig. 8. In all these tetrads the maximum number of neighboring bases is either two or three. The authors of the pseudoknot structure (PDB code: 437d) have pointed out that the base C8 forms a tetrad with G12, A25 and C26 and that this is the first report of one base H-bonded to three other bases.⁵⁷ A detailed analysis however shows that C8 is only linked to A25 and G12 via H-bonds of an appropriate geometry. The shortest distance between one of the two hydrogen atoms of the exocyclic amino group in C26 and O22 of C8 is 2.77 Å and the N4-H...O2 angle is 54°. Hence, both geometrical parameters are outside the adopted H-bond limits and C8 cannot be considered H-bonded to C26.

As already mentioned in the introduction tetrad structures are also found in structures used in supra-molecular chemistry.⁷⁰⁻⁷⁴ The topologies found in the structures analyzed and shown in Table 1 are IV.1, IV.2, IV.3 and IV.4.

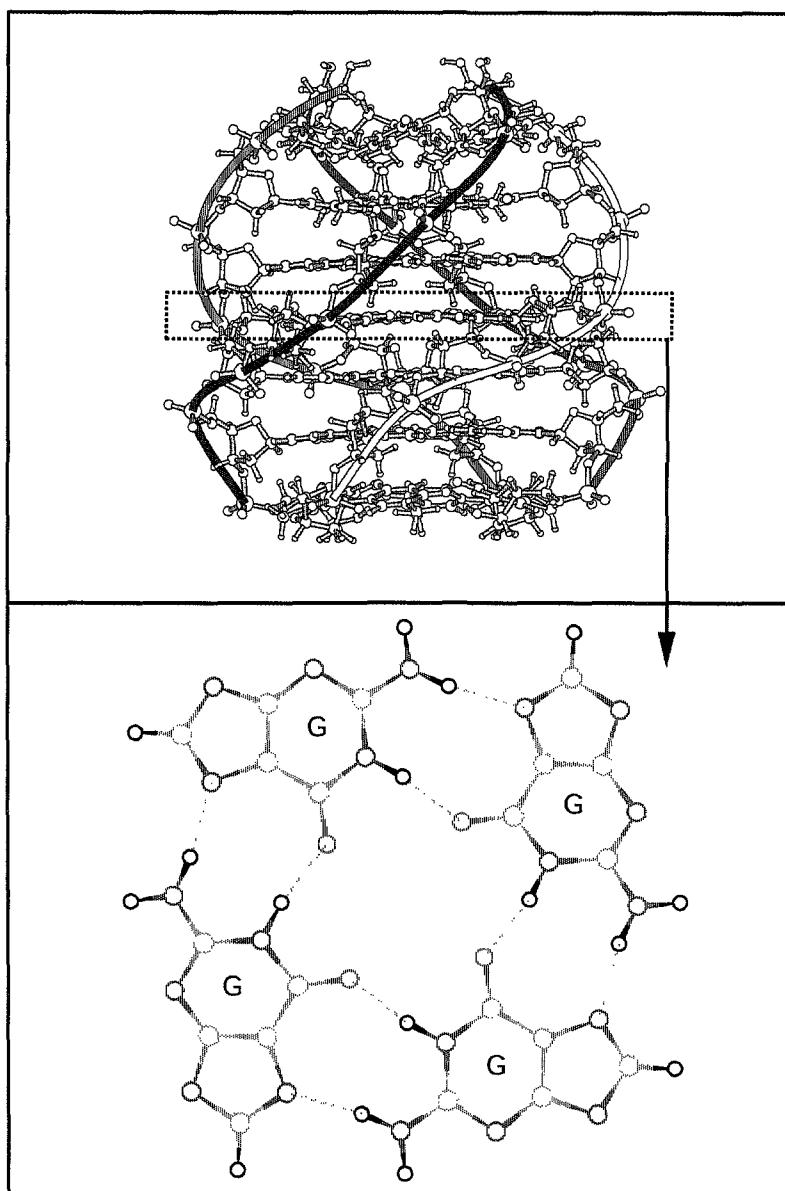


Fig. 6. Parallel-stranded DNA tetraplex formed from the *Tetrahymena* telomeric sequence d(TTGGGGT) solved by NMR spectroscopy (PDB code: 139d).⁵¹ The structure contains four stacked G-tetrads in the center and additional T-tetrads. The latter ones are underdefined, however. The G-tetrads have a cyclic topology IV.2(24).

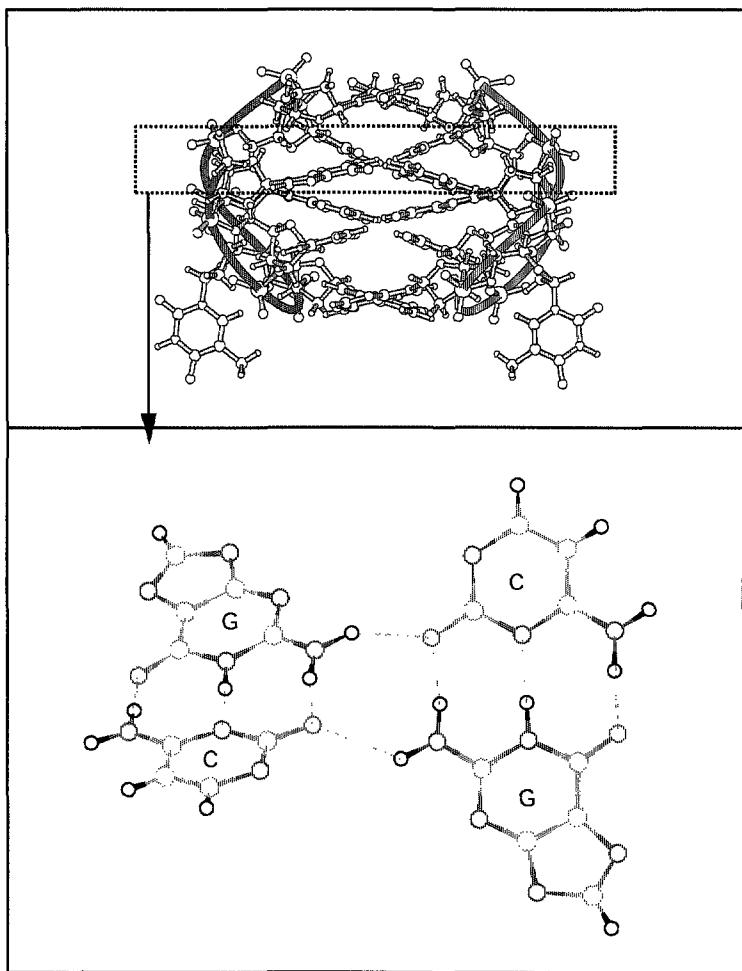


Fig. 7. Non-planar GCGC tetrads in a DNA tetraplex structure of d(GCATGCT) (PDB code: 184d).⁹⁸ The tetrad adopts the topology IV.1(2₄).

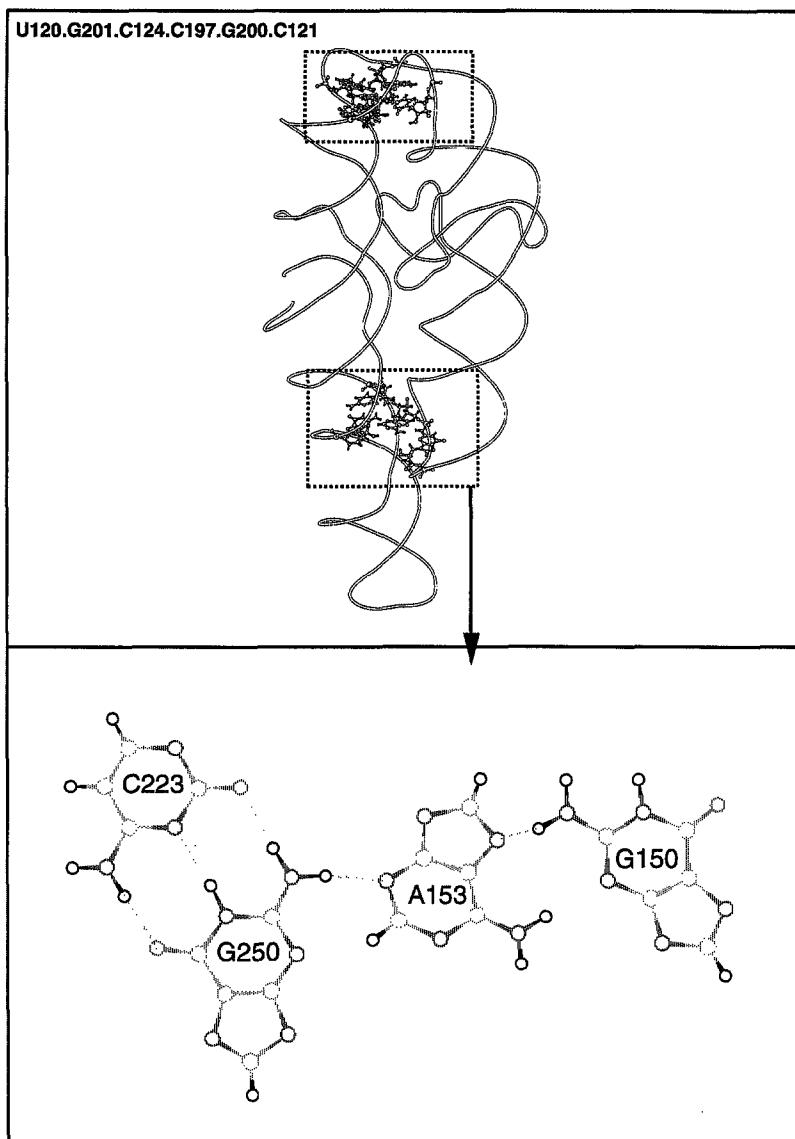


Fig. 8. Base tetrad and hexad in the crystal structure of the P4-P6 domain of the *Tetrahymena thermophyla* group I intron ribozyme (PDB code: 1gid).⁵⁰ Only chain A is shown. The tetrad has a non-cyclic unbranched topology IV.1(2₂1₂).

4.3. Pentads, hexads, heptads, octads

Base polyads can be comprised of more than four bases. To our knowledge, the first case of a cyclic pentad structure has been reported in 1999.⁹⁹ Cesium ions were found to induce pentameric assembly of DNA sequences d(T₈iG₄T) and d(T₄iG₄T) bearing the non-standard nucleobase isoguanine iG. Thus far there is however no high-resolution study of this pentaplex. The structure information was only indirectly derived from autoradiograms of polyacrylamide gels and additional quantum-chemical calculations. The tentative topology of the pentads in this pentaplex structure is V.2(2₅). Recently, it has been reported that isoguanines do also form tetraplex structures.¹⁰⁰

Interestingly a pentad (PDB code: 1jjp),⁵⁴ a hexad (PDB code: 1eeg),⁵⁵ a heptad (not yet deposited at PDB)⁵⁶ and even an octad (PDB code: 1j8g)⁵⁸ has been found in DNA or RNA tetraplex structures. This clearly shows that the geometry of tetraplexes exhibits a substantial structural diversity and is not that regular. In Figs. 9 and 10 a heptad and an octad occurring in an DNA and an RNA tetraplex respectively are shown. In both cases the central G-tetrad is approximately planar, whereas the additional bases exhibit a marked deviation from planarity.

Among the structurally more diverse RNA structures large polyads are also not unlikely, in particular when different polyad planes are linked by bases. Table 2 includes, for example, base hexad motifs in the P4-P6 group I intron ribozyme domain and the vitamin B12 aptamer. A heptad motif occurs in the ribosomal frameshifting pseudoknot structure shown in Fig. 11. Finally, it should be noted that a base hexad has also been designed with metal-modified bases.⁷³ In this case a central GG pairs is linked on both sides to AU pairs. This leads to the topology VI.10(3₂2₄).

In summary, except for the latter structure for which a high-resolution structure is available but not deposited at the PDB and for the cyclic pentad for which no high-resolution structure is available, Table 2 includes one pentad (PDB code: 1jjp, V.3), three hexads (PDB code: 1eeg, VI.8; PDB code: 1gid, VI.1; PDB code: 1ddy, VI.12), two heptads (PDB code: 437d, VII.1; not yet deposited at the PDB) and one octad (VIII.3; PDB code: 1j8g). As shown in Table 1 these polyad topologies are not confined to the simple non-cyclic or cyclic unbranched motifs where the maximum number of neighboring bases is two. Therefore, an increasing topological diversity of base polyads has to be expected.

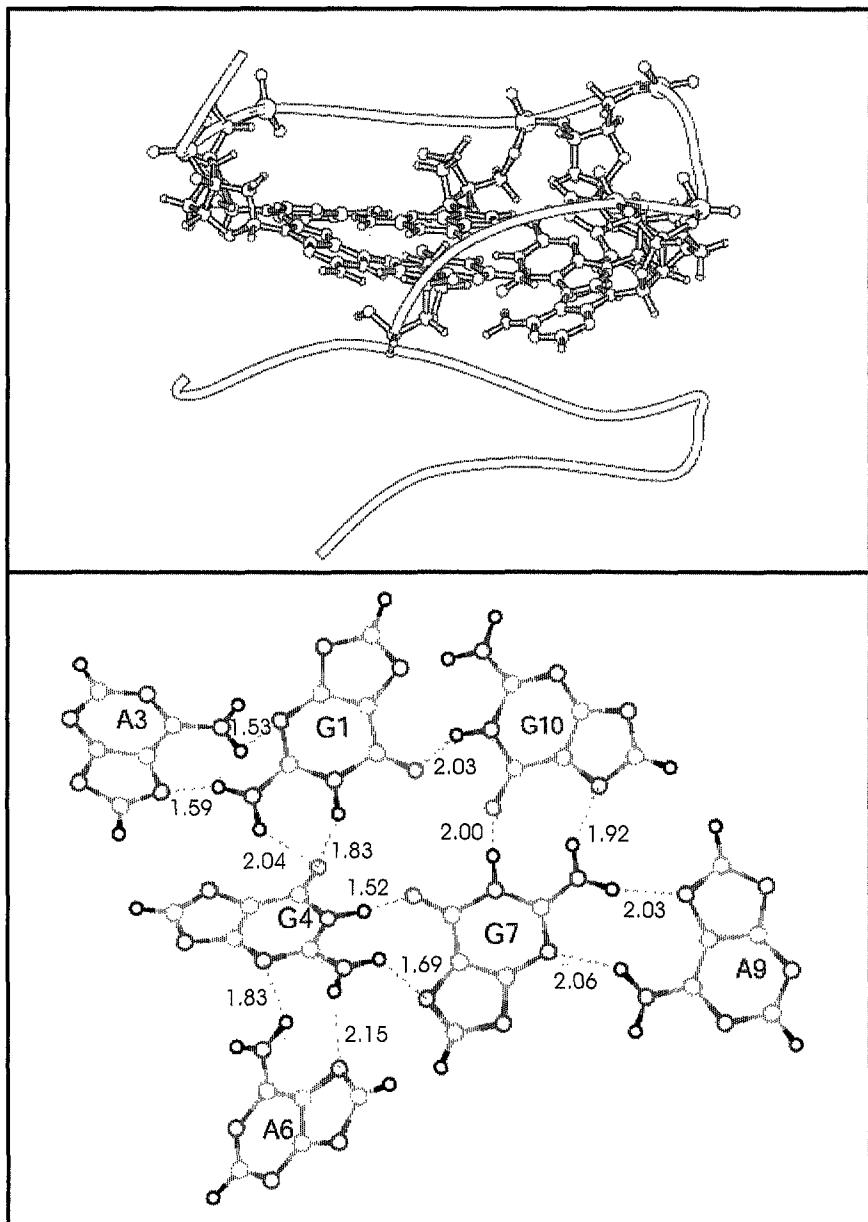


Fig. 9. Base heptad in a DNA tetraplex structure.⁵⁶ The heptad has the topology VII.3(3₃2₁3). At the dotted lines the H-bond distances are indicated.

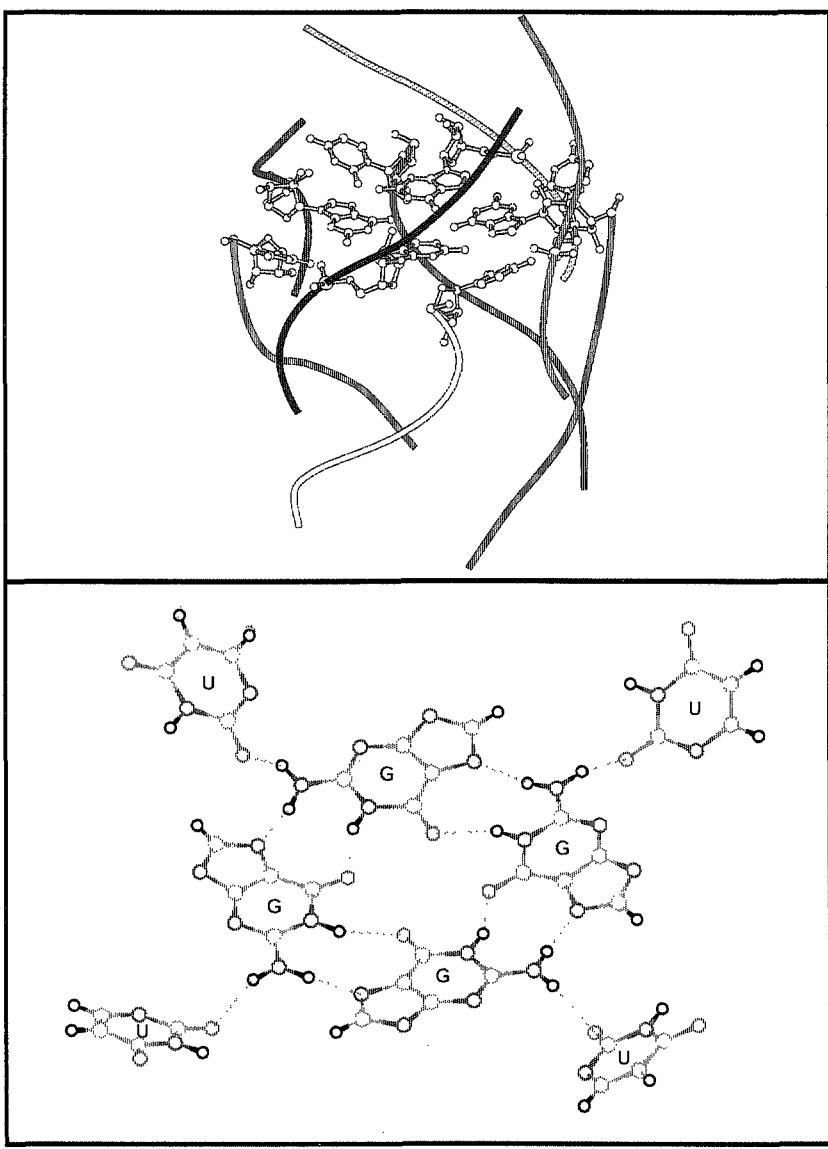


Fig. 10. Base octad in an RNA tetraplex structure (UGGGGU)₄ (PDB code: 1j8g).⁵⁸ Two tetraplexes stack on each other. The 5'-end uridines rotate around the backbone and form octads with the neighboring G-tetrad. The octad topology is VIII.3(3₄1₄).

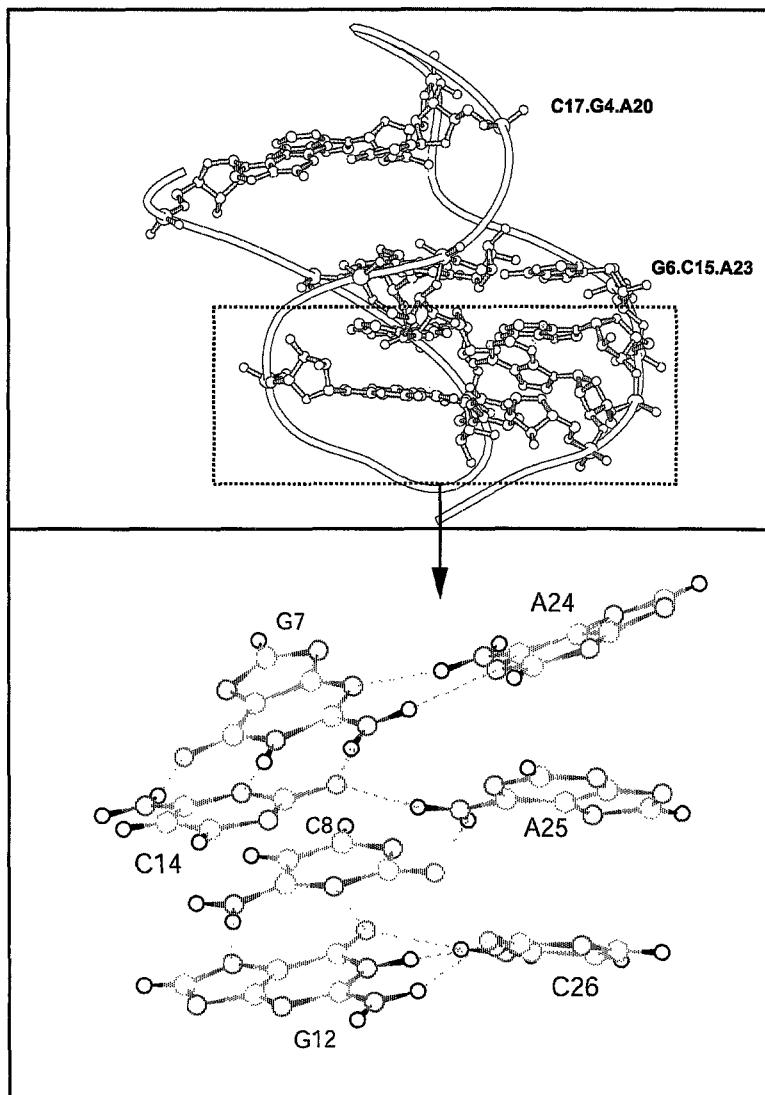


Fig. 11. Base triads and a heptad in the crystal structure of a pseudoknot from beet western yellow virus (BWYV) involved in frameshifting (PDB code: 437d).⁵⁷ The heptad is formed from two triads linked by A25. The topology is VII.1(2₂1₅).

4.4 Taking into account base-backbone and backbone-backbone H-bond interactions

The H-bond analysis in this study has so far been restricted to base-base interactions. For a full understanding of nucleic acid structure and function all possible H-bonds have to be taken into account. For regular conformations there are no other H-bond types than base-base interactions. In more irregular DNA structures and in particular in RNA structures however, the situation is completely different. The 2'-OH group in the RNA backbone is an additional H-bond donor/acceptor and in many cases base-backbone or backbone-backbone H-bond interactions are very likely. In these cases a comprehensive H-bond analysis may not only identify base triads to heptads but much larger clusters of nucleotides linked by H-bonds. In addition, the maximum number of neighboring nucleotides may increase. An impressive example is provided by the crystal structure of the P4-P6 domain of the group I intron ribozyme (PDB code: 1gid).⁵⁰ A164 is part of the base triad A139.A164.A178 and has thus two base neighbors linked by base-base H-bonds. If, however, all H-bonds are taken into account the number of nucleotide neighbors increases to six and now also includes C165, G181, G176 and A186. As a further analysis shows, this local motif is part of a much larger cluster of 50 nucleotides that forms an uninterrupted H-bond network.

In a next step C-H...O and C-H...π interactions could be taken into account. It has already been found that backbone-backbone C2'-H...O4' interactions in DNA and RNA and C5'-H...O2' interactions in the RNA backbone do frequently occur.^{19,101} If these interactions are considered in addition to the conventional H-bonds, the size of nucleotide clusters is likely to increase further.

5. Computational studies on base polyads

The experimental studies on biopolymer structures are increasingly supplemented by computational approaches.¹⁰²⁻¹⁰⁶ First, it has to be realized that computation is a sine qua non for experimental structure determination by diffraction methods and NMR spectroscopy themselves. In addition, independent computational studies can provide useful information on structure and dynamics of biopolymers not accessible, at least currently, by experiments. With regard to base polyads there are three fields that have to be mentioned here primarily: quantum-chemical studies of nucleic building blocks, MD simulations of medium-sized nucleic acids and structural bioinformatics.

Within the last decade important progress has been made in the reliability of MD simulations of solvated nucleic acids using improved force fields and, in particular, a better treatment of electrostatics by the particle-mesh Ewald method. For the first time unrestrained simulations have become possible. Starting out from experimental geometries it is now possible to explore the conformational space in the vicinity of the starting geometry and to study conformational transitions.^{102,104,105}

In contrast to the empirical force fields used in MD simulations of medium-sized biopolymers methods of ab initio quantum chemistry require no empirical information. In fact, quantum-chemical calculation are the most important tool for parametrizing empirical force fields. Unfortunately, ab initio studies that take into account electron correlation are rather time-consuming and therefore restricted to biopolymer building blocks. Nucleic acid base pairs have been a preferred object of these studies.¹⁰⁶⁻¹⁰⁸

A third area is the statistical analysis of experimental structure sets. Even though the number of experimentally known three-dimensional nucleic acid structures is still rather small as compared to proteins the statistical analysis and search for structural motifs becomes more and more important also for nucleic acids and nucleic acid/protein complexes.^{19,109}

It is believed that the synergistic interplay of these three approaches can be an especially effective computational strategy for a better understanding of nucleic acid structure and function.

5.1. Quantum-chemical studies

Quantum-chemical ab initio studies have been performed on triads¹¹⁰⁻¹¹⁶ on tetrads¹¹⁷⁻¹²⁷ and on a pentad⁹⁹ structure. Except for one recent paper¹²⁸ we are not aware of further semiempirical calculations on base polyads. Even though semiempirical approaches yield occasionally sensible results we know from studies on nucleic acid base pairs that a consistent theoretical picture of their structure and energetics did only emerge when ab initio calculations with a consistent treatment of electron correlation have become feasible.^{106,107} Hence, it is highly recommended to adopt non-empirical approaches for studies on base polyads. A detailed comparison of non-empirical, semiempirical and empirical approaches for the description of DNA base interactions has been performed by Hobza *et al.*¹²⁹

The-quantum-chemical methods follow mainly two lines, the Hartree-Fock (HF) approach possibly supplemented by the inclusion of electron correlation via second order Møller-Plesset perturbation theory (MP2) or alternatively a density functional theory (DFT) approach.^{106,130,131} Within the conventional scheme geometry optimization is usually performed at the HF level and in a

subsequent single point calculation electron correlation is taken into account. Full geometry optimizations at the MP2 level have occasionally been reported for base pairs²⁰ but were not yet applied to base polyads. Because MP2 scales with n^5 , where n is the number of basis functions, the size of the system under study is a critical quantity that currently prevents studies on larger systems at this level of theory. This is even more true for Møller-Plesset perturbation theory of higher order or for still other methods for the treatment of electron correlation. Local MP2 calculations,¹³² that are a special case of so-called linear scaling methods,¹³³ may be an efficient alternative for larger systems. However, their applicability to biopolymer building blocks remains to be shown.

Within the alternative DFT approach the exchange term of the HF method is replaced by a more general exchange-correlation functional including the exchange term and contributions to the correlation energy. This leads to a computationally less expensive method as compared to the HF/MP2 approach. In view of the computational demand of quantum-chemical ab initio methods taking into account electron correlation, it is not surprising that the density functional theory (DFT) has become very popular for calculations on large systems.

There are different types of exchange-correlation energy functionals. In the most simple case a local density approximation (LDA) is used. More advanced gradient-corrected methods take into account the density and its gradient. Becke has introduced functionals consisting of a mixture of Hartree-Fock theory and DFT combined with a gradient-corrected correlation. In general, gradient-corrected or hybrid calculations give more accurate results for hydrogen bonded systems. A series of comparisons indicate that DFT calculations lead to similar results for H-bonded base pairs as MP2 or MP2/HF calculations of nucleic acids.^{22,22,106,134}

A serious disadvantage of DFT methods is that the dispersion energy is not properly taken into account. This prevents the method to be applied to van-der-Waals complexes and this also means that base stacking cannot be studied with the DFT approach. A promising workaround is to combine the DFT method with an empirical dispersion energy term.¹³⁵ Interestingly, it has recently been shown that a non-local density functional formalism with a generalized gradient approximation and a Slater-type basis set, that is suitable for the description of long-range interactions, can reproduce the distance and angular dependence of the cytosine dimer stacking energy obtained from MP2 calculations.¹³⁶

In general, it can be stated that within the last years detailed information on the performance of different types of DFT approaches to biological systems has been accumulated. This makes the method a rather reliable tool and

enables studies on larger systems that were currently not possible with the conventional ab initio approach.

For the polyad studies described below the hybrid type DFT calculations with the Becke 3 term and the Lee, Yang and Parr exchange functional^{137,138} and also the BLYP method¹³⁹ have been used. In addition to the standard gaussian-type basis sets such as 6-31G(d) or 6-311G(d,p) the DZVP and TZVP basis sets optimized for DFT calculations have also been applied.¹⁴⁰ Finally, for polyad complexes with metal ions pseudo potentials have been adopted.¹⁴¹

For the calculation of interaction energies the basis set superposition error (BSSE) has to be taken into account. It leads to a lowering of the energy of the polyad with respect to the energies of the individual bases. BSSE is identically zero in the limit of a complete basis set and usually it is corrected by the counterpoise method. Details regarding BSSE may be found in other reviews.¹⁴²

Almost all of the calculations on polyads were performed for isolated base complexes thus neglecting effects of nucleic backbone, solvent and also of entropic contributions. Experiments most directly related to quantum-chemical results are gas phase investigations on base pairs.¹⁴³ Nevertheless, it makes also sense to compare the results of quantum-chemical studies on base pairs or polyads to complete three-dimensional nucleic acid structures. One should realize, however, that there is a long way to go from results of quantum-chemical studies on building blocks to nucleic acid structure and function. The power of this approach is that it can help to separate the effects of different parts of the nucleic acid, such as backbone and bases, on the overall structure. In addition, it provides useful information on charge distribution and electrostatic potentials relevant for the determination of interaction sites with cations such as K⁺, the most abundant cation in cells. Finally, the interaction energies can be studied in great detail. For example non-additive contributions can be determined.

To our knowledge the first quantum chemical ab initio study of a base triad was published in 1994 (minimal STO-3G basis set).¹¹⁰ In this study substantial positive or negative cooperative effects have been found. Almost all triads with Watson-Crick GC base pairs showed a positive cooperativity, whereas base triads containing Watson-Crick AT base pairs mostly exhibited a negative cooperativity. The study on triads and on metal ion/triad complexes taking into account electron correlation did appear in 1997.^{111,112} In contrast to the 1994 work it was found that metal-free and neutral triads AAT, GAT, CGC do not exhibit significant cooperativity, the GGC Hoogsteen triad being an exception, however. A marked non-additivity of the interaction energy was found for the triad/metal ion complexes and also for protonated metal-free

triads. Geometry optimization of base triads with various force fields resulted in many cases in T-shaped or stacked structures. In a 1999 paper on base triads a similar ordering of interaction energies was found, the protonated base triads being the most stable ones.¹¹³ In addition, it has been suggested that the neutral form of 8-thioadenine may be an alternative protonated cytosine for efficient triplex formation. A further paper on triads has addressed the question whether GC Hoogsteen wobble pairs can contribute to the stability of d(GCC) triplexes even though this pair forms only one H-bond between the exocyclic amino group of C and the acceptor atom N7 located in the Hoogsteen edge of G.¹¹⁴ The quantum-chemical DFT studies showed that the wobble pair is quite stable. From MD and statistical mechanics calculations it turned out, however, that the wobble pair can be included in triplex structures if neutral GCC triads are mixed with protonated GCC⁺ ones. A similar computational approach that combines quantum chemistry with MD and thermodynamic integration has also been used to study the effect of amino groups in the 2 and 8 positions of hypoxanthine.¹¹⁶ Hypoxanthine a guanine without the exocyclic amino group. Surprisingly, the calculations showed that an amino group in position 2 stabilizes both the duplex and the triplex. On the other hand, the presence of an amino group in position 8 (at the five-ring) stabilizes the triplex but destabilizes the duplex. Finally, the DFT approach has also been used to determine the Fermi contact contributions to various coupling constants and the isotropic chemical shifts in H-bonding regions of Hoogsteen-Watson-Crick base triads TTA and C+GC.¹¹⁵

Five years after the first quantum-chemical ab initio investigation on a base triad the first tetrad paper was published.¹¹⁷ In the work described a guanine tetrad without a metal ion has been studied. Now, additional reports on G-tetrads^{119,122,123} and further studies on A-,^{121,125} U-,^{122,124} C-,^{125,127} T-tetrads,^{124,125} as well as on the mixed GCGC,^{120,125} ATAT¹¹⁸ and AGAG tetrads¹¹⁸ have been reported. In some of these studies the influence of metal ions has been investigated. Finally, very recently the first study on a sandwich structure of two guanine tetrads with intercalating Na⁺ or K⁺ ions has been published.¹²⁶

All the tetrads studied thus far have a cyclic topology corresponding to the notation IV.2. Therefore, symmetry constraints have been applied in the calculations which substantially reduce the computational effort. For tetrads consisting of one base type only likely symmetries of the tetrads without metal ions are C_{4h} for planar systems and either C_4 or S_4 in non-planar structures. Note, however, that applying symmetry constraints may prevent the detection of relevant structures. A further problem of the tetrad studies is that not in all cases the minima obtained have been checked by frequency calculations.

Because of the size of the structures the DFT method seems to be the method of choice. The results are often compared to results obtained from conventional HF calculations without taking into account electron correlation. For the G-tetrad sandwich structure only HF results are available. The preferred DFT approach is the B3LYP method and a widely used basis set is 6-311G(d,p). However, for the G-tetrad the dependence of the results on both the functional used (B3LYP, BLYP, HTH) and on basis sets has also been investigated.^{122,123}

In the following we first discuss polyad properties without taking into account metal ions and proceed then with the effects of metal ions. In the very first tetrad paper the G-tetrad has been studied by means of Hartree-Fock (HF) and density functional (DFT) approaches at the HF/6-311G(d,p) and B3LYP/6-311G(d,p) levels of theory.¹¹⁷ The interaction energies taking into account the basis set superposition error are -62 and -66 kcal/mol for the HF and DFT approaches respectively. The calculated distance between the exocyclic oxygen atoms of diagonal bases is about 6 Å. The electrostatic potential is significantly negative in the central tetrad. Therefore, a cation placed here is expected to contribute to the further stabilization of the tetrad. Interestingly, the bases are linked by bifurcated H-bonds between N1-H, N2-H and O4. This H-bond pattern differs from the Hoogsteen pairing found in experimental structures. In a subsequent study the interaction of metal ions with the G-tetrad has been analyzed.¹¹⁹ In contrast to the metal-free tetrad in this case the Hoogsteen pairing seen in experimental structures, see for example Fig. 6, has been found. From these facts it has been concluded that the metal ions change the H-bonding pattern in G-tetrads. Surprisingly, in an independent study on G- and U-tetrads and on G-tetrad/metal ion complexes such an alteration has not been found.¹²² Therefore, a further study has tried to resolve this obvious discrepancy.¹²³ It turned out that the bifurcated structure does not correspond to a minimum at the HF level of theory whereas the DFT calculations yield minima for both the bifurcated and Hoogsteen H-bond pattern. However, the energy difference between the two conformations is very small and thus the preference for one or the other structure is dependent on the basis set adopted. At the HF/6-311G(d,p) level there is only one minimum corresponding to the bifurcated base pairing, whereas DFT-B3LYP and DFT-BLYP calculations yield additional minima for the base tetrad with the Hoogsteen-type H-bond pattern. The energy differences between the two tetrad conformations are very small, however. At the B3LYP/DZVP level the bifurcated S_4 -symmetric structure is the most stable one. At other levels of theory such as B3LYP/6-311G+(d,p)//B3LYP/6-311G(d,p) the Hoogsteen S_4 -structure is preferred and finally at B3LYP/6-311+G(2d,p)//B3LYP/6-311G(d,p) a planar bifurcated structure is most stable. In addition to the more

technical aspects these results show that the energy differences between different conformations are so small that forces exerted by the nucleic acid environment, ions and solvent can easily affect the H-bond pattern of the G-tetrad. The studies on G-tetrads have further shown that cooperativity contributes significantly to the total interaction energy.

Studies on non-G-tetrads comprised of one base type only have thus far been performed on A-,^{121,125} U-,^{122,124} C-^{125,127} and T-tetrads.^{124,125} The C-tetrad and the U-tetrad with an C-H...O base-base link are planar, but for the T-tetrad, the U-tetrad linked by an N3-H3...O4 interaction and A-tetrads with either N6-H6...N7 and N6-H6...N1 H-bonds non-planar structures are more stable than planar ones.

As a general rule, S_4 -symmetric polyads are somewhat more stable than C_{4h} -symmetric structures for the tetrads mentioned above. The interaction energy is dependent on the number and type of H-bonds linking the bases and on planarity. This means that the interaction energies for the G- and C-tetrads are substantially larger than for the A-, T- and U-tetrads. In all cases cooperativity plays an important role. The effect is especially large in planar structures where cooperativity may contribute more than 15% to the total interaction energy.

Except for A-tetrads, the electrostatic potential in the tetrad centre is negative with marked variations, however. The most electronegative potential is observed for G-, C- and U-tetrads. Cation binding is, however, not only affected by this potential but also by cavity size and, as will be discussed below, by hydration energy.

Studies on mixed tetrads have been performed for GCGC, TATA and GAGA.^{118,120} From the V-shaped structure of the AGAG tetrad it has been concluded that this tetrad may not be important in tetraplexes. The Watson-Crick type and Hoogsteen type ATAT tetrads are also non-planar the plane angle between diagonal bases being approximately 140°. A local energy minimum of the GCGC tetrad corresponds to an almost planar structure.

Various studies have explored the effect of different ions on G-tetrads.^{119,122,126} Small ions attract the oxygen atoms and lead to a significant deviation from planarity. Na^+ seems to fit perfectly into an (almost) planar G-tetrad and for K^+ it has been found that a geometry with the ion located above the tetrad plane is more stable than a structure with the ion in the tetrad center.¹²²

A further question that has been addressed is the sequence of binding energies for different metal ions.^{119,122} From calculations on the G-tetrad sandwiches with Na^+ and K^+ it has been found that the binding energy between the metal ions and the base tetrads is larger for Na^+ than for K^+ . However, after the hydration energy correction this ordering is reversed.¹²⁶ This results

confirms an idea originally suggested by Hud *et al.* that the relative energies of hydrating are the dominating effect for ion selectivity.¹⁴⁴

An experimental study on DNA pentaplex formation has been supplemented by quantum-chemical calculations on iso-guanine pentad complexes with K⁺ or Cs⁺ restraining the base pentad to a planar geometry.⁹⁹ The calculations show that K⁺ is located in the plane center, whereas Cs⁺ is sandwiched between two pentad planes. The latter structure seems to be the largest base complex ever studied by ab initio quantum chemistry.

In summary, the quantum-chemical studies on base polyads have provided many interesting results on geometries, energies and other properties of these complexes. Because of the size of base polyads there seems to be a preference for DFT studies. The results can either directly be used in combination with experimental approaches such as NMR spectroscopy, for example, or they can lead to a better understanding of the underlying principles of nucleic acid structure.

5.2. Molecular dynamics simulations

Quantum-chemical calculations can provide a detailed picture of the geometric, energetic and charge distribution properties of base polyads. However, for a full understanding of the function of polyad-containing nucleic acids the complete nucleic acid structure has to be taken into account. Therefore, MD simulations of complete medium-sized molecules in a solvent are a very useful complementary tool. More and more structurally diverse RNA structures are studied by means of this approach. One example is a MD study of the flavin-mononucleotide RNA aptamer (PDB code: 1fmn) which contains one base triad.¹⁴⁵ The following discussion is focused on DNA triplex and tetraplex structures, however. First, it should be noted that restrained MD is usually an integral part of NMR structure refinement. Work in this field is not described below. In addition, there are simulations that usually start out from experimental structure but then attempt to study properties currently not available from experimental approaches. Furthermore, simulations may be related to other experiments than structure determination. We review in the following work in the latter two fields on triplexes¹⁴⁶⁻¹⁶⁵ and on tetraplexes.^{39,40,166-180} A few earlier model building and molecular mechanics studies on tetraplexes are compiled by Špačkova *et al.*¹⁶⁸

From a methodological point of view it has to be noted that the reliability of MD simulations on biopolymers, and in particular on nucleic acids, has been substantially improved since the particle-mesh Ewald summation for an appropriate treatment of long-range electrostatics has become available in the second half of the nineties.¹⁸¹ A second point that is worth mentioning refers

to the role of metal ions that are vital to tetraplex structures. These ions should exert significant polarization effects that are not properly accounted for in standard force fields. So, in this regard the results obtained have to be seen with caution.

The first MD simulation of a DNA triplex structure we are aware of has been reported in 1992.¹⁴⁶ At that time no high-resolution triplex structures were known and therefore structural information for building the initial structure was taken from fiber diffraction data. The simulation yielded insight into some conformational transitions such as the repuckering of the sugars in the purine strand that could be related to low-resolution NMR data on triplexes. For historical reasons it is also interesting to note that the total simulation time of 20 ps required 170 h of central processing unit time on a 3-processor Alliant FX40/3 computer.

Later MD studies on DNA triplexes were often devoted to the stability and the H-bond pattern of the base triads, to the conformational properties of the sugar-phosphate backbone, to the effect of modified nucleotides and to the hydration pattern.¹⁵³ Modified nucleotides include structures derived from 2-aminopyridine¹⁵⁴ and 8-aminoguanine¹⁶¹ a triplex with the phosphodiester linkage in two strands replaced by S-methylthiourea linkers,¹⁶⁰ 2'-aminoethoxy-substituted riboses¹⁶² and a -CH₂-lengthening of the internucleotide linkage.¹⁶⁴

As already noted, metal ions play a vital role in DNA tetraplex structures. The first MD study that did address this point was published in 1994.¹⁶⁶ It was found that the simulation could reproduce a few properties of the structure but could not correctly describe the metal ion nucleic acid interaction. This is not surprising as at that time the particle-mesh Ewald approach was not yet available. Later simulations do not suffer from this deficiency.

A nanosecond MD simulation of d(T₄G₄T₄) from 1999 has clearly shown a destabilization of the system when removing the metal ions from the central ion channel.¹⁶⁸ Interestingly, a very stable alternative conformation involving a guanine triad has been found in the simulations. This observation can be related to the fact that experimental DNA tetraplex structures may not only contain tetrads but also triads, pentads, hexads, heptads and octads. The two inner G-tetrads of the four stacked tetrad planes exhibit contrary to the experimental structure a bifurcated H-bond pattern.

As for triplexes the effect of structural modifications has also been studied for DNA-tetraplexes.¹⁷⁵ The simulations on structures containing inosine, 6-thioguanine and 6-thiopurine indicate that a nucleic acid stem formed from G-tetrads can easily incorporate inosine without any marked effects on structure and stability. On the other hand, 6-thioguanine and 6-thiopurine have more or less adverse effects. They cause a destabilization of the tetraplex structure

since the sulfur leads to steric conflicts and expels ions from the central ion channel, even though thiogroups interact favourably with Na^+ , in general. An incorporation of a single thioguanine does not cause major perturbations in the tetraplex structures, whereas thioguanine tetraplexes collapse immediately. In the absence of cations the inosine tetraplexes are disrupted immediately in contrast to guanine tetraplex structures, since inosine tetrads are stabilized only by a single H-bond between neighbor bases and guanine tetrads are linked by two.

In a recent very extensive study 20 independent simulations with a total simulation time of 90 ns have been performed for DNA tetraplexes with all-guanine and mixed guanine-cytosine tetrads.¹⁷⁶ The metal ion/tetrad interaction is the dominating force stabilizing the tetraplex structures containing all-guanine tetrads, whereas this interaction is weaker in the mixed tetrad. The metal ions have a clear influence on the GCGC tetrad conformation. In the presence of K^+ the sheared conformation is formed that has no H-bonds between the two GC pairs. On the other hand, Na^+ induces a closed conformation with a H-bond between the two Watson-Crick pairs. This observation is in agreement with experimental data.⁹³ According to the simulations, two cations are required for the stabilization of a tetraplex consisting of two G- and GCGC tetrads, respectively. Both cations are able to pass through the cavity of the tetrad planes without destabilizing the structures significantly. Finally, the thymidine loops of the tetraplex structure formed by d(GGGCTTTGGGC)₂ do not have the ability to form stable associations with cations.

In other nanosecond MD simulations on parallel¹⁷³ and antiparallel¹⁷⁹ G-tetrads the effect of coordinated cations on structure, flexibility and stability has been studied. A further MD study has started with G-tetraplex structure without any ions in the initial structure.¹⁷⁷ The tetraplex structure is not disrupted but undergoes small structural changes, which allow the Na^+ ions to move into the empty central ion channel. Even though not all potential ion binding sites were occupied during the time of simulation the structure remained stable.

It has already been noted that G-tetraplex structures are drug targets. Therefore, a few MD simulations have also been performed for G-tetraplex ligand complexes.^{39,40,174,178} For telomerase inhibitors of the acridine type it has been shown, for example, that the calculated relative binding energies are in agreement with experimental equilibrium binding constants.³⁹ MD simulations have been used to study an external stacking versus an intercalated binding mode of cationic porphyrins with an antiparallel d(AG₃[T₂AG₃]₃) tetraplex.¹⁷⁸ The external mode is driven by ΔH° , whereas intercalated binding is governed

by the entropy term. Taken together, external binding is preferred over intercalation.

In summary, MD simulations represent a very useful tool for analyzing nucleic acid structures with base polyad motifs. With increasing simulation times the sampling of the conformational space becomes more and more complete and this will greatly improve the reliability of the simulations. The extended timescale will hopefully also allow the identification of relatively slow conformational transitions.

6. Conclusions

A survey of the occurrence of base polyads in experimentally determined nucleic acid structures shows that these structural motifs exhibit a great variability in topology and geometry. The number of bases involved in polyads is ranging from three to eight and the topologies found include non-cyclic and cyclic unbranched patterns as well as branched patterns with cyclic and non-cyclic parts. In currently known structures the maximum number of neighboring bases linked to one base via H-bonds is three. The polyad motifs play an important role in the relative regular triplexes and tetraplexes as well as in structurally more diverse RNA structures ranging from small aptamers, pseudoknots and ribozymes with twenty to fifty nucleotides to the ribosome with several thousand subunit nucleotides. Computational studies on these motifs include quantum-chemical calculations on triads, tetrads and pentads as well as molecular dynamics simulations on DNA triplexes and tetraplexes. The quantum-chemical calculations yield interesting results on the intrinsic geometries and stabilities of polyads and on the role of metal ions. They can thus help to define in more detail the role of base pairs and base polyads as compared to effects that are due to the backbone. A challenge for future quantum-chemical studies is the treatment of larger systems such as complete dinucleotide steps taking into account both bases and the backbone. Molecular dynamics simulations can contribute to a better understanding of the dynamics and of solvation effects.

Appendix

Table 2. A listing of experimental nucleic acid structures containing H-bonded base polyads.^a

Molecule: Polyad	Topo- logy	Method	Year ^b	PDB/NDB Code
<i>tRNA(Phe):</i> A9.A23.U12 m2G10.C25.G45 C13.G22.m7G46	III.1 III.1 III.1	x-ray (2.50 Å) x-ray (1.93 Å) x-ray (2.00 Å)	1978 2000 2000	4tna/trna10 ⁸² 1ehz/tr0001 ¹⁸² 1evv/tr0002 ¹⁸³
<i>tRNA(Asp):</i> U8.A14.A21 A9.A23.U12 G10.U25.G45 PSU13.G22.A46	III.1 III.1 III.1 III.1	x-ray (3.00 Å)	1988	2tra/trna07 3tra/trna08 ¹⁸⁴
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G	IV.2	<i>x-ray</i> (2.30 Å)	1992	1d59/udl018 ⁸³
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G	IV.2	NMR	1992/ 1994 ^c	156d ^{84,85}
<i>RNA tetraplex:</i> G.G.G.G U.U.U.U	IV.2 IV.2	<i>NMR</i>	1992	1rau ⁹²
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G T.T.T.T	IV.2 IV.2	<i>NMR</i>	1993	139d ⁵¹
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G	IV.2	<i>NMR</i>	1993	143d ¹⁸⁵
<i>DNA tetraplex/protein complex (aptamer):</i> G.G.G.G	IV.2	x-ray (2.90 Å)	1993	1hut/pde013 ¹⁸⁶
<i>DNA triplex (RRY):</i>		NMR	1993	134d, 135d,

G.G.C T.A.T	III.1 III.1			136d ¹⁸⁷
DNA tetraplex (<i>aptamer</i>): G.G.G.G	IV.2	NMR	1994	148d ¹⁸⁸
DNA tetraplex (<i>telomeric sequence</i>): G.G.G.G	IV.2	NMR	1994	186d ¹⁸⁹
DNA triplex (YRY): T.A.T C+G.C G.T.A	III.1 III.1 III.1	NMR	1994	149d ¹⁹⁰
DNA triplex (YRY): G.T.A T.C.G	III.1 III.1	NMR	1994	177d ¹⁹¹
tRNA(Ser)/protein complex: G9.G13.A22 U8.A14.A21 DHU20.G15.C48 ^d	III.1 III.1 III.1	x-ray (2.90 Å)	1994	1ser/ptr004 ¹⁹²
DNA tetraplex : G.G.G.G	IV.2	x-ray (1.20 Å)	1994	244d/udf036 ¹⁹³
DNA tetraplex: G.C.G.C	IV.2	x-ray (1.80 Å)	1995	184d/udg028 ⁹⁸
DNA tetraplex (<i>telomeric sequence</i>): G.G.G.G	IV.2	NMR	1995	201d ¹⁹⁴
DNA tetraplex (<i>fragile X syndrome repeat</i>): G.G.G.G G.C.G.C	IV.2 IV.2	NMR	1995	1a6h ¹⁹⁵
DNA (<i>tetraplex</i>): G.G.G.G	IV.2	NMR	1995	1fqp ¹⁹⁶
TAR RNA/ <i>peptide complex</i> : U21.A10.U7	III.1	NMR	1995	1mnb ¹⁹⁷

<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G G.G.G.I	IV.2 IV.2	NMR	1995	230d ¹⁹⁸
<i>DNA peptide nucleic acid triplex:</i> CPN.A.TPN ^e CPN.G.TPN IPN.A.TPN	III.1 III.1 III.1	x-ray (2.50 Å)	1995	1pnn/pna001 ⁸¹
<i>RNA/FMN aptamer:</i> G10.U12.A25	III.1	NMR	1996	1fmn ¹⁹⁹
<i>RNA P4-P6 group I ribozyme domain:</i> A,B:U120.G201.C124.C197. G200.C121 A,B:G150.A153.G250.C223 additional triads	VI.1 IV.1 -	x-ray (2.50 Å)	1996	1gid/urx053 ⁵⁰
<i>DNA tetraplex:</i> G.G.G.G	IV.2	NMR	1996	1qdf, 1qdh, 1qdk ²⁰⁰
<i>DNA tetraplex/protein complex:</i> G.G.G.G	IV.2	x-ray (2.80 Å)	1996	1hap/pde0118 ²⁰¹ 1hao/pde0119
<i>DNA triplex:</i> T.A.T C+.G.C	III.1 III.1	NMR	1996	1wan ²⁰²
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G T.A.A	IV.2 III.1	NMR	1997	1aff ²⁰³
<i>DNA tetraplex (i-motif), centromeric sequence:</i> T.G.G.T	IV.1	NMR	1997	1c11 ³¹
<i>DNA tetraplex:</i> G.G.G.G	IV.2	x-ray (0.95 Å)	1997	352d/udf062 ⁸⁶
<i>DNA triplex (YRY):</i> C+.G.C T.A.T ⁷ G.G.C ^f	III.1 III.1 III.1	NMR	1997	1gn7 ²⁰⁴

<i>DNA triplex (YRY):</i> T.A.T C+G.C	III.1 III.1	NMR	1997	1at4 ²⁰⁵
<i>DNA tetraplex:</i> G.G.G.G	IV.2	NMR	1998	1a8w ⁹³
<i>DNA tetraplex:</i> G.G.G.G G.C.G.C	IV.2 IV.2	NMR	1998	1a8n ⁹⁴
<i>DNA triplex (YRY):</i> T.A.T C+G.C	III.1 III.1	NMR	1998	1bcb ²⁰⁶
<i>DNA triplex (YRY):</i> C+G.C T.A.T	III.1 III.1	NMR	1998	1bce ²⁰⁷
<i>S8 rRNA binding site:</i> U8.A18.A17	III.1	NMR	1998	1bgz ²⁰⁸
<i>DNA triplex (YRY):</i> T.A.T C+G.C	III.1 III.1	NMR	1998	1d3x ⁴⁸
<i>DNA triplex (YRY):</i> P.A.T ^g C+G.C	III.1 III.1	NMR	1998	1p3x ²⁰⁹
<i>DNA/RNA triplex (YRY):</i> U.A.T C+G.C	III.1 III.1	NMR	1998	1r3x ²¹⁰
<i>RNA sarcin/ricin loop:</i> G10.U11.A20 G2655.U2656.A2665	III.1 III.1	x-ray (2.10 Å) x-ray (1.11 Å)	1998/ 1999	430d/ur0002 ²¹¹ 483d/ur0007
<i>DNA tetraplex (i-motif):</i> A.A.T	III.1	x-ray (2.20 Å)	1999	1bqj/ud0003 ²¹²
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G T.T.T.T C.C.C.C	IV.2 IV.2 IV.2	NMR	1999	1emq ⁸⁹

<i>DNA tetraplex (telomeric sequence):</i> A.A.A.A G.G.G.G T.T.T.T	IV.2 IV.2 IV.2	NMR	1999	1evm, 1evn ⁹⁰
<i>RNA/peptide complex (aptamer):</i> C6.G25.A27 G8.C23.C20	III.1 III.1	NMR	1999	1exy ²¹³
<i>DNA tetraplex (telomeric sequence):</i> G.G.G.G	IV.2	NMR	1999	1k4x ²¹⁴
<i>DNA triplex (YRY) with triplex-duplex junctions:</i> C+.G.C T.A.T	III.1 III.1	NMR	1999	1bwg ²¹⁵
<i>DNA triplex with triplex-duplex junctions:</i> C+.G.C BU.A.BU ^h	III.1 III.1	x-ray (1.80 Å)	1999	1d3r/bd0017 ²¹⁶
<i>ribosomal RNA/protein: complex :</i> C,D: C150.G141.G121 C,D: G149.C142.C122 C,D: U151.A140.A139 C,D: C154.G105.A135	III.2 III.2 III.1 III.1	x-ray (2.80 Å)	1999	1qa6/pr0015 ⁴⁹
<i>RNA pseudoknot :</i> A5.U18.A24	III.1	NMR	1999	2tpk ²¹⁷
<i>RNA pseudoknot:</i> C17.G4.A20 G6.C15.A23 C26.G12.C8.A25.C14.G7.A24	III.1 III.1 VII.1	x-ray (1.60 Å)	1999	437d/ur0004 ⁵⁷
<i>RNA (HIV-1 dimerization initiation site):</i> G7(A).C17(B).G9(A).A16(B) strongly non-planar	IV.1	x-ray (2.30 Å)	1999	462d/ur0005 ⁹⁷
<i>DNA tetraplex (thrombin-binding aptamer):</i> G.G.G.G	IV.2	NMR	2000	1c32, 1c34, 1c35, 1c38 ²¹⁸
<i>16S rRNA/protein complex:</i> C53.G18.G51	III.1	x-ray (2.80 Å)	2000	1dk1/tr0005 ²¹⁹
<i>DNA tetraplex:</i> A.G.G.G.G.A G.G.G.G	VI.8 IV.1	NMR	2000	1eeg ⁵⁵

<i>RNA (HIV-1 psi site):</i> A.U.A	III.1	NMR	2000	1esy ²²⁰
<i>DNA tetraplex (bi-loop):</i> G.C.G.C	IV.2	NMR	2000	1eu2 ⁹⁶
<i>DNA tetraplex (bi-loop):</i> A.T ⁱ A.T	II.1 II.1	NMR	2000	1eu6 ⁹⁶
<i>DNA tetraplex (telomeric sequence :</i> C.C.C.C G.G.G.G T.T.T.T	IV.2 IV.2 IV.2	NMR	2000	1evo ⁹¹
<i>RNA (malachite green aptamer :</i> C10.G23.A27 A22.+U11.A26 G24.G29.(C7).A31	III.1 III.1 IV.3	x-ray (2.80 Å)	2000	1flt/ur0011 ⁵³
<i>RNA pseudoknot (aptamer):</i> C4(A).G19(A).A23(B) G6(A).C17(A).A26(B) G13(A).C29(B).C8(A)	III.1 III.1 III.1	x-ray (1.30 Å)	2000	1f27/dr0005 ²²¹
<i>DNA tetraplex:</i> G.G.G.G C.G.A	IV.2 III.1	NMR	2000	1f3s ²²²
<i>RNA/protein complex (UUCG loops):</i> C53.G18.G51	III.1	x-ray (2.80 Å)	2000	1f7y/rr0010 ²²³
<i>DNA tetraplex :</i> G.G.G.G A.T.A	IV.2 III.1	NMR	2000	1d6d ²²⁴
<i>RNA (vitamin B12 aptamer):</i> A:C18.G28.G10 C:C222.G207.U223 C:G210.C218.G228 C:U215.U227.C229 E:G410.C418.G428 C:(U215.U227.C229).A217 G:(U615.U627.C629).A617 A:A17.C29.(U15).U27.(U15). C11.G30 E:A417.C429.(U415),U427. (U415).C411.G430	III.1 III.1 III.2 III.2 III.2 IV.4 IV.4 VI.12 VI.12	x-ray (3.00 Å)	2000	1ddy/ur0008 ⁵²

<i>RNA/protein complex (large ribosomal subunit): many triads</i>		<i>x-ray</i> (2.40 Å)	2000	1ffk/rr0011 ⁶⁸
<i>RNA/protein complex (30 S ribosomal subunit): many triads, one tetrad</i>		<i>x-ray</i> (3.00 Å) <i>x-ray</i> (3.05 Å)	2000	1fjg/rr0016 ⁶⁶ 1fjf/rr0015 ⁶⁵
<i>16S rRNA/protein complex: C754(E).G654(D).G752(E)^k</i>	III.1	<i>x-ray</i> (2.60 Å)	2000	1g1x/rr0019 ²²⁵
<i>DNA tetraplex: G.G.G.G T.G.T</i>	IV.2 III.1	<i>NMR</i>	2001	1i34 ²²⁶
<i>DNA/protein complex (telomeric sequence): G.G.G.G</i>	IV.2	<i>x-ray</i> (1.86 Å)	2001	1jb7/pd0218 ²²⁷
<i>DNA tetraplex: G.G.G.G G.G.G.G.A</i>	IV.2 V.3	<i>NMR</i>	2001	1jjp ⁵⁴
<i>DNA tetraplex: A.T.A.T G.C.G.C</i>	IV.2 IV.2	<i>NMR</i>	2001	1jvc ²²⁸
<i>RNA tetraplex: U.U.U.U G.G.G.G G(U).G(U).G(U).G(U)</i>	IV.2 IV.2 VIII.3	<i>x-ray</i> (0.61 Å)	2001	1j8g/ur00114 ⁵⁸
<i>DNA/RNA: U.A.U</i>	III.1	<i>x-ray</i> (1.70 Å)	2001	1fuf/ar0029 ²²⁹

DNA tetraplex: G.G.A G.(A).G.(A).G(.A).G	III. VII.	NMR	2001	Not yet deposited at PDB ⁵⁶
--	--------------	-----	------	---

^a This table takes only structures into account for which atomic coordinates have been deposited at the Protein Data Bank. Given polyads occur in several chain this indicated before the polyad. If polyads include bases from different chains the chain identifier is indicated in parenthesis. The numbering of bases involved in polyads is not given for regular triplex/tetraplex structures. The topology corresponds to the classification presented in Table 1. H-bonds between bases are indicated by dots. No difference is made between Watson-Crick and non-canonical base-base interactions. The base sequence follows the H-bond interaction pattern. For example, in the non-cyclic triad C.G.A C is bound to G and G to A. In more complex cases parentheses have been used. The H-bond analysis of nucleic acid structures has been performed with HBExplore.²³⁰

^b Year of publication. Given the data base entry provides no reference the data base release date is used.

^c This structure was originally published in 1992 and later refined.

^d DHU: dihydrouridine

^e CPN: 2-aminoethylglycine-carbonylmethylene-cytosine, TPN: 2-aminoethylglycine-carbonylmethylene-thymine, IPN: 2-aminoethylglycine-carbonylmethylene-iodouracil

^f ⁷G: N7 glycosylated guanine

^g P: 1-propynyl-deoxyuridine (P)

^h BU: 5-bromouracil

ⁱ There is no H-bond interaction between the two A.T pairs.

^k An analogous triad is found in chains I,J

Acknowledgement

We would like to thank our co-workers M. Brandl and C. Schneider who were involved in our computational studies on nucleic acids. J. Reichert has helped with the topologies. We are grateful to M. Katahira who provided the coordinates of a heptad-containing DNA tetraplex structure prior to PDB deposition. The Thüringer Ministerium für Wissenschaft, Forschung und Kunst has funded part of the work.

References

1. J. D. Watson, F. H. C. Crick (1953) *Nature* **171**, 737.
2. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2179.
3. S. Arnott (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 1.

4. F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi (1977) *J. Mol. Biol.* **112**, 535.
5. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000) *Nucleic Acids Res.* **28**, 235.
6. H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan and B. Schneider (1992) *Biophys. J.* **63**, 751.
7. <http://pdb.weizmann.ac.il:8500/oca-docs/>
8. <http://www.rnabase.org>
9. R. A. Laskowski (2001) *Nucleic Acids Res.* **29**, 221.
10. J. Reichert and J. Sühnel (2002) *Nucleic Acids Res.* **30**, 253.
11. B. Masquida and E. Westhof (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 532.
12. J. Nowakowski and I. Tinoco (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 567.
13. J. G. Arnez and D. Moras (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 603.
14. C. C. Correll, B. Freeborn, P. B. Moore and T. A. Steitz (1997) *Cell* **91**, 705.
15. M. Brandl, M. Meyer and J. Sühnel (1999) *J. Am. Chem. Soc.* **121**, 2605.
16. M. Brandl, M. Meyer and J. Sühnel (2000) *J. Phys. Chem. A* **104**, 11177.
17. C. Schneider, M. Brandl and J. Sühnel (2001) *J. Mol. Biol.* **305**, 659.
18. M.C. Wahl, S.T. Rao and M. Sundaralingam (1996) *Nat. Struct. Biol.* **3**, 24.
19. M. Brandl, K. Lindauer, M. Meyer and J. Sühnel (1999) *Theor. Chem. Acc.* **101**, 103.
20. M. Brandl, M. Meyer and J. Sühnel (2001) *J. Biomol. Struct. Dyn.* **18**, 545.
21. E. T. Kool (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1.
22. M. Meyer and J. Sühnel (1997) *J. Biomol. Struct. Dyn.* **15**, 619.
23. M. C. Wahl and M. Sundaralingam (1997) *Trends Biochem. Sci.* **22**, 97.
24. M. S. Weiss, M. Brandl, J. Sühnel, D. Pal and R. Hilgenfeld (2001) *Trends Biochem. Sci.* **26**, 521.
25. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
26. H. Moine and J.-L. Mandel (2001) *Science* **294**, 2487.
27. A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott (1991) *Nature* **350**, 718.
28. W. Guschlbauer, J. F. Chantot and D. Thiele (1990) *J. Biomol. Struct. Dyn.* **8**, 491.
29. J. R. Williamson (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703.
30. T. Simonsson, P. Pecinka and M. Kubista (1998) *Nucleic Acids Res.* **26**, 1167.
31. J. Gallego, S. H. Chou and B. R. Reid (1997) *J. Mol. Biol.* **273**, 840.
32. H. Arthanaria and P. H. Bolton (2001) *Chem. Biol.* **8**, 221.

33. T. Simonsson (2001) *Biol. Chem.* **382**, 621.
34. B. P. Casey and P. M. Glazer (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **67**, 163.
35. M. P. Knauert and P. M. Glazer (2001) *Hum. Mol. Genet.* **10**, 2243.
36. P. D. Richardson, B. T. Kren and C. J. Steer (2001) *Curr. Opin. Mol. Ther.* **3**, 327.
37. H. Han and L. H. Hurley (2000) *Trends Pharmacol. Sci.* **21**, 136.
38. S. M. Kerwin (2000) *Curr. Pharm. Des.* **6**, 441.
39. M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland and S. Neidle (2001) *Proc. Natl. Acad. Sci. USA* **98**, 4844.
40. M. A. Read, A. A. Wood, J. R. Harrison, S. M. Gowan, L. R. Kelland, H. S. Dosanjh and S. Neidle (1999) *J. Med. Chem.* **42**, 4538.
41. S. Neidle and M. A. Read (2001) *Biopolymers* **56**, 195.
42. P. J. Perry, R. P. John and T. C. Jenkins (2001) *Expert Opin. Investig. Drugs* **10**, 2141.
43. I. Tinoco Jr. (1993) in Appendix 1 of *The RNA World*, R. F. Gesteland and J. F. Atkins (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, p. 603.
44. G. Dirheimer, G. Keith, P. Dumas and E. Westhof (1995) in *tRNAs Structure, Biosynthesis, and Function*, D. Söll and U. RajBhandary (eds.), American Society for Microbiology, Washington, p. 93.
45. G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, Berlin, 1991, p. 247).
46. N. B. Leontis and E. Westhof (2001) *RNA* **7**, 499.
47. J. Sühnel (2002) *Biopolymers* **61**, 32.
48. M. Tarköy, A. K. Phipps, P. Schultze and J. Feigon (1998) *Biochemistry* **37**, 5810.
49. G. L. Conn, D. E. Draper, E. E. Lattman and A. G. Gittis (1999) *Science* **284**, 1171.
50. J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech and J. A. Doudna (1996) *Science* **273**, 1678.
51. Y. Wang and D. J. Patel (1993) *J. Mol. Biol.* **234**, 1171.
52. D. Sussman, J. C. Nix and C. Wilson (2000) *Nat. Struct. Biol.* **7**, 53.
53. C. Baugh, D. Grate and C. Wilson (2000) *J. Mol. Biol.* **310**, 117.
54. N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin and D. J. Patel (2001) *J. Mol. Biol.* **311**, 1063.
55. A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones and D. J. Patel (2000) *J. Mol. Biol.* **297**, 627.
56. A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi and M. Katahira (2001) *J. Mol. Biol.* **313**, 255.
57. L. Su, L. Chen, M. Egli, J. M. Berger and A. Rich (1999) *Nat. Struct. Biol.* **6**, 285.
58. J. Deng, Y. Xiong and M. Sundaralingam (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13665.

59. D. E. Gilbert and J. Feigon (1999) *Curr. Opin. Struct. Biol.* **9**, 305.
60. J. Feigon (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 355.
61. D. J. Patel (1999) in *Oxford Handbook of Nucleic Acid Structure*, S. Neidle (ed.), Oxford University Press, Oxford, p. 389.
62. A. C. Lane and T. C. Jenkins (2001) *Curr. Org. Chem.* **5**, 845.
63. M. A. Keniry (2001) *Biopolymers* **56**, 123.
64. V. Ramakrishnan and P. B. Moore (2001) *Curr. Opin. Struct. Biol.* **11**, 144.
65. B. T. Wimberly, D. E. Brodersen, W. M. Clemons, Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch and V. Ramakrishnan (2000) *Nature* **407**, 327.
66. A. P. Carter, W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly and V. Ramakrishnan (2000) *Nature* **407**, 340.
67. W. M. Clemons Jr., D. E. Brodersen, J. P. McCutcheon, J. L. May, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly and V. Ramakrishnan (2001) *J. Mol. Biol.* **310**, 827.
68. N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz (2000) *Science* **289**, 905.
69. M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate and H. F. Noller (2001) *Science* **292**, 883.
70. S. Forman, J. C. Fettinger, S. Pieraccini, G. Gottarelli and J. T. Davis (2000) *J. Am. Chem. Soc.* **122**, 4060.
71. F. W. Kotch, J. C. Fettinger and J. T. Davis (2000) *Org. Lett.* **2**, 3277.
72. M. S. Lüth, E. Freisinger, B. Lippert (2001) *Chem. Eur. J.* **7**, 2104.
73. R. K. Sigel, S. M. Thompson, E. Freisinger, F. Glahe and B. Lippert (2001) *Chem. Eur. J.* **7**, 1968.
74. E. Mezzina, P. Mariani, R. Itri, S. Masiero, S. Pieraccini, G. P. Spada, F. Spinazzi, J. T. Davis and G. Gottarelli (2001) *Chem. Eur. J.* **7**, 3888.
75. C. C. Hardin, A. G. Perry and K. White (2001) *Biopolymers* **56**, 147.
76. U. Nagaswamy, N. Voss, Z. Zhang and G. E. Fox (2000) *Nucleic Acids Res.* **28**, 375.
77. J. Weil, T. Min, C. Yang, S. Wang, C. Sutherland, N. Sinha and C. Kang (1999) *Acta Crystallogr. D Biol. Crystallogr. (Pt 2)*, 422.
78. A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin and D. J. Patel. (2000) *J. Mol. Biol.* **301**, 129.
79. V. Kuryavyi, A. Kettani, W. Wang, R. Jones and D. J. Patel (2000) *J. Mol. Biol.* **295**, 455.
80. V. Kuryavyi, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones and D. J. Patel (2001) *J. Mol. Biol.* **310**, 181.
81. L. Betts, J.A. Josey, J. M. Veal and S. R. Jordan (1995) *Science* **270**, 1838.
82. B. Hingerty, R. S. Brown and A. Jack (1978) *J. Mol. Biol.* **124**, 523.
83. C. Kang, X. Zhang, R. Ratliff, R. Moyzis and A. Rich (1992) *Nature* **356**, 126.
84. F. W. Smith and J. Feigon (1992) *Nature* **356**, 164.
85. P. Schultze, F. Smith and J. Feigon (1994) *Structure* **2**, 221.

86. K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley and B. Luisi (1997) *J. Mol. Biol.* **273**, 171.
87. V. Hud and M. Polak (2001) *Curr. Opin. Struct. Biol.* **11**, 293.
88. T. Hermann and E. Westhof (1998) *Structure* **6**, 1303.
89. P. K. Patel and R. V. Hosur (1999) *Nucleic Acids Res.* **27**, 2457.
90. P. K. Patel, A. S. Koti and R. V. Hosur (1999) *Nucleic Acids Res.* **27**, 3836.
91. P. K. Patel, N. S. Bhavesh and R. V. Hosur (2000) *Biochem. Biophys. Res. Commun.* **270**, 967.
92. C. Cheong and P. B. Moore (1992) *Biochemistry* **31**, 8406.
93. S. Bouaziz, A. Kettani and D. J. Patel (1998) *J. Mol. Biol.* **282**, 637.
94. A. Kettani, S. Bouaziz, A. Gorin, H. Zhao, R. A. Jones and D. J. Patel (1998) *J. Mol. Biol.* **282**, 619.
95. M. Gueron and J. L. Leroy (2000) *Curr. Opin. Struct. Biol.* **10**, 326.
96. N. Escaja, E. Pedroso, M. Rico and C. Gonzalez (2000) *J. Am. Chem. Soc.* **122**, 12732.
97. E. Ennifar, M. Yusupov, P. Walter, R. Marquet, B. Ehresmann, C. Ehresmann and P. Dumas (1999) *Structure Fold. Des.* **7**, 1439.
98. G. A. Leonard, S. Zhang, M. R. Peterson, S. J. Harrop, J. R. Helliwell, W. B. Cruse, B. L. d'Estaintot, O. Kennard, T. Brown and W. N. Hunter (1995) *Structure* **3**, 335.
99. J. C. Chaput and C. Switzer (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10614.
100. F. Seela and R. Kröschel (2001) *Bioconjugate Chem.* **12**, 1043.
101. P. Auffinger and E. Westhof (1997) *J. Mol. Biol.* **274**, 54.
102. P. Auffinger and E. Westhof (1998) *Curr. Opin. Struct. Biol.* **8**, 227.
103. F. Major and R. Griffey (2001) *Curr. Opin. Struct. Biol.* **11**, 282.
104. M. Zacharias (2000) *Curr. Opin. Struct. Biol.* **10**, 311.
105. T. E. Cheatham and P. Kollman (2000) *Annu. Rev. Phys. Chem.* **511**, 435.
106. P. Hobza and J. Šponer (1999) *Chem. Rev.* **99**, 3247.
107. J. Šponer, J. Leszczynski and P. Hobza (2001) *J. Mol. Struct. (THEOCHEM)* **573**, 43.
108. C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders and E. J. van Baerends (2000) *J. Am. Chem. Soc.* **122**, 4117.
109. M. Treger and E. Westhof (2001) *J. Mol. Recognit.* **14**, 199.
110. S. P. Jiang, R. L. Jernigan, K. L. Ting, J. L. Syi and G. Raghunathan (1994) *J. Biomol. Struct. Dyn.* **12**, 383.
111. J. Šponer, J. V. Burda, P. Mejzlik, J. Leszczynski and P. Hobza (1997) *J. Biomol. Struct. Dyn.* **14**, 613.
112. J. Šponer, M. Sabat, J. V. Burda, A. M. Doody, J. Leszczynski and P. Hobza (1998) *J. Biomol. Struct. Dyn.* **16**, 139.
113. D. Venkateswarlu and J. Leszczynski (1999) *J. Phys. Chem. A* **103**, 3489.
114. R. Soliva, F. J. Luque and M. Orozco (1999) *Nucleic Acids Res.* **27**, 2248.
115. M. Barfield, A. J. Dingley, J. Feigon and S. Grzesiek (2001) *J. Am. Chem. Soc.* **123**, 4014.
116. E. Cubero, R. Güimil-Garcia, F. J. Luque, R. Eritja and M. Orozco (2001) *Nucleic Acids Res.* **29**, 2522.

117. J. Gu, J. Leszczynski and M. Bansal (1999) *Chem. Phys. Lett.* **311**, 209.
118. J. Gu and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 1898.
119. J. Gu and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 6308.
120. J. Gu and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 7353.
121. J. Gu and J. Leszczynski (2001) *Chem. Phys. Lett.* **335**, 465.
122. M. Meyer, T. Steinke, M. Brandl and J. Sühnel (2001) *J. Comput. Chem.* **22**, 109.
123. M. Meyer, M. Brandl and J. Sühnel (2001) *J. Phys. Chem. A* **105**, 8223.
124. J. Gu and J. Leszczynski (2001) *J. Phys. Chem. A* **105**, 10366.
125. M. Meyer, C. Schneider, M. Brandl and J. Sühnel (2001) *J. Phys. Chem. A* **105**, 11560.
126. J. Gu and J. Leszczynski (2002) *J. Phys. Chem. A* **106**, 529.
127. J. Gu and J. Leszczynski (2002) *Chem. Phys. Lett.* **351**, 403-409.
128. S. X. Liu, Y. Li, H. Cao, X. G. Xie and C. Q. Liu (2001) *Chinese J. Struct. Chem.* **20**, 313-318.
129. P. Hobza, M. Kabeláč, J. Šponer, P. Mejzlík and J. Vondrášek (1997) *J. Comp. Chem.* **18**, 1136.
130. P. M. W. Gill (1998) in *Encyclopedia of Computational Chemistry*, P. von R. Schleyer (ed.), John Wiley & Sons, New York, p. 678.
131. A. St-Amant (1996) in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd (eds.), Vol. 7, VCH Publishers, New York, p. 217.
132. S. Saeboe and P. Pulay (2001) *J. Chem. Phys.* **115**, 3975.
133. S. Goedecker (1999) *Rev. Mod. Phys.* **71**, 1085.
134. J. Šponer, J. V. Burda, J. Leszczynski and P. Hobza (1997) *J. Biomol. Struct. Dyn.* **17**, 61.
135. M. Elstner, P. Hobza, T. Frauenheim, S. Suhai and E. Kaxiras (2001) *J. Chem. Phys.* **114**, 5149.
136. N. Kurita, M. Araki, K. Nakano and K. Kobayashi (2000) *Int. J. Quantum. Chem.* **76**, 677.
137. A. D. Becke (1993) *J. Chem. Phys.* **98**, 5648.
138. C. Lee, G. Yang and G. Parr (1998) *Phys. Rev. B* **37**, 785.
139. A. D. Becke (1988) *Phys. Rev. A* **38**, 3098.
140. N. Godbout, D. R. Salahub, J. Andzelm and E. Wimmer (1992) *Can. J. Chem.* **70**, 560.
141. L. F. Pacios and P. A. Christiansen (1985) *J. Chem. Phys.* **82**, 2644.
142. N. R. Kestner (1999) in *Reviews of Computational Chemistry*, K. B. Lipkowitz and D.B. Boyd (eds.), Vol. 13, VCH Publishers, New York, p. 99.
143. E. Nir, K. Kleinermanns and M. S. de Vries (2000) *Nature* **408**, 949.
144. N. V. Hud, P. Schultz and J. Feigon (1998) *J. Am. Chem. Soc.* **120**, 6403.
145. C. Schneider and J. Sühnel (1999) *J. Biomol. Struct. Dyn.* **50**, 287.
146. C. A. Laughton and S. Neidle (1992) *J. Mol. Biol.* **223**, 519.
147. C. Y. Sekharudu, N. Yathindra and M. Sundaralingam (1995) *J. Biomol. Struct. Dyn.* **11**, 225.

148. S. Weerasinghe, P. E. Smith and B. M. Pettitt (1995) *Biochemistry* **34**, 16269.
149. L. Yang, S. Weerasinghe, P. E. Smith and B. M. Pettitt (1995) *Biophys J.* **69**, 1519.
150. M. R. Kiran and M. Bansal (1995) *J. Biomol. Struct. Dyn.* **13**, 493.
151. M. Ouali, M. Bouziane, C. Ketterle, J. Gabarro-Arpa, C. Auclair and M. Le Bret (1996) *J. Biomol. Struct. Dyn.* **13**, 835.
152. M. R. Kiran and M. Bansal (1997) *J. Biomol. Struct. Dyn.* **15**, 333.
153. M. Ouali, H. Gousset, F. Geinguenaud, J. Liquier, J. Gabarro-Arpa, M. Le Bret and E. Taillandier (1997) *Nucleic Acids Res.* **25**, 4816.
154. S. A. Cassidy, P. Slickers, J. O. Trent, D. C. Capaldi, P. D. Roselt, C. B. Reese, S. Neidle and K. R. Fox (1997) *Nucleic Acids Res.* **25**, 4891.
155. M. R. Kiran and M. Bansal (1998) *J. Biomol. Struct. Dyn.* **16**, 511.
157. R. Soliva, C. A. Laughton, F. J. Luque and M. Orozco (1998) *J. Am. Chem. Soc.* **120**, 11226.
158. L. Yang, M. Liu, W. Deng, C. Wang, C. Bai and L. S. Kan (1999) *Biophys. Chem.* **76**, 25.
159. R. Soliva, F. J. Luque and M. Orozco (1999) *Nucleic Acids Res.* **27**, 2248.
160. J. Luo and T. C. Bruice (2000) *J. Biomol. Struct. Dyn.* **17**, 629.
161. R. Soliva, Garcia R. Guimil, J. R. Blas, R. Eritja, J. L. Asensio, C. Gonzalez, F. J. Luque and M. Orozco (2000) *Nucleic Acids Res.* **28**, 4531.
162. T. Carlomagno, M. J. Blommers, J. Meiler, B. Cuenoud and C. Griesinger (2001) *J. Am. Chem. Soc.* **123**, 7364.
163. C. Giancola, L. Petraccone, M. Pieri and G. Barone (2001) *Biophys. Chem.* **94**, 23.
164. J. Hanuš, I. Barvík, K. Ruszová-Chmelová, J. Štěpánek, P. Y. Turpin, J. Bok, I. Rosenberg and M. Petrová-Endová (2001) *Nucleic Acids Res.* **29**, 5182.
165. E. Cubero, R. Guimil-Garcia, F. J. Luque, R. Eritja and M. Orozco (2001) *Nucleic Acids Res.* **29**, 2522.
166. W. S. Ross and C. C. Hardin (1994) *J. Am. Chem. Soc.* **116**, 6070.
167. G. D. Strahan, R. H. Shafer and M. A. Keniry (1994) *Nucleic Acids Res.* **24**, 5447.
168. N. Špačkova, I. Berger and J. Šponer (1999) *J. Am. Chem. Soc.* **121**, 5519.
169. D. Mohanty and M. Bansal (1995) *Biophys. J.* **69**, 1046.
170. J. Tohl and W. Eimer (1997) *J. Mol. Modeling* **2**, 327.
171. J. Tohl and W. Eimer (1997) *Biophys. Chem.* **67**, 177.
172. G. D. Strahan, M. A. Keniry and R. H. Shafer (1998) *Biophys. J.* **75**, 968.
173. S. Chowdhury and M. Bansal (2000) *J. Biomol. Struct. Dyn.* **18**, 11.
174. M. A. Read and S. Neidle (2000) *Biochemistry* **39**, 13422.
175. R. Stefl, N. Špačkova, I. Berger, J. Koca and J. Šponer (2001) *Biophys. J.* **80**, 455.
176. N. Špačkova, I. Berger and J. Šponer (2001) *J. Am. Chem. Soc.* **123**, 3295.
177. S. Chowdhury and M. Bansal (2001) *J. Phys. Chem. B* **105**, 7572.

178. H. Han, D. R. Langley, A. Rangan and L. H. Hurley (2001) *J. Am. Chem. Soc.* **123**, 8902.
179. S. Chowdhury and M. Bansal (2001) *J. Biomol. Struct. Dyn.* **18**, 647.
180. S. Redon, S. Bombard, M. A. Elizondo-Riojas and J. C. Chottard (2001) *Biochemistry* **40**, 8463.
181. D. M. York, W. Yang, H. Lee, T. Darden, L. G. Pedersen (1995) *J. Am. Chem. Soc.* **117**, 5001.
182. H. Shi and P. B. Moore (2000) *RNA* **6**, 1091.
183. L. Jovine, S. Djordjevic and D. Rhodes (2000) *J. Mol. Biol.* **301**, 401.
184. E. Westhof, P. Dumas and D. Moras (1988) *Acta Crystallogr. A* **44** (Pt. 2), 112.
185. Y. Wang and D. J. Patel (1993) *Structure* **1**, 263.
186. K. Padmanabhan, K. P. Padmanabhan, J. D. Ferrara, J. E. Sadler and A. Tulinsky (1993) *J. Biol. Chem.* **268**, 17651.
187. I. Radhakrishnan and D. J. Patel (1993) *Structure* **1**, 135.
188. P. Schultze, R. F. Macaya and J. Feigon (1994) *J. Mol. Biol.* **235**, 1532.
189. Y. Wang and D. J. Patel (1994) *Structure* **2**, 1141.
190. I. Radhakrishnan and D. J. Patel (1994) *Structure* **2**, 17.
191. I. Radhakrishnan and D. J. Patel (1994) *J. Mol. Biol.* **241**, 600.
192. V. Biou, A. Yaremcchuk, M. Tukalo and S. Cusack (1994) *Science* **263**, 1404.
193. G. Laughlan, A. I. Murchie, D. G. Norman, M. H. Moore, P. C. Moody, D. M. Lilley and B. Luisi (1994) *Science* **265**, 520.
194. Y. Wang and D. J. Patel (1995) *J. Mol. Biol.* **251**, 76.
195. A. Kettani, R. A. Kumar and D. J. Patel (1995) *J. Mol. Biol.* **254**, 638.
196. M. A. Keniry, G. D. Strahan, E. A. Owen and R. H. Shafer (1995) *Eur. J. Biochem.* **233**, 631.
197. J.D. Puglisi, L. Chen, S. Blanchard and A.D. Frankel (1995) *Science* **270**, 1200.
198. F. W. Smith, P. Schultze and J. Feigon (1995) *Structure* **3**, 997.
199. P. Fan, A. K. Suri, R. Fiala, D. Live and D. J. Patel (1996) *J. Mol. Biol.* **258**, 480.
200. V. M. Marathias, K. Y. Wang, S. Kumar, T. Q. Pham, S. Swaminathan and P. H. Bolton (1996) *J. Mol. Biol.* **260**, 378.
201. K. Padmanabhan and A. Tulinsky (1996) *Acta Crystallogr. D* **52**, 272.
202. E. Wang, K. M. Koshlap, P. Gillespie, P. B Dervan and J. Feigon (1996) *J. Mol. Biol.* **257**, 1052.
203. A. Kettani, S. Bouaziz, W. Wang, R. Jones and D. J. Patel (1997) *Nat. Struct. Biol.* **4**, 382.
204. K. M. Koshlap, P. Schultze, H. Brunar, P. B. Dervan and J. Feigon (1997) *Biochemistry* **36**, 2659.
205. J. P. Bartley, T. Brown and A. N. Lane (1997) *Biochemistry* **36**, 14502.
206. J. L. Asensio, A. N. Lane, J. Dhesi, S. Bergqvist and T. Brown (1998) *J. Mol. Biol.* **275**, 811.
207. J.L. Asensio, T. Brown and A. N. Lane (1998) *Nucleic Acids Res.* **26**, 3677.

208. K. Kalurachchi and E. P. Nikonowicz (1998) *J. Mol. Biol.* **280**, 639.
209. A. K. Phipps, M. Tarköy, P. Schultze and J. Feigon (1998) *Biochemistry* **37**, 5820.
210. C.H. Gotfredsen, P. Schultze and J. Feigon (1998) *J. Am. Chem. Soc.* **37**, 4281.
211. C. C. Correll, A. Munishkin, Y. L. Chan, Z. Ren, I. G. Wool and T. A. Steitz (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13436.
212. J. Weil, T. Min, C. Yang, S. Wang, C. Sutherland, N. Sinha and C. Kang (1999 Pt 2) *Acta Crystallogr. D, Biol. Crystallogr.* **422**.
213. F. Jiang, A. Gorin, W. Hu, A. Majumdar, S. Baskerville, W. Xu, A. Ellington and D. J. Patel (1999) *Structure Fold. Des.* **7**, 1461.
214. P. Schultze, N. V. Hud, F. W. Smith and J. Feigon (1999) *Nucleic Acids Res.* **27**, 3018.
215. J. L. Asensio, T. Brown and A. N. Lane (1999) *Structure Fold. Des.* **7**, 1.
216. S. Rhee, Z. Han, K. Liu, H. T. Miles and D. R. Davies (1999) *Biochemistry* **38**, 16810.
217. J.A. Holland, M. R. Hansen, Z. Du and D. W. Hoffman (1999) *RNA* **5**, 257.
218. V. M. Marathias and P. H. Bolton (2000) *Nucleic Acids Res.* **28**, 1969.
219. A. Nikulin, A. Serganov, E. Ennifar, S. Tishchenko, N. Nevskaya, W. Shepard, C. Portier, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov and P. Dumas (2000) *Nat. Struct. Biol.* **7**, 273.
220. G. K. Amarasinghe, R. N. De Guzman, R. B. Turner and M. F. Summers (2000) *J. Mol. Biol.* **299**, 145.
221. J. Nix, D. Sussman and C. Wilson (2000) *J. Mol. Biol.* **296**, 1235.
222. A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin and D. J. Patel (2000) *J. Mol. Biol.* **301**, 129.
223. E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov and P. Dumas (2000) *J. Mol. Biol.* **304**, 35.
224. V. Kuryavyi, A. Kettani, W. Wang, R. Jones and D. J. Patel (2000) *J. Mol. Biol.* **295**, 455.
225. S. C. Agalarov, G. Sridhar Prasad, P. M. Funke, C. D. Stout and J. R. Williamson (2000) *Science* **288**, 107.
226. V. Kuryavyi, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones and D. J. Patel (2001) *J. Mol. Biol.* **310**, 181.
227. M. P. Horvath and S. C. Schultz (2001) *J. Mol. Biol.* **310**, 367.
228. N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin and D. J. E. Patel (2001) *J. Mol. Biol.* **312**, 1073.
229. J. Deng, Y. Xiong, C. Sudarsanakumar, K. Shi and M. Sundaralingam (2001) *RNA* **7**, 1425.
230. K. Lindauer, C. Bendic and J. Sühnel (1996) *Comput. Appl. Biosci.* **12**, 281.

Chapter 5: Model Calculations of Radiation Induced Damage in DNA Constituents Using Density Functional Theory

David M. Close

*Department of Physics, East Tennessee State University
P.O. Box 70652, Johnson City, TN, 37614*

Abstract

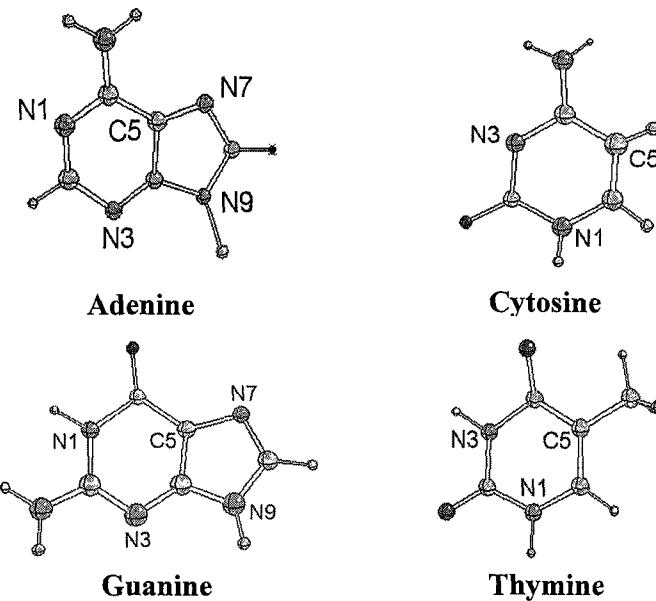
The aim of the present review is first to summarize what is known about the primary radiation induced defects in nucleic acid constituents from detailed EPR/ENDOR experiments, then to explore the use of theoretical calculations to assist in making free radical assignments. The theoretical calculations include estimates of spin densities and isotropic and anisotropic hyperfine couplings which can be compared with the experimental results. In many cases the theoretical and experimental results agree rather well. In other cases there are discrepancies between the theoretical and experimental results. The successes and failures of DFT to calculate spin densities and hyperfine couplings of more than twenty primary radiation induced radicals observed in the nucleobases are discussed. Some unsolved problems and suggestions for future work are also discussed.

1. Introduction

The study of radiation damage to nucleic acids holds a central place in radiation biology. It is from the study of the free radical chemistry of nucleic acids that one may begin to understand the lethal effects of ionizing radiation. The results of detailed electron paramagnetic resonance/electron nuclear double resonance (EPR/ENDOR) experiments on nucleic acid constituents have played a major role in understanding the primary radiation effects (radical cations and radical anions) produced by ionizing radiation. The aim of the present review is to: 1) briefly summarize what is known experimentally about the primary radiation induced defects in nucleic acid constituents (for a complete review see ref. [1]), 2) to see how theoretical calculations are currently being used to assist in making free radical assignments, and finally, 3) to look at unsolved problems and make suggestions for future work.

Several recent papers have reported Density Functional Theory (DFT) calculations on the primary oxidation and reduction products observed in irradiated single crystals of the common nucleobases Thymine², Cytosine³, Guanine⁴, and Adenine.⁵ The theoretical calculations include estimates of spin densities and isotropic and anisotropic hyperfine couplings which can be compared with experimental results (obtained from detailed EPR/ENDOR experiments). In many cases the theoretical and experimental results agree rather well. In other cases there are discrepancies between the theoretical and experimental results. Herein the successes and failures of using DFT to calculate spin densities and hyperfine couplings of the primary radiation induced free radicals observed in the nucleobases will be discussed.

A few years ago the calculation of spin densities and hyperfine couplings on even small molecules was a very challenging task. Colson *et al.* reported the spin densities, computed at the 6-31G*/3-21G level, for the anions and cations of the four DNA bases.⁶ Their results correctly indicated the major regions of spin density. For example, the major sites of spin density for the adenine reduction product were computed to be $\rho(C2) = 0.71$, $\rho(C8) = 0.03$, and $\rho(N3) = 0.08$. While these are the sites of spin density expected for an adenine reduction product, these results are not very close to the experimentally determined spin densities of $\rho(C2) = 0.41$, $\rho(C8) = 0.14$, and $\rho(N3) = 0.12$.⁷ The numbering scheme used here is as shown.



More recent calculations have been made with the inclusion of electron correlation through the second-order Møller-Plesset (MP2) theory.^{8,9} The MP2 calculations require the transformation of two-electron integrals over atomic orbital basis functions to the molecular orbital basis. Computation and storage of such integrals is very time consuming for the nucleobases (with 10 or 12 heavy atoms), so currently calculations are done with DFT which provides an efficient method for studying large model systems at a correlated level. In cases where DFT does not give very good answers, one question to be considered is whether other approaches give more reliable answers for these free radical systems.¹⁰

It should also be pointed out that the calculation of accurate hyperfine coupling constants is rather difficult. Two factors are involved: the isotropic component (A_{iso}) and the anisotropic component (T_{xx} , T_{yy} , T_{zz}). One must have a good description of electron correlation and a well defined basis set in order to calculate accurate isotropic hyperfine couplings. This is not easy to do with molecules the size of the DNA bases. Even when the computational demands are met, the theoretical calculations may deviate more than 20% from the experimental results. Recently it has been shown that the calculation of anisotropic hyperfine couplings for hydrogens are often within 5-10% of the experimental values.³ The goal is to

make comparisons of calculated and experimental isotropic and anisotropic hyperfine couplings a useful guide in identifying radiation induced free radicals.

The present level of theoretical work can be seen in a recent paper by Sevilla's group. They have employed DFT to study the details of proton transfer reactions in Guanine:Cytosine base pairs.¹¹ Using the DFT functional B3LYP with the 6-31+G(d) basis set on the entire G:C molecule (19 heavy atoms) results are presented which show it is energetically favorable for the N1 proton of guanine to transfer to the N3 of cytosine after reduction of cytosine or after oxidation of guanine.

Several additional problems must also be considered. All of the calculations discussed here are computationally challenging. They involve a single point calculation on the optimized structure using triple-zeta plus polarization functions (B3LYP/6-311G(2df,p)) in order to compute spin densities. Also, the detailed DFT calculations discussed here were performed on isolated molecules, whereas some of the experimental results reported involve free radical formation in the solid state, mainly in single crystals. Therefore the theoretical calculations are ignoring the electrostatic environment of the radicals discussed, in particular the intricate hydrogen bonding structure that the free radicals are imbedded in. As discussed below, this often leads to non-planar radicals which may or may not represent what is believed to be observed experimentally in the solid state.

Another problem has to do with radical stability. Since radiation scatters electrons from different molecular orbitals at random, one might expect to see a great variety of damaged products. Usually this is not the case. Theoretical calculations are useful here in ranking the energies of the various oxidation and reduction products. It is often possible therefore to predict which products will be observed in a particular system.

Bernhard has considered the stability of radicals in various crystalline environments.¹² For example, after irradiation, 1-MethylCytosine (1-MeC) is known to have a very low radical yield, so it is argued that a large percentage of the initial radicals formed by the ionizing radiation must recombine. Looking at the hydrogen bonding network of 1-MeC one sees that the network does not favor long range proton displacements. Consequently there are no energetically favorable paths which would promote the separation of unpaired spin and charge, leaving the initial sites prone to recombination. On the other hand, in many of the systems considered here, there are efficient pathways for returning ionization sites to their original charge states, thereby effectively inhibiting recombination. As a con-

sequence, many of the radiation induced defects reported are not the primary radiations induced events, i.e. native cations or anions, but rather neutral products, (deprotonated cations or protonated anions) which are less susceptible to recombination.

These ideas have been brought together in a recent study of the co-crystalline complex of 1-MethylCytosine:5-FluoroUracil.¹³ Using model calculations it was shown how the hydrogen bonding network of the crystal is able to sustain a proton shuttle which leads to the selective formation of certain radicals. Calculations were able to predict that the site of reduction would be the cytosine base (yielding the N3 protonated cytosine anion), while the uracil base would be the site of oxidation (yielding the N1 deprotonated uracil cation). These are indeed the primary radiation induced species observed experimentally.

Geimer and co-workers in Leipzig have written a series of articles recently which report precise hyperfine couplings for oxidation and reduction products in pyrimidine bases.¹⁴⁻¹⁶ These studies involve *in situ* photolysis experiments in liquids. From the highly resolved FT EPR experiments the authors have obtained accurate hyperfine coupling constants of even the small proton couplings that are usually hidden in the linewidth of solid-state EPR spectra. Results are now available for 1-MethylCytosine (1-MeC), Thymine, and 1-MethylThymine (1-MeT), and will be discussed herein. The authors have also performed DFT calculations on the oxidation and reduction products observed in the pyrimidine bases.^{17,18} These results will also be discussed.

1.1 Experimental Considerations

The EPR/ENDOR measurements described here have been performed on single crystals which are accurately oriented with an x-ray precession camera, x-ray irradiated, and observed at *ca.* 10 K. Anisotropic proton hyperfine couplings are measured by rotating the crystals in the external magnetic field. From hundreds of angular measurements various proton hyperfine couplings are obtained. From the direction cosines associated with each coupling, comparisons can be made with specific molecular directions known from the x-ray crystal structure, in particular enabling the identification of the major sites of unpaired spin density.

Furthermore, in the experiments, EPR spectra are also recorded at every orientation, from which one can make good estimates of anisotropic nitrogen hyperfine couplings which are not normally detected in the ENDOR experiments. In most cases complicated single crystal (and even powder) EPR spectra can be

faithfully reproduced with the accurate proton couplings obtained from the ENDOR experiments and the nitrogen hyperfine couplings obtained from the EPR spectra.¹⁹

The purpose of obtaining all this information is to present reasonable free radical models for the primary oxidation and reduction products observed in the irradiated crystals. This begins with, and is usually based on, the precise information about major sites of spin density. There are however, some problems in dealing with all of the small hyperfine couplings obtained from the ENDOR data. This could mean, for example, that one may have problems with establishing precisely what the protonation state of a given model is. One procedure used to solve this problem is to repeat the entire experiment with partially deuterated single crystals to learn which of the many small hyperfine couplings are at exchangeable bonding sites. As shown here, one may also use theoretical calculations to aid in making suitable free radical assignments.

Typical results for the cytosine reduction product observed in irradiated single crystals of 1-MethylCytosine:5-FluoroUracil are shown in Table I. First one notes the three principal values of the hyperfine coupling tensor. For an ordinary π -electron radical with unit spin density on the central carbon, the principal values are known to be *ca.* 91, 61, and 29 MHz.²⁰ One sees that the principal values of the $>\text{C}_6\text{-H}_\alpha$ hyperfine coupling tensor listed in Table I are approximately 60% of these numbers, reflecting the fact that the unpaired spin density at C6 is approximately 60%. It is important to note that the proportional relationship between A_{\max} , A_{mid} , and A_{\min} for these two couplings is the same, and when this is so it indicates that there is no rocking motion at the radical site. This is good evidence therefore that the radical site is essentially planar.

Table I: Hyperfine Coupling Parameters for the 1-MeC Reduction Product

Matrix (Coupling)	Principal Values	Isotropic Value	Dipolar Values	Direction Cosines		
				$\langle a \rangle$	$\langle b \rangle$	$\langle c \rangle$
1-MeC ^a	-62.47		-23.9	0.5140	-0.4883	-0.7052
C6-H _α	-34.58	-38.60	4.0	0.7847	-0.0643	0.6165 ^b
	-18.74		19.9	0.3464	0.8703	-0.3502 ^c

All hyperfine couplings in MHz.

^aSee reference 13.

^bAngle that the direction of A_{mid} makes with the perpendicular to the ring plane is $2.0 \pm 0.5^\circ$.

^cAngle that the direction of A_{\min} makes with the C6-H bond direction is $2.8 \pm 0.3^\circ$.

The best evidence for radical planarity comes from the analysis of the direction cosines associated with each principal values of the hyperfine coupling tensor. The direction of A_{\min} (Table I) is known to be associated with the direction of the >C-H bond, while the direction associated with the A_{mid} indicates the direction of the π -electron orbital. These directions are easily calculated from the crystal structure, and are included in Table I. One sees that the direction associated with A_{mid} deviates only 2.0° from the computed perpendicular to the ring plane, while the direction of A_{\min} deviates only 2.8° from the computed direction of the C6-H bond. The errors listed on these values are at the 95% confidence level. This is very clear evidence that the radical shown here is planar in the solid state. Any torsional motion of the C6-H would lead to asymmetries of the hyperfine coupling tensor, and would not produce the observed agreement between the direction cosines and the known directions obtained from the crystal structure.

One can gain even more confidence in the experimental results by attempting to simulate the EPR spectra from the hyperfine coupling tensors. Examples of successful simulations are presented in a recent paper on the primary radicals observed in 5'-dCMP.²¹

1.2 Theoretical Details

The theoretical results usually involve two sets of calculations. First the B3LYP hybrid functional in conjunction with Pople's split valance 6-31G(d,p) basis set was used to obtain the optimized geometry. Frequency calculations were performed at the same level of theory to ensure that the systems represent true minima on the potential energy surfaces. Single point calculations were subsequently carried out at the B3LYP level with Pople's triply split valance and polarized 6-311G(2df,p) basis set for obtaining the isotropic and anisotropic hyperfine coupling constants. All calculations were performed with the Gaussian 98 suite of programs.²²

As mentioned above, this technique produces some very impressive agreements between the experimentally determined spin densities and hyperfine coupling constants. However, in some cases the agreement is not very satisfactory. Attempts have therefore been made to obtain better theoretical models. In some cases it seems that at the present level, DFT is not adequate to give proper agreement between with the experimentally determined spin densities and hyperfine coupling constants. All of the results discussed here are summarized in a final Table which contains comments on the "goodness of fit" between the experimental and the theoretical results.

2. Experimental and Theoretical Results

2.1 Pyrimidines

2.1.1 Cytosine Reduction Products

Cytosine has received a great deal of attention recently. The results of Bernhard²³ and Sevilla²⁴ suggest that cytosine may be the primary reduction site in DNA. Also, a dissimilarity in the EPR spectrum from monomers and oligomers of cytosine has focused attention on the protonation state of the cytosine reduction product.²⁵

Reduction of cytosine produces a radical with sites of unpaired spin density at C2, C4 and C6. The hyperfine coupling of the unpaired spin with the C6-H_a produces a ~1.4 mT doublet which is the main feature of the "cytosine anion" EPR signal which has been observed in various cytosine derivatives.²⁶ The problem here is to determine the charge state of this reduction product.

The cytosine anion is a strong base ($pK_a > 13$)²⁷ and is therefore expected to rapidly protonate in solution. Hissung and von Sonntag²⁸ have shown by conductance techniques that the radical anion of cytosine is protonated by water (most likely at N3 or O2) in less than 4 ns.

In the solid-state it is rare to observe charged radicals. After trapping an electron the cytosine anion would have a net negative charge. Such a radical would be unstable because it would attract positive charge and therefore could be subject to recombination. Most reduction products observed in pyrimidines in the solid state are actually protonated anions.¹ These protonated anions are of course neutral, and are more stable since they are less likely to suffer recombination. The point being that if one is able to separate charge from spin, the resulting neutral product is more likely to be a stable product.

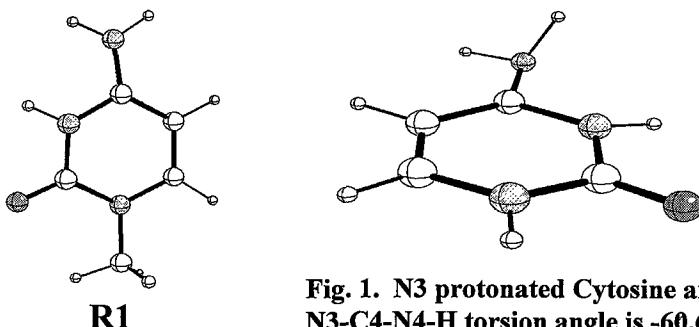
From studies of cytosine monohydrate single crystals irradiated at 10 K Sagstuen *et al.*¹⁹ concluded that the primary reduction product is the N3 protonated anion (Radical R1). ENDOR experiments detected the C6-H_a hyperfine coupling, the N3-H_a hyperfine coupling and one of the small couplings to the N4-H₂ protons (see Table II).²⁹

Table II: Hyperfine Coupling Parameters for Cytosine Reduction Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Isotropic	Results Dipolar	(Ref.)
—Cytosine (N1-H) N3 protonated anion (Fig. 1)—							
Cm	(19)	-62.5		-24.8		-26.1	(3)
C6-H _α		-35.5	-37.8	2.3	-42.9	2.4	
		-15.3		22.5		23.7	
		-11.3		-5.8		-6.2	
N3-H _α		-7.0	-5.5	-1.5	-7.7	-1.4	
		1.7		7.3		7.6	
		-9.2		-4.7		-5.6	
N4-H _{α1}		-6.7	-4.5	-2.2	-6.7	-2.3	
		2.3		6.9		7.9	
						-7.2	
N4-H _{α2}	not detected				-8.4	-1.7	
						8.9	
—Cytosine (N1-R) N3 protonated anion (R1)—							
1-MeC	(13)	-62.5		-23.9		-25.4	(3)
C6-H _α		-34.6	-38.6	4.0	-42.2	2.8	
		-18.7		19.9		22.6	
		-11.5		-5.8		-6.9	
N3-H _α		-6.4	-5.6	-0.8	-8.3	-1.8	
		1.0		6.6		8.7	
						-6.0	
N4-H _{α1}	not detected				-7.5	-1.3	
						7.3	
						-5.3	
N4-H _{α2}	not detected				-6.5	-2.3	
						7.6	

All hyperfine couplings in MHz.

Wetmore *et al.*³ report that their theoretical spin densities on Radical R1 are in close agreement with those computed from the experimental data. However they mention that their calculations indicate a 55 MHz hyperfine coupling from one of the -N4-H₂ protons. They give no explanation as to why a proton confined to the pyrimidine ring plane (which it certainly is in the single crystal) should give such a large hyperfine coupling. The authors have kindly provided the xyz coordinates of their model, which is shown in Fig 1. Here we see the 55 MHz isotropic hyperfine coupling results form the proton on the exocyclic nitrogen being 60.6° out of the molecular plane.



**Fig. 1. N3 protonated Cytosine anion.
N3-C4-N4-H torsion angle is -60.6°**

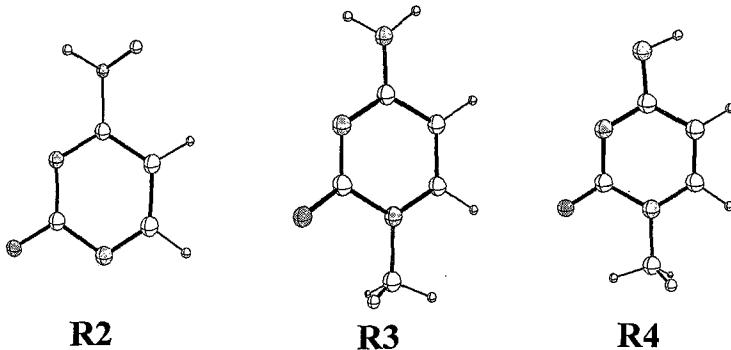
Calculations on a planar Radical R1 are shown in Table II. For the radical in cytosine (N1-H) and in 1-MeC(N1-CH₃) the theoretical results for the isotropic hyperfine couplings and the dipolar couplings are pretty good for the C6-H_a and N3-H_a protons. The theoretical results predict that there will be two additional small couplings to the N4-H_a protons. One of these additional hyperfine couplings was observed in cytosine monohydrate, but not for 1-MeC. Nor were the N4-H_a proton couplings reported in 3'-CMP³⁰ or 5'-dCMP.²¹ In spite of this lack of agreement, the overall comparison between the experimental and theoretical results is rated as fair since the disagreement is mainly with the small hyperfine couplings. As discussed later on, in these experiments it is very difficult to sort out all of the small hyperfine couplings observed.

If the native cytosine bases is protonated, as in 3'-CMP and 5'-dCMP, reduction does not result in further protonation. For example, in 5'-dCMP, the result of reduction produces a radical which is equivalent to the N3 protonated anion (R1) discussed above.²¹

2.1.2 Cytosine Oxidation Products

Oxidation of cytosine produces a radical with sites of unpaired spin density at N1, N3 and C5. The cytosine cation has a pK_a < 4 and in solution deprotonates at NH₂.²⁷ In the solid state Sagstuen *et al.*¹⁹ assigned the primary oxidation radical observed in cytosine monohydrate as the N1 deprotonated cation Radical R2. It is known from the ENDOR experiment that $\rho(\text{N}1) = 0.30$ and $\rho(\text{C}5) = 0.57$. Furthermore there are two small exchangeable N-H couplings whose angular variations correlate well with the exo-cyclic N4-H's.

Wetmore *et al.*³ report DFT calculations on four different deprotonated cations of cytosine. They say that the N1 deprotonated cation (**R2**) best fits the experimental results ($\rho(C5) = 0.49$, $\rho(N1) = 0.29$, and $\rho(O2) = 0.35$). However they add that the C5-H isotropic hyperfine coupling they calculate is -31.5 MHz whereas the experimental value is -41.5 MHz. This, together with a lack of N4 spin density gives them cause to reject the N1 deprotonated cation model, in spite of the fact that their calculations show this model to be the lowest lying radical formed by net hydrogen removal.



One must ask if there are other possible models with hyperfine couplings that give a better match to the experimental data? Wetmore *et al.*³ claim a radical formed by net OH addition to C5 of cytosine produces isotropic hyperfine couplings of $A(N1\text{-H}) = -10.7$ MHz, $A(C5\text{-H}) = 104.9$ MHz and $A(C6\text{-H}) = -37.3$ MHz. No experimental hyperfine coupling of *ca.* 105 MHz was detected. Also, the hyperfine coupling of -37.3 MHz is associated with carbon C6, while the experimental results indicate that the major site of spin density is on C5. It is easy to differentiate between the angular variations of a C5-H_a coupling and a C6-H_a in the analysis of the ENDOR data.

Here then is a case where the theoretical calculations are not in agreement with the model derived from the experimental results which indicate that the oxidation product in cytosine is the N1 deprotonated cation. The agreement between the theoretical calculations and the experimental results is therefore rated here as poor. This problem will be discussed later on.

In some nucleotides, where the cytosine moiety is protonated at N3, oxidation of the base results in an N3 deprotonated cation (Radical **R3**). For example in 5'-

dCMP, the N3 deprotonated cation is characterized by spin densities $\rho(C5) = 0.60$, $\rho(N4) = 0.17$, $\rho(N1) = 0.30$.²¹ Wetmore *et al.*³ have performed calculations on a cytosine cation which appears to be equivalent in structure to the N3 deprotonated cation observed experimentally in 5'-dCMP. They report spin densities $\rho(C5) = 0.33$, $\rho(N3) = 0.24$, $\rho(O2) = 0.45$ which are not very close to those cited above. However, as noted in the Introduction, the accurate calculations of spin densities is very difficult.

Let's look at the computed anisotropic hyperfine couplings for Radical **R3** (Table III). For the primary site of unpaired spin, C6-H_a, the computed isotropic hyperfine is too small, though the computed dipolar couplings are in good agreement with the experimental values. The theoretical calculations nicely reproduce the large N1-C1'-H_b hyperfine coupling indicating significant spin density on N1. The theoretical calculations do not however reproduce the small C4-NH_a couplings determined experimentally. Overall the agreement between the experimental and theoretical results is rated as fair.

A recent paper by Geimer and co-workers reports observing cytosine oxidation products in solution.¹⁶ In cytosine they observe the N1 deprotonated cation (**R2**) with hyperfine couplings $A_{iso}(C5-H) = 42.79$ MHz and $A_{iso}(N1) = 14.15$ MHz which agree with the values determined for the same species observed in single crystals of cytosine monohydrate discussed above. There are some differences for the smaller hyperfine coupling, in particular for the C4-NH₂ couplings. These differences may merely reflect the differences between the radical at room temperature in solution versus the radical in a single crystal at 10 K.

For 1-MeC Geimer *et al.*¹⁶ report a cation deprotonated at the exocyclic amino nitrogen (**R4**). In some of the crystals of cytosine discussed above the base is protonated at N3. In these systems the cation often deprotonates at N3. An aminyl radical has been observed in cytosine hydrochloride.³¹ Here the cytosine base is protonated at N3. The aminyl radical has a rather different spin density distribution in this positive species $\rho(C5) = 0.57$ and $\rho(N1) = 0.16$, giving rise to isotropic hyperfine couplings of C5-H_a = -41.3 MHz and N4-H_a = -12.6 MHz. Calculations on the aminyl radical reported by Geimer *et al.*¹⁶ are shown in Table III. The agreement between the theoretical and experimental isotropic hyperfine couplings is pretty good for the N4-H_a coupling, but is not very good for the other small hyperfine couplings.

Table III: Hyperfine Coupling Parameters for Cytosine Oxidation Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Isotropic	Results Dipolar	(Ref.)
Cytosine N1 deprotonated cation (R2)							
Cm	(19)	-62.4		-21.0		-19.7	
C5-H _α		-42.2	-41.4	-0.8	-30.7	-0.4	
		-19.6		22.8		20.1	
		-23.6		-9.3		-1.3	
C4-NH _{α1}		-16.1	-14.3	-1.8	-1.1	-0.8	
		-3.2		11.1		2.1	
		-19.2		-6.2		-1.9	
C4-NH _{α2}		-16.6	-13.0	-3.6	-0.9	-1.5	
		-3.3		9.8		3.4	
Cytosine(N1-R) N3 deprotonated cation (R3)							
5'-dCMP	(21)	-62.6		-21.4		20.4	
C5-H _α		-42.9	-41.2	-1.7	-32.9	-1.4	
		-18.0		23.1		21.8	
		46.8		-2.4		-3.4	
N1-C1'-H _β		39.5	41.9	-2.4	40.6	0.7	
		39.5		4.8		4.1	
		-18.6		-6.2		-1.3	
C4-NH _{α1}		-16.4	-12.4	-4.0	-0.9	-1.0	
		-2.3		10.2		2.3	
		-24.5		-10.0		-1.9	
C4-NH _{α2}		-16.8	-14.5	2.3	0.1	-1.8	
		-2.3		12.3		3.7	
Cytosine (N1-R) Amino deprotonated cation (R4)							
N1	(16)		5.7			3.1	
N3			12.2			8.4	
N4			29.5			18.7	
N1-CH ₃			12.2			8.0	
C5-H _α			7.4			-5.6	
C6-H _α			3.9			3.7	
N4-H _α			48.9			44.3	

All hyperfine couplings in MHz.

2.1.3 Thymine Reduction Products

The thymine anion radical has been identified in duplex DNA when it is exposed to ionizing radiation at temperatures of 77 K or below.³² Evidence for this radical

assignment comes from measurements that demonstrate the conversion of an EPR doublet associated with the thymine anion into a readily identifiable eight-line spectrum of the 5,6-dihydrothymine-5-yl radical. This conversion is due to the protonation of the thymine anion at C6. While this process seems rather straightforward, questions remain about the fraction of the free radical population assignable to the thymine anion (reduction will also occur at cytosine), and whether or not the thymine anion protonates at C4=O prior to protonation at C6.

The thymine anion is only a weak base ($pK_a = 6.9$).²⁷ This means that protonation of the anion may depend on the specific environment. The primary reduction product observed in the solid state in thymine derivatives is the C4-OH protonated anion Radical R5.¹ This species exhibits significant spin density at C6 and O4. Here one must distinguish between two different situations (see Table IV). In single crystals of thymidine the C4-OH _{β} proton is out of the molecular plane which gives rise to an additional 33.1 MHz isotropic hyperfine coupling.³³ A similar situation is observed in single crystals of anhydrous thymine.³⁴ In 1-MeT however the C4-OH _{β} proton is in the molecular plane and consequently the proton coupling is very small.³³

Wetmore *et al.*² claim that theoretical studies indicate that "the O4-H proton moves back into the molecular plane resulting in a very small hyperfine coupling". This is indeed what happens if one attempts to optimize the structure with O4-H initially out of the ring plane. There is no doubt however from the EPR/ENDOR results from thymidine that the C4-OH _{β} proton is out of the molecular plane (its position correlates with an O4•••O3' hydrogen bond determined from the x-ray crystallographic structure).

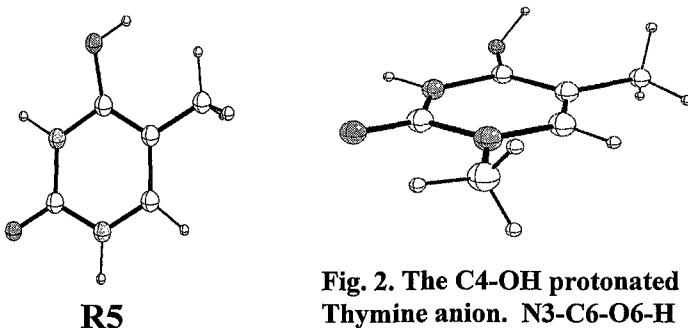


Fig. 2. The C4-OH protonated Thymine anion. N3-C6-O6-H torsion angle here is 110.4°

Three sets of calculations are included in Table IV for the C4-OH protonated anion. For 1-MeT the OH proton is in the molecular plane. Here the calculations nicely reproduce the experimental hyperfine couplings for the two major sites of spin density. For the crystals of anhydrous thymine and for thymidine, the OH is constrained by hydrogen bonding to be out of the molecular plane. Here the calculations are able to reproduce the hyperfine couplings for the C6-H_a, the N3-H_a and for the C4-OH_b protons (with the OH constrained to be out of the molecular plane).

Lü *et al.*¹⁸ have recently reported a study of the thymine anion in solution. They claim that reduction of thymine yields the thymine radical anion. Their results are shown in Table IV. There is a striking similarity between these results and the results for the thymine reduction product observed in a single crystal of thymidine in that both have the same C6-H_a isotropic hyperfine coupling. As shown above, there is ample evidence that the reduction product observed in thymidine in the solid state is protonated at C4=O.

DFT calculations by Lü *et al.*¹⁸ were used to examine the thymine anion assignment. The authors state that a planar radical geometry produces a 48 MHz C6-H_a isotropic hyperfine coupling using B3LYP/6-31G(d). To produce a lower isotropic hyperfine coupling the authors had to consider a non-planar radical geometry with the C6-H_a proton ~10° out of the molecular plane. Previous theoretical work by Wetmore *et al.*² revealed two minima on the potential energy surface of the thymine anion. One structure is nearly planar (C_s) and the other is of C₁ symmetry lying 1.9 kcal/mole below the planar structure. This is a very small energy difference, so the anion observed experimentally could actually include a variety of vibrational states centered about the planar geometry.

The problem however is that in the solid state one knows that the radical geometry is essentially planar. It is not possible for the C6-H_a proton to be ~10° out of the radical plane according to the analysis of the ENDOR experimental data. It is very interesting to note that if one merely uses a higher level basis set to compute the EPR parameters (B3LYP/6-311+G(2df,p) one has a -31.06 MHz isotropic hyperfine coupling with the C6-H_a proton only ~5° out of the radical plane (see Table XI), which is close to the error limits of the direction cosines determined from the ENDOR experiment.³⁵ Of course the EPR/ENDOR experiments were conducted on a single crystal at 10 K so it may not be fair to compare these results with experiments at room temperature in aqueous solution. However it is necessary to explore different basis sets and various means of computing hyperfine couplings before one could conclude that the radical site is bent more than just a few degrees.

Table IV: Hyperfine Coupling Parameters for Thymine Reduction Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Isotropic	Results Dipolar	(Ref.)
Native Thymine (N1-H) anion							
N1	(18)		0.0		-1.4		(2)
N1-H _α			2.3		-6.0		
N3			0.6		1.4		
N3-H _α			2.3		-4.5		
C5-CH ₃			3.9		2.7		
C6-H _α			33.0		-40.3		
Thymine (N1-H) C4-OH protonated anion (R5)							
Ta	(34)	-62.7		-23.4		-24.8	(2)
C6-H _α		-37.0	-39.8	3.7	-40.3	5.4	
		-19.6		19.7		19.4	
		48.8		14.4		9.4	
C4-OH _β		27.4	34.4	-7.0	29.9	-5.7	
		27.0		-7.4		-3.7	
		-12.9		-6.9		-8.4	
N3-H _α		-7.9	-6.0	-1.9	-9.9	-2.7	
		2.7		8.8		11.1	
Thymine (N1-R) C4-OH protonated anion (Fig. 2)							
1-MeT	(33)	-61.1		-23.3		-25.8	(2)
C6-H _α		-33.9	-37.8	3.9	-40.7	2.8	
		-18.4		19.4		23.0	
		-12.0		-5.6		-7.3	
N3-H _α		-7.6	-6.4	-1.3	-8.7	-2.2	
		0.5		6.9		9.5	
TdR ^a	(33)	-55.7		-22.6		-23.2	(2)
C6-H _α		-30.3	-33.1	2.8	-38.6	2.3	
		-13.2		19.8		20.9	
		42.6		-9.5		-5.4	
C4-OH _β		29.2	33.1	3.9	29.1	-3.5	
		27.6		5.6		8.9	
		-11.9		-6.4		-8.4	
N3-H _α		-7.6	-5.4	-2.2	-9.6	-2.6	
		3.2		8.6		11.0	

All hyperfine couplings in MHz.

^aFor this calculation the C4-OH proton is 69.6° out of the molecular plane.

Another interesting part of this study has to do with native anions and protonated anions. As discussed above, Wetmore *et al.* have shown that the radical anion with C₁ symmetry is very non-planar.² In fact the C4-C5-C6-H_α torsion angle is 24°. If one starts with this non-planar geometry with a proton on C4=O, a full optimization produces a nearly planar structure.

2.1.4 Thymine Oxidation Products

There is not much discussion of thymine oxidation products since they are viewed as unimportant in the radiation chemistry of DNA. The feeling being that in DNA most of the oxidation will occur on the purines. However when model systems are used, there are several known pathways that involve oxidation of the thymine base. When a thymine base is ionized, the resulting thymine cation is an acid with $pK_a = 3.6$ for deprotonation in solution.²⁷ The thymine cation will likely deprotonate at N3 though one must look for alternative routes for the cation to eliminate excess charge if N3 is not hydrogen bonded to a good proton acceptor. One could have reversible deprotonation of the thymine cation at N3, or irreversible deprotonation at the C5-CH₃.

In all thymine derivatives studied so far in the solid state there is always a significant concentration of a radical formed by net H abstraction from the >C5-CH₃ group.¹ This allyl-like radical (**R6**) is present at helium temperatures. From studies of frozen thymine solutions it can be shown that the precursor of the allyl-like radical is the thymine cation.³⁶ There are of course other ways to form this neutral product, for example by homolytic dissociation of a C-H bond on the C5-methyl group from an excited state.

The theoretical calculations on radical **R6** are in good agreement with the experimental results (Table V). The two methylene isotropic hyperfine couplings are -42.9 MHz and -45.1 MHz while the theoretical isotropic couplings are -41.4 and -38.6 MHz. The C6-H_a coupling is -28.5 MHz experimentally compared with the theoretical value of -29.9 MHz. The comparison between the experimental and theoretical hyperfine couplings is rated as excellent.

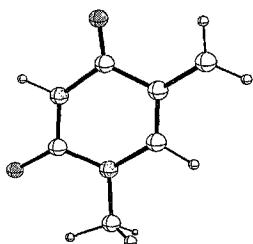
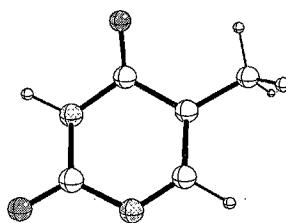
**R6****R7**

Table V: Hyperfine Coupling Parameters for Thymine Oxidation Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Results Isotropic	Computational Results Dipolar	(Ref.)
—Thymine (N1-R) radical (R6)—							
TdR	(33)	-41.8		-13.3		-15.8	(2)
C6-H _α		-29.7	-28.5	-1.2	-29.9	-1.1	
		-14.0		14.5		16.9	
		-65.4		-22.5		-25.9	
C5-CH ₁		-42.9	-42.9	0.0	-41.4	0.6	
		-20.4		22.5		25.3	
		-70.2		-25.1		-23.2	
C5-CH ₂		-42.8	-45.1	2.3	-38.6	-0.6	
		-22.2		22.8		23.8	
—Thymine (N1-H) deprotonated cation (R7)—							
T	(14)						(17)
N1			12.9		10.9		
N3			2.1		-2.5		
N3-H _α			0.9		-0.6		
C5-CH ₃			55.0		49.1		
C6-H _α			1.7		5.4		
—Thymine (N1-R) native cation —							
1-MeT	(15)						(15)
N1			17.7		15.1		
N3			2.0		-2.4		
N3-H _α			22.4		29.2		
C5-CH ₃			58.0		62.0		
C6-H _α			2.0		-0.3		
—Thymine (N1-R) N3-deprotonated cation (R8)—							
1-MeT	(15)			R8a	R8b		(15)
N1			17.7	19.1	-1.9		
N3			1.0	-0.9	-25.8		
N1-CH ₃			22.7	32.7	2.1		
C5-CH ₃			57.8	55.0	8.6		
C6-H _α			1.0	1.9	-0.6		

All hyperfine couplings in MHz.

Geimer *et al.*^{14,15} have looked at oxidation products in thymine and 1-MeT. Experimentally they were able to detect the N1-H deprotonated cation in thymine

(R7), and the native cation in 1-MeT at pH = 1.0. At pH = 7.0 two radical species were detected in 1-MeT, the radical cation, and the N3 deprotonated successor radical. These two radicals have very common spectral characteristics. Both have sites of major spin density at N1 and C5 as expected. These two radicals have not been observed in the solid state.

DFT calculations on these structures have been performed by Naumov *et al.*¹⁷ Their attempts to calculate the structure of the N3 deprotonated structure produced a non-planar structure (with an N1-C2-N3-C4 dihedral angle of *ca.* 40°) in the gas phase. In a dielectric medium their calculations (using the Onsager model) produced a nearly planar structure. Attempts to produce a non-planar N3 deprotonated cation led to a new radical structure shown here (Radical R8a). Also shown here is the non-planar radical structure R8b.

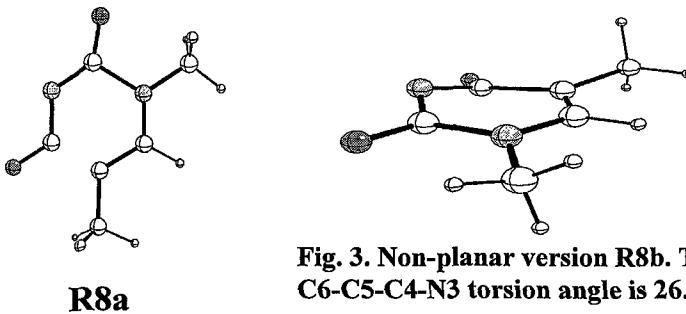


Fig. 3. Non-planar version R8b. The C6-C5-C4-N3 torsion angle is 26.2°

These new calculations were done in the gas phase. No dielectric medium was imposed. Tabulated results show this new structure is slightly lower in energy than that reported by Naumov *et al.*¹⁷ Also the calculated hyperfine couplings seem to agree slightly better with the experimental results (Table V). One needs to comment on the structure of this radical. The N1-C2 distance is nearly 1.6 Å. Drawing programs show radical R8a as a ring-opened radical. However even at this unusually long separation, there is still some bonding character between N1 and C2. A second column in Table V for the N3 deprotonated cation is for the non-planar structure R8b. The computed isotropic hyperfine couplings are very different from those of radical R8a, and are not even close to the reported experimental hyperfine couplings.

2.2 Purines

There has been continued interest in the radiation chemistry of the purines since early reports on oriented DNA by Gräslund *et al.*³⁷ which suggest that the main site of oxidation in DNA is the guanine base. Crystals of purines are difficult to prepare and often give weak EPR signals. Some of the experiments discussed below involve crystallization with HCl which results in protonation (for example at N7 in Guanine). In the study of radiation induced defects in model systems one observes both reduction and oxidation products. It is a remarkable fact that in aqueous solution the electron adducts of the purine nucleosides and nucleotides undergo irreversible protonations at carbon with a rate constant two orders of magnitude higher than that for carbon protonation of the electron adduct in thymidine.³⁸ It is therefore important to know the properties of the various purine reduction products, and to ask why they haven't been observed in irradiated DNA.

2.2.1 Guanine Reduction Products

In single crystals of 5'-dGMP the native molecule is not protonated at N7. EPR/ENDOR experiments detected a narrow doublet whose hyperfine coupling correlates with a C8-H_a interaction. The computed spin density was $\rho(C8) = 0.11$. This radical was unstable on warming above 10 K, and therefore it was proposed that the radical responsible was the pristine radical anion.³⁹ Wetmore *et al.*⁴ report that the theoretical spin densities of the radical anion are $\rho(C8) = 0.08$, $\rho(C2) = 0.57$ and $\rho(N2) = 0.12$. There are significant distortions in this model with the amino group out of the molecular plane. In fact the authors report that one of the amino protons is actually perpendicular to the molecular plane which gives rise to a large hyperfine coupling (Figure 4). The computed isotropic hyperfine couplings are C8-H_a = 7.0 MHz, N2-H_{a1} = 89.4 MHz and N2-H_{a2} = 4.5 MHz. The large N2-H_{a1} coupling was not observed experimentally. A C_s geometry of the anion was obtained through a constrained optimization which is 17.9 kcal/mole higher in energy than the non-planar structure. The computed isotropic hyperfine coupling for the two N2-H_a protons are now *ca.* 7.0 MHz in much better agreement with the experimental results. The authors do comment that the computations produced two additional couplings to the N1 and N9 protons that are not observed experimentally. A calculation on the C_s geometry of the anion yields a significant N2 = 70.6 MHz isotropic hyperfine coupling (Table VI). This nitrogen coupling was also not observed experimentally. This leads the authors to question whether the radical anion has actually been observed.

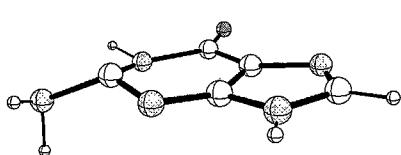


Fig. 4. Native guanine anion. The N3-C2-N2-H torsion angle is -66.7°

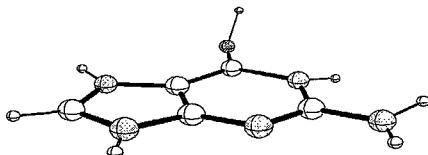


Fig. 5. C6-OH protonated anion. The N1-C6-O6-H torsion angle is -74.9°

Since nitrogen and oxygen have greater electron affinity than carbon, it is reasonable to assume that the negative charge in the radical anion resides mainly on the heteroatoms. These heteroatoms should be the sites of protonation in guanine after reduction. It is important to consider if the experimental data could be explained by a C6-OH protonated anion. Calculations on the C6-OH protonated guanine anion are given in Table VI. For the planar structure there seems to be rather good agreement with the experimental data originally assigned to the native anion in 5'-dGMP. Basically the C8-H_a isotropic hyperfine coupling is pretty good, and there are no significant nitrogen hyperfine couplings that would tend to broaden the EPR spectra. This may also explain the instability observed when the reduction product in 5'-dGMP is warmed. It is conceivable that at 10 K the native anion protonates at C6-O. The protonation reverses upon warming leaving the original anion which is now subject to recombination. Here then is a case where the theoretical calculations point out problems with a radical assignment based on the experimental results, and furthermore provide a more suitable radical assignment.

In several experimental studies the native molecule was protonated at N7.⁴⁰ In these crystals reduction leads to protonation at O6. The O6 protonated anion is characterized by $\rho(C8) = 0.28$, $\rho(N1) = 0.15$ and $\rho(N7) = 0.10$. Calculations on the O6 protonated guanine anion by Wetmore *et al.*⁴ produce a non-planar structure. The N1-C6-O6-H torsion angle is $> 70^\circ$ out of the molecular plane which results in a very large isotropic hyperfine coupling (61.7 MHz) from the OH proton (Fig. 5). The hydrogen on N1 is also slightly out of the plane and this gives rise to an $A_{iso}(N1-H) = 19.9$ MHz. Neither of these couplings were observed experimentally.

Table VI: Hyperfine Coupling Parameters for Guanine Reduction Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Isotropic	Results Dipolar	(Ref.)
Guanine native anion (Fig. 4)							
5'-dGMP	(39)			Planar	Non-planar		(4)
N1			23.6		4.3		
N3			14.9		19.1		
N2			70.6		55.1		
N2-H1			-4.7		4.1		
N2-H2			-0.1		88.0		
N3-H _α			-2.7		-4.1		
C8-H _α		-8.5		-0.4	-5.3		
N9-H _α			-1.4		-3.4		
Guanine C6-OH protonated anion							
5'-dGMP	(39)			Planar	Non-planar		(4)
N1			-0.1		6.0		
N3			1.2		2.7		
N2			-0.8		-0.6		
N2-H1			-0.4		0.0		
N2-H2			-0.5		-1.1		
C8-H _α		-8.5		-7.0	-6.4		
N9-H _α			-5.4		-1.4		
Guanine N7-H C6-OH protonated anion (Fig. 5)							
GMP	(40)			Planar	Non-planar		(4)
N1				1.2	5.4		
N3				2.1	2.7		
N1-H _α		-17.6		-7.3		-7.5	
		-12.0	-10.3	-1.7	-8.6	-0.6	-2.8
		-1.2		9.0			10.3
N2-H1				0.0	-0.5		
N2-H2				-0.1	-0.1		
N7-H _α		-13.9		-4.6		-4.9	
		-12.1	-9.3	-2.8	-8.0	-5.9	-3.3
		-2.0		7.4			8.2
		-30.1		-9.9			-19.6
C8-H _α		-21.2	-20.2	-1.0	-35.3	-32.6	1.7
		-9.3		10.9			17.9
N9-H _α				2.5	2.5		
O6-H _β		5.5		4.3		-5.9	
		1.4	1.2	0.2	4.7	60.5	-3.5
		-3.4		-4.5			9.4

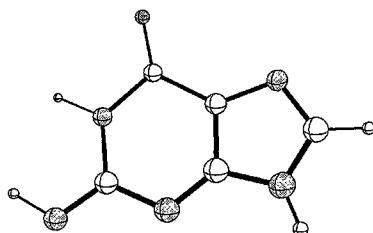
All hyperfine couplings in MHz.

It is therefore necessary to perform calculations with the O6 protonated anion constrained to C_s symmetry (this radical lies only 1.7 kcal/mole above the energy of the non-planar radical). For the planar radical the computed C8-H isotropic hyperfine coupling is in good agreement with the experimental values (-7.0 vs. -8.5 MHz). The other calculated parameters are also in good agreement with the experimental values as seen in Table VI.

2.2.2 Guanine Oxidation Products

The guanine cation is a weak acid ($pK_a = 3.9$).²⁷ Therefore deprotonation will depend on the environment. Bachler and Hildenbrand have studied the guanine oxidation product in aqueous solution of 5'-dGMP.⁴¹ The best fit to their EPR spectra seems to be from the radical cation (guanine remains protonated at N1).

In the solid state, oxidation of 5'-dGMP at 10 K leads to deprotonation at the exo-cyclic nitrogen (**R9**) which is characterized by $\rho(C8) = 0.175$ and $\rho(N2) = 0.33$.³⁹ The same radical was detected in crystals of 3',5'-cyclic guanosine 5'-monohydrate. In this second study, the N3 spin density was determined to be 0.31.⁴²



R9

Wetmore *et al.*⁴ report calculations on the N2 deprotonated cation with spin densities $\rho(C8) = 0.19$, $\rho(N2) = 0.35$ and $\rho(N3) = 0.37$. Here one sees excellent agreement with the same parameters determined experimentally (Table VII). This table also contains the dipolar tensors. There is also rather good agreement between the dipolar couplings determined from the experimental data, and the corresponding dipolar couplings determined from the DFT calculations.

Table VII: Hyperfine Coupling Parameters for Guanine Oxidation Products

Matrix (Coupling)	(Ref.) Principal Values	Isotropic Value	Dipolar Values	Computational Results Isotropic	(Ref.) Dipolar
—Guanine native cation—					
G:HCl:H ₂ O (43)					(4)
N1				-2.2	
N3		16.8		6.9	
N7				-1.3	
N9				-4.1	
N2		10.0		3.4	
N2-H1		12.1		-8.2	
N2-H2		12.1		-7.1	
	-21.0		-6.5		-6.5
C8-H _a	-14.0	-14.5	0.5	-22.7	-1.6
	-8.4		6.0		8.1
N9-H _a				0.6	
—Guanine N2-H deprotonated cation (R9) —					
5'-dGMP (39)					(4)
N1				-0.8	
N3				11.9	
N7				-2.1	
N9				-1.1	
	-46.2		-19.2		-20.6
N2-H1	-29.6	-27.0	-2.6	-22.7	-3.1
	-5.1		21.8		23.7
	-21.1		-7.3		-9.3
C8-H1	-14.5	-13.8	-0.7	-16.1	-0.7
	-5.7		8.0		10.0
N9-H _a				-1.6	

All hyperfine couplings in MHz.

Wetmore *et al.*⁴ have performed theoretical calculations on a guanine molecule that was originally protonated at N7. Electron loss by this molecule leads to deprotonation at N7 yielding a radical which is equivalent to the guanine cation. The resulting spin densities are $\rho(N3) = 0.21$, $\rho(N2) = 0.10$, $\rho(C4) = 0.17$, $\rho(C5) = 0.29$, and $\rho(C8) = 0.28$. The experimental results from the guanine cation are listed in Table VII. Basically one has $\rho(C8) = 0.18$, $\rho(N2) = 0.17$ and $\rho(N3) = 0.28$.⁴³

The experimental results on the guanine oxidation product are as follows. From the ENDOR spectra there is a small C8-H_a coupling with $A_{iso} = -14.5$ MHz. To simulate the EPR spectra one needs two nitrogen couplings and two small proton couplings chosen as N3 = 16.8 MHz and N2 = 10 MHz.⁴³ The easiest way to produce a radical that could have these couplings would be via N7 deprotonation. Wetmore *et al.*⁴ report calculations on four other dehydrogenated radicals which don't seem to provide any better models for the guanine cation. At least one can say that while the calculations don't reproduce the experimental results very well, they do provide a model which has features in common with the model suggested from the experimental results.

In their summary of the calculations on guanine radicals Wetmore *et al.*⁴ make an observation about the situation in the crystals where the native molecule is protonated at N7. In these crystals the experimentally assigned reduction product has the C6-O protonated structure as discussed above. The authors note that the computed hyperfine couplings for this species are very similar to those expected for the guanine cation. The question is, can one mix up a oxidation product assignment with a reduction product assignment from the analysis of the experimental data. Basically one would have to look at all the radicals observed at 10 K. Since an irradiated crystal must remain with no net charge, one needs an equal number of electron loss and electron gain events. Take guanine:HCl:H₂O for example. There are essentially two distinct EPR signals. An EPR doublet is easily assigned to the C6-OH protonated anion, while the broad EPR signal resulting from extensive nitrogen hyperfine couplings is assigned to the N7 deprotonated cation. One would not want to assign the oxidation product as a reduction product if this means upsetting the balance between electron loss and electron gain products.

2.2.3 Adenine Reduction Products

Wetmore *et al.*⁵ have computed the structure of the native adenine anion. The pyrimidine ring is planar, with the amino group out of the ring plane. However the imidazole ring is puckered at the C8 position. A C_s geometry of the anion was obtained through a constrained optimization. This structure was 4.7 kcal/mole higher in energy than the non-planar radical anion. To date the native adenine anion has not been detected in the solid state, so there are no experimental results to compare the computed hyperfine couplings with.

The adenine anion has a pK_a = 3.5.²⁷ After electron capture the negative charge of the adenine radical anion resides mainly on N1, N3, and N7 and therefore

protonation likely occurs at one of these nitrogens. The results in Table VIII show that in single crystals examined at 10 K that reduction of adenine leads to the N3 protonated adenine anion (**R10**) with spin densities of *ca.* $\rho(C2) = 0.41$, $\rho(C8) = 0.14$, and $\rho(N3) = 0.12$.⁷

Table VIII: Hyperfine Coupling Parameters for Adenine Reduction Products

Matrix (Coupling)	(Ref.) (7)	Principal Values	Isotropic Value	Dipolar Values	Computational Results Isotropic	Computational Results Dipolar	(Ref.)
—Adenine N3-H protonated anion (R10)—							
AR				Planar	Non-planar		(5)
N1				-8.1	-8.9		
N3				4.1	12.5		
N6				1.4	10.9		
		19.2		9.4		13.5	
N3-H _a		11.6	-9.8	1.8	-12.2	-46.1	-3.0
		-1.4		-11.2			-10.5
		-46.0		-16.1			-23.8
C2-H _a		-27.8	-29.8	2.0	-42.1	-36.0	1.8
		-15.7		14.1			22.0
		-18.4		-7.0			-7.0
C8-H _a		-10.4	-11.4	1.0	-12.2	-6.9	0.6
		-5.5		6.0			6.4
N9-H _a				0.1	0.6		

All hyperfine couplings in MHz.

Recent theoretical calculations on the N3 protonated anion yield spin densities of $\rho(N3) = 0.11$, and $\rho(C2) = 0.49$.⁵ These results are due to distortions with the amino group puckered and both hydrogens out of the molecular plane (Fig. 6). Calculations on a planar model yield spin densities of $\rho(N3) = 0.12$, $\rho(C8) = 0.13$ and $\rho(C2) = 0.53$. These agree nicely with the experimentally determined results. The fully optimized C_s geometry is only 1.7 kcal/mole above the non-planar form.

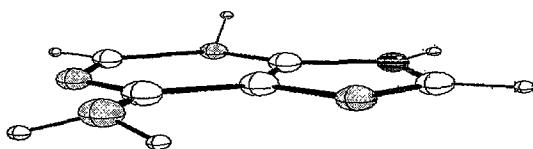


Fig. 6. The N3 protonated adenine anion (R10).
The N6-H2 protons are out of the plane. The N1-C2-N3-H torsion angle is -142°.

These are very satisfying results. The detailed DFT calculations performed on an isolated N3 protonated adenine anion produced an N3-H_a hyperfine coupling of 46.1 MHz, in poor agreement with the experimentally determined value of -9.8 MHz. The theoretical calculations are ignoring hydrogen bonding network that the free radical is imbedded in. When the calculation is repeated with the N3-H_a proton confined to the molecular plane, the hyperfine coupling is computed to be -12.2 MHz, in good agreement with the experimentally observed N3-H_a isotropic hyperfine coupling (Table VIII).

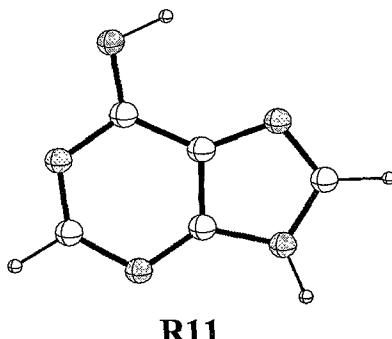
2.2.4 Adenine Oxidation Products

The adenine cation was observed in a single crystal of adenine hydrochloride hemihydrate.⁴⁴ In this crystal the adenine is protonated at N1. After electron loss the molecule deprotonates at N1. This produces a radical that is structurally equivalent to the cation of the neutral adenine molecule with spin density on C8 and N6 ($p(C8) = 0.17$ and $p(N6) = 0.25$). Theoretical calculations on the cation of the neutral adenine molecule have $p(C8) = 0.18$ and $p(N6) = 0.25$ which are seen in Table IX to be in excellent agreement with the experimental results. One also notes here rather satisfying agreement between the computed and experimental dipolar couplings.

Table IX: Hyperfine Coupling Parameters for Adenine Oxidation Products

Matrix (Coupling)	(Ref.)	Principal Values	Isotropic Value	Dipolar Values	Computational Results Isotropic	(Ref.)
—Adenine native cation—						
A:HCl:1/2H ₂ O						(5)
N1	(44)				-0.3	
N3		6.0			5.5	
N6		3.6			9.6	
C2-H _α				-6.7		
		-32.4		-12.7		-15.0
N6-H1		-23.3	-19.7	-3.6	-18.6	-3.6
		-3.4		16.3		18.6
		-35.2		-14.7		-13.3
N6-H2		-23.6	-20.5	-3.0	-18.9	-4.0
		-2.8		17.7		17.3
		-20.6		-6.9		-8.9
C8-H _α		-14.8	-13.7	-1.1	-15.4	-1.6
		-5.7		8.0		10.5
N9-H _α				-1.0		
—Adenine N6-H deprotonated cation (R11)—						
Adenosine						(5)
N1	(7)				4.8	
N3					8.1	
N7				-2.2		
N9				-1.5		
		-55.9		-22.0		-31.3
N6-H1		-37.4	-33.9	-3.5	-35.8	-5.3
		-8.4		25.5		36.6
		-18.8		-6.3		-6.1
C8-H _α		-12.1	-12.4	0.4	-10.4	-0.5
		-6.5		5.9		6.6
N9-H _α				-1.5		

All hyperfine couplings in MHz.



The adenine cation is strongly acidic ($pK_a < 1$).²⁷ This strong driving force makes the reaction independent of environmental conditions. In single crystals of adenosine⁷ and anhydrous deoxyadenosine⁴⁵ the N6 deprotonated cation (**R11**) is observed which is characterized by $\rho(C8) = 0.16$ and $\rho(N6) = 0.42$. In single crystals of deoxyadenosine⁴⁶ the site of oxidation seems to the deoxyribose moiety. Theoretical calculations on the N6 deprotonated cation yield spin densities of $\rho(N10) = 0.59$, $\rho(N1) = 0.17$, and $\rho(N3) = 0.23$.⁵ The experimental isotropic hyperfine couplings are $N6-H_a = 33.9$ MHz, and $C8-H_a = 12.4$ MHz while the calculated couplings are $N6-H_a = 35.8$ MHz and $C8-H_a = 10.4$ MHz, showing satisfactory agreement (Table IX).

3. Comparisons of Experimental and Theoretical Results

Nineteen different radiation induced defects were presented in Tables II-IX. A summary on the ability of DFT to reproduce the observed hyperfine couplings and spin densities is presented in Table X. There are separate entries for the main sites of spin density, and the secondary sites that result in numerous small hyperfine couplings.

Table X: Summary of Calculated Results

Crystal System	Radical	Computational Results		Overall Rating
		Prim. Sites	Sec. Sites	
Cytosine:H ₂ O	N3-protonated anion (Fig. 1)	+	+	Good
1-MeCytosine	N3-protonated anion R1	+	-	Fair
Cytosine:H ₂ O	N1-deprotonated cation R2	-	-	Poor
5'-dCMP	N3-deprotonated cation R3	-	-	Fair
Cytosine	Amino deprotonated cation R4	+	-	Good
Thy. anhyd.	C4-OH protonated anion R5	+	+	Good
Thymine	Native anion	-	+	Fair
1-MeThymine	C4-OH prot. anion (Fig. 2)	+	+	Good
TdR	Allyl-like radical R6	+	+	Excellent
Thymine	N1-H deprotonated cation R7	+	+	Good
1-MeThymine	N3-H deprotonated cation R8	+	+	Good
5'-dGMP	Native anion (Fig. 4)	-	-	Poor
5'-dGMP	C6-OH protonated anion	+	0	Good
GMP	N7-H C6-OH prot. anion (Fig. 5)	-	+	Fair
G:HCl:H ₂ O	Native cation	-	-	Poor
5'-dGMP	N2-H deprotonated cation R9	+	+	Good
AR	N3-H protonated anion R10	+	0	Good
A:HCl:1/2H ₂ O	Native cation	+	0	Excellent
Adenosine	N6-H deprotonated cation R11	+	0	Excellent

Offhand the agreement shown in Table X looks pretty good. Twelve out of nineteen ratings are either good or excellent. However it must be remembered that calculations on many of the negative ions produced distorted molecules with protons out of the molecular plane. This led to some very large hyperfine couplings which were not observed experimentally. A first order solution was to repeat the calculation on a molecule with planar symmetry. While this led to some

improvements, there are still some disagreements for many of the small N-H_α hyperfine couplings. Also it must be recalled that calculations on the dominant C6-H_α hyperfine interactions observed in the thymine reduction products were not in very good agreement with the experimental couplings for some of the protonated anions. Basically, gas phase calculations on the protonated anions produce non-planar molecules. However the experimental data comes from the analysis of free radicals in single crystals. In single crystals the molecules are constrained to nearly planar geometries by the hydrogen bonding environment. How then is one to reconcile these differences?

First of all it is important to note that there are cases where planarity is found in a crystalline environment and nonplanarity in solution. This is so for the biphenyl molecule. Also, there is still a chance for a radical in the solid state to be slightly non-planar. In Tables I and II there are experimental results on the cytosine reduction products. Here the experimental data indicates (Table I) a molecule planar at the C6-H_α radical site. Also one notes (Table II) pretty good agreement between the computed and experimental C6-H_α isotropic hyperfine coupling. The same cannot be said for the reduction products in thymine. The C6-H_α isotropic hyperfine coupling for the radical anion is only -33 MHz, while most calculations for this radical have the C6-H_α coupling more like -40 MHz. A gas phase calculation has the C6-H proton 24.3° out of the plane. This is difficult to reconcile with the experimental hyperfine coupling tensor for this radical. However, suppose one looks at the situation somewhere in the region between the planar and highly distorted radical (Table XI). Here one sees reasonable agreement between the experimental and calculated C6-H_α isotropic hyperfine coupling for a small non-planar bending (less than 5°). This small non-planarity is easy to reconcile with the experimental data. This idea may also explain other problems. Some time ago Bernhard and Patrzalek showed a spread of C6-H_α isotropic hyperfine couplings for various thymine anions in various LiCl glasses.³² It may be that some of the smaller C6-H_α isotropic hyperfine couplings are from molecules that are slightly non-planar at the radical site.

Table XI: Hyperfine Couplings for the Non-Planar Thymine Anion

Experimental Coupling ^a	Isotropic Value ^a	Torsion Angle ^b	Total Energy ^c
Thymine			
	-40.31	0.0°	-454.157367
	-31.06	5.0°	-454.157731
C6-H _a	-21.59	10.0°	-454.158108
	-33.1	15.0°	-454.158432
	-2.55	20.0°	-454.158648
	+5.23	24.3°	-454.158711

^aAll hyperfine couplings in MHz.

^bThis angle is the C4-C5-C6-H torsion angle.

^cEnergy units in Hartrees. The first entry is for the planar C₂ structure. The other entries are for the non-planar C_s structures, with the last entry being the energy minimum.

There have been calculations on smaller systems (the ethynyl radical) by a hybrid DFT + Molecular Dynamics scheme. The vibrationally averaged motions modify the hyperfine structure of the radical and are shown to agree much better with the experimental data than do the optimized vacuum geometry results.⁴⁷ It would be fascinating to extend these methods to some of the radicals discussed here where vibrations at the radical site seem to be present. In cases where the theoretical and experimental results are not too different, perhaps it is time to work together on alternative models which embody the best parts of the experimental results in concert with the very useful trends indicated by the theoretical results. Perhaps some compromise here might actually produce models which show even better agreement with the experimental results.⁴⁸

At the top of Table X something seems amiss with the cytosine results. One sees rather poor agreement when DFT is used to calculate the small NH₂ hyperfine couplings. The addition of d-polarization functions lead to small sp³ pyramidalization of the DNA base amino groups when calculations are done in the gas phase. Of course the experimental data are derived from single crystals where the NH₂ are involved in hydrogen bonds. It is assumed that this leads to basically planar structures. No one knows however if this is indeed the case because it is so difficult to detect hydrogens with x-ray diffraction techniques. Furthermore, who

knows what the actual structure of a free radical actually is? It was shown above that there is good evidence that some cytosine reduction products are essentially planar at the C6-H_a site. It is still an open question however as to whether or not there may be some non-planarity in the cytosine amino groups. Also one must ask what is so different about cytosine? In Table III one sees further problems with calculating the N-H_a couplings for two cytosine oxidation products. It is interesting to note that for the native adenine cation (Table IX), there is excellent agreement between the theoretical calculations and the small N4-H₂ experimental hyperfine couplings.

What is needed here is a thorough review of the actual experimental data. When there are no reports of small NH₂ hyperfine couplings, one needs to know exactly why. Are these couplings actually absent, or was their presence masked by the interference of other small hyperfine couplings? Basically one needs to know what limits can be placed on the experimental results. From the theoretical side, one needs to know a whole lot more about the limits of pyramidalization of the DNA base amino groups. One already knows what happens in the gas phase, where the β-protons are way out of the molecular plane. This does not occur in the solid state, and most likely doesn't occur in DNA. So one needs to know what acceptable limits can be placed on the non-planarity of these small β-protons.

The next problem seen in Table X involves the reduction product in neutral guanine. The discussion above points out that the original radical assignment has to be questioned. It seems more likely that the second entry in Table VI, the C6-OH protonated anion, is a better model for the reduction product observed in 5'-dGMP.

The last problem in Table X involves the oxidation product in N7 protonated guanine. The original model for this defect was the N7 deprotonated cation. Wetmore *et. al.* reject this model since their DFT calculations don't match the experimental hyperfine coupling very well.⁴ One notes that two large nitrogen hyperfine couplings are needed to reproduce the EPR spectrum. Here it seems proper while questioning the interpretation of the experimental results to also ask how good one expects the computed results to be. Perhaps this is a situation where the lack of the electrostatic environment in the calculations is a problem. The experimental data is taken from a crystal grown in HCl. One needs to consider whether or not the inclusion of these counter ions affects the observed hyperfine couplings.

4. Relevance to DNA

The studies of the model compounds discussed here were directed at understanding the radiation chemistry of DNA. This is a very challenging problem since it is already evident here that one does not yet completely understand the radiation chemistry of the individual DNA bases. In DNA one is faced with the problem of understanding the radiation chemistry of the DNA base pairs. Single crystals of dinucleotides have been grown that are large enough for x-ray diffraction studies, but they have never been grown large enough for detailed EPR/ENDOR experiments. This is a severe limitation since it is necessary to have the high resolution of the ENDOR experiment to sort out the small hyperfine couplings so as to be able to determine the protonation states of the radiation induced products in DNA.

One needs to step back here and ask what are the relevant questions, and can theoretical calculations be of use in answering these questions. Certainly one would like to know, at the molecular level, how DNA is damaged by ionizing radiation. This would include an understanding of what causes DNA strand breaks. While it is believed that one electron oxidation of guanine is a likely intermediate in this process, the steps proceeding from the oxidation of guanine to strand cleavage are not at all clear. Since bigger and faster computers are now available, it should now be possible to accurately model radiation damage pathways in DNA. As a first step, consider the problem of oxidation of a nucleotide.

Sevilla and co-workers have looked at phosphorylated deoxyribose.⁴⁹ They optimized the geometry of 2'-deoxyribose 5'-phosphate at the ROHF/3-21 G and found the hole localized in the deoxyribose ring oxygen (O4'). When they performed a single point calculation on the cation in 3'-dCMP they found the resulting spin density entirely localized on the cytosine. A review of this work on assessing radiation damage to the sugar phosphate backbone of DNA has appeared.⁵⁰

To update this work a new calculation has been performed on 5'-dCMP. The molecule was optimized at HF/6-31 G(d,p) with only the glycosidic torsion angle frozen to keep the phosphate group away from the N3 protonated cytosine. Spin densities were calculated on the optimized structure using UB3LYP/6-311 G(2df,p). These calculations show the hole localized at the P-OH oxygen ($\rho(O19) = 0.832$). The phosphorous isotropic hyperfine couplings is calculated to be $A_{iso}(P) = -102.2$ MHz. These features are characteristics of a PO_4^{2-} radical.⁵¹ For a pure PO_4^{2-} radical $A_{iso}(P) = -81$ MHz, and can increase to >100 MHz as the hole be-

comes more localized on the protonated oxygen. A calculation on the nucleoside (with the phosphate removed and replaced with the proton from N3) yields the cytosine cation radical (**R3**) as seen in Table III. A search is currently underway to look for the hole localized on the deoxyribose ring oxygen using the higher level basis sets.

5. Summary

This review has focused on the primary radiation induced damage to the DNA bases. The emphasis has been on using theoretical calculations to assist in making free radical assignments. In many cases there is rather good agreement between the experimental and theoretical results. In other cases the agreement between the experimental and theoretical results is not very good. In some of these later cases the following events may lead to incorrect theoretical results. Most of the figures presented here were chosen to show the results of theoretical calculations that resulted in non-planar free radical structures. In many cases these structures contained out-of-plane protons that give rise to large hyperfine couplings that were not observed experimentally. Basically what is happening here is that nuclear screening is altered after electron loss or electron addition. It would be highly desirable to relax the initial configuration so that the optimization proceeded to the nearest potential minimum. In the actual calculations however, the nuclei pass right through this minima into another configuration. In these cases the predicted geometry of the free radicals are not in good agreement with the experimental results, and so these theoretical geometries must be used with caution in assessing the experimental results.

6. Acknowledgements

This work is supported by PHS Grant R01 CA36810-15 awarded by the National Cancer Institute, DHHS.

7. References

1. D. M. Close, *Radiat. Res.*, 1993, **135**, 1.
2. S. D. Wetmore, R. J. Boyd, and L. A. Eriksson, *J. Phys. Chem. B*, 1998, **102**, 5369.
3. S. D. Wetmore, F. Himo, R. J. Boyd, and L. A. Eriksson, *J. Phys. Chem. B*, 1998, **102**, 7484.
4. S. D. Wetmore, R. J. Boyd, and L. A. Eriksson, *J. Phys. Chem. B*, 1998, **102**, 9332.
5. S. D. Wetmore, R. J. Boyd, and L. A. Eriksson, *J. Phys. Chem. B*, 1998, **102**, 10602.
6. A. O. Colson, B. Besler, D. M. Close, and M. D. Sevilla, *J. Phys. Chem.*, 1992, **96**, 661.
7. D. M. Close and W. H. Nelson, *Radiat. Res.*, 1989, **117**, 367.
8. J. Sponer and P. Hobza, *Int. J. Quant. Chem.*, 1996, **57**, 959.
9. M. D. Sevilla, B. Besler, and A. O. Colson, *J. Phys. Chem.*, 1995, **99**, 1060.
10. Batra *et al.* (*J. Am. Chem. Soc.* 1996, **100**, 18371) have calculated isotropic hyperfine couplings in a variety of organic radicals. They claim that BLYP and B3LYP calculations give isotropic hyperfine couplings that are within *ca.* 10% from the experimental values. Jensen *et al.* (*J. Phys. Chem.* 1996, **100**, 954) have compared DFT and MP2 calculations of spin densities for tryptophan radicals. For these calculations the MP2/6-31++G spin densities do not agree with the experimental spin densities as well as those obtained by B3LYP/TZ2P calculations. While this is interesting, it still remains to be determined exactly what are the best methods to use for calculating hyperfine couplings in the primary radiation induced radicals observed in the nucleobases.
11. X. Li, Z. Cai, and M. D. Sevilla, *J. Phys. Chem. B*, 2001, **105**, 10115.
12. W. A. Bernhard, J. Barnes, K. M. Mercer, and N. Mroczka, *Radiat. Res.*, 1994, **140**, 199.
13. D. M. Close, L. A. Eriksson, E. O. Hole, E. Sagstuen, and W. H. Nelson, *J. Phys. Chem. B*, 2000, **104**, 9343.
14. J. Geimer, O. Brede, and D. Beckert, *Chem. Phys. Letts.*, 1997, **276**, 411.
15. J. Geimer and D. Beckert, *J. Phys. Chem. A*, 1999, **103**, 3991.
16. J. Geimer, K. Hildenbrand, S. Naumov, and D. Beckert, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4199.

17. S. Naumov, A. Barthel, J. Reinhold, F. Dietz, J. Geimer, and D. Beckert, *Phys. Chem. Chem. Phys.*, 2000, **2**, 4207.
18. J. M. Lü, J. Geimer, S. Naumov, and D. Beckert, *Phys. Chem. Chem. Phys.*, 2001, **3**, 952.
19. E. Sagstuen, E. O. Hole, W. H. Nelson, and D. M. Close, *J. Phys. Chem.*, 1992, **96**, 8269.
20. H. M. McConnell, C. Heller, T. Cole, and R. W. Fessenden, *J. Amer. Chem. Soc.*, 1960, **82**, 766.
21. D. M. Close, E. O. Hole, E. Sagstuen, and W. H. Nelson, *J. Phys. Chem.*, 1998, **102**, 6737.
22. Gaussian 98, Revision A.7, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.
23. W. A. Bernhard, in 'Early Effects of Radiation on DNA', NATO ASI Series H, Cell Biology, eds. E. M. Fielden and P. O'Neill, Springer-Verlag, Berlin, Heidelberg, 1991, p. 141.
24. M. D. Sevilla, D. Becker, M. Yan, and S. R. Summerfield, *J. Phys. Chem.*, 1991, **95**, 3409.
25. J. Barnes, W. A. Bernhard, and K. R. Mercer, *Radiat. Res.*, 1991, **126**, 104.
26. J. P. Barnes and W. A. Bernhard, *J. Phys. Chem.*, 1994, **98**, 887.
27. S. Steenken, *Free Radical Res. Comm.*, 1992, **16**, 349.
28. A. Hissung and C. von Sonntag, *Int. J. Radiat. Biol.*, 1979, **35**, 449.
29. A word about the format of the Tables. Two sets of references are included. The first is to the experimental results (usually the detailed EPR/ENDOR experiment on an irradiated single crystal). The second reference is to the original theoretical results. All of the calculations from

the open literature were repeated for the present work. This was necessary in order to have the final xyz coordinates for the radical structures shown in the Figures. In most cases the new calculations agree with the published results.

30. H. C. Box, W. R. Potter, and E. E. Budzinski, *J. Chem. Phys.*, 1975, **62**, 3476.
31. E. O. Hole, W. H. Nelson, E. Sagstuen, and D. M. Close, *Radiat. Res.*, 1998, **149**, 109.
32. W. A. Bernhard and A. Z. Patrzalek, *Radiat. Res.*, 1989, **117**, 379.
33. E. O. Hole, E. Sagstuen, W. H. Nelson, and D. M. Close, *J. Phys. Chem.*, 1991, **95**, 1494.
34. E. Sagstuen, E. O. Hole, W. H. Nelson, and D. M. Close, *J. Phys. Chem.*, 1992, **96**, 1121.
35. D. M. Close, *Phys. Chem. Chem. Phys.*, 2001, **3**, 43.
36. M. D. Sevilla, *J. Phys. Chem.*, 1971, **75**, 626.
37. A. Gräslund, A. Ehrenberg, and A. Rupprecht, *Int. J. Radiat. Biol.*, 1977, **31**, 145.
38. L. P. Candeias, P. Wolf, P. O'Neill, and S. Steenken, *J. Phys. Chem.*, 1992, **96**, 10302.
39. E. O. Hole, W. H. Nelson, E. Sagstuen, and D. M. Close, *Radiat. Res.*, 1992, **129**, 119.
40. E. Sagstuen, E. O. Hole, W. H. Nelson, and D. M. Close, *Radiat. Res.*, 1988, **116**, 196.
41. V. Bachler and K. Hildenbrand, *Radiat. Phys. Chem.*, 1992, **40**, 59.
42. E. O. Hole, E. Sagstuen, W. H. Nelson, and D. M. Close, *Radiat. Res.*, 1992, **129**, 1.
43. D. M. Close, E. Sagstuen, and W. H. Nelson, *J. Chem. Phys.*, 1985, **82**, 4386.
44. W. H. Nelson, E. Sagstuen, E. O. Hole, and D. M. Close, *Radiat. Res.*, 1992, **131**, 272.
45. W. H. Nelson, E. Sagstuen, E. O. Hole, and D. M. Close, *Radiat. Res.*, 1998, **149**, 75.
46. D. M. Close, W. H. Nelson, E. Sagstuen, and E. O. Hole, *Radiat. Res.*, 1994, **137**, 300.
47. L. A. Eriksson and A. Laaksonen, *J. Chem. Phys.*, 1996, **105**, 8195.
48. Several of the problems with the agreements between published theoretical results and experimental results have recently been discussed in the literature. See for example, Comment on "Effects of Ionizing Radiation on

- Crystalline Cytosine Monohydrate”, by Close *et al.*, *J. Phys. Chem. B*, 1999, **103**, 3049, immediately followed with a Reply by Wetmore *et al.*, *ibid.* 3051, and Comment on “A Fourier transform EPR study of uracil and thymine radical anions in aqueous solution”, by Close, *Phys. Chem. Chem. Phys.*, 2002, **4**, 43 immediately followed with a Reply by Naumov and Beckert, *ibid.* 45.
- 49. A. O. Colson, B. Besler, and M. D. Sevilla, *J. Phys. Chem.*, 1993, **97**, 8092.
 - 50. A. O. Colson and M. D. Sevilla, in ‘Computational Molecular Biology, Theoretical Computational Chemistry’, vol. 8., ed. J. Leszczynski, Elsevier Science, 1999, p. 245.
 - 51. S. Subramanian, M. C. R. Symons, and H. W. Wardale, *J. Chem. Soc. (A)*, 1970, 1239.

Chapter 6: Excited States of Nucleic Acid Bases

M. K. Shukla and Jerzy Leszczynski

Computational Center for Molecular Structure and Interactions

Department of Chemistry, Jackson State University

1400 J.R. Lynch Street, Jackson, MS 39217

USA

Abstract

This chapter is devoted to a comprehensive analysis of theoretical and experimental investigations dealing with photophysical properties of nucleic acid bases (NABs) and base pairs under UV-irradiation. Experimental determination of ground state geometries of NABs were performed long ago. However, geometries of complex molecular systems such as NABs in the electronic excited states cannot be determined by experimental methods; only some limited information about the nature of excited geometries can be obtained. Experimental methods in some cases have indicated nonplanar electronic excited state geometries, which have been validated from recent theoretical investigations. Modes of interaction of water molecules with NABs are generally found to be significantly different in the electronic excited states and especially in the $n\pi^*$ states than in the corresponding ground state. In this chapter, the main emphasis has been given to the electronic singlet excited state properties e.g. transition energies, transition dipole moments, charges, excited state geometries, phototautomerism of NABs and base pairs and their interaction with water molecules.

1. Introduction

Genetic information contained in nucleic acids is encoded in the form of specific sequences of hydrogen bonded purine and pyrimidine bases. It

critically depends on bases existing in the natural (keto-amino) tautomeric forms.¹ Adenine and guanine belong to purine bases, while cytosine and thymine belong to pyrimidine bases (Fig. 1). In RNA, the thymine base is replaced with uracil which has a hydrogen at the C5 site instead of a methyl group that is present in thymine (Fig. 1). In DNA and RNA, adenine is hydrogen-bonded with thymine (uracil in case of RNA), while guanine is hydrogen-bonded with cytosine (Fig. 2). Under certain conditions, bases may convert to minor tautomeric forms through proton transfer which eventually may lead to mutation due to the alteration of hydrogen bonding patterns.¹ Such formation of rare tautomeric forms by proton transfer may be either in the ground state or in higher energy states through photoexcitation. High level theoretical studies performed on some popular proton transfer model species predicted the proton transfer barriers in the lowest singlet $\pi\pi^*$ excited state to be significantly reduced with respect to the corresponding ground state values.^{2,3} The proton transfer process in electronic excited state is quite important since we are exposed to different types of radiation. The continuous increment of the ultraviolet (UV) exposure on earth due to ozone depletion is posing a dangerous challenge to the living world.^{4,5} Ultraviolet radiation also alters DNA due to the formation of pyrimidine dimers between adjacent thymine bases on the same strand. The formation of these dimers is the most common UV-induced DNA damage.^{6,7} In spite of these facts, the principle of survival of the fittest governs evolution. It appears that nature has adopted an efficient mechanism to release extra energy obtained due to the excitation of nucleic acid bases (NABs) under UV-irradiation through ultrafast nonradiative internal conversion relaxation processes which does not permit enough time for the occurrence of different excited state reactions. It has long been believed that nonradiative decay in natural NABs is of the order of picoseconds. Recent state-of-the-art experimental studies have demonstrated that an internal conversion process occurs on a subpicosecond time scale and that the inclusion of bulky groups in the NABs increases the time scale.⁸⁻²⁰

The fluorescence quantum yields for all NABs are very low in aqueous solutions at room temperature, the excitation energy being dissipated through the nonradiative decay channels.²¹⁻²³ On the other hand protonated purines show fluorescence at room temperature²⁴⁻²⁶ as well as after being absorbed on a chromatographic paper.²⁷ To obtain insight into the excited state properties of the neutral bases and nucleotides in polar solvents, different low temperature experiments were performed.^{22,28-31} The first low temperature work on nucleic acids was reported in 1960,²⁸ while the phosphorescence of nucleic acids was first reported for adenine derivatives in 1957.³² The first

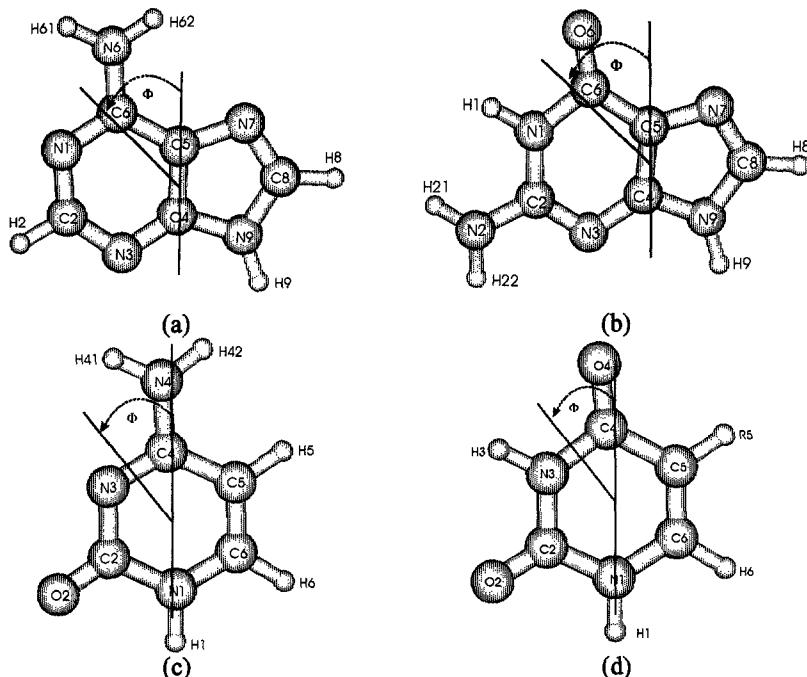


Figure 1. Structure and atomic numbering schemes of nucleic acid bases, (a) adenine (N9H), (b) guanine (keto-N9H), (c) cytosine (keto-N1H), (d) uracil (R5=H) and thymine (R5=CH3). The Φ represents the transition moment direction according to the DeVoe-Tinoco convention [23].

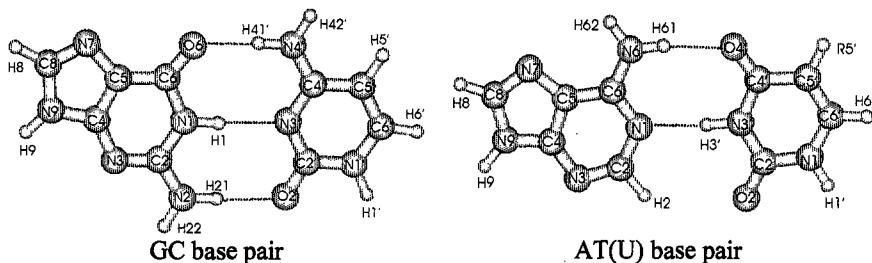


Figure 2. Structure and atomic numbering schemes of the GC, AT (R5'=CH3) and AU (R5'=H) base pairs.

results on isolated monomers were obtained in 1962 by Longworth³³ and in 1964 by Bersohn and Isenberg.²⁹ Initially low temperature measurements were performed using a frozen aqueous solution, but due to the inherent problems in such matrices, most investigations were turned to polar glasses such as ethylene or propylene glycols usually mixed with equal volumes of water.²² Recently, some jet-cooled high level spectroscopic investigations were performed on nucleic acid bases, base pairs and their interactions with water molecules. These studies included resonance enhanced multi photon ionization (REMPI/R2PI) and spectral hole burning (SHB) studies of adenine (A), guanine (G), their alkyl derivatives, other substituted purines, guanine-cytosine (GC), guanine-guanine (GG) base pairs and their nucleosides by Nir et al.,³⁴⁻³⁸ REMPI, laser-induced fluorescence (LIF) and SHB studies of guanine and its hydrated form by Piuzzi et al.,^{39a} REMPI and IR-UV depletion spectroscopic studies of guanine and methylguanine by Mons et al.,^{39b} REMPI studies of adenine and 9-methyladenine by Luhrs et al.,⁴⁰ REMPI and LIF studies of adenine and hydrated clusters by Kim et al.,^{41,42} REMPI studies of uracil and thymine by Brady et al.⁴³ and those by Tsuchiya et al.⁴⁴ and Fujii et al.⁴⁵ on uracil and thymine. These studies have been devoted to extract information on the excited state structures, tautomerism, spectral origin of the 0-0 transitions of nucleic acid bases and base pairs and their interactions with water molecules in excited states.

It is well known that some substituted purines also have large fluorescence quantum yields and are used to monitor the structures and dynamics of nucleic acid polymers.⁴⁶⁻⁴⁹ The classic example in this context is 2-aminopurine (2AP) which differs from the nucleic acid purine base adenine only with respect to the position of an amino group; the amino group is at the C6 position in adenine while it is at the C2 position in 2AP. The quantum yield for fluorescence is about 0.0003 for adenine and 0.5 for 2AP.^{23,50} The lowest energy absorption band for 2AP is significantly red-shifted compared to the corresponding band of adenine, and this property of 2AP has been utilized as an excitation energy trap.^{23,50-52} Santhosh and Mishra⁵² have reported that the fluorescence decay of 2AP in water has two components with lifetimes of 2.1 and 24.6 ns. The fluorescence originates from two tautomeric species, and the phenomenon of fluorescence reabsorption was also reported.⁵²

Obtaining explicit information concerning the structure-function relationships of nucleic acid bases and their complexes has always been an important and difficult task in biomolecular sciences. Various spectroscopic techniques are used to study conformations of polynucleotides in different environments.^{53,54} Absorption spectroscopy is one of the oldest and most

common techniques used to elucidate the structure in chemical science. Since an absorption peak arises due to a vertical transition the knowledge about energy differences between the ground state and excited states lying vertically above the same enables us explain absorption spectra. Explanation of fluorescence and phosphorescence requires knowledge concerning relaxed singlet and triplet excited states respectively. Ground state geometries of nucleic acid bases have long been studied experimentally.⁵⁵ An experimental determination of the excited state geometries of a molecule requires a rotational analysis of their electronic absorption or fluorescence excitation spectra at high resolution followed by an interpretation of the observed changes in the rotational constants in terms of changes in the molecular geometry consequent to excitation.⁵⁶ This approach would lead to unambiguous results for molecules with highly symmetrical excited state geometries while for asymmetric polyatomic molecules, it involves many difficulties. Therefore, it is not surprising that quantitative information concerning excited state geometries of nucleic acid bases is not available. However, some spectroscopic studies such as supersonic jet-cooled⁴³ and resonance Raman⁵⁷ studies have indicated nonplanar excited state geometries that are supported by theoretical studies.⁵⁸

With the advent of state-of-the-art hardware and advanced algorithms, quantum chemical methods are now routinely used to study ground state properties of nucleic acid bases and related molecules at a high level of accuracy.⁵⁹⁻¹¹¹ However, such a level of affordability is still far away for excited state calculations. Certain ab initio calculations of electronic spectra, transition moments and excited state geometries of nucleic acid bases and related molecules are reported.^{58,112-137} However, excited state studies are far less in numbers than those dealing with the ground state properties of nucleic acid bases.

This article is devoted to a comprehensive review of theoretical and experimental results on excited state properties (electronic transitions, excited state geometries, charges, interaction etc.) of the natural nucleic acid bases and base pairs. It is well known that there are three important components of DNA. (i) aromatic N-heterocyclic purine and pyrimidine bases, (ii) sugar, and (iii) phosphate group. The sugar and phosphate groups do not affect the excited state properties of the bases appreciably except affecting prototropic tautomerism involving the protons at the N9 and N1 sites of the purine and pyrimidine bases, respectively (Fig. 1). Such tautomeric processes are blocked in nucleosides and nucleotides.

2. Methodology

To study electronically excited states of molecules the most commonly and widely used single excitation based method is the configuration interaction-singles (CIS) method¹³⁸ which is also known as the Tamm-Dancoff approximation. In the CIS approach, the excited state wave function Ψ_{cis} is expressed as a linear combination of all possible singly excited determinants from some reference configuration, generally taken to be the one corresponding to the HF ground state. Thus

$$\Psi_{\text{cis}} = \sum_i^{\text{Nocc}} \sum_a^{\text{Nvirt}} c_{ia} \psi_{ia} \quad (1)$$

where, Nocc and Nvirt stand for the number of occupied and virtual orbitals in the ground state. Given a reference wave function

$$\Psi_{\text{HF}} = |\varphi_1^2 \varphi_2^2 \cdots \varphi_i^2 \cdots \varphi_{\text{Nocc}}^2\rangle \quad (2)$$

The excited determinants are of the form

$$\psi_{ia} = |\varphi_1^2 \varphi_2^2 \cdots \varphi_i \cdots \varphi_{\text{Nocc}}^2 \varphi_a\rangle \quad (3)$$

The configuration interaction coefficients can be obtained as a normalized eigenvector of the Hamiltonian matrix with elements given by

$$\langle \psi_{ia} | H | \psi_{jb} \rangle = [E_{\text{HF}} + \varepsilon_a - \varepsilon_b] \delta_{ij} \delta_{ab} - (ja||ib) \quad (4)$$

where, $\varepsilon_a, \varepsilon_b$ are orbital energies. The total energy for a CIS excited state, which is an eigenvalue of the matrix shown in equation (4), can be written as

$$E_{\text{CIS}} = E_{\text{HF}} + \sum_i \sum_a c_{ia}^2 (\varepsilon_a - \varepsilon_i) - \sum_{ij} \sum_{ab} c_{ia} c_{jb} (ja||ib) \quad (5)$$

More details of the CIS method can be found in an article by Foresman et al.¹³⁸ The CIS method is the simplest level of approximation of response methods to study excited states. It is often regarded as the Hartree-Fock analogue for excited states.¹³⁸ Other methods, such as the equation-of-motion couple cluster (EOMCC) method¹³⁹⁻¹⁴¹ and the time-dependent density functional theory (TDDFT)¹⁴²⁻¹⁴⁶ based on higher levels of response theories treat electron correlation in better ways through the effective Hamiltonian in the space of single excitation CI from the ground state. Analytical gradients and Hessians are available for the CIS method in the Gaussian program package.¹⁴⁷ Computations of vertical transition energies and oscillator strengths at the TDDFT level are implemented in the Gaussian-98 program package.^{147a} The complete active space self-consistent field (CASSCF) method

is the popular multi-configuration SCF (MCSCF) method used for highly accurate calculations on varieties of molecules.¹⁴⁸⁻¹⁵² In the CASSCF method, one divides orbitals into three parts: inactive, active and secondary orbitals. The inactive orbitals are always doubly occupied, while secondary orbitals are always unoccupied. Active orbitals consist of some occupied and some virtual orbitals. A full CI is taken within the active orbitals known as the active space. A proper selection of appropriate orbitals in the active space, depending upon the nature of the problem, is necessary in the CASSCF method. The CASSCF excitation energies are usually higher than the experimental ones but with the inclusion of dynamic electron correlation, the CASSCF excitation energies agree with experimental ones within an accuracy of about 0.2 eV.¹⁴⁸⁻¹⁵² However, the practical applicability of the CASSCF method for excited state calculations is limited to small sized systems (currently smaller than the pairs of DNA bases), which is especially true for excited state geometry optimizations.

3. Ground State Properties

It is well known that DNA bases can occur in more than one tautomeric form; the equilibrium between them depends on the nature of the environment. Interest in tautomeric equilibria of DNA bases has been inspired to a large extent by the theory of spontaneous point mutations.¹⁵³⁻¹⁵⁶ An occurrence of an NAB in its rare form could result in the formation of a base pair with a noncomplementary partner (e.g., guanine might form hydrogen bonds with thymine, and a rare form of adenine would match cytosine, etc.). Consequently, such a mispairing may cause spontaneous mutation. In this case the frequency of the mutation event is governed by the concentration of the rare tautomeric forms. The experimentally observed mutation rates¹ are of the order 10^{-8} — 10^{-11} per base pair synthesized which covers the mutationally significant concentrations of minor tautomers to 10^{-4} — 10^{-5} . Since such concentrations fall below the detection limits of available experimental techniques, accurate quantum mechanical predictions of tautomeric equilibria are needed. A significant amount of experimental and high level quantum chemical theoretical studies are devoted to the investigation of ground state properties of nucleic acid bases and related molecules.^{55,59-110,157-174} Discussion concerning ground state geometries, tautomerism, transition states corresponding to proton transfer from the canonical form to the rare tautomeric form, base pair formation, stacking interactions, interactions with

metal ions and hydration have been the main issue of some very recent review articles.⁵⁹⁻⁶⁶ Therefore, we will discuss here the ground state tautomeric and other important properties of natural nucleic acid bases and base pairs only very briefly.

It has long been believed that adenine exists mainly in two tautomeric forms, N9H and N7H, and this belief was supported by observations of the existence of these tautomers in polar solvents.¹⁷⁵⁻¹⁷⁸ Experimental studies performed in low temperature matrices also support the existence of these two tautomers, but the equilibrium was found to be shifted towards the N9H tautomeric form.¹⁷⁹⁻¹⁸⁴ This conclusion is also supported from different ab initio calculations.^{77,184} Guanine exhibits both keto-enol and prototropic tautomerisms.^{59-66,78,90,91,157-161} The ratio of the keto-enol forms has been found to be sensitive to the molecular environment.¹⁵⁷⁻¹⁶¹ The two forms of guanine may occur in nearly equal abundance in an inert argon matrix, but the keto form is dominantly present in polar media.^{90,91,157-161} A recent jet-cooled spectroscopic study has suggested the existence of four tautomers (keto-N9H, keto-N7H, enol-N9H, and enol-N7H) of guanine in the gas phase.^{39b} High level ab initio calculations predict numerous tautomers with close energies; the relative stability was also found to be basis set dependent.^{60,78,91} At the MP2/6-31G(d) level, the keto-N7H form of guanine is slightly more stable than the keto-N9H form and is followed by the enol-N9H form.^{60,78,91} However, at the MP2/6-311++G(df,pd)/MP2/6-31G(d,p) level, the enol-N9H form of guanine is found to be the most stable tautomer in the gas phase.⁷⁸ Under aqueous solvation the tautomeric equilibrium is shifted towards the keto-N9H form.^{90,91} Consequently, the keto-N9H tautomer is found to be most stable. These calculations are in good correspondence with the experimental studies of isolated tautomers in argon and nitrogen matrices.¹⁵⁷⁻¹⁶¹

It is generally believed that among the pyrimidine bases, uracil and thymine mainly exist in the keto form. However, the existence of a small amount of the enol form of these compounds has also been indicated in some experiments. Hauswirth and Daniels¹⁸⁵ have explained the observed deviation in the excitation spectrum of thymine from the corresponding absorption spectrum in terms of the possibility of emission from the enol tautomer. On the other hand, Vigny and Duquesne¹⁸⁶ observed that both absorption and fluorescence excitation spectra of thymine are in resemblance with each other. Suwaiyan et al.¹⁸⁷ have suggested the existence of a small amount of the enol tautomer in aqueous solutions of 5-chlorouracil at room temperature. The existence of keto-enol tautomerism in uracil, thymine and their derivatives has also been suggested in supersonic jet-cooled spectroscopic studies.^{44,45}

However, Brady et al.⁴³ have shown that the observed sharp features in the excitation spectra of Fujii et al.^{44,45} are due to the formation of some impurity produced in the oven. Recently, Morsy et al.¹⁸⁸ have suggested the existence of a trace amount of the enol form of thymine in aqueous solutions. However, electron diffraction,¹⁷⁴ X-ray crystallography,⁵⁵ NMR,¹⁸⁹ UV,¹⁹⁰ and IR¹⁹¹⁻¹⁹⁴ data and theoretical calculations^{60,79} strongly suggest the existence of only the keto form.

The pyrimidine base cytosine also exhibits several tautomeric structures under different environmental conditions.^{59-66,195-200} At low temperature, in argon and nitrogen matrices, it exists as a mixture of the amino-hydroxy and amino-oxo (N1H) tautomeric forms in comparable amounts, the equilibrium being shifted slightly towards the former tautomeric form.^{195,196} The existence of the imino-oxo tautomeric form has been also indicated in the matrix isolation study of 1-methylcytosine.¹⁹⁷ The microwave study has yielded rotational constants for three tautomers, namely amino-oxo, imino-oxo, and amino-hydroxy tautomers of cytosine.¹⁹⁸ Furthermore, while in aqueous solutions¹⁹⁹ both of the amino-oxo forms (N1H and N3H) are present; under the crystalline environments cytosine exists mainly in the N1H amino-oxo form.²⁰⁰ High level quantum chemical theoretical calculations yielded small energy differences between cytosine tautomers, and such energy differences are found to be extremely sensitive to the levels of approach, giving rise to somewhat contradictory results in some cases.^{94,103,201-203} For example, in order to predict the correct relative stability of different tautomers at the coupled-cluster level, an approach with single, double and triple excitations (CCSD(T)) is necessary.²⁰¹ Generally the MP2 and CCSD(T) levels reveal similar trends,^{94,201} but a definite conclusion cannot be reached even at different levels of MPn treatment.⁹⁴ Although the DFT method provides excellent agreement with the experimental rotational constants and vibrational frequencies, its results differ with regard to the computed relative stability of different tautomers from those obtained at the CCSD(T) and MPn levels of theory.^{103,201-203}

The reference geometry has a significant influence on the predicted properties of the molecules. Until the early 1990s, isolated NABs were believed to have perfectly planar geometries including those of the amino group. Recent ab initio calculations with reasonable basis set (at least the 6-31G(d) basis set; polarization functions on the amino nitrogen are necessary) suggest that the amino group of guanine, adenine and cytosine is pyramidal (Table 1).^{58-66,119,120} Such pyramidal character is largest for guanine. Recently, additional support for nonplanar geometries of NABs has been provided.²⁰⁴⁻²⁰⁸ Ab initio calculations performed at the HF, MP2 and DFT(B3LYP) levels

demonstrate that the pyrimidine ring in the NABs possesses high conformational flexibility.²⁰⁴⁻²⁰⁷

Table 1. Dihedral angles ($^{\circ}$) showing nonplanarity of amino groups of isolated DNA bases.⁶⁰

Nucleobase/ Method	XCNHn2 ^a	XCNHn1 ^b	Σ HNY ^c	ΔE^d (kcal/mol)
Cytosine				
MP2/6-31G(d)	-26.2	14.1	348.8	-0.38
MP2/6-311G(2df,p)	-21.4	12.6	351.9	-0.15
Adenine				
MP2/6-31G (d)	-21.1	18.7	349.3	-0.34
MP2/6-311G(2df,p)	-15.3	16.5	352.9	-0.13
Guanine				
MP2/6-31G(d)	-11.8	43.2	338.1	-1.63
MP2/6-311G(2df,p)	-13.3	39.2	339.6	-1.12

^aX=C5 of cytosine and adenine, and N3 of guanine; ^bX=N3 of cytosine, N1 of adenine and guanine; n in XCNHn corresponds to 6 for adenine, 2 for guanine and 4 for cytosine (Fig. 1);

^cSum of amino angles, this sum equals 360° for a planar molecule and is less than 360° if the amino group is pyramidal, see Fig.1 for details; ^dThe inversion barrier for pyramidalization.

Although the amino group is pyramidal for isolated bases, the geometries of Watson-Crick (WC) base pairs (GC, AT and AU) are planar including the amino group.^{60,66,113,114} The structures of different reverse Watson-Crick (RWC), Hoogsteen (H) and reverse Hoogsteen (RH) base pairs have also been investigated, and the geometries of some of them have been found to be nonplanar.^{60,66,209} It is well known that the Watson-Crick GC base pair is characterized by three hydrogen bonds, while the Watson-Crick AT and AU base pairs are characterized by two hydrogen bonds. In the case of the AU base pair, the existence of the attractive interaction between the C2H site of adenine and the C2O2 group of uracil has been speculated by Leonard et al.²¹⁰ and supported by Starikov and Steiner.²¹¹ This attractive interaction is suggested to contribute to the stability of the AU base pair.²¹⁰ However, theoretical calculations at the MP2 level do not support the existence of this type of so-called hydrogen bond in the AU base pair.²¹²

4. Excited State Properties

In general, excited state charge distributions are appreciably different from those in the ground state. Consequently, excited state properties are appreciably different from the corresponding ground state properties.^{21-23,58,112-137}

Generally the ground state geometries of NABs are planar except the amino group (in adenine, guanine, and cytosine).⁵⁹⁻⁶⁶ However, electronically excited state geometries are generally nonplanar.^{43,57,58,115,119-123,125} In some cases e.g. the N7H tautomer of adenine in the $n\pi^*$ excited state, the structure of the molecule resembles that in an intramolecular charge transfer state with a small charge transfer.^{120,122} The excited state structural deformation for NABs has been suggested as a possible nonradiative decay channel.^{119,122} The modes of interaction of NABs with water molecules are also found to be different in the excited states and especially in the $n\pi^*$ excited states as compared to the ground state. The hydrogen bond accepting sites under the $n\pi^*$ excitations, the excitation being localized on them provide repulsive potential, consequently base pairs are destabilized under such excitations.^{113,114}

It is well known that CIS-computed transition energies are much larger than the experimental ones due to the lack of electron correlation, and some scaling factors are needed to make the computed excitation energies comparable to the experimental ones. For the nucleic acid bases, the scale factor of 0.72 is found suitable.^{113,122,124} Our theoretical results regarding electronic transitions and excited state geometries presented in this chapter were computed at the CIS level of the theory¹³⁸ using the Gaussian 94 suite of programs.^{147b} Further, the effect of aqueous solvation was studied using the supermolecular approach in which three water molecules were considered in hydration of NABs.

4.1 Adenine

4.1.1 Electronic Transitions

4.1.1.1 Experimental Data

It is now well established that the main absorption band (in the 260 nm region) of adenine consists of two $\pi\pi^*$ transitions, one being strong while the second is considerably weaker; the relative positions of the two transitions are dependent on the environment.²³ Thus while the absorption spectra of adenine

in the vapor phase exhibit two peaks near 252 and 207 nm (4.92 and 5.99 eV), in water and in TMP (trimethyl phosphate) the main absorption peak (252 nm) is found to be red-shifted to 260 nm (4.77 eV).^{213,214} It was Mason in 1954²¹⁵ who first suggested that the main absorption band of adenine actually consists of two electronic transitions, the weaker one is short axis polarized while the more intense one is long axis polarized. In water solution, these bands of adenine are resolved, but the stronger band which appears at 261 nm (4.75 eV) is short axis polarized, and a weak band appearing as a shoulder near 267 nm (4.64 eV) is long axis polarized.²¹⁶ Similar results were also found for 9-methyladenine, partially oriented in stretched polymer poly(vinyl alcohol) films.¹²⁴ Stewart and Davidson²¹⁶ have recorded polarized absorption spectra of single crystals of 1-methylthymine, 9-methyladenine and their 1:1 hydrogen bonded complexes (Hoogsteen type adenine-thymine base pair) and have shown that the main absorption band of 9-methyladenine is composed of two transitions. The intense transition near 275 nm (4.51 eV) is short axis polarized, while the weaker one near 255 nm (4.86 eV) is long axis polarized.²¹⁶ Thus, the splitting between these two transitions is appreciably increased in a crystal environment as compared to solution.²¹⁶ Similar results were found in the photoacoustic spectra of the evaporated film of adenine.²¹⁷ In this case, four absorption peaks were found in the 300—180 nm region. The components of the 260 nm (4.77 eV) band were found near 270 nm (4.59 eV) (strong) and 290 nm (4.28 eV) (weak) suggesting that in this case the transitions are significantly red-shifted and separated by about 20 nm.²¹⁷ Although the splitting of the 260 nm band is observed generally in all experiments, linear dichroism (LD),^{218,219} magnetic circular dichroism (MCD),^{220,221} single crystal absorption,^{216,222} fluorescence polarization,²²³ substituent effects,²²⁴ however, this splitting is not observed in the CD spectra.²²⁵⁻²²⁷

The results for the transition moment directions are not mutually consistent in different experimental observations. The polarized absorption spectra of 9-methyladenine single crystals have shown that the strong component of the 260 nm band appearing at 275 nm makes an angle of -3° with the C4C5 bond (DeVoe—Tinoco convention, Fig. 1a).^{216,222} Matsuoka and Norden²¹⁹ from the film dichroism study have suggested that this angle is 9° (transition being at 263 nm). Chen and Clark²²⁸ have shown that in crystals of adenine hydrochloride (where adenine is in the protonated form) the strong transition near 257 nm (4.82 eV) makes an angle of 100° , while the weak transition near 273 nm (4.54 eV) is -28° with the C4C5 direction. They have also assigned transition moment directions of other bands.²²⁸ Recently, Clark^{229,230} has

performed extensive and comprehensive investigations on the polarized spectra of crystals of 9-methyladenine and 6-(methylamino)purine to model electronic spectra of adenine in the extended range 350—130 nm. He has assigned eight bands along with their transition moment directions and oscillator strengths. The strong transition (265 nm) of the main UV absorption band was shown to be polarized at 25° with respect to the C4C5 bond, while the weaker transition (near 275 nm) was found to be polarized close to the long molecular axis. Recently Holmen et al.¹²⁴ have measured the transition moment directions of several transitions of 9-methyl and 7-methyl adenine (9MA and 7MA) samples oriented in stretched polymer films. The measured transition moment directions for the first two transitions are generally in agreement with those suggested by Clark.^{229,230} However, the transition moment directions for higher energy transitions are different from those obtained by Clark.²³⁰ Furthermore, these authors¹²⁴ have predicted a new $\pi\pi^*$ transition near 5.38 eV for 9MA, which was not previously observed. The existence of a transition near 230 nm (5.39 eV) has been predicted by earlier semiempirical calculations²³ and it was also indicated in the MCD²²⁰ and CD^{225,226} spectra. It has been assigned as one of $n\pi^*$ type on the basis of the semiempirical calculations.²³ The experimental electronic transitions of adenine and its derivatives are summarized in Table 2.

Clark²³¹ has tentatively assigned the existence of $n\pi^*$ transitions near 244 and 204 nm (5.08 and 6.08 eV, respectively) in the 2'-deoxyadenosine. The possibility of the existence of such $n\pi^*$ transitions is also supported from a recent theoretical study.¹¹³ The existence of an $n\pi^*$ transition near the first $\pi\pi^*$ absorption transition has also been assigned in the stretched polymer film work on adenine derivatives¹²⁴ and in the molecular beam of hydrated adenine clusters.⁴² The existence of the $n\pi^*$ transition as the first transition (the energy is very close to the first $\pi\pi^*$ transition) in adenine in the gas phase is predicted by Mennucci et al.¹²³ using the time dependent density functional theory (TDDFT) and the multi-reference perturbation configuration interaction method (known as CIPSI). However, this prediction is not supported by CIS and CASSCF/CASPT2 calculations.^{113,120,126,136} Kim et al.⁴¹ have recently performed REMPI and fluorescence study of jet-cooled adenine, and suggested that the first transition of adenine has a $n\pi^*$ character with the 0-0 band located at 35503 cm^{-1} (~281.7 nm, ~4.40 eV), while the corresponding band of the first $\pi\pi^*$ transition is located at 36108 cm^{-1} (~276.9 nm, ~4.48 eV). Luhrs et al.⁴⁰ have performed a similar study of adenine and 9MA but their results do not support the assignment of the $n\pi^*$ transition suggested by

Table 2. Summary of experimental transition energies (ΔE , eV) of adenine and its derivatives. The f represents oscillator strength, and Φ represents transition moment direction ($^\circ$) according to the DeVoe-Tinoco convention (Figure 1a).

Transitions						References		
Absorption Spectra								
ΔE	4.92		5.99			Adenine, vapor [213]		
ΔE	4.77		5.96			Adenine, TMP [213]		
ΔE	4.81		5.85			9MA, MCH [213]		
ΔE	4.77		5.90			9MA, TMP [213]		
ΔE	4.77		5.99			Adenine, water [214]		
ΔE	4.59	4.77		5.90		Adenine, water [232]		
ΔE	4.63	4.77		6.05		Adenine, water [220]		
ΔE		4.77		6.02		Adenosine, water [220]		
ΔE	4.59		5.90		6.81	7.75	Adenine sublimed film [276]	
ΔE	4.51	4.68		5.82	6.08	6.81	7.75	9MA, crystal [230]
f	0.1	0.2		0.25	0.11	0.30	0.23	
Φ	83	25		-45	15	72	6	
LD Spectra								
ΔE	4.55	4.81	5.38	5.80	5.99		9MA, stretched film [124]	
f	0.047	0.24	0.027	0.14	0.12			
Φ	66	19	-15	-21	-64			
CD Spectra								
ΔE		4.63		5.93	6.36		Adenines, water [225]	
ΔE		4.77		5.74	6.36	6.63	Adenines, water [226]	
ΔE		4.68		5.51			Adenosine, water [220]	
MCD Spectra								
ΔE	4.59	4.92		5.90			Adenine, water [220]	
ΔE	4.56	4.90		5.77			Adenosine, water [220]	
Photo Acoustic Spectra								
ΔE	4.28	4.59			6.20	6.89	Adenine, film [217]	
Electron Scattering								
ΔE	4.53		5.84		6.50	7.71	Adenine, film [226]	

TMP: trimethylphosphate, 9MA: 9-methyladenine, MCH: methylcyclohexane, adenines: adenine derivatives; for details see relevant references.

Kim et al.⁴¹ Luhrs et al.⁴⁰ have speculated that the $n\pi^*$ peak observed by Kim et al.⁴¹ may be due to the formation of other tautomers of adenine since the latter study involved the use of higher temperatures in heating the sample.

Luhrs et al.⁴⁰ have observed the 0-0 band of the first $\pi\pi^*$ transition of adenine and 9MA at 36105 cm^{-1} ($\sim 277\text{ nm}$, $\sim 4.48\text{ eV}$) and 36136 cm^{-1} ($\sim 276.7\text{ nm}$, $\sim 4.48\text{ eV}$), respectively, and these results are in accordance with the observation made by Kim et al.⁴¹ Similar results were also found from the REMPI study by Nir et al.³⁵ who used the laser desorption technique instead of heating the samples.

Mishra and coworkers²³² have reported spectral oscillation in the absorption spectra of adenine solution in water when air was passed through the solution and exposed to UV-radiation intermittently over a long period of time. The main absorption peak near 260 nm (4.77 eV) was shifted to near 300 nm (4.13 eV) region which was explained in terms of the formation of the N3H tautomer of adenine,^{120,233,234} the process of which is closely related to the tautomerization of 7-azaindole.^{171,235,236} This assertion (the formation of the N3H tautomer in the excited state) is also supported from a recent theoretical study of adenine tautomers at the MCQDPT2/CASSCF levels of theory.¹²¹

4.1.1.2 Theoretical Data

There are several semiempirical calculations dealing with the excited state properties of adenine.^{23,237-240} The first ab initio calculations of the electronic transitions of adenine were carried out by Petke et al.¹²⁷ They applied the multi-reference configuration interaction (MRCI) and random phase approximation (RPA) methods using the ground state self-consistent field orbitals with double- ζ /polarization/diffuse gaussian basis set.¹²⁷ The molecular geometry was assumed planar and taken from the experimental data. They found that the computed transition energies were higher by $1.48-1.86\text{ eV}$ compared to the experimental transition energies, and linear scaling was needed for comparison with experimental data.¹²⁷ The most accurate calculations concerning the electronic transitions of adenine were performed by Roos and coworkers¹²⁶ applying the complete active space self-consistent field (CASSCF) method and the complete active space multiconfigurational second-order perturbation theory (CASPT2) using a large ANO-type basis set. However, they have considered the planar form of the molecule. Although the CASSCF method gave appreciably higher excitation energies, the application of CASPT2 resulted in a tremendous improvement in the excitation energies of adenine tautomers.¹²⁶ The experimental excitation energies were explained mainly in terms of the N9H tautomer; the possibility of some contribution from the N7H tautomer was also suggested.¹²⁶

Table 3. Vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies (ΔE , eV), oscillator strengths (f), transition moment directions (Φ ,°) and dipole moments (μ , Debye) of the N9H and N7H tautomers of adenine in the isolated and hydrated forms at the CIS/6-311G(d,p)//HF(6-311G(d,p)) level.

CIS						Experimental Data ^a					
Isolated			Hydrated			CASPT2/CASSCF ^b			Abs Crystal LD		
ΔE	f	Φ	μ^c	ΔE^d	Φ	ΔE	Φ	$\Delta E^e/\Delta E^f/\Phi/\mu$	ΔE	$\Delta E/f/\Phi$	$\Delta E/\Phi$
N9H											
<i>$\pi\pi^*$ Transitions</i>											
6.61	0.394	60	2.85	4.76	6.61	0.440	50	4.76	5.20/6.48/0.37/37/2.30	4.77	4.68/0.2/25
6.65	0.024	-6	3.40	4.79	6.59	0.038	-66	4.74	5.13/5.73/0.07/23/2.37	4.59	4.51/0.1/83
8.20	0.398	-38	0.83	5.90	8.09	0.342	-31	5.82	6.24/7.80/0.851/-57/2.13	5.90	5.82/0.25/-45
8.58	0.447	15	2.02	6.18	8.50	0.423	19	6.12			5.80/0.14/-21
8.67	0.547	-87	3.14	6.24	8.57	0.589	-77	6.17			
9.39	0.232	29	2.65	6.76	9.41	0.375	23	6.78	6.99/8.77/0.565/27/3.42		6.81/0.30/72
<i>$n\pi^*$ Transitions</i>											
7.19	0.001	-	2.47	5.18	7.47	0.000	-	5.38	6.15/6.43/0.001/-2.14		

N7H		$n\pi^*$ Transitions		$n\pi^*$ Transitions	
7.66	0.002	-	0.93	5.52	7.99
7.97	0.014	-	1.62	5.74	8.29
				0.001	0.001
				-	-
				5.75	6.86/7.16/0.001/-1.93
				5.97	
$n\pi^*$ Transitions		$n\pi^*$ Transitions		$n\pi^*$ Transitions	
6.38	0.162	35	6.83	4.59	6.36
6.78	0.051	3	6.51	4.88	6.84
8.06	0.766	81	6.87	5.80	7.99
8.27 ^a	0.163	-44	1.75	5.95	8.44
8.47	0.377	-12	6.00	6.10	8.55
6.87	0.012	-	4.73	4.95	7.24
7.33	0.002	-	6.32	5.28	7.61
7.78	0.014	-	4.24	5.60	8.11
				0.004	0.004
				-	-
				5.48	5.21
				-	-
				5.84	5.48

^aAbs: absorption in aqueous medium [232]. Crystal: based on the polarized spectra of single crystals of 6-(methylamino)purine and 9-methyladenine [230]. LD: LD spectra of 9-methyladenine and 7-methyladenine oriented in stretch poly(vinyl alcohol) film [124]. ^b ΔE^1 corresponds to CASPT2 and ΔE^2 corresponds to CASSCF transition energies [126]. ^cground state dipole moments of the N9H and N7H tautomers at the HF6-311G(d,p) level are 2.51 and 6.83 Debye, respectively, ^dscaled (scaling factor 0.72) excitation energies, ^eAverage of transitions at 6.18 and 6.24 eV, ^fRydberg contamination.

Our recent study revealed excited state properties of adenine tautomers and their interaction with water molecules.^{120,120a} Our computed vertical singlet $\pi\pi^*$ and $n\pi^*$ transition energies, transition moment directions and dipole moments of the N9H and N7H tautomers of adenine and their hydrated forms with three water molecules each obtained at the CIS/6-311G(d,p) level are presented in Table 3. The super molecular approach considering three water molecules in hydration of adenine tautomers was used to model aqueous solvation. The ground state geometries optimized at the HF/6-311G(d,p) level and found to be a minima at the respective potential energy surfaces, were used in the transition energy calculations. The N9H tautomer was found to be about 0.41 eV more stable than the N7H tautomer, and the relative stability of the former over the latter was increased to 0.47 eV after hydration with three water molecules. The CIS computed transitions were assigned as those of $\pi\pi^*$, $n\pi^*$ and $\pi\sigma^*/\text{Rydberg}$ ^{3,241,242} types. Because of uncertainty in assignment and possibly significant basis set dependence, we are not giving $\pi\sigma^*$ Rydberg type transitions in Table 3. Considering excitation energies in the increasing order, the third transition of each of the N9H and N7H tautomers are found to be of the $n\pi^*$ type and localized at the ring nitrogens (Table 3). The CASSCF and CASPT2 calculations also suggest that the third singlet transition is of the $n\pi^*$ type.¹²⁶ In the N9H tautomer of adenine, the first $\pi\pi^*$ transition is stronger, while the second $\pi\pi^*$ transition is much weaker. After hydration the transition energy of the weaker transition is decreased (Table 3). It should be noted that in an aqueous solution, a weak shoulder near 270 nm (4.59 eV) and a strong peak near 260 nm (4.77 eV) are observed.²³² Thus the calculated transitions of the hydrated N9H tautomer are in a qualitative agreement with the experimental data,²³² although the computed splitting between the two transitions is too small (Table 3). Other experimental transitions shown in Table 3 can be explained within an accuracy of 0.2 eV in terms of the scaled computed transitions of the hydrated N9H tautomer. Table 3 suggests that two transitions computed at 6.18 and 6.24 eV for the isolated N9H tautomer and at 6.12 and 6.17 eV for its hydrated form would contribute to the 6.2 eV experimental region (Table 2). It should also be noted that the computed transition moment directions of these transitions are approximately perpendicular to each other (Table 3). The MCD results reveal that the UV-absorption band in the 200 nm (6.2 eV) region is composed of two transitions, and their transition moment directions are not parallel to each other.²²¹ Therefore, our computed results may correspond to the MCD observation in this regard. Although the predicted weak transition near 5.38 eV in the LD

spectra of 9MA¹²⁴ is not obtained in the calculations, it was found in the planar form of adenine.¹¹³ The CIS calculation predicts three nπ* transitions near 5.18, 5.52 and 5.74 eV (scaled values) for the N9H tautomer of adenine, the corresponding values for the hydrated form are 5.38, 5.75 and 5.97 eV, respectively (Table 3). The computed first nπ* transition may be related to that indicated in the MCD²²⁰ and CD^{225,226} spectra in the 230 nm (5.39 eV) region as discussed earlier. It can also be suggested as the possible source of the first nπ* transition indicated near 244 nm (5.08 eV) in the 2'-deoxyadenosine.²³¹ Although, it is not possible to relate the second computed nπ* transition, however, the third computed transition near 5.97 eV (hydrated form) can be related to the 204 nm (6.08 eV) transition of 2'-deoxyadenosine.²³¹ The agreement between the CIS computed ππ* transitions of the N7H tautomer and those obtained by CASPT2 calculations¹²⁶ and LD technique¹²⁴ is exceptionally good for the first two transitions; however, such an agreement is not shown for higher energy transitions (Table 3). Due to close proximity of the computed transitions of the N7H tautomer and its hydrated form to those of the N9H tautomer and its hydrated form, contributions to the observed spectra of adenine from the N7H form cannot be ruled out. It is known from different experimental and theoretical studies that the N7H tautomer is present along with the N9H form under different environmental conditions.¹⁷⁵⁻¹⁸⁴

4.1.2 Geometries and Charges

Table 4 shows the geometrical parameters of the N9H and N7H tautomers and their hydrated forms in the ground and lowest singlet ππ* and nπ* excited states. All geometries were found to be minima at their respective potential energy surfaces. Here, it should be noted that the excited state assignments and numbering (in the ascending energy order) correspond to isolated molecules. The ground state geometries of the N9H and N7H tautomers are almost planar, except for the amino group which is pyramidal. The pyramidal character of the amino group in the N7H tautomer is appreciably more than that in the N9H form (Table 4).

In the S₁(ππ*) excited state of the N9H tautomer, the ring geometry is almost planar and the amino group is slightly more nonplanar than that in the ground state. In the S₃(nπ*) excited state, the amino group pyramidalization is similar to that in the ground state, but the ring geometry is nonplanar especially around the N1C2N3 portion of the ring with the C2H group being appreciably out-of-plane (Fig. 3a). In the S₁(ππ*) excited state of the N7H tautomer, the ring geometry is nonplanar including the amino group.

Table 4. Ground and singlet excited state optimized bond lengths (\AA), bond angles ($^{\circ}$) and dihedral angles ($^{\circ}$) of the N9H and N7H tautomers of adenine and their hydrated forms.^a

Parameters	N9H			N7H		
	S_0	$S_1(\pi\pi^*)$	$S_2(\pi\pi^*)$	S_0	$S_1(\pi\pi^*)$	$S_2(\pi\pi^*)$
N1C2	1.327	1.324	1.320	1.329	1.377	1.335
C2N3	1.312	1.310	1.402	1.401	1.399	1.303
N3C4	1.327	1.334	1.278	1.281	1.320	1.331
C4C5	1.375	1.373	1.428	1.426	1.377	1.382
C6N1	1.325	1.339	1.310	1.313	1.306	1.313
C5C6	1.400	1.409	1.459	1.470	1.427	1.392
C5N7	1.383	1.385	1.318	1.326	1.374	1.384
N7C8	1.278	1.282	1.335	1.330	1.276	1.353
C4N9	1.362	1.358	1.383	1.384	1.356	1.373
C8N9	1.371	1.362	1.366	1.356	1.384	1.286
N6C6	1.344	1.327	1.348	1.328	1.340	1.372
H61N6	0.992	0.996	0.995	0.999	0.992	0.996
H62N6	0.992	0.998	0.996	1.000	0.992	0.995
H2C2	1.076	1.071	1.070	1.070	1.075	1.075
H8C8	1.072	1.071	1.069	1.069	1.070	1.072
H7/9N7/9	0.993	0.999	0.992	0.999	0.993	0.991
N1C2N3	128.7	128.6	126.9	126.2	111.6	128.3
					+3W	+3W

H2C2N1C6	179.9	179.7	-180.1	179.5	177.1	179.9	179.9	180.0	-175.7	179.7	180.0
A4N3C2N1	0.1	0.2	0.4	0.9	-39.2	-1.1	0.2	-5.0	0.2	-18.6	180.0
C5C4N3C2	-0.1	0.1	-0.3	0.0	21.8	0.2	-0.1	5.0	0.1	-60.9	-60.9
N9C4N3C2	179.7	179.9	179.5	179.8	-162.5	179.2	179.9	175.0	179.7	179.7	180.0
G6C5C4N3	-0.1	-0.3	-0.2	-0.8	0.8	1.0	-0.2	-6.0	-0.3	0.0	0.0
V7C5C4N3	179.8	179.8	179.8	179.7	177.4	179.8	-179.9	171.5	179.8	180.0	180.0
S5C6N1C2	-0.4	0.1	-0.6	-0.2	-17.0	0.7	-0.4	-5.3	0.0	0.0	0.0
N6C6N1C2	178.7	-179.9	177.2	179.9	163.6	178.5	178.4	172.5	179.7	179.7	180.0
N1C6C5C4	0.4	0.2	0.7	0.8	-3.9	-1.4	0.4	6.0	0.2	0.0	0.0
N6C6C5C4	178.7	-179.9	-177.2	-179.2	175.6	179.2	-178.3	-171.6	-179.4	180.0	180.0
S8N7C5C4	0.1	0.0	0.1	0.0	0.0	0.8	-0.1	0.0	-0.2	0.0	0.0
N9C8N7C5	0.0	0.0	-0.1	0.1	-0.7	-1.2	0.1	0.3	0.2	0.0	0.0
H8C8N7C5	180.0	180.0	179.8	180.0	179.7	179.0	-179.9	174.1	179.8	180.0	180.0
S8N9C4N3	179.7	-179.8	-179.8	-179.7	177.4	179.1	180.0	-170.3	179.6	180.0	180.0
H9N9C8N7	180.0	179.8	179.7	-179.7	177.3						
H7N7C5C4						170.6	-178.9	-158.8	-178.6	180.0	
H61N6C6N1	10.1	0.2	11.7	-0.4	-11.2	12.6	10.2	-11.7	2.1	-60.9	
H61N6C6C5	-170.8	-179.7	-170.6	179.7	169.3	-169.7	-171.1	166.0	-178.2	119.1	
H62N6C6N1	170.4	-179.6	163.3	-179.9	174.2	146.7	167.4	-150.2	177.2	60.8	
H62N6C6C5	-10.5	0.5	-18.9	0.2	6.4	-35.6	-13.9	27.5	-3.1	-119.1	

^aGround and excited state parameters are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The S₁, S₃ represent singlet excited states in the ascending energy order (Table 3). The “+3W” represents the corresponding hydrated species with three water molecules.

Interestingly, the hydrogen atom at the N7 position shows an appreciable out-of-plane displacement in this state. An interesting feature is observed in the $S_3(n\pi^*)$ state of the N7H tautomer. The molecule has C_s symmetry in this state and the amino group is rotated such that the hydrogens of this group form a dihedral angle of $\pm 61^\circ$ with respect to the ring plane, and the plane containing the amino nitrogen and hydrogens are perpendicular to the ring plane (Table 4, Fig. 3b). This structure is reminiscent of the twisted intramolecular charge transfer states of different molecules.^{122,245-249} However, in the present case no significant intramolecular charge transfer was found.

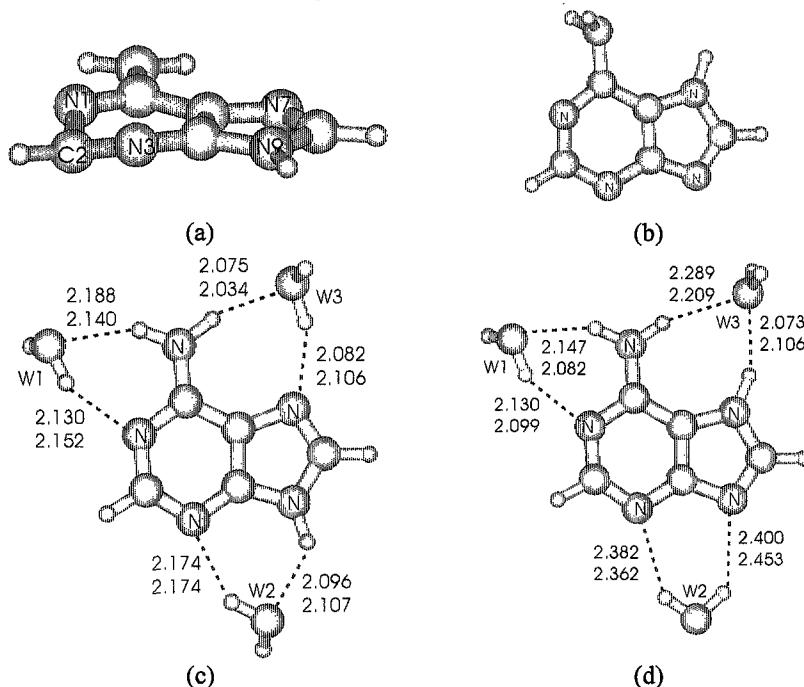


Figure 3. Structure of adenine tautomers: (a) N9H tautomer in the $S_3(n\pi^*)$ state, (b) N7H tautomer in the $S_3(n\pi^*)$ state, (c) hydrated form of the N9H tautomer in the ground state (top indices) and $S_1(\pi\pi^*)$ state (bottom indices), (d) hydrated form of the N7H tautomer in the ground state (top indices) and $S_1(\pi\pi^*)$ state (bottom indices).

The hydrated structures of the N9H and N7H tautomers showing hydrogen bond lengths in the ground and lowest singlet $\pi\pi^*$ excited states are presented in Figs. 3c and 3d, respectively. Hydration of adenine tautomers with three

Table 5. Ground and singlet excited state charges at the atomic sites of the N9H and N7H tautomers of adenine.^a

N9H						N7H					
Mulliken	S ₀	S ₁ (ππ*)	S ₂ (ππ*)	S ₃ (ππ*)	S ₀	S ₁ (ππ*)	S ₂ (ππ*)	S ₃ (ππ*)	S ₀	S ₁ (ππ*)	S ₂ (ππ*)
N1	-0.521	-0.458	-0.447	-0.514	-0.415	-0.452	-0.502	-0.521	-0.524	-0.423	-0.413
C2	0.249	0.148	0.122	0.178	0.093	0.035	0.226	0.230	0.248	0.128	0.178
N3	-0.486	-0.438	-0.456	-0.426	-0.425	-0.428	-0.445	-0.430	-0.441	-0.377	-0.369
C4	0.534	0.515	0.532	0.456	0.588	0.585	0.400	0.345	0.371	0.342	0.377
C5	-0.049	0.037	0.010	0.013	-0.053	-0.083	0.063	0.163	0.142	0.130	0.145
C6	0.551	0.448	0.479	0.495	0.530	0.585	0.547	0.483	0.467	0.461	0.427
N7	-0.448	-0.432	-0.424	-0.449	-0.423	-0.430	-0.533	-0.518	-0.516	-0.494	-0.520
C8	0.259	0.221	0.206	0.278	0.218	0.236	0.283	0.215	0.168	0.241	0.233
N9	-0.503	-0.496	-0.500	-0.487	-0.497	-0.497	-0.419	-0.383	-0.364	-0.381	-0.402
N6	-0.550	-0.516	-0.502	-0.521	-0.564	-0.527	-0.542	-0.517	-0.507	-0.512	-0.552
H61	0.239	0.242	0.238	0.242	0.236	0.242	0.234	0.236	0.242	0.228	0.222
H62	0.248	0.247	0.244	0.249	0.246	0.256	0.208	0.206	0.214	0.176	0.195
H2	0.110	0.106	0.115	0.108	0.101	0.112	0.112	0.118	0.123	0.112	0.115
H8	0.116	0.121	0.128	0.124	0.114	0.116	0.120	0.120	0.125	0.124	0.115
H79	0.251	0.255	0.255	0.254	0.251	0.250	0.248	0.253	0.252	0.245	0.249
CHELPG											
N1	-0.834	-0.751	-0.724	-0.803	-0.350	-0.364	-0.822	-0.854	-0.813	-0.676	-0.429
C2	0.662	0.510	0.438	0.581	-0.093	-0.281	0.645	0.711	0.666	0.461	0.119

N3	-0.812	-0.748	-0.745	-0.752	-0.294	-0.811	-0.827	-0.774	-0.656	-0.273	-0.302	
C4	0.665	0.690	0.695	0.642	0.401	0.420	0.912	0.881	0.786	0.776	0.501	0.518
C5	-0.102	0.126	0.138	0.033	0.079	0.087	-0.326	-0.105	-0.071	-0.138	-0.029	-0.149
C6	0.783	0.611	0.610	0.674	0.518	0.598	0.798	0.663	0.621	0.630	0.438	0.507
N7	-0.576	-0.602	-0.587	-0.630	-0.516	-0.600	-0.500	-0.531	-0.491	-0.498	-0.500	-0.302
C8	0.361	0.354	0.314	0.474	0.235	0.332	0.431	0.348	0.240	0.383	0.302	0.217
N9	-0.551	-0.661	-0.645	-0.665	-0.487	-0.519	-0.697	-0.686	-0.615	-0.649	-0.572	-0.578
N6	-0.859	-0.817	-0.786	-0.831	-0.877	-0.861	-0.839	-0.815	-0.766	-0.806	-0.837	-0.886
H61	0.388	0.398	0.388	0.394	0.393	0.389	0.365	0.383	0.377	0.369	0.365	0.339
H62	0.378	0.363	0.345	0.379	0.390	0.412	0.346	0.328	0.319	0.303	0.325	0.339
H2	0.035	0.037	0.057	0.024	0.124	0.219	0.038	0.022	0.034	0.039	0.108	0.144
H8	0.087	0.085	0.100	0.073	0.103	0.092	0.078	0.083	0.099	0.078	0.086	0.113
H7/9	0.375	0.405	0.402	0.407	0.374	0.370	0.382	0.399	0.388	0.384	0.396	0.313

^aGround and excited state charges are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The S₁, S₃ represent singlet excited states in the ascending energy order (Table 3). The “Vert” represents the vertical and “Opt” represents optimized state.

water molecules induces planarity in the base geometry including the amino group in the ground state and the lowest singlet $\pi\pi^*$ excited state (Table 4). Consequently, the tautomers are almost planar in the ground and lowest singlet $\pi\pi^*$ excited states (Table 4). Furthermore, in the lowest singlet $\pi\pi^*$ excited state of the hydrated tautomers, some of the hydrogen bond lengths are increased, while some are decreased compared to the corresponding ground state values. But the overall trend is such that the adenine tautomers-water molecules interaction would be stronger in the lowest singlet $\pi\pi^*$ excited states of the tautomers compared to the corresponding ground states (Table 4).

The computed Mulliken and electrostatic potential fitted CHELPG charges^{243,244} of the N9H and N7H tautomers in the ground and excited states are shown in Table 5. In going from the ground state to vertical singlet $\pi\pi^*$ excited states of the adenine tautomers, generally a small change in charge distribution takes place. However, in going from the ground state to the vertical lowest singlet $n\pi^*$ excited state, a significant change in charge distribution occurs and this change is quite pronounced in CHELPG charges. Further, such change is mainly localized in the six-membered ring. In both the N9H and N7H tautomers of adenine, the N1 and N3 atoms lose while the C2 atom gains a significant amount of electronic (CHELPG) charge in going from ground state to vertical lowest singlet $n\pi^*$ excited state. The C4 atom in the N7H tautomer also gains electronic (CHELPG) charge in going from the ground state to the singlet vertical lowest $n\pi^*$ excited state, the corresponding gain for the C4 atom in the N9H tautomer being relatively smaller. The charge distributions in the geometrically relaxed excited states are generally similar to those in the corresponding vertical excited states (Table 5).

4.2 Guanine

4.2.1 Electronic Transitions

4.2.1.1 Experimental Data

Different investigations on the electronic spectra of guanine and its derivatives have suggested the existence of five transitions in the UV region.^{23,126} In most of the measurements, the first (I) band lies near 275 nm (4.51 eV) and the second (II) band appears near 250 nm (4.96 eV), the intensity of the second band being larger than that of the first one.^{23,126,220,224,250-254} The third band in the 225 nm (5.51 eV) region is weak (the oscillator strength is in the range of

0.01-0.03) and is not often observed. Evidence for the existence of such a band is found in the CD spectra,^{226,227} crystal spectra of guanine and 9-ethylguanine,²⁵⁰ and in aqueous solutions of protonated guanine.²⁵⁰ CD spectra have suggested that this band may be due to a weak $\pi\pi^*$ or $n\pi^*$ transition.²²⁶ The fourth and fifth transitions are strong and located near 204 nm (6.08 eV) and 188 nm (6.59 eV), respectively.^{23,224,250,251} Various measurements of the transition moment directions of these bands gave different results. The crystal field is found to have a profound effect on the transition moment directions.^{216,255} Callis and coworkers²⁵⁶ have estimated the angle between the I and II bands to be about $61 \pm 10^\circ$, while it was found to be 71° by Clark.²⁵⁰ Now it is almost certain that the long wavelength band (I) is polarized along the short axis (C4C5), while the short wavelength band (II) is polarized along the long axis.^{218,219,250,251} More detailed discussions concerning the transition moment directions can be found in an excellent review by Callis²³ and in a recent article by Clark.²⁵¹

Clark has suggested the existence of three $n\pi^*$ transitions near 238, 196, and 175 nm (5.21, 6.32, and 7.08 eV, respectively) in guanine but is uncertain about their assignment.²⁵¹ Some interesting observations were made by Mishra and coworkers²⁵²⁻²⁵⁴ in photophysical studies of aqueous solutions of guanine. A slow damped oscillation of absorption peak intensities was found to be set up as the result of ultraviolet irradiation of an oxygen-rich aqueous solution of guanine.²⁵² The spectral range covered by the fluorescence peak of guanine was also found to be strongly broadened covering the region from around 332 nm to 450 nm when air or oxygen was passed through the solution.²⁵³ It was found that the effect of spectral oscillation and the broadening of fluorescence can be partly and temporarily reversed by passing nitrogen gas through the aqueous solution that would remove some of the dissolved oxygen from the solution.^{252,253} The interaction of guanine with oxygen following the aforementioned ultraviolet irradiation eventually produced a stable reaction product.²⁵⁴ A summary of the experimental transitions of guanine and its derivatives are presented in the Table 6.

It is well known that while guanine in the ground state in an aqueous environment abundantly exists in the keto-N9H tautomeric form, the major part of fluorescence originates from the keto-N7H tautomer of the molecule.^{23,119,257} The normal fluorescence of guanine in aqueous solution has a peak near 332 nm (3.73 eV) and a weak shoulder near 450 nm (2.76 eV).^{253,258} However, the main fluorescence region gets extended appreciably from around 332 nm to 450 nm after air is passed through the solution.²⁵³

Table 6. Summary of experimental transition energies (ΔE , eV) of guanine and its derivatives. The f represents oscillator strength, and Φ represents transition moment direction ($^\circ$) according to the DeVoe-Tinoco convention (Figure 1b).

Transitions						References
Absorption Spectra						
ΔE	4.46	5.08		6.20	6.57	Guanine, model [251]
f	0.15	0.24		0.40	0.48	
Φ	-12	80		70	-10	
ΔE	4.51	5.04		6.33		Guanine, water [214]
ΔE	4.56	5.04		6.19	6.67	Guanosine, water [251]
f	0.15	0.24		0.40	0.48	
Φ^a	-24	88		86	-8 to 44	
ΔE	4.56	4.98		6.02	6.63	9EtG, water [250]
f	0.14	0.21		0.38	0.42	
ΔE	4.51	4.84		6.11	6.52	9EtG, TMP [224]
ΔE	4.51	4.92		6.05	6.59	9EtG, water [224]
ΔE	4.46	4.88	5.46	6.08	6.56	9EtG, crystal [250]
f	0.16	0.25	<0.05	0.41	0.48	
Φ	-4	-75	-75	-9		
ΔE	4.35	5.00		6.23	6.70	Guanine, sublimed film [276]
LD Spectra						
ΔE	4.43	5.00				Guanine, stretched film [219]
Φ	4	-88				
CD Spectra						
ΔE	4.51	4.92	5.51	6.20	6.63	dGMP, water [226]
ΔE			5.06	5.77		Guanosine, water [220]
MCD Spectra						
ΔE	4.46	5.00				Guanosine, water [220]

^aBased on polarized absorption spectra of crystalline guanosine [251], 9EtG: 9-ethylguanine, TMP: trimethylphosphate; for details see relevant reference.

Recently some advanced spectroscopic studies of laser-desorbed, jet-cooled guanine and its substituted analogs and guanine-guanine and guanine-cytosine base pairs have been performed.^{34,36-39} These investigations included REMPI studies on guanine³⁸ and guanosines,³⁷ REMPI and spectral hole burning (SHB) studies on guanine, methyl guanine,³⁴ guanine-guanine and guanine-cytosine base pairs,³⁶ guanine and hydrated guanine^{39a} and REMPI and IR-UV depletion spectroscopic studies on guanine and methyl guanine.^{39b} In these studies,^{34,36-39} the spectral origin (0-0 transition) of the S_1 excited state

and some lower vibrational frequencies corresponding to the same were determined, and the existence of different tautomers of guanine was investigated.^{34,39} In some cases, the excited state vibrational frequencies were compared with the ground state vibrational frequencies, and it has been concluded that the geometry of guanine in the S₁ excited state is similar to that in the ground state.^{34,36-38} The conclusions of these studies^{36,37} are based on similarity between the observed low frequencies of guanosines and base pairs (GC and GG) with the calculated frequencies, but no comparison with the frequencies of guanine itself was made. These low frequencies correspond to either intermolecular hydrogen bonds (in GC and GG base pairs) or to the vibration of the sugar moiety with respect to the base. Thus, the adopted approach is itself not convincing since it will not give proper information concerning the geometry of guanine unless, at least, a comparison is made between the observed and computed frequencies of guanine. Excited state geometry optimization studies show that the S₁ geometry of guanine tautomers is appreciably nonplanar, and the excited state vibrational frequencies are highly red-shifted compared to ground state vibrational frequencies.^{119,125} Investigations by Nir et al.³⁴ and Piuzzi et al.³⁹ suggest the existence of three tautomers of guanine, namely, keto-N9H, keto-N7H and enol-N9H (Piuzzi et al.³⁹ are not sure if it is enol-N9H or enol-N7H) in gas phase, but they differ with regard to the spectral origins of these tautomers. Recently, Mons et al.^{39b} have suggested the existence of four tautomers of guanine in gas phase, namely, enol-N7H (22864 cm⁻¹), keto-N7H (33269 cm⁻¹), keto-N9H (33910 cm⁻¹) and enol-N9H (34755 cm⁻¹); the values in parentheses correspond to the 0-0 transition of the corresponding tautomer.

4.2.1.2 Theoretical Data

The first ab initio quantum chemical calculations on the excited states of guanine were carried out by Petke et al.¹²⁷ using the random phase approximation (RPA) and multireference configuration interaction (MRCI) methods. In these calculations, the molecular geometry was taken from the experimental data and molecule was assumed to be planar. The computed transition energies were found to be higher by 1.48–1.86 eV as compared to the experimental ones, and linear scaling was used for comparison with experimental data.¹²⁷ The most accurate theoretical calculations on the electronic spectra of guanine were reported by Roos and coworkers¹²⁶ using the CASSCF method with dynamic correlation to CASSCF energies at the CASPT2 level employing the ANO-type basis set. Only the keto-N9H tautomer of guanine was considered and it was assumed to be planar. The

Table 7. Vertical singlet $\pi\pi^*$ and nn^* excitation energies (ΔE, eV), oscillator strengths (f), transition moment directions (Φ), and quadrupole moments (μ , Debye) of the keto-N9H and keto-N7H tautomers of guanine in the isolated and hydrated forms at the CIS/6-311G(d,p)/HF/6-311G(d,p) level.

CIS	Experimental Data ^a											
	Hydrated						CASPT2/CASSCF ^b					
	ΔE	f	Φ	μ^c	ΔE^d	ΔE^e	f	Φ	ΔE^d	$\Delta E^f/\Delta E^g/\mu$	$\Delta E^h/\Phi$	$\Delta E^i/\Phi$
keto-N9H												
<i>nr*</i> * Transitions												
6.39 0.282 -42	5.96	4.60	6.45	0.245	-254.64	4.76	6.08/0.113/-15/7.72	4.73/0.154/4	4.51	4.56/0.15/-24	4.51	4.4-4.6
7.25 0.516 66	7.74	5.22	7.18	0.567	70.5.17	5.09	6.99/0.231/73/6.03	5.11/0.242/75	5.04	5.04/0.24/88	4.92	4.8-5.1
8.32 0.104 51	6.18	5.99	8.27	0.089	59.5.95	5.96	7.89/0.023/7/5.54	5.98/0.021/6			5.51	5.4-5.5
9.25° 0.113 79	5.52	6.66	9.13	0.512	-896.57	6.65	8.60/0.161/-80/10.17	6.49/0.287/-85	6.33	6.19/0.40/86	6.20	6.0-6.3
9.26 0.356 81	6.17	6.67										
<i>nr*</i> * Transitions												
6.6.97 0.001 -	4.71	5.02	7.28	0.001	-	5.24	5.79/6.22/10 ⁴ /-/4.31					
7.82 0.010 -	5.84	5.63	8.01	0.010	-	5.77	6.60/8.05/0.013/-4.63					

8.58	0.003	-	7.24	6.18	8.89	0.002	-	6.40	6.63/7.97/0.002/-2.64
keto-N7H									
<i>ππ*</i> Transitions									
6.16	0.222	-3	1.71	4.44	6.11	0.277	114.40		
7.52	0.232	64	1.60	5.41	7.44	0.095	515.36		
8.03	0.614	86	1.35	5.78	7.86	0.796	895.66		
8.63	0.239	-63	3.98	6.21	8.65	0.338	-716.23		
9.68	0.111	-32	2.71	6.97	9.48	0.234	-286.83		
<i>ππ*</i> Transitions									
7.05	0.001	-	4.66	5.08	7.53	0.001	-	5.42	
8.06	0.058	-	3.07	5.80	8.18	0.004	-	5.89	
8.24	0.017	-	1.18	5.93	8.57	0.001	-	6.17	
9.04	0.015	-	1.89	6.51	9.05	0.015	-	6.52	

^aAbs¹: absorption of guanine in water [214], Abs²: absorption of guanosine in water and Φ values are based on polarized absorption spectra of crystalline guanosine [251], CD: CD spectra in aqueous solution of deoxy guanosine 5'-phosphate (dGMP) [226], ^bΔE¹ corresponds to CASPT2, and ΔE² corresponds to CASSCF transition energies in the gas phase, ΔE³/f/Φ corresponds to results in water [126], ^cground state dipole moments of the keto-N9H and keto-N7H tautomers at the HF/6-311G(d,p) level are 6.77 and 1.78 Debye, respectively, ^dscaled (scaling factor 0.72) excitation energies, ^eRydberg contamination, ^fAverage of transitions at 6.66 and 6.67 eV.

effect of the aqueous solvent on electronic transitions was considered using the self-consistent reaction field (SCRF) model. The computed CASPT2 transition energies were found to be in reasonably good agreement (with an accuracy of 0.3 eV) with experimental data, while the CASSCF transition energies were much larger.¹²⁶

Our computed vertical singlet $\pi\pi^*$ and $n\pi^*$ transition energies, transition moment directions and dipole moments of the keto-N9H and keto-N7H tautomers of guanine and their hydrated forms (with three water molecules) obtained at the CIS/6-311G(d,p) level are presented in Table 7 along with some experimental and the CASSCF and CASPT2 results¹²⁶ for comparison. Because of uncertainty in assignment and possibly significant basis set dependence, we are not giving $\pi\sigma^*$ Rydberg type transitions in this table. It should be noted here that the excited state assignment and numbering corresponds to isolated species. The ground state geometries optimized at the HF/6-311G(d,p) level and found minima at the respective potential energy surfaces were used in the transition energy calculations. Here it should be noted that the keto-N9H tautomer in the gas phase is found to be about 0.86 kcal/mol more stable than the keto-N7H form. Under hydration with three water molecules the latter tautomer is found to be about 3.19 kcal/mol more stable than the former one. As the value of molecular dipole moment of the keto-N7H form is smaller than the keto-N9H form (Table 7), the greater stability of keto-N7H tautomer would be due to the local nature of hydrogen bonding interactions rather than due to molecular dipole moment stabilization. This conclusion has been confirmed by the fact that adding more water molecules does not change the order of relative stabilities of both the tautomers. However, given the small total energy difference between the tautomers, it is clear that both forms of guanine will be present in the gas phase and in aqueous solutions.^{60,91,119}

Before comparing the CIS computed transitions with experimental data and the CASSCF/CASPT2 results, it may be mentioned that $n\pi^*$ transition energies are increased (blue-shifted) after hydration (Table 7). Consequently, while the second singlet excited state is of $n\pi^*$ type for both the keto-N9H and keto-N7H tautomers in gas phase, after hydration the first singlet $n\pi^*$ transition becomes the third transition in the complexes (Table 7). The CASSCF calculation also predicted the second transition to be of $n\pi^*$ type for the keto-N9H tautomer in gas phase, while it is the third transition according to the CASPT2 method.¹²⁶ The CIS/6-311G(d,p) and CASPT2/CASSCF results are in agreement with respect to the assignment of the first $n\pi^*$

transition as being due to the excitation of the carbonyl group lone pair. The computed oscillator strength of the first $\pi\pi^*$ transition is lower than that of the second $\pi\pi^*$ transition (Table 7). The order in transition intensity agrees with the solution spectra of guanine and its derivatives in which the first transition (near 275 nm) appears as a weak peak in comparison to the stronger peak near 250 nm region.²³ Table 7 shows that there is good correspondence between the computed transitions (scaled) of the keto-N9H tautomer (and its hydrated form) and the CASPT2 results (solvation included), in particular when comparison is made with the transition of the hydrated tautomer. Further, the agreement is better in the lower energy region than in the higher energy region. The third $\pi\pi^*$ transition computed at 5.95 eV of the hydrated keto-N9H tautomer has the lowest oscillator strength among all the $\pi\pi^*$ transitions shown in Table 7. This calculated transition can be considered for explanation of 5.5 eV band in experimental data. As discussed earlier, this transition has been suggested as a weak $\pi\pi^*$ or $n\pi^*$ transition on the basis of CD spectra.²²⁶ Two almost degenerate transitions near 6.66 and 6.67 eV (scaled values) of the keto-N9H tautomer in gas phase correspond to a single transition at 6.57 eV for the hydrated form which explains the 6.0–6.3 eV experimental region of guanine (Table 7). The existence of two transitions near 200 nm (6.2 eV) is revealed by the MCD spectra of guanosine 5'-diphosphate, which shows that the 200 nm band is composed of two transitions.²²¹

The transition energy of the first $\pi\pi^*$ transition of the keto-N7H tautomers and its hydrated forms is lower than the corresponding transition energies of the keto-N9H tautomer and its hydrated form (Table 7). This prediction is in accordance with the experimental result that the first absorption band of 7-methylguanine is red-shifted by about 10 nm relative to that of guanosine monophosphate (GMP).²⁵⁹ The computed $n\pi^*$ transitions of the hydrated keto-N9H tautomer are located at 5.24, 5.77 and 6.40 eV (Table 7). These results support the findings of Clark²⁵¹ with regard to the existence of $n\pi^*$ transitions near the 5.21 and 6.32 eV in guanine. It appears that in the 5.5 eV region of guanine, weak $\pi\pi^*$ and $n\pi^*$ transitions are present and are responsible for the ambiguous assignment of transitions in that region.

4.2.2 Geometries and Charges

Geometrical parameters of the keto-N9H and keto-N7H tautomers and their hydrated forms in the ground and lowest singlet $\pi\pi^*$ and $n\pi^*$ excited states are shown in Table 8. All geometries were found to be minima at their respective potential energy surfaces. The ground state geometries of the keto-N9H and

Table 8. Ground and excited state optimized bond lengths (Å), bond angles (°) and dihedral angles (°) of the keto-N9H and keto-N7H tautomers of guanine and their hydrated form.^a

Parameters	keto-N9H		keto-N7H			
	S ₀	S ₁ (ππ*)	S ₀	S ₂ (ππ*)	S ₁ (ππ*)	S ₂ (ππ*)
N1C2	1.357	1.354	1.397	1.399	1.364	1.366
C2N3	1.286	1.293	1.393	1.387	1.282	1.277
N3C4	1.355	1.357	1.284	1.287	1.367	1.364
C4C5	1.367	1.369	1.431	1.432	1.361	1.367
N1C6	1.417	1.403	1.429	1.420	1.434	1.397
C5C6	1.436	1.428	1.455	1.450	1.463	1.424
C5N7	1.378	1.379	1.354	1.352	1.371	1.377
N7C8	1.276	1.278	1.293	1.301	1.283	1.339
C4N9	1.352	1.350	1.376	1.374	1.356	1.359
C8N9	1.375	1.370	1.360	1.349	1.368	1.298
O6C6	1.188	1.201	1.184	1.196	1.282	1.195
N2C2	1.362	1.355	1.370	1.363	1.367	1.369
H21N2	0.994	0.994	1.006	1.003	0.996	0.995
H22N2	0.995	0.993	0.996	0.995	0.995	0.996
H1N1	0.996	1.003	0.998	1.003	0.995	0.996
H8C8	1.071	1.071	1.072	1.071	1.072	1.072
H7/9N7/9	0.993	0.999	0.992	0.999	0.993	1.001
N1C2N3	123.9	124.0	118.4	119.0	125.7	124.5
C2N3C4	112.7	112.8	109.1	109.4	112.8	114.4
N3C4C5	128.9	128.3	125.1	124.4	128.6	124.4

C2N1C6	126.3	111.2	123.7	125.1	114.0	116.9	121.6
C4C5C6	118.6	118.5	119.3	118.0	122.4	121.5	121.5
N1C6C5	109.6	110.6	109.9	110.3	110.4	109.1	110.5
C4C5N7	110.5	110.2	110.1	109.8	110.8	105.7	104.5
C6C5N7	130.8	131.2	128.7	129.6	131.2	131.9	132.6
C5N7C8	104.8	104.7	105.2	105.1	104.8	105.5	105.3
N3C4N9	125.7	126.0	130.9	131.1	126.0	125.4	129.2
C5C4N9	105.4	105.6	104.0	104.3	105.3	110.2	109.4
N7C8N9	112.8	113.0	114.4	114.6	112.6	113.8	114.3
C4N9C8	106.4	106.4	106.2	106.2	106.5	104.7	104.5
O6C6N1	119.2	119.3	121.7	122.0	116.2	121.4	120.4
O6C6C5	131.2	130.2	128.4	127.7	116.4	129.5	129.1
N2C2N1	116.2	116.0	116.7	117.1	114.4	115.1	115.6
N2C2N3	119.8	120.1	120.0	120.0	119.9	120.3	120.6
H2IN2C2	117.9	118.8	115.3	116.7	116.3	116.8	118.2
H22N2C2	113.8	115.7	112.7	114.4	112.9	112.5	115.1
H2IN2H22	115.0	116.3	111.4	113.3	114.1	113.8	116.5
H1N1C2	119.5	119.7	113.9	114.4	116.3	119.5	119.7
H8C8N7	125.7	125.4	124.2	123.8	125.6	121.8	121.7
H79N7/9C8	127.7	128.9	127.1	129.0	127.6	127.7	128.9
N3C2N1C6	-0.6	0.2	-64.0	-64.8	-1.6	-0.5	0.2
N3C2N1H1	176.0	-177.0	170.9	169.9	157.7	174.9	-177.8

-145.5

-167.5

-169.8

-177.8

-199.8

-200.8

-201.8

-202.8

-203.8

-204.8

-205.8

-206.8

-207.8

C4N3C2N1	0.8	0.8	44.2	42.4	-5.5	0.9	0.4	22.9	13.6	0.3
C4N3C2N2	177.2	-177.5	-161.4	-160.5	171.7	177.0	-177.6	-165.3	-171.0	177.9
C5C4N3C2	-1.0	-1.2	-2.4	0.0	5.9	-1.2	-0.6	7.3	9.0	-6.6
N9C4N3C2	179.5	179.2	174.5	173.7	170.8	179.3	180.0	176.4	176.6	177.2
C6C5C4N3	0.9	0.5	-18.5	-18.8	0.9	1.1	0.3	-14.4	-10.8	7.4
N7C5C4N3	179.6	-179.7	175.9	173.2	178.3	179.6	-179.4	170.3	169.3	178.1
C5C6N1C2	0.3	-0.8	36.2	38.9	8.1	0.3	-0.5	32.6	29.0	-4.0
O6C6N1C2	179.4	179.2	-145.1	-142.8	-127.2	179.4	179.6	-147.5	-152.8	147.9
N1C6C5C4	-0.4	0.5	-0.6	-3.3	-7.4	-0.6	0.3	-7.0	-8.9	-1.8
N1C6C5N7	179.8	-179.2	162.0	162.1	171.6	179.6	179.9	166.7	171.0	174.5
O6C6C5C4	179.3	-179.6	179.2	178.5	127.8	179.1	-179.8	173.1	173.1	150.3
O6C6C5N7	-0.1	0.7	-16.6	-16.2	-53.2	0.0	-0.2	-13.1	-7.0	37.0
C8N7C5C4	0.1	0.0	-0.1	0.4	0.7	0.1	0.0	0.9	0.8	1.4
N9C8N7C5	0.0	0.1	2.0	1.4	0.0	0.0	-0.1	-0.8	-0.9	-0.9
H8C8N7C5	180.0	-179.9	178.5	-179.1	179.9	180.0	179.9	179.7	179.3	178.3
H7N7C5C4					179.7	179.8	177.8	179.9	179.9	176.9
H9N9C8N7	179.8	-179.7	-175.6	-177.2	178.1					
H21N2C2N1	30.4	23.3	42.3	34.8	-39.6	36.6	25.6	32.3	17.0	39.9
H21N2C2N3	-151.4	-158.2	-112.5	-122.7	142.9	-145.3	-156.2	-140.1	-158.7	141.6
H22N2C2N1	169.4	169.0	171.8	170.4	174.3	170.9	169.7	170.2	171.4	171.7
H22N2C2N3	-12.4	-12.6	16.9	12.9	8.1	-11.0	-12.2	-2.3	-4.3	-9.9

^aGround and excited state parameters are at the HF/6-31G(d,p) and CIS/6-311G(d,p) levels, respectively. The S₁, S₂ represent singlet excited state in the ascending energy order (Table 7). The “+3W” represents the corresponding hydrated species with three water molecules.

keto-N7H tautomers are almost planar except for the amino group which is pyramidal. The pyramidalization of the amino group in the keto-N7H tautomer is larger than that in the keto-N9H form (Table 8). In the $S_1(\pi\pi^*)$ excited state of the keto-N9H tautomer, the pyrimidine ring is highly distorted especially around the C6N1C2N3 region. The N1 and C6 atoms are significantly out-of-plane, and pyramidalization of the amino group is also appreciably increased in this state. The geometry of the molecule in this state is shown in Fig. 4a. In the $S_2(n\pi^*)$ excited state, the ring geometry is slightly nonplanar, and the amino group is also more pyramidal compared to that in the ground state. The most significant change is found to be the large increase of C6O6 bond length (by about 0.1 Å) compared to the ground state value. The O6 and H1 atoms are displaced away from the ring plane on opposite sites in the $S_2(n\pi^*)$ state. In the case of the keto-N7H tautomer, geometrical changes in the excited states are similar but usually smaller than those in the keto-N9H tautomer. The hydration with three water molecules in the ground and excited states induces planarity in the ring geometry and in the amino group of the guanine tautomers. The computed hydrogen bond lengths of the hydrated forms in the ground and $S_1(\pi\pi^*)$ excited state are also shown in Fig. 4. A similar trend of change in geometrical parameters was found in a recent study on guanine tautomers¹²⁵ where geometries of the molecules were also optimized in the ground and excited states in aqueous media; the effect of solvation being treated using the integral equation formalism of the polarizable continuum model (IEF-PCM).^{260,261} Mennucci et al.¹²⁵ have found larger changes in the geometrical parameters in going from gas phase to the aqueous medium, although the excited state gas phase results are approximately the same despite the fact that a different basis set was used (the CIS method was used in the excited state geometry optimization). This change appears to be due to the difference in approach used to incorporate the effect of hydration.

The computed Mulliken and electrostatic potential fitted CHELPG charges^{243,244} of the keto-N9H and keto-N7H tautomers of guanine in the ground and singlet excited states are presented in Table 9. The charge distributions of guanine tautomers are generally similar in the ground and vertical singlet $S_1(\pi\pi^*)$ excited state, while these are modified in the relaxed $S_1(\pi\pi^*)$ excited state (Table 9). In going from the ground state to the vertical $S_2(n\pi^*)$ excited state of guanine tautomers the C6 atom gains while the O6 atom loses significant amounts of electronic charge. The loss of electronic charge by the O6 atom is in agreement with the fact the $n\pi^*$ transition involves the excitation of the carbonyl (C6O6) lone pair. The charge distributions in the

Table 9. Ground and singlet excited state charges at the atomic sites of the keto-N9H and keto-N7H tautomers of guanine.^a

Atom	keto-N9H			keto-N7H			keto-N7H		
	S ₀	S ₁ (ππ*)	S ₂ (ππ*)	S ₀	S ₁ (ππ*)	S ₂ (ππ*)	S ₀	S ₁ (ππ*)	S ₂ (ππ*)
Mulliken	Vert	Opt	Vert	Vert	Opt	Vert	Vert	Opt	Opt
N1	-0.586	-0.600	-0.449	-0.579	-0.573	-0.584	-0.478	-0.577	-0.570
C2	0.743	0.658	0.356	0.743	0.719	0.709	0.677	0.517	0.708
N3	-0.566	-0.476	-0.415	-0.580	-0.570	-0.529	-0.463	-0.420	-0.539
C4	0.504	0.418	0.544	0.443	0.460	0.372	0.295	0.362	0.310
C5	-0.138	-0.076	-0.060	-0.023	-0.060	-0.025	0.074	0.069	0.076
C6	0.613	0.579	0.536	0.354	0.337	0.623	0.562	0.499	0.399
N7	-0.390	-0.390	-0.384	-0.430	-0.407	-0.486	-0.495	-0.506	-0.498
C8	0.236	0.291	0.289	0.223	0.227	0.275	0.225	0.259	0.224
N9	-0.491	-0.506	-0.505	-0.482	-0.479	-0.427	-0.391	-0.389	-0.402
O6	-0.455	-0.428	-0.416	-0.159	-0.167	-0.485	-0.468	-0.469	-0.195
N2	-0.550	-0.535	-0.531	-0.553	-0.551	-0.548	-0.528	-0.521	-0.551
H21	0.223	0.209	0.201	0.215	0.223	0.216	0.217	0.222	0.208
H22	0.243	0.235	0.226	0.243	0.239	0.243	0.243	0.245	0.243
H1	0.253	0.246	0.220	0.239	0.252	0.255	0.249	0.217	0.242
H8	0.114	0.127	0.133	0.105	0.112	0.119	0.122	0.127	0.115
H7/9	0.247	0.248	0.255	0.241	0.244	0.261	0.265	0.266	0.245
CHELPG									
N1	-0.932	-0.856	-0.529	-0.923	-0.452	-0.955	-0.884	-0.743	-0.922

^aGround and excited state charges are at the HF(6-31G(d,p) and CIS(6-31G(d,p) levels, respectively. The S₁, S₂ represent singlet excited states in the ascending energy order (Table 7). The “Vert” represents the vertical and “Opt” represents optimized state.

relaxed excited states are also noticeably different from that in the corresponding vertical state. This difference between the charge distributions of the vertical and relaxed excited states would be partly due to nonplanarity of geometry in the $\pi\pi^*$ and $n\pi^*$ excited states as discussed earlier (Table 8).

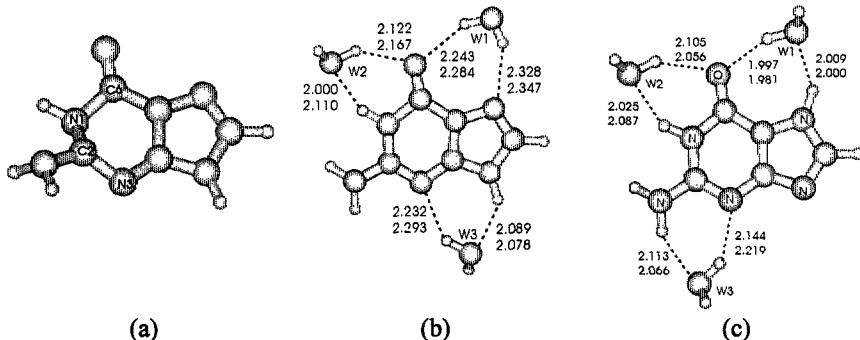


Figure 4. Structure of guanine tautomers: (a) geometry of the keto-N9H tautomer in the $S_1(\pi\pi^*)$ state, (b) hydrated form of keto-N9H tautomer in the ground state (top indices) and $S_1(\pi\pi^*)$ state (bottom indices) (c) hydrated form of keto-N7H tautomer in the ground state (top indices) and $S_1(\pi\pi^*)$ state (bottom indices).

4.2.3 Phototautomerism

The energy level diagram of the keto-N9H and keto-N7H tautomers of guanine and their hydrated forms obtained at the CIS/6-311G(d,p) level are shown in Fig. 5. The energy level diagram reveals that the vertical and relaxed lowest singlet $\pi\pi^*$ excited states of the keto-N7H tautomer and its hydrated form lie lower than those of the keto-N9H tautomer and its hydrated form, except the relaxed $S_1(\pi\pi^*)$ state of the keto-N9H tautomer which also lies at the same height as that of the keto-N7H tautomer. Therefore, in an aqueous medium, fluorescence would mainly originate from the keto-N7H tautomer of guanine.^{23,119,257} However, in the gas phase, there would be strong coupling between the relaxed $S_1(\pi\pi^*)$ states of both tautomers due to their closed proximity. Thus, on electronic excitation of guanine, energy transfer from the lowest singlet $\pi\pi^*$ excited state of the keto-N9H tautomer to the corresponding state of the keto-N7H form of the molecule would take place. The energy absorbed by the former species would be mainly liberated as fluorescence by the latter (keto-N7H) and partly as its own fluorescence. Obviously, this energy transfer process in the excited state would give rise to phototautomerization of guanine from the keto-N9H form to the keto-N7H

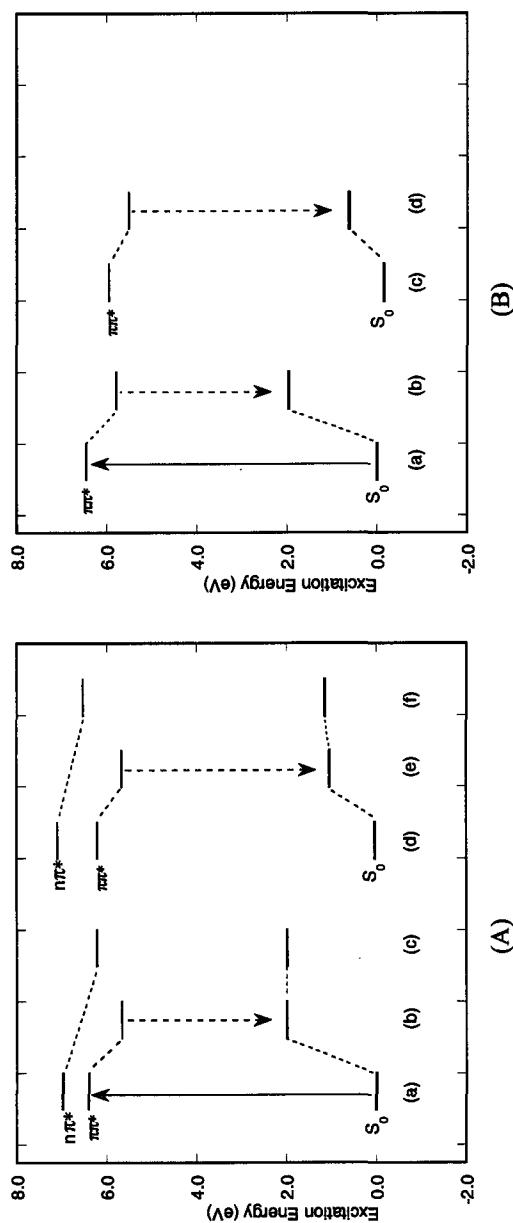


Figure 5. Energy level diagrams of guanine tautomers, (A) isolated form; (B) hydrated form. In the case of isolated tautomers, (a), (b) and (c) show vertical, optimized lowest singlet $\pi\pi^*$ and nr^* states, respectively, of the keto-N9H tautomer, while (d), (e) and (f) show vertical, optimized lowest singlet $\pi\pi^*$ and nr^* states, respectively, of the keto-N7H tautomer. For hydrated forms (a) and (b) show vertical and optimized lowest singlet $\pi\pi^*$ state, respectively, of the hydrated keto-N9H tautomer, while (c) and (d) show vertical and optimized lowest singlet $\pi\pi^*$ state, respectively, of the hydrated keto-N7H tautomer. The upward arrows correspond to absorption and downward arrows correspond to emission.

form. Since the $S_1(\pi\pi^*)$ excited state geometry of the keto-N9H tautomer is much more nonplanar than that of the corresponding state of the keto-N7H tautomer, it would cause a strong nonradiative decay of excitation energy of the former species. Therefore, only a weak fluorescence is expected from the keto-N9H form, and a major part of fluorescence would originate from the keto-N7H form of guanine.¹¹⁹ The energy differences between the relaxed singlet $S_1(\pi\pi^*)$ excited states of the hydrated keto-N7H and keto-N9H tautomers and the corresponding ground states lying vertically below them are 4.89 and 3.82 eV, respectively. Therefore, the corresponding scaled values would be 3.52 and 2.75 eV respectively. The observed fluorescence peaks in an aqueous solution lie near 3.73 (332 nm) and 2.75 eV (450 nm).^{23,253} Thus the main fluorescence near 332 nm would originate from the N7H tautomer, and weak fluorescence near 450 nm would correspond to the keto-N9H tautomer.^{23,119,257} Since the ground state corresponding to the relaxed $S_1(\pi\pi^*)$ excited state of the keto-N9H tautomer of guanine lies about 2 eV above the corresponding total energy minimum, it would be in a very reactive state. Therefore, in an oxygen rich aqueous solution, it may react with oxygen, and thus may be the reason for the formation of a new reaction product.^{119,254} The observed oscillation of UV absorption intensities at 275 and 250 nm in an irradiated aqueous solution of guanine²⁵² was explained as arising due to an oscillation in the relative populations of the keto-N9H and keto-N7H tautomers of the same.¹¹⁹ This process is facilitated by the complexation of the tautomers of guanine with oxygen which eventually leads to formation of the reaction product.^{119,254}

4.3 Uracil and Thymine

4.3.1 Electronic Transitions

4.3.1.1 Experimental Data

Thymine being 5-methyluracil, electronic spectra of uracil and thymine are similar. Thus, both uracil and thymine show three absorption bands near 260, 205 and 180 nm (4.77, 6.05 and 6.89 eV, respectively). The first and third bands in thymine are generally slightly red- and blue-shifted, respectively, with respect to the corresponding bands in uracil.^{23,213,214,217,220,224,262,263} The CD spectra reveal a composite nature of the 205 nm band corresponding to peaks near 215 and 195 nm (5.77 and 6.36 eV, respectively).²²⁶ The CD spectra

indicated the presence of a band near 240 nm (5.17 eV) that was assigned as $n\pi^*$ type.^{23,226} Hug and Tinoco²³⁸ have predicted an $n\pi^*$ transition at 250 nm (4.96 eV) which has been suggested as the possible source of the 240 nm band observed in the CD spectra.^{23,226}

The existence of an $n\pi^*$ transition within the 260 nm envelope is confirmed by different investigations.^{23,45,58,217,263} Backer and coworkers²⁶³⁻²⁶⁶ have performed a series of experiments on the absorption and emission properties of uracil, thymine and their derivatives in polar protic and aprotic solvents at low and room temperatures. It has been found that uracil, thymine and thymidine exhibit strong phosphorescence in polar aprotic solvents (2-methyltetrahydrofuran (2-MTHF)) while a relatively stronger fluorescence was observed in a polar protic solvent (EtOH-MeOH). It has been concluded that in an aprotic solvent, the lowest singlet excited state of each of uracil, thymine and thymidine has a predominantly $n\pi^*$ character. However, in a protic solvent or under 1,3-dimethylation, the lowest singlet excited state character is changed to $\pi\pi^*$.²⁶³ Although the $n\pi^*$ transition is not assigned in the photoacoustic spectra of thymine, the existence of another transition within the 270 nm (4.59 eV) envelope was suggested.²¹⁷ The relative position of the $n\pi^*$ transition is solvent dependent and in gas phase and in an aprotic solvent, the $n\pi^*$ state lies the lowest while in a protic environment, it lies higher than the $\pi\pi^*$ state. Consequently, the $n\pi^*$ state becomes the second state in protic environment.^{23,45,58,263}

The polarized absorption^{23,216,267} and reflection experiments^{23,268} suggest that the transition moment direction for the first band is close to 0° for uracil and -20° for thymine with respect to the N1C4 direction (Fig. 1d). Regarding the second band transition moment direction, Novros and Clark²⁶⁹ have suggested two values -53° or $+59^\circ$ with respect to the N1C4 direction and favored the latter being consistent with the LD spectra of uracil.²¹⁹ However, Anex et al.²⁷⁰ have suggested it to be -31° with respect to the N1C4 direction. Eaton and Lewis have estimated that polarization of the I and II bands are approximately perpendicular to each other.²⁶⁷ Holmen et al.²⁷¹ have found an angle of 35° with respect to the N1C4 direction for the second transition of 1,3-dimethyluracil. Summary of the experimental transitions obtained for uracil and thymine are shown in Tables 10 and 11, respectively alongwith their substituted analogs.

Table 10. Summary of experimental transition energies (ΔE , eV) of uracil and its derivatives. The Φ represents the transition moment direction ($^\circ$) according to the DeVoie-Tinoco convention (Figure 1d).

Transitions				References
Absorption Spectra				
ΔE	5.08	6.05	6.63	Uracil, vapor [213]
ΔE	4.84	6.05	6.63	1,3-Dimethyluracil, vapor [213]
ΔE	4.68	6.08	6.63	1,3-Dimethyluracil, water [213]
ΔE	4.81	6.11		Uracil, water [220]
ΔE	4.75	6.05		Uridine, water [220]
ΔE	4.81	6.11	6.85	Uracil, TMP [224]
ΔE	4.79	6.14	6.85	Uracil, water [224]
ΔE	4.70	6.02	6.74	1,3-Dimethyluracil, TMP [224]
ΔE	4.73	6.11	6.81	1,3-Dimethyluracil, MCH [224]
ΔE	4.51	5.82		1-Methyluracil, crystal [269]
Φ	-9	59		
ΔE	4.66	6.08	6.97	Uracil, sublimed film [276]
CD Spectra				
ΔE	4.73	5.77	6.36	Uridine, water [226]
ΔE	4.63	5.71		Uridine, water [220]
ΔE	4.68	5.82	6.26	Uridine, water [225]
MCD Spectra				
ΔE	4.86	5.85		Uracil, water [220]
ΔE	4.77	5.71		Uridine, water [220]
Electron Scattering				
ΔE	4.70	5.93	6.93	Uracil, film [262]

TMP: trimethylphosphate, MCH: methylcyclohexane; for details see relevant reference.

4.3.1.2 Theoretical Data

Ab initio calculations on the electronic transitions of uracil and thymine were performed by Petke et al.¹²⁹ using the MRCI and RPA methods and by Roos and coworkers^{131,128} at the CASSCF and CASPT2 levels of the theory. The CIS method was also applied to this problem,^{58,136} one of them being mainly devoted to the excited state geometries of molecules.⁵⁸ The X-ray crystallographic geometry²⁷² was used by Petke et al.¹²⁹ and the computed transition energies were generally higher within the range of 1.5-1.9 eV as compared to the experimental ones, and a linear scaling was used in order to

compare the calculated results with experimental data. The linear scaling method was also used in the CIS calculations, and scaled transition energies were found to be in reasonably good agreement with experiment.^{58,136} In the CASSCF and CASPT2 calculations,¹²⁸ the average experimental geometries from Taylor and Kennard²⁷³ were used. The NH and CH bond lengths were assumed to be 1.01 and 1.08 Å, respectively. Although the CASSCF energies are appreciably higher, the CASPT2 level of dynamic correlation correction to CASSCF energies yielded transition energies that were in agreement with experimental data.¹²⁸

Table 11. Summary of experimental transition energies (ΔE , eV) of thymine and its derivatives. The f represents oscillator strength.

Transitions	References			
Absorption Spectra				
ΔE 4.68	6.08			Thymine, water [220]
ΔE 4.64	6.05			Thymidine, water [220]
ΔE 4.54	5.99			1-methylthymine, water [216]
f 0.19	0.28			
ΔE 4.64	5.88	7.04		Thymine, sublimed film [276]
CD Spectra				
ΔE 4.68	5.77	6.36	7.00	Thymidine, water [226]
ΔE 4.54	5.69			Thymidine, water [220]
ΔE 4.63	5.85	6.42		Thymidine, water [225]
MCD Spectra				
ΔE 4.71	5.77			Thymine, water [220]
ΔE 4.73	5.64			Thymidine, water [220]
Photo Acoustic Spectra				
ΔE 4.59	5.90	7.08		Thymine, film [217]
Electron Scattering				
ΔE 4.66	5.94	7.08	8.82	Thymine, film [262]

Our computed vertical singlet $\pi\pi^*$ and $n\pi^*$ transition energies, transition moment directions and dipole moments of uracil, thymine and their hydrated forms (with three water molecules) using the CIS/6-311G(d,p)/HF/6-311G(d,p) level are shown in Tables 12 and 13. Because of uncertainty in assignment and possibly significant basis set dependence, we are not giving $\pi\sigma^*$ Rydberg type transitions in these tables. Further, in these tables, the

excited state assignment corresponds to isolated species. For the sake of the comparison, some experimental and CASPT2/CASSCF transition energies are also shown in Tables 12 and 13. Among all $n\pi^*$ transitions of uracil and thymine (Tables 12 and 13), the first $n\pi^*$ transition is localized at the C4O4 group. The second $n\pi^*$ transition is localized at the C2O2 group, while the third $n\pi^*$ transition is of mixed type with contribution from both of the C4O4 and C2O2 groups. Assignments of these transitions are in agreement with the MRCI, RPA¹²⁹ and CASSCF/CASPT2 results.¹²⁸ Upon hydration, the $n\pi^*$ transition energies are increased (blue-shifted). Consequently, in gas phase, the first singlet vertical excited state of uracil and thymine has a $n\pi^*$ character. After hydration, the $\pi\pi^*$ state is lowered. This change in the nature of the excited state is in accordance with experimental observations where, in gas phase or in an aprotic solvent, uracil and thymine were found to have a $n\pi^*$ state as the lowest singlet excited state. In protic solvent, the $\pi\pi^*$ state is the lowest one.^{23,263} The computed first $n\pi^*$ transition after hydration has the energy of 6.79 eV for both uracil and thymine. The corresponding scaled value will be 4.89 eV (Tables 12 and 13). This transition can be correlated with the $n\pi^*$ transition near 250 nm (4.96 eV) of uracil and thymine in an aqueous medium, the relative position of which is solvent dependent.^{23,128,220,225,226} The computed second $n\pi^*$ transition in uracil and thymine is predicted near 5.8 eV (Tables 12 and 13). The existence of an $n\pi^*$ transition near 217 nm (5.71 eV) has been previously suggested in 1-methyluracil²⁷⁴ which is in an excellent agreement with the CIS prediction. The computed first singlet $\pi\pi^*$ transition energy of thymine is lower than the corresponding transition energy of uracil (Tables 12 and 13). This result is in agreement with the experimental observation that the first absorption band of thymine is red-shifted with respect to the corresponding band of uracil.^{23,213,214,217,220,224,262,263} The data shown in Tables 12 and 13 suggest that there is good correspondence between the scaled CIS-computed excitation energies and the CASPT2 excitation energies except for the second transition in uracil and the second and third transitions in thymine.¹²⁸ The composite nature of the 205 nm (6.05 eV) band observed in CD spectra²²⁶ of these molecules is also revealed in the CIS computations although the splitting is small. Furthermore, the observed CD spectra in an aqueous solution of thymidine and uridine and the aqueous absorption spectra of uracil and thymine can be easily explained in terms of the computed (scaled) transitions of hydrated uracil and thymine with an accuracy of about 0.2 eV except for the second transition for which the error is somewhat higher.

Table 12. Vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies (ΔE , eV), oscillator strengths (f), transition moment directions (Φ, γ) and dipole moments (μ , Debye) of the keto tautomer of uracil in the isolated and hydrated forms at the CIS/6-311G(d,p)/HF/6-311G(d,p) level.²⁵

CIS								Experimental Data ^a								
Isolated				Hydrated				CASPT2/CASSCF ^b				Abs ¹	Abs ²	CD	Crystal	Range
ΔE	f	Φ	μ^c	ΔE^d	ΔE	f	Φ	ΔE^d	$\Delta E^1/\Delta E^2/f/\Phi \mu$	ΔE	ΔE	ΔE	$\Delta E/\Phi$	ΔE		
<i>$\pi\pi^*$ Transitions</i>																
6.83	0.446	-7	5.07	4.92	6.74	0.447	-6	4.85	5.00/6.88/0.19/-7/6.3	5.08	4.79	4.73	4.51/-9	4.6-5.1		
8.89	0.123	36	3.48	6.40	8.73	0.140	46	6.29	5.82/7.03/0.08/-29/2.4	6.05	6.14	5.77	5.82/59	5.8-6.1		
9.29	0.386	-66	4.99	6.69	9.12	0.439	-57	6.57	6.46/8.35/0.29/23/6.9	6.63	6.36	6.36		6.3-6.6		
10.0	0.322	-14	2.43	7.20	9.93	0.251	-15	7.15	7.00/8.47/0.76/-42/3.7		6.85	7.00		6.7-7.0		
<i>$n\pi^*$ Transitions</i>																
6.51	0.000	-	2.82	4.69	6.79	0.001	-	4.89	4.54/4.78/-/-3.4							
7.98	0.000	-	5.10	5.75	8.11	0.000	-	5.84	6.00/6.31/-/-4.8							
9.96 ^c	0.006	-	7.06	7.17	9.97	0.009	-	7.18	6.37/7.80/-/-8.7							

^aAbs¹: absorption in the gas phase [213], Abs²: absorption in aqueous medium [224], CD: CD spectra of uridine in an aqueous medium [226], Crystal: transition energy/transition moment direction [269]; Range: range of transitions observed in different experiments, ^bΔE¹ represents CASPT2, and ΔE² represents CASSCF transition energies; for the f values of nr* transitions see original paper [128], "ground state dipole moment at the HF/6-31G(d,p) level is 4.67 Debye, ^dscaled (scaling factor 0.72) excitation energies, ^eRydberg contamination.

Table 13. Vertical singlet $\pi\pi^*$ and nr* excitation energies (ΔE , eV), oscillator strengths (f), transition moment directions (Φ , °) and dipole moments (μ , Debye) of thymine in the isolated and hydrated forms at the CIS/6-311G(d,p)/HF/6-311G(d,p) level.

CIS								Experimental Data ^a											
Isolated				Hydrated				CASPT2/CASSCF ^b				Abs		PA		CD		Range	
ΔE	f	Φ	μ^c	ΔE^d	ΔE	f	Φ	ΔE^d	$\Delta E^e/\Delta E^2/f/\Phi/\mu$	ΔE^f	ΔE^g	ΔE							
<i>nr*</i> Transitions																			
6.70	0.455	-10	5.06	4.82	6.61	0.450	-8	4.76	4.88/6.75/0.17/15/6.5	4.68	4.59	4.68	4.5-4.9						
8.77	0.162	67	5.14	6.31	8.57	0.188	75	6.17	5.88/7.15/0.17/-19/1.5	6.08	5.90	5.77	5.8-6.1						
9.16	0.436	54	3.56	6.59	9.03	0.493	48	6.50	6.10/8.33/0.15/67/7.8					6.36	6.3-6.4				
10.1	0.211	-18	2.31	7.27	10.0	0.216	-10	7.20	7.13/8.62/0.85/-25/3.1					7.00	6.7-7.1				
<i>nr*</i> Transitions																			
6.52	0.000	-	3.22	4.69	6.79	0.000	-	4.89	4.39/5.22/-1/3.2										
8.04	0.000	-	4.67	5.79	8.19	0.000	-	5.90	5.91/6.77/-1/4.6										

9.96 0.000 - 4.66 7.17 9.94 0.004 - 7.16 6.15/8.14/-/8.6

^aAbs: absorption of thymine in water [220], PA: photoacoustic spectra of thymine [217], CD: CD spectra of thymidine in water [226], Range: range of transitions observed in different experiments, ^b ΔE^1 represents CASPT2, and ΔE^2 represents CASSCF transition energies; for the f values of mr^* transitions see original paper [128], ^cground state dipole moment at the HF/6-311G(d,p) level is 4.56 Debye, ^dscaled (scaling factor 0.72) excitation energies.

Table 14. Ground and excited state optimized bond lengths (Å), bond angles (°) and dihedral angles (°) of uracil tautomers and their hydrated forms. Excited state numbering corresponds to isolated species.^a²⁷⁵

	U-keto		U-enol		U-O2H1	
	S_0	$S_1(mr^*)$	S_0	$S_1(mr^*)$	S_0	$S_1(mr^*)$
	+3W	+3W	+3W	+3W	+3W	+3W
C2N1	1.373	1.370	1.371	1.370	1.313	1.344
N3C2	1.370	1.366	1.367	1.357	1.320	1.324
C4N3	1.391	1.380	1.417	1.415	1.308	1.315
C5C4	1.463	1.458	1.465	1.467	1.397	1.405
C6N1	1.371	1.365	1.388	1.380	1.330	1.336
C6C5	1.328	1.331	1.326	1.328	1.368	1.361
O2C2	1.188	1.196	1.191	1.201	1.319	1.314
O4C4	1.188	1.199	1.280	1.284	1.319	1.301
H1N1	0.993	1.001	0.992	0.998		
H3N3	0.997	1.010	0.994	1.004		
					0.997	1.011
						0.999
						1.009

H5C5	1.070	1.070	1.071	1.070	1.071	1.074	1.075	1.071	1.073	1.076
H6C6	1.074	1.074	1.073	1.074	1.076	1.076	1.071	1.075	1.072	1.071
H1O2					0.944	0.954	0.945	0.958	0.946	0.959
H3O4					0.945	0.962	0.945	0.963		
N1C2N3	113.6	114.8	114.6	115.7	127.2	126.9	125.3	126.3	125.0	124.4
C2N3C4	127.8	126.5	125.6	124.9	116.0	116.9	111.7	112.2	123.2	123.0
N3C4C5	113.8	114.8	115.2	115.4	123.1	121.7	124.1	122.2	112.0	113.2
C2N1C6	123.2	122.7	123.3	122.8	115.5	115.3	118.2	116.6	115.3	115.5
C4C5C6	119.2	118.9	118.7	118.3	114.7	115.5	116.5	116.8	119.4	118.8
N1C6C5	122.3	122.3	121.9	122.1	123.5	123.7	106.1	107.2	125.1	125.1
N1C2O2	122.8	122.3	122.6	121.9	117.5	118.8	116.6	117.1	121.3	121.8
N3C2O2	123.6	122.9	122.7	122.4	115.3	114.3	118.0	116.6	113.6	113.8
N3C4O4	120.7	121.1	115.4	115.5	117.7	119.7	115.9	118.4	120.0	120.7
C5C4O4	125.5	124.1	118.3	116.9	119.3	118.6	120.0	119.4	127.9	126.1
C2N1H1	115.6	115.7	115.0	115.4					128.1	127.5
C6N1H1	121.2	121.6	120.7	121.1						
C2N3H3	115.7	115.9	115.0	116.9						
C4N3H3	116.5	117.6	119.1	118.2						
C4C5H5	118.3	118.5	119.3	119.5	121.7	121.1	121.2	120.2	118.1	118.6
C6C5H5	122.5	122.7	122.0	122.2	123.6	123.5	121.9	121.9	122.6	122.8
N1C6H6	115.2	115.2	115.6	115.5	115.9	115.8	118.7	118.5	114.5	114.7
C5C6H6	122.5	122.5	122.5	122.3	120.6	120.5	121.2	121.7	120.4	122.5

C2O2H1		108.0	110.8	109.4	111.4	108.1	110.4	109.0	111.0
C4O4H3	0.0	0.7	7.0	-4.6	108.4	113.0	108.3	113.0	
N3C2N1C6	0.0	-1.0	0.8	-3.7	0.0	0.1	29.8	25.5	0.0
C4N3C2N1	0.0	0.8	-6.3	8.4	0.0	-0.6	4.7	9.2	0.0
C5C4N3C2	0.0	0.8	4.3	-5.1	0.0	0.4	-14.4	-16.2	0.0
C6C5C4N3	0.0	-0.2	-9.1	7.8	0.0	-0.4	-8.0	-8.9	0.0
C5C6N1C2	0.0	-0.3	0.0	2.8	-2.4	0.0	-47.9	-47.2	0.0
N1C6C5C4	0.0	0.0	179.2	-171.7	173.9	180.0	179.9	-153.0	39.8
O2C2N1C6	180.0	180.0	179.2	137.1	-133.0	180.0	179.4	-155.9	180.0
O4C4N3C2	180.0	180.0	179.0	175.5	-175.5	166.6	163.4	179.6	179.8
N3C2N1H1	180.0	180.0	179.0	166.1	-169.0	180.0	179.4	167.1	144.9
C5C4N3H3	180.0	180.0	179.9	-174.6	174.2	180.0	179.8	178.7	-151.1
H5C5C4N3	180.0	179.9	-176.0	175.8	180.0	180.0	176.8	179.1	-151.3
H6C6C5C4	180.0	179.9	0.0	0.0	1.3	2.7	0.8	0.0	24.6
H1O2C2N1					-1.8	-1.2	-3.3	0.2	-36.6
H3O4C4N3					0.0	0.0	0.0	0.0	8.1

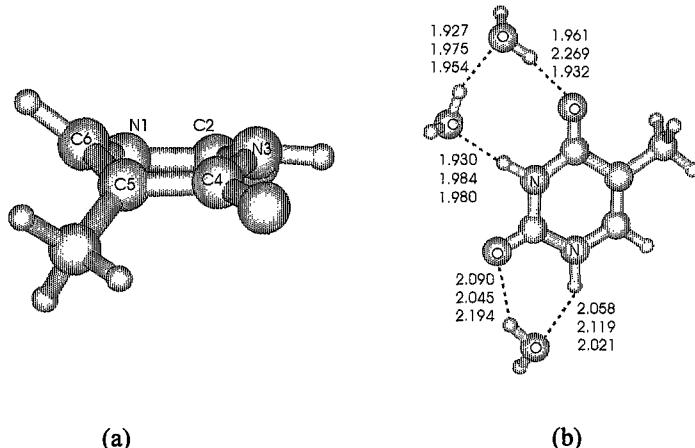
^aGround and excited state parameters are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The “+3W” represents the corresponding hydrated species with three water molecules.

However, the computed transitions of these compounds are generally within the observed ranges shown in Tables 12 and 13.

4.3.2 Geometries and Charges

The ground and excited state optimized geometries of uracil, thymine and their hydrated complexes are presented in Tables 14 and 15. All geometries were found to be minima at their respective potential energy surfaces. Although the ground state geometries of thymine and uracil are planar in the gas phase and under hydration, the lowest singlet $n\pi^*$ excited state geometry is slightly nonplanar, and the C4O4 bond, the length of which is increased by about 0.1 Å, lies appreciably out of the approximate ring plane. Such deformation is consistent with the fact that the excitation is localized at the C4O4 group. The geometry of thymine in the lowest singlet $\pi\pi^*$ excited state is highly nonplanar and adopts a boat type of structure in which the N1, C2, C4 and C5 atoms are approximately in a plane, while the N3 and C6 atoms are out of this plane (Fig. 6a). The $\pi\pi^*$ excited state of hydrated thymine was optimized at the 6-311⁺⁺G(d,p) basis set since the 6-311G(d,p) basis set gave convergence problems. The hydrated species was predicted to have a slightly more planar structure for thymine although it was of the boat type. The hydrogen bond lengths for the hydrated structure of thymine in the ground and excited states are shown in Fig. 6b. The lengths of hydrogen bonds involving the C4O4 group are increased in the $n\pi^*$ excited state. Similar results are found for the hydrated complex of uracil in the $n\pi^*$ state (Fig. 7a). The geometry optimization of uracil and its hydrated complex in the $\pi\pi^*$ excited state were not successful. As the geometry optimization of uracil and its hydrated form in the lowest singlet $\pi\pi^*$ excited state proceeded, large deformation in the ring occurred. In particular, the elongation and twisting of the C5C6 bond took place such that the H5 and H6 atoms attached to the C5 and C6 sites, respectively were found to be displaced on opposite sides to each other with respect to the approximate ring plane. Excitation energy corresponding to the $\pi\pi^*$ transition was also found to decreased continuously and at about 0.2 eV of transition energy, the calculations were aborted for the lack of convergence. On the basis of the CIS calculation, it has been earlier suggested that the lowest singlet $\pi\pi^*$ excited state of uracil may be dissociative due to the large twisting of the C5C6 bond of the molecule.⁵⁸ Further, the presence of the methyl group at the C5 position in thymine has been suggested to provide more stability to it in comparison to uracil in the excited state, although geometrical deformation in this case is also large.⁵⁸ The present results support

the earlier ones.⁵⁸ In jet-cooled studies, the geometrical deformation of the $\pi\pi^*$ excited state of thymine and uracil has been suggested to be responsible for the diffuseness of spectra of these compounds.^{43,57}



(a)

(b)

Figure 6. (a) Geometry of thymine in the singlet $S_2(\pi\pi^*)$ excited state, (b) hydrogen bond lengths of hydrated thymine in the ground, $S_1(\pi\pi^*)$ and $S_2(\pi\pi^*)$ states (from top to bottom indices, respectively). All parameters are at the 6-311G(d,p) basis set except the hydrated species in the $S_2(\pi\pi^*)$ excited state for which 6-311++G(d,p) basis set was used. Excited state numberings and assignments correspond to the isolated thymine.

The computed Mulliken and CHELPG charges of uracil and thymine in the ground and excited states are shown in Table 16. In the case of uracil, in going from the ground state to the vertical singlet $n\pi^*$ excited state, the O4 atom loses and the C4 and C6 atoms gain electronic charge. However, in the vertical singlet $\pi\pi^*$ excited state, the C4 and C6 atoms gain and the C5 atom loses electronic charge (Table 16). The CHELPG charge distributions is also modified in the relaxed $n\pi^*$ excited state as compared to the corresponding vertical state and such changes in charge distribution would be associated with nonplanarity of the molecule in the excited state (Table 16). In the case of the thymine, the C4 and C6 atoms gain while C5 atom loses electronic charges in going from the ground state to the vertical singlet $\pi\pi^*$ and $n\pi^*$ excited states (Table 16). Reduction in electronic charges at the O4 atom in the vertical singlet $n\pi^*$ excited state as compared to the ground and singlet vertical $\pi\pi^*$ states of thymine is also revealed. Further, the predicted difference of charge

distributions between the vertical and corresponding relaxed excited state would be associated with nonplanarity in the geometry of the corresponding excited states (Tables 14, 15). The reduction in the electronic charge at the O4 site of uracil and thymine in the singlet $n\pi^*$ excited state is in conformity with characterization of the state as due to excitation of the C4O4 lone pair electron.

Table 15. Ground and singlet excited state optimized bond lengths (\AA), bond angles ($^\circ$) and dihedral angles ($^\circ$) of isolated and hydrated thymine. Excited state numbering corresponds to isolated thymine.^a

parameters	S_0	$S_1(n\pi^*)$		$S_2(\pi\pi^*)$	
	+3W	+3W	+3W	+3W ^b	
N1C2	1.368	1.364	1.366	1.364	1.403
C2N3	1.371	1.367	1.367	1.358	1.371
N3C4	1.387	1.376	1.417	1.414	1.417
C4C5	1.472	1.466	1.474	1.477	1.441
C6N1	1.377	1.371	1.393	1.386	1.346
C5C6	1.328	1.331	1.326	1.328	1.439
O2C2	1.189	1.198	1.192	1.202	1.184
O4C4	1.190	1.201	1.279	1.282	1.196
H1N1	0.993	1.000	0.992	0.998	1.001
H3N3	0.997	1.010	0.994	1.004	0.997
C'5C5	1.502	1.502	1.504	1.504	1.497
H6C6	1.074	1.074	1.074	1.074	1.080
H51C'5	1.085	1.085	1.085	1.085	1.083
H52C'5	1.085	1.085	1.085	1.086	1.090
H53C'5	1.084	1.085	1.084	1.084	1.093
N1C2N3	113.3	114.5	114.3	115.4	113.9
C2N3C4	127.7	126.5	125.5	124.9	123.6
N3C4C5	114.9	115.9	116.1	116.3	112.5
C2N1C6	123.4	122.8	123.4	122.9	120.4
C4C5C6	117.6	117.2	117.0	116.7	116.9
N1C6C5	123.1	123.1	122.8	123.1	114.5
N1C2O2	123.2	122.7	123.1	122.3	121.0
N3C2O2	123.4	122.8	122.6	122.3	125.0
N3C4O4	120.6	121.0	114.9	115.0	119.4
C5C4O4	124.5	123.2	118.6	117.3	127.8
C2N1H1	115.7	115.8	115.1	115.5	113.7
C6N1H1	120.9	121.4	120.4	120.8	122.7
C2N3H3	115.8	116.0	115.1	117.0	114.2
C4N3H3	116.5	117.6	118.9	118.0	114.7
C4C5C'5	118.0	118.3	119.2	119.5	119.7

C6C5C'5	124.5	124.4	123.8	123.8	120.9	120.2
N1C6H6	114.7	114.8	115.1	115.0	118.9	119.6
C5C6H6	122.2	122.2	122.1	121.9	126.6	125.5
C5C'5H51	110.7	110.8	111.1	111.3	109.8	110.5
C5C'5H52	110.7	110.9	110.8	110.8	111.2	110.5
C5C'5H53	110.7	110.7	110.4	110.4	111.1	110.8
N1C6C5C4	0.0	0.0	-3.5	-3.4	41.3	-34.5
N3C2N1C6	0.0	0.8	-7.5	-5.0	5.5	-6.6
C4N3C2N1	0.0	-1.0	-0.3	-3.5	28.5	-24.9
C5C4N3C2	0.0	0.8	5.9	8.2	-24.8	24.0
C5C6N1C2	0.0	-0.3	9.9	8.7	-39.0	35.2
C6C5C4N3	0.0	-0.2	-3.7	-4.4	-10.8	6.4
O2C2N1C6	180.0	179.1	171.3	173.7	-171.1	170.8
C4N3C2O2	180.0	178.9	-179.2	177.8	-155.0	157.8
O4C4N3C2	180.0	179.2	-138.7	-134.6	148.9	-152.8
C6C5C4O4	180.0	179.7	139.5	137.5	176.1	176.9
C'5C5C4N3	180.0	179.9	175.2	174.7	151.3	164.0
N1C6C5C'5	180.0	179.9	177.6	177.5	-120.5	136.1
N3C2N1H1	180.0	180.0	-175.4	175.5	165.8	170.8
C5C6N1H1	180.0	179.5	177.1	178.7	162.5	-161.9
C5C4N3H3	180.0	179.0	-165.8	-169.0	-172.7	176.7
H3N3C2N1	180.0	178.8	171.6	173.6	176.5	178.1
H6C6N1C2	180.0	179.6	-169.2	169.9	138.3	150.1
H6C6C5C4	180.0	179.9	175.5	175.0	-135.7	151.1
H51C'5C5C4	59.4	2.4	64.0	-0.1	10.0	-5.9
H51C'5C5C6	-120.6	177.5	-117.1	178.4	171.3	171.3
H52C'5C5C4	-59.4	11.3	-55.7	-83.0	130.6	7.0
H52C'5C5C6	120.6	169.7	123.2	134.6	-68.0	-169.5
H53C'5C5C4	180.0	-2.3	-176.1	-2.5	-111.0	-4.6
H53C'5C5C6	0.0	177.0	2.8	173.2	50.3	168.6

^aGround and singlet excited state parameters are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The S₁, S₂ represent singlet excited states in the ascending energy order (Table 13), ^bAt the 6-311++G(d,p) basis set. The “+3W” represents the corresponding hydrated species with three water molecules.

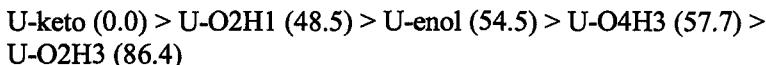
Table 16. Ground and singlet excited state charges at the atomic sites of uracil and thymine.^a

Uracil (keto form)				Thymine						
S ₀	S ₁ (nπ*)		S ₂ (ππ*)	S ₀	S ₂ (nπ*)		S ₂ (ππ*)		Vert	Opt
	Vert	Opt	Vert		Vert	Opt	Vert	Opt	Vert	Opt
Mulliken										
N1	-0.538	-0.551	-0.546	-0.421	-0.537	-0.550	-0.545	-0.430	-0.421	
C2	0.704	0.715	0.710	0.670	0.701	0.711	0.706	0.672	0.680	
N3	-0.569	-0.564	-0.559	-0.571	-0.567	-0.565	-0.562	-0.572	-0.540	
C4	0.612	0.374	0.318	0.537	0.638	0.398	0.336	0.570	0.584	
C5	-0.381	-0.273	-0.288	-0.325	-0.454	-0.316	-0.325	-0.402	-0.354	
C6	0.273	0.168	0.206	0.159	0.296	0.188	0.227	0.176	0.132	
O2	-0.470	-0.479	-0.484	-0.443	-0.475	-0.484	-0.489	-0.446	-0.443	
O4	-0.452	-0.162	-0.144	-0.455	-0.453	-0.170	-0.149	-0.471	-0.464	
H1	0.269	0.259	0.262	0.280	0.266	0.257	0.259	0.277	0.254	
H3	0.276	0.262	0.267	0.272	0.274	0.260	0.266	0.269	0.264	
R5	0.136	0.126	0.129	0.154	-0.136	-0.167	-0.163	-0.148	-0.186	
H6	0.140	0.125	0.129	0.143	0.128	0.114	0.118	0.133	0.144	
H51					0.117	0.113	0.114	0.141	0.134	
H52					0.117	0.113	0.114	0.141	0.110	
H53					0.085	0.098	0.093	0.090	0.106	
CHELPG										
N1	-0.566	-0.411	-0.191	-0.438	-0.577	-0.436	-0.259	-0.433	-0.510	
C2	0.959	0.852	0.609	0.850	0.945	0.830	0.641	0.846	0.975	
N3	-0.762	-0.649	-0.202	-0.576	-0.743	-0.609	-0.244	-0.590	-0.820	
C4	1.000	0.522	-0.103	0.640	0.875	0.379	-0.109	0.535	0.773	
C5	-0.586	-0.452	-0.029	-0.268	-0.222	-0.084	0.233	0.073	-0.327	
C6	0.214	-0.035	-0.208	0.044	0.054	-0.166	-0.287	-0.109	0.174	
O2	-0.665	-0.668	-0.635	-0.626	-0.666	-0.666	-0.644	-0.625	-0.640	
O4	-0.655	-0.235	-0.075	-0.598	-0.621	-0.204	-0.088	-0.583	-0.601	
H1	0.347	0.309	0.263	0.366	0.360	0.326	0.286	0.372	0.340	
H3	0.379	0.396	0.278	0.343	0.382	0.394	0.296	0.353	0.417	
R5	0.198	0.221	0.121	0.159	-0.193	-0.174	-0.328	-0.259	0.132	
H6	0.137	0.150	0.172	0.104	0.172	0.182	0.198	0.127	0.080	
H51					0.086	0.079	0.130	0.105	0.014	
H52					0.086	0.079	0.086	0.105	-0.008	
H53					0.062	0.070	0.089	0.083	0.001	

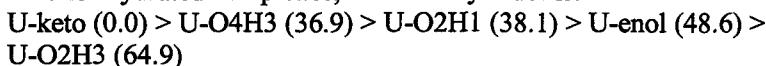
^aGround and singlet excited state charges are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The "Vert" represents the vertical and "Opt" represents the optimized state, R5=H5 for the uracil and C'5 for the thymine.

4.3.3 Evidence for Phototautomerism in Uracil (Thymine)²⁷⁵

Recently, Morsy et al.¹⁸⁸ have performed detailed spectroscopic studies of aqueous solutions of thymine at different pH. These authors¹⁸⁸ have found that the fluorescence peak for aqueous solutions of thymine at a neutral pH lies near 325 nm when excited in the range of 260-270 nm. A much stronger band of fluorescence with a peak near 405 nm is observed with the excitation of the sample at 295 or 300 nm. These observations have been explained in terms of keto-enol tautomerism; the peak of fluorescence near 325 nm is explained due to the emission from the keto tautomer, while the 405 nm fluorescence has been attributed to an emission from the enol form of thymine which absorbs in the longer wavelength region. The concentration of the enol form is suggested to be about 2%.¹⁸⁸ The phenomenon of phototautomerism in uracil has been recently theoretically investigated²⁷⁵ and the results of such a study can be applied to thymine as well. The geometries of different tautomers of uracil were optimized at the HF/6-311G(d,p) level of theory. The effects of aqueous solvation were considered using the supermolecule approach in which three water molecules were complexed with the uracil tautomers, and their geometries were also optimized at the HF/6-311G(d,p) level. The selection of three water molecules for hydration was based on an experimental observation by Chahinian et al.¹⁷³ using Overhauser spectroscopy suggesting that the first solvation shell of uracil indeed includes three water molecules. The relative stability of the isolated tautomers follows the order:



While for hydrated complexes, the stability order is:



Where the numbers shown in the parentheses are the relative energies in kJ/mol. The nomenclature for the different tautomers are such that U-enol represents the dienol tautomer, while U-OmHn represents the mono enol tautomer in which the Hn hydrogen atom attached to the Nn site of the keto form of the uracil ring is attached to the Om carbonyl oxygen atom (Fig. 1d). Low-lying vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies of different tautomers of uracil in gas phase and after hydration are presented in Table 17, while the corresponding energy level diagrams are shown in Fig. 8. The

computed geometrical parameters are shown in Table 14, while the hydrated complexes and hydrogen bond lengths are shown in Fig. 7.

Table 17. Vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies (ΔE , eV), oscillator strengths (f), transition moment directions ($\Phi, {}^\circ$) and dipole moments (μ , Debye) of the U-enol, U-O2H1, U-O4H3 and U-O2H3 tautomers of uracil and their hydrated forms at the CIS/6-311G(d,p)//HF/6-311G(d,p) level.²⁷⁵

Isolated					Hydrated				
ΔE	f	Φ	μ^a	ΔE^b	ΔE	f	Φ	ΔE^b	
U-enol									
<i>$\pi\pi^*$ Transitions</i>									
6.65	0.180	-4	1.08	4.79	6.62	0.213	-5	4.77	
7.60	0.076	-88	1.75	5.47	7.63	0.081	-74	5.49	
<i>$n\pi^*$ Transitions</i>									
7.04	0.007	-	1.61	5.07	7.30	0.006	-	5.26	
U-O2H1									
<i>$\pi\pi^*$ Transitions</i>									
6.36	0.316	-27	3.09	4.58	6.40	0.342	-23	4.61	
8.25	0.276	39	3.43	5.94	8.12	0.281	41	5.85	
<i>$n\pi^*$ Transitions</i>									
6.75	0.000	-	1.31	4.84	7.02	0.000	-	5.05	
U-O4H3									
<i>$\pi\pi^*$ Transitions</i>									
6.27	0.190	26	4.37	4.51	6.46	0.202	23	4.65	
8.33	0.208	-65	5.89	6.00	8.19	0.123	-77	5.90	
<i>$n\pi^*$ Transitions</i>									
7.08	0.002	-	2.39	5.10	7.26	0.002	-	5.23	
U-O2H3									
<i>$\pi\pi^*$ Transitions</i>									
6.91	0.199	-14	6.08	4.98	6.88	0.222	-12	4.95	
8.11	0.403	10	6.71	5.84	8.02	0.331	11	5.77	
<i>$n\pi^*$ Transitions</i>									
6.04	0.000	-	3.45	4.35	6.42	0.000	-	4.62	

^aGround state dipole moments of U-enol, U-O2H1, U-O4H3 and U-O2H3 tautomers at the HF/6-311G(d,p) level are 1.30, 3.45, 5.19 and 6.71 Debye, respectively, ^bscaled (scaling factor 0.72) excitation energies.

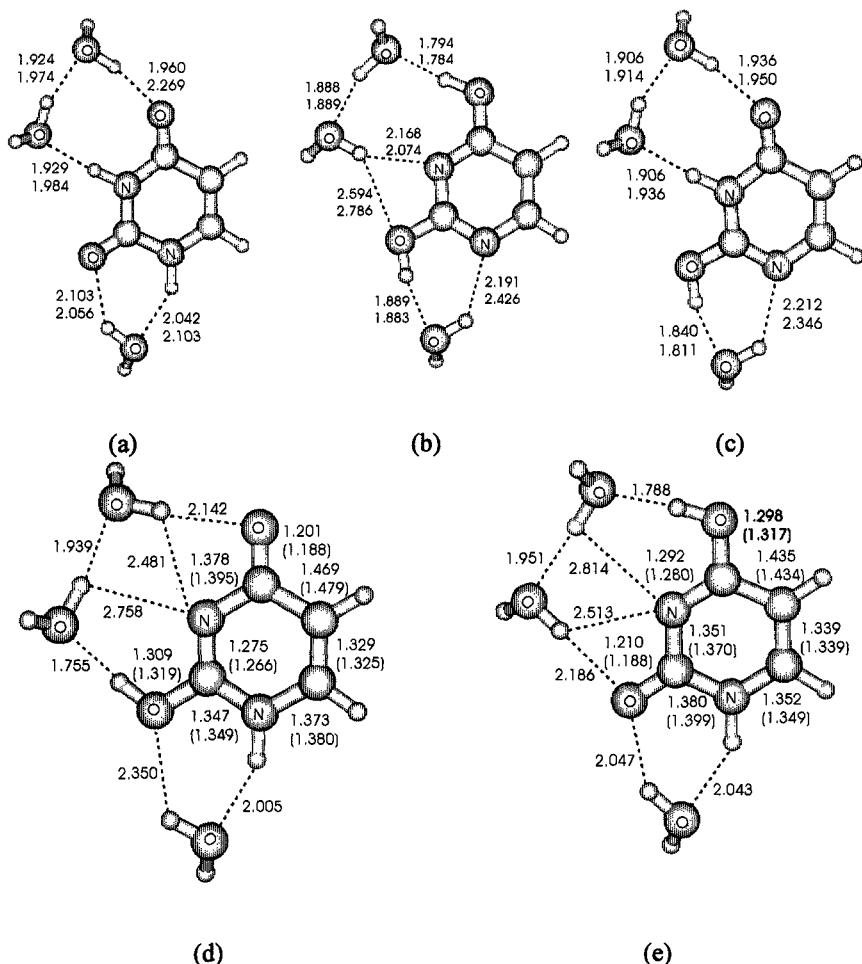


Figure 7. Hydrogen bond lengths of hydrated uracil tautomers: (a) U-keto form in the ground (top indices) and lowest singlet $n\pi^*$ excited state (bottom indices), (b) U-enol tautomer in ground (top indices) and the lowest singlet $n\pi^*$ excited state (bottom indices), (c) U-O2H1 tautomer in ground (top indices) and the lowest singlet $n\pi^*$ excited state (bottom indices). Bond lengths of hydrated tautomers of uracil in the ground state: (d) U-O2H3 tautomer and (e) U-O4H3 tautomer. The numbers in parentheses in (d) and (e) correspond to corresponding isolated species. Excited state assignments correspond to the corresponding isolated tautomers.

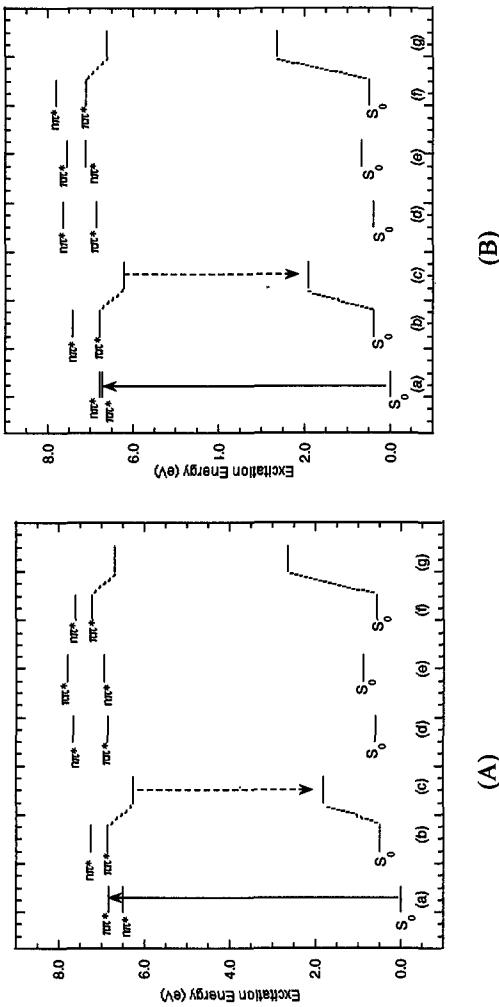


Figure 8. Energy level diagrams of uracil tautomers: (A) isolated form and (B) hydrated form. In the case of isolated tautomers (a), (b), (c), (d), (e) and (f) represent vertical states of U-keto, U-O2H1, U-O4H3, U-O4H3 and U-enol tautomers, respectively. The optimized state shown in (c) corresponds to the U-O2H1 tautomer while in (g) corresponds to the U-enol tautomer. In the case of hydrated tautomers (a), (b), (c) and (f) correspond to vertical states of U-keto, U-O2H1, U-O4H3, U-O4H3 and U-enol tautomers, respectively. The optimized state shown in (c) corresponds to the U-O2H1 tautomer while in (g) corresponds to the U-enol tautomer. The upward arrows correspond to the absorption while downward arrows correspond to the emission.

The energy level diagrams suggest that, among the vertical lowest singlet $\pi\pi^*$ states of different tautomers, the state corresponding to the U-O2H1 and U-O4H3 tautomers are higher in energy by about 0.05 and 0.11 eV, respectively, than the corresponding state of the uracil (keto) form. The corresponding states of the U-enol and U-O2H3 tautomers are appreciably higher (Fig. 8). Under hydration, the lowest singlet vertical $\pi\pi^*$ excited states of the U-O2H1 and U-O4H3 tautomers become closer to the corresponding state of uracil (keto form) (Fig. 8). Therefore, there may be a strong interaction among the lowest vertical singlet $\pi\pi^*$ states of the keto, U-O2H1 and U-O4H3 tautomers of uracil. Thus, upon electronic excitation of uracil, the energy transfer from the lowest singlet $\pi\pi^*$ state of uracil to the corresponding states of U-O2H1 and U-O4H3 forms may take place. In other words, there is strong probability that uracil will phototautomerize to the U-O2H1 and U-O4H3 enol tautomers through a transfer of energy and through necessary structural modifications. However, the rate of formation of these enol tautomers will also depend on the lifetimes of the $S_1(\pi\pi^*)$ states of uracil (keto), U-O2H1 and U-O4H3 tautomers. In the case of thymine, the excited state lifetime of the keto form is found to be much smaller than the corresponding lifetime of the enol tautomers.¹⁸⁸ It would lead to the formation of a small amount of the enol tautomers. It is also interesting to note that the scaled vertical first singlet $\pi\pi^*$ transition energies for the U-O2H1 and U-O4H3 tautomers are 4.58 and 4.51 eV, respectively (the corresponding values for hydrated forms being 4.61 and 4.62 eV, respectively), which are close to the 295 nm (4.20 eV) value. The excitation of thymine at 295 nm in a neutral aqueous solution gives rise to a peak of fluorescence near 405 nm, which was assigned as due to the emission from the enol tautomer.¹⁸⁸ The relaxed lowest singlet excited $\pi\pi^*$ state for the U-O2H1 tautomer in gas phase and after hydration is also shown in the energy level diagrams. It is clear that fluorescence due to enol tautomers would originate from the relaxed lowest singlet $\pi\pi^*$ states of the U-O2H1 and U-O4H3 tautomers. The computed energy difference (4.29 eV) corresponds to the vertical transition between the relaxed lowest singlet $\pi\pi^*$ excited state of the hydrated U-O2H1 tautomer and corresponding ground state. The corresponding scaled value amount to 3.09 eV which is in good agreement with the 405 nm (3.06 eV) fluorescence observed in thymine after excitation at 295 nm.¹⁸⁸ Since we were not able to optimize the lowest singlet $\pi\pi^*$ excited state of the U-O4H3 tautomer, it is difficult to judge its contribution to fluorescence. The reason for the strong fluorescence peak of the enol tautomer of thymine as compared to that of the normal form may be due to the longer

excited state lifetime of the former species.¹⁸⁸ A similar situation is observed for adenine and 2-aminopurine where adenine has a very short lifetime and a very weak fluorescence, while 2-aminopurine has a longer lifetime and a very strong fluorescence.^{35,52,122}

4.4 Cytosine

4.4.1 Electronic Transitions

4.4.1.1 Experimental Data

In general, features of the cytosine spectrum are found to be solvent dependent.^{23,224} The aqueous absorption spectrum of cytosine shows two broad peaks near 266 and 197 nm (4.66, and 6.29 eV, respectively) and two weak peaks or shoulders near the 230 and 212 nm (5.39 and 5.85 eV, respectively).^{23,214,224,226,276-280} The first absorption band of cytidine in aqueous media is red-shifted to about 271 nm (4.57 eV) compared to the corresponding band of cytosine.^{214,278} A significant red-shift in this band is also observed in the absorption spectrum of 3-methylcytosine in aqueous media which shows a peak near 289 nm (4.29 eV).²⁷⁹ The 197 nm band of cytosine is suggested to be composite corresponding to the 202 nm (6.14 eV) and 189 nm (6.56 eV) transitions observed in CD spectra.^{226,281,282} The CD spectra of different cytosine derivatives reveal that even though the first CD transition agrees well with the corresponding transition in absorption spectra, the second and third CD bands are significantly red-shifted.^{226,281} Table 18 contains the summary of experimental transitions of cytosine and its derivatives in different environments. The Clark's group have assigned the transition moment direction for the first three transitions explicitly and gave two possible values for the fourth transition of cytosine with the help of polarized reflection spectroscopy of single crystals of cytosine monohydrate.²⁷⁷ The predicted angle of 6° with respect to the N1C4 direction (Fig. 1c) for the first transition²⁷⁷ is in accordance with the 9° value of Callis and Simpson²⁸³ and the 10° value of Lewis and Eaton.²⁸⁴ These authors²⁷⁷ have suggested that the predicted discrepancy in the transition moment direction obtained in LD studies of cytosine and related compounds^{219,285} dissolved in stretched poly(vinyl alcohol) films can be removed if the orientation axis for the stretched film is rotated. Under such a rotation, the LD results were found to be in agreement with the crystal results.²⁷⁷ The Clark and coworkers²⁷⁷ have

suggested that the transition moment directions for the II, III and IV bands would be -46°, 76° and -27° or 86°, respectively with respect to the N1C4 direction (Fig. 1c). Such a conclusion regarding the transition moment directions for the second and third transitions was based on the study of polarized fluorescence of 5-methylcytosine²⁶⁸ which resulted in the angle between the transition moments of the I and II bands as 40±15°. Similar results were also found for cytosine monophosphate.²⁸⁶

Table 18. Summary of experimental transition energies (ΔE , eV) of cytosine and its derivatives. The f represents oscillator strength, and Φ represents the transition moment direction (°) according to the DeVoe-Tinoco convention (Figure 1c).

Transitions	References					
Absorption Spectra						
ΔE	4.66	5.39	5.85	6.29		Cytosine, water [277]
f	0.14	0.03	0.13	0.36		
Φ^a	6	-46	76	-27 or 86		
ΔE	4.64		6.31			Cytosine, water [214]
ΔE	4.57	5.39	6.26			Cytidine, water [214]
ΔE	4.48	5.23	6.08	6.63		Cytosine, TMP [224]
ΔE	4.59	5.28	5.74	6.26		dCMP, water [226]
ΔE	4.64	5.21	5.83	6.46		Cytosine, water [279]
f	0.096	0.100	0.211	0.639		
ΔE	4.57	5.34	5.77	6.26		Cytidine, water [278]
ΔE	4.57			6.17		Cytosine, sublimed film [276]
ΔE	4.54	5.40	6.07	6.67	7.35	Cytosine, sublimed film [280]
f	0.058	0.073	0.115	0.072	0.072	
LD Spectra						
ΔE	4.63	5.17				Cytosine, polymer film [219]
Φ	25±3	6±4				
	or	or				
	-46±4	-27±3				
CD Spectra						
ΔE	4.59	5.27	5.74	6.14	6.56	7.38
ΔE	4.59	5.02	5.64	6.36		dCMP, water [226]
ΔE	4.59	5.17	5.64	6.36		Cytosine nucleosides [281]
						Cytidine ^b [281]

^aBased on polarized spectra of cytosine crystal [277], TMP: trimethylphosphate, for details see the relevant reference; ^bbased on the CD and absorption measurements of cytosine nucleosides in different solvents (water, acetonitrile, dioxane, 1,2-dichloroethane).

The stretched film work^{219,285,287} indicated that the angle between the transition moments of the I and III bands is larger than that between the I and II bands. Furthermore, the Clark's group²⁷⁷ have questioned the selection of the axis of orientation by Bott and Kurucsev²⁸⁵ and suggested that the orientation axis should be -9° in contrast to the -38° selected by Bott and Kurucsev. Under such circumstances, the predicted values by Bott and Kurucsev²⁸⁵ have been shown to be in accordance with the results obtained by the Clark's group.²⁷⁷ With regard to the nπ* transitions, there is definite experimental^{520,277,278,288} and theoretical^{58,130,136} evidence to suggest the existence such transition in the 5.3 eV (232 nm) region in cytosine. Zaloudek et al.²⁷⁷ have suggested the existence of another nπ* transition near 5.6 eV (220 nm).

4.4.1.2 Theoretical Data

Petke et al.¹²⁹ have computed electronic transition energies and transition moment directions at the MRCI and RPA levels of theory using the experimental X-ray crystallographic geometry of cytosine.²⁰⁰ In this case also, the computed transition energies were appreciably higher than the experimental ones, and linear scaling was needed to make them comparable to the experimental data. The lowest two singlet nπ* transitions are characterized by contributions from both of the N3 and carbonyl oxygen lone pairs. The CASSCF and CASPT2 calculations for electronic transitions of cytosine were carried out by Fulscher and Roos¹³⁰ using an average experimental geometry²⁷³ employing the ANO type basis set. Although the CASSCF transition energies were higher than the experimental values, the CASPT2 transition energies were found to be in agreement with the latter. Both the CASSCF and the CASPT2 methods predicted the first nπ* transition to be the second singlet valence excitation of cytosine. However, in the computations of nπ* transitions, the carbonyl oxygen lone pair orbital was not included in the active space.¹³⁰

As discussed earlier, cytosine may exist in different tautomeric forms in various environments; however, in an aqueous medium, the N1H tautomer would be most stable.^{59-66,195-203} In the electronic transition calculations on cytosine by Petke et al.¹²⁹ and by Fulscher and Roos,¹³⁰ only the keto-N1H tautomer was considered. We have optimized the ground state geometries of the keto-N1H, keto-N3H, enol and imino tautomeric forms of cytosine and their complexes with three water molecules at the HF/6-311G(d,p) level.²⁸⁹ The supermolecule approach considering three water molecules was used to

model the effects of hydration. The relative total energies of these tautomers (keto-N1H, keto-N3H, enol and imino) with respect to the keto-N1H form in the gas phase were found to be 0.0, 29.41, -5.38 and 3.89 kJ/mol, while those for the hydrated complexes were 0.0, 23.6, 11.29 and 19.16 kJ/mol, respectively. Therefore, the computed results are in agreement with the experimental and theoretical observations that in gas phase, the enol form is most stable, while in aqueous media, the keto-N1H form is most stable.^{195,196,199,201-203}

The computed vertical singlet $\pi\pi^*$ and $n\pi^*$ transition energies, transition moment directions and dipole moments of cytosine tautomers in the gas phase and after hydration²⁸⁹ at the CIS/6-311G(d,p)//HF/6-311G(d,p) level are shown in Tables 19 and 20 along with some experimental data and CASPT2/CASSCF results.¹³⁰ Because of uncertainty in assignment and possibly significant basis set dependence, we are not giving $\pi\sigma^*$ /Rydberg type transitions in the table. For all tautomers, the lowest singlet transition is predicted to be of $\pi\pi^*$ type, and the second singlet transition is of $n\pi^*$ type. The first $n\pi^*$ transition of the keto-N1H and keto-N3H tautomers are characterized mainly by excitation of the N1 and N3 lone pair electron, respectively while the second $n\pi^*$ transition of the tautomers are localized mainly at the carbonyl group. However, these results are not completely in agreement with the findings of Petke et al.¹²⁹ which suggest the participation of lone pairs of both of the N3 atom and carbonyl oxygen in the lowest two singlet $n\pi^*$ transitions of the keto-N1H tautomer of the cytosine. In the case of the enol tautomer, both of $n\pi^*$ transitions are localized at the N1 and N3 sites, while in the imino tautomer such transitions are localized at the carbonyl oxygen and imino nitrogen.

The data shown in Table 19 suggest that the observed spectral transitions of cytosine in water can be explained in terms of the scaled transition energies of the hydrated keto-N1H tautomer within the accuracy of about 0.2 eV except the second transition. However, the error becomes smaller if a comparison is made between the computed transitions of the hydrated keto-N1H tautomer and the observed transitions of the cytosine sublimed film.²⁸⁰ Such a difference between the computed and experimental data is not unexpected since the general features of the cytosine spectrum are found to be solvent dependent.^{23,224} It appears that the absorption peak of cytosine in water^{277,279} near 212 nm (5.85 eV) is shifted to 6.07 eV in the sublimed film experiment.²⁸⁰ The suggested splitting of the 197 nm (6.29 eV) transition in the form of 202 and 189 nm (6.14 and 6.56 eV, respectively) bands found in CD experiments^{226,281,282} is neither revealed in the CIS calculations nor in the

Table 19. Vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies (ΔE , eV), oscillator strengths (f), transition moment directions ($\Phi, {}^\circ$) and dipole moments (μ , Debye) of the keto-N1H tautomers of cytosine in the isolated and hydrated forms at the CIS/6-311G(d,p)//HF/6-311G(d,p) level.²⁸⁹

CIS						Experimental Data ^a					
Isolated			Hydrated			CASPT2/CASSCF ^b			Abs ^c		
ΔE	f	Φ	μ^c	ΔE^d	Φ	ΔE	f	Φ	$\Delta E^1/\Delta E^2/f/\Phi/\mu$	$\Delta E/f$	$\Delta E/f/\Phi$
<i>$\pi\pi^*$ Transitions</i>											
6.32	0.171	24	5.60	4.55	6.46	0.221	15	4.65	4.39/5.18/0.061/60.6/4.7	4.64 /0.096	4.66/0.146
7.84	0.382	-20	6.73	5.64	7.75	0.365	-19	5.58	5.36/6.31/0.108/-1.5/7.0	5.21/0.100	5.39/0.03/-46
8.42	0.686	-48	7.06	6.06	8.41	0.649	-49	6.06	6.16/7.30/0.863/-39.7/6.2	5.83/0.211	5.85/0.137/6
9.49	0.167	17	6.74	6.83	9.23	0.293	19	6.65	6.74/7.82/0.147/14.9/9.3	6.46 /0.639	6.29 /0.36/X
<i>$n\pi^*$ Transitions</i>											
6.97	0.003	-	4.28	5.02	7.46	0.007	-	5.37	5.00/5.13/0.005/-4.7		
7.50	0.001	-	3.73	5.40	7.93	0.000	-	5.71	6.53/7.14/0.001/-6.4		

^aAbs¹: absorption of cytosine in water [279]; Abs²: absorption of cytosine in water; Φ values are based on polarized spectra of cytosine crystal and X= -2 or 86 [277]; Abs^c: absorption of cytidine in water [278]; Abs^d: absorption of sublimed cytosine [280], ^b ΔE^1 represents CASPT2, and ΔE^2 represents CASSCF transition energies [130], ^cground state dipole moment at the HF/6-311G(d,p) level is 6.98 Debye, ^dscaled (scaling factor 0.72) excitation energies.

CASPT2/CASSCF computations.¹³⁰ The computed nπ* transitions of the keto-N1H tautomer lie near 5.02 and 5.4 eV. The corresponding transitions for the hydrated species are near 5.37 and 5.71 eV (Table 19). The existence of the nπ* transition near 5.3 and 5.6 eV in cytosine has been indicated in different experimental and theoretical investigations.^{58,130,136,220,277,278,288}

Table 20. Vertical singlet ππ* and nπ* excitation energies (ΔE, eV), oscillator strengths (f), transition moment directions (Φ, °) and dipole moments (μ , Debye) of the keto-N3H, imino and enol tautomers of cytosine and their hydrated forms at the CIS/6-311G(d,p)//HF/6-311G(d,p) level.²⁸⁹

CIS								Experimental Data ^a		
Isolated				Hydrated				Abs		
ΔE	f	Φ	μ^b	ΔE ^c	ΔE	f	Φ	ΔE ^c	ΔE	f
keto-N3H Tautomer : ππ* Transitions										
5.83	0.308	-2	8.12	4.20	5.90	0.351	-6	4.25	4.29	0.198
7.96 ^d	0.152	-87	5.57	5.73	7.89	0.303	-86	5.68	5.47	0.275
nπ* Transitions										
6.24	0.004	-	5.36	4.49	6.66	0.003	-	4.80		
7.31	0.001	-	5.53	5.26	7.84	0.021	-	5.64		
Imino Tautomer : ππ* Transitions										
6.62	0.560	-7	4.74	4.77	6.57	0.250	-6	4.73	4.64	0.17
8.29	0.143	17	2.97	5.97	8.18	0.196	15	5.89	5.59	0.39
Enol Tautomer : ππ* Transitions										
6.53	0.202	-9	3.55	4.70	6.47	0.250	-12	4.66		
7.41	0.088	47	3.38	5.34	7.47	0.130	35	5.38		
nπ* Transitions										
7.00	0.008	-	1.12	5.04	7.32	0.007	-	5.27		
7.92	0.001	-	3.14	5.70	8.20	0.002	-	5.90		

^aAbs: absorption transitions of cytosine in water; experimental data for keto-N3H and imino tautomers correspond to 3-methylcytosine and 3-methylcytidine, respectively, obtained in aqueous solution at pH 11 [279], ^bground state dipole moments of keto-N3H, imino and enol tautomers at the HF/6-311G(d,p) level are 8.15, 5.12 and 3.34 Debye, respectively, ^cscaled (scaling factor 0.72) excitation energies, ^dRydberg contamination.

A comparison of the lowest singlet $\pi\pi^*$ transition of cytosine tautomers presented in Tables 19 and 20 suggests that, with respect to the keto-N1H tautomer, the transition in the keto-N3H tautomer is appreciably red-shifted for both the isolated and hydrated forms. While in the case of the enol and imino tautomers, the isolated forms show a blue-shift with respect to the isolated form of the keto-N1H tautomer; for hydrated species, the imino form shows a slight blue-shift, and the enol form does not show any shift (Tables 19 and 20). The observed red-shift in the first $\pi\pi^*$ transition of the keto-N3H tautomer is in accordance with the experimental fact that the absorption spectrum of 3-methylcytosine in aqueous media shows a significant red-shift with a peak near 289 nm (4.29 eV) compared to the corresponding transition of cytosine observed near 266 nm (4.66 eV).²⁷⁹ The observed transitions of 3-methylcytosine near 4.29 and 5.47 eV can be explained in terms of the computed transitions of hydrated forms of the keto-N3H tautomer near 4.25 and 5.68 eV, respectively (Table 20). The computed transitions of the imino form can be compared with the observed transitions of 3-methylcytidine in a water solution.²⁷⁹ Table 20 shows that the observed transition near 4.64 and 5.59 eV of 3-methylcytidine can be explained in terms of the computed transitions of the imino form of cytosine within an error of about 0.3 eV. Therefore, it appears that the absorption spectra of aqueous solutions of cytosine would be mainly dominated by the keto-N1H tautomer and contributions from other tautomers will not be significant.

4.4.2 Geometries and Charges

The ground and excited state optimized geometrical parameters of the cytosine tautomers and their hydrated forms are shown in Table 21. The computed hydrogen bond lengths in the ground and excited states of the hydrated complexes are shown in Fig. 9. All geometries were found to be minima at their respective potential energy surfaces. The geometry optimizations of the lowest singlet $\pi\pi^*$ excited state of the keto-N1H and keto-N3H tautomers and their hydrated forms at the CIS level of the theory¹³⁸ gave convergence problems with the 6-311G(d,p) and other basis sets. However, geometry optimizations were successful when the 4-31G basis set was used for all atoms except the amino nitrogen for which we employed the 6-311+G(d) basis set.²⁸⁹ Such a combination of basis sets was found to be suitable for reproducing the ground state amino group nonplanarity at the HF level and for optimizing the lowest singlet $\pi\pi^*$ and $n\pi^*$ excited state geometries of the keto-N1H tautomer of cytosine at the CIS level.⁵⁸ Further, in the case of the keto-N3H tautomer

Table 21. Ground and excited state optimized bond lengths (\AA), bond angles ($^\circ$) and dihedral angles ($^\circ$) of the keto-N1H, keto-N3H and enol tautomers of cytosine and their hydrated forms^a₂₈₉

parameters	keto-N1H		keto-N3H		enol	
	S ₀	S ₁ ($\pi\text{-}\pi^*$) ^b	S ₀	S ₁ ($\pi\text{-}\pi^*$) ^c	S ₀	S ₁ ($\pi\text{-}\pi^*$) ^d
C2N1	1.403	1.388	+3W	1.375	+3W	+3W
N3C2	1.362	1.352	1.336	1.355	1.366	1.352
C4N3	1.294	1.309	1.348	1.399	1.389	1.410
C5C4	1.445	1.444	1.392	1.450	1.447	1.346
C6N1	1.347	1.348	1.350	1.398	1.403	1.290
C6C5	1.338	1.338	1.426	1.337	1.338	1.411
O2C2	1.191	1.206	1.218	1.182	1.202	1.189
N4C4	1.347	1.330	1.386	1.407	1.395	1.358
H1N1	0.994	1.000	0.995	0.994	1.000	0.996
H3N3						
O2H2						
H5C5	1.071	1.071	1.072	1.074	1.075	1.071
H6C6	1.074	1.074	1.068	1.073	1.072	1.079
H41N4	0.991	0.991	1.000	1.005	1.002	0.994
H42N4	0.993	1.003	0.999	0.998	1.002	0.993
N1C2N3	116.4	117.7	117.6	108.6	110.1	116.4
C2N3C4	120.5	120.8	121.4	119.8	118.8	124.8
N3C4C5	124.0	122.4	121.0	110.8	110.2	117.8
C2N1C6	123.2	122.2	119.5	121.8	121.2	118.4

H2O2C2N1	-179.7	-179.4	-173.9	-154.6	-154.3	179.8	-179.6	171.8	-140.3	177.3	169.1	-179.7	179.8	-159.9	153.7
H5C5C4N3	180.0	-179.9	158.4	179.2	179.7	180.0	-179.5	174.7	-176.5	146.8	-170.3	180.0	-179.1	176.2	
H6C6C5C4	-8.2	-4.0	-14.9	32.9	18.6	24.7	-12.8	29.8	-7.0	49.4	8.4	-11.8	2.5	-6.7	0.6
H41N4C4N3	172.8	176.2	176.1	166.9	178.2	157.3	168.2	145.0	177.6	134.8	171.6	169.4	177.5	174.0	-178.0
H41N4C4N5	-167.7	-176.2	-142.7	-91.8	-109.0	-167.6	-171.1	-165.1	-157.3	-179.2	163.6	-164.7	180.0	-167.3	173.7
H42N4C4N3	13.3	4.1	43.3	68.4	54.2	14.3	9.9	9.7	27.4	5.0	-16.4	16.5	0.0	13.4	7.7

^aGround and singlet excited state parameters are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively. The S₁, S₂ represent singlet excited states in the ascending energy order (Tables 19 and 20). ^bAt the 4-31G and 6-311+G(d) mixed basis sets, see text for detail.^cUsing the read-window (RW) option and 4-31G and 6-311G and 6-311+G(d) mixed basis sets, see text for detail [289]. The “+3W” represents the corresponding hydrated species with three water molecules.

and its hydrated form, the molecular orbitals from 29 to 105 and 44 to 144, respectively were used in the CI calculations.²⁸⁹ Unfortunately, geometry optimizations of the lowest singlet $\pi\pi^*$ excited states of the imino tautomer, its hydrated form and hydrated form of the keto-N1H tautomer were not successful. As the geometry optimizations proceeded, the excitation energy went down to about 0.02 eV and calculation aborted due to the lack of the convergence.

For the keto-N1H tautomer in gas phase, the amino group is pyramidal both in the ground and lowest singlet $\pi\pi^*$ excited state ($S_1(\pi\pi^*)$). The ground state ring geometry is planar while the $S_1(\pi\pi^*)$ excited state ring geometry is nonplanar mainly around the N1C6C5C4 fragment. In the lowest singlet $n\pi^*$ excited state ($S_2(n\pi^*)$), the N3 atom lies appreciably out-of-plane, and this deformation is similar to that obtained earlier using smaller basis set.⁵⁸ The amino group in this state is considerably rotated. It is possible that the out-of-plane deformation of the N3 atom in the $S_2(n\pi^*)$ state may be due to the fact that the state is mainly characterized by the excitation of N3 lone pair and contribution from the carbonyl oxygen is negligible. This consideration is also supported by the fact that as compared to the ground state, no substantial increase in the C2O2 bond length takes place on geometrical relaxation in this excited state ($S_2(n\pi^*)$). Due to a strong nonplanarity of the species, some mixing between the π and σ orbitals would also occur. Some contribution from the amino group nitrogen lone pair is also revealed in the relaxed $S_2(n\pi^*)$ excited state and it may be responsible for the out-of-plane deformation of the amino group. Under hydration with three water molecules, the amino group is more planar (Table 21, Fig. 9a). In going from the ground state to the $S_2(n\pi^*)$ excited state, the mode of water complexation is modified (Table 21, Fig. 9b). The N3 site provides a repulsive potential to hydrogen bonding. However, the repulsive nature of the N3 atom in this state is justified since the lone pair electron is promoted to antibonding orbitals.

In the case of the keto-N3H tautomer in the ground state, the ring geometry is planar, the amino group is more pyramidal than in the corresponding state of the keto-N1H tautomer. In the $S_1(\pi\pi^*)$ excited state, the keto-N3H tautomer adopts a boat type of structure where, the N1C2C4C5 atoms are approximately in one plane and the N3 and C6 atoms are out of this plane (Table 21). In the $S_2(n\pi^*)$ excited state, the structure is twisted around the N1C2 bond, the N1 atom being highly displaced from the approximate ring plane. The amino group remains pyramidal in both the excited states. In the $S_1(\pi\pi^*)$ excited state, after hydration, geometrical distortion exists around the

C₂N₃ bond, and these atoms (especially N3H) are appreciably out of the plane. The hydrogen bond distance between water hydrogen and the N1 atom is also increased (Fig. 9c). Such an increase may be partly due to mixing of the $\pi\pi^*$ transition with the $n\pi^*$ transition localized at the N1 lone pair and partly may also be due to the limited number of orbitals considered in the configuration. In the S₂(n π^*) excited state, the hydrated structure is completely modified (Fig. 9d). Such changes in the hydrated structures in the n π^* state are in accordance with the established fact that hydrogen bonds are largely destabilized under n π^* excitations.^{42,113,236,290}

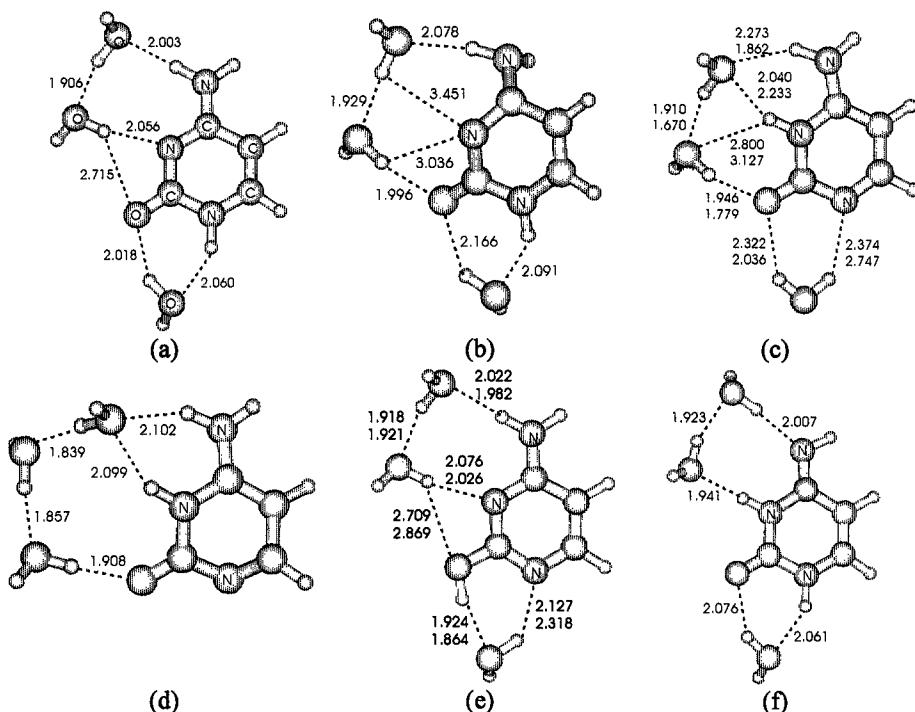


Figure 9. Structure and hydrogen bond lengths of hydrated cytosine tautomers: (a) keto-N1H tautomer in the ground state, (b) keto-N1H tautomer in the S₂(n π^*) excited state, (c) keto-N3H tautomer in the ground state (top indices) and S₁($\pi\pi^*$) excited state (bottom indices), (d) keto-N3H tautomer in the S₂(n π^*) excited state, (e) enol tautomer in the ground state (top indices) and S₁($\pi\pi^*$) excited state (bottom indices), and (f) imino tautomer in the ground state.

In the case of the enol tautomer in the ground state, the amino group pyramidalization is more pronounced than that in the keto-N1H tautomer but it is less pronounced than that in the keto-N3H tautomer. In the $S_1(\pi\pi^*)$ excited state, the ring geometry is largely distorted, especially around the C5C6 bond, and the C5C6 bond length is increased by about 0.092 Å. Hydration of the enol tautomer significantly reduces the amino group pyramidal character in the ground state (Table 21). The geometrical deformation of the hydrated enol form in the $S_1(\pi\pi^*)$ excited state is similar to that in the isolated tautomer.

Table 22. Ground and vertical singlet excited states charges at the atomic sites of the keto-N1H and keto-N3H tautomers of cytosine.^a

Atom	keto-N1H			keto-N3H		
	Mulliken	S_0	$S_1(\pi\pi^*)$	$S_2(\pi\pi^*)$	S_0	$S_1(\pi\pi^*)$
N1	-0.523	-0.454	-0.541	-0.479	-0.435	-0.321
C2	0.622	0.604	0.683	0.635	0.605	0.691
N3	-0.547	-0.531	-0.401	-0.582	-0.578	-0.607
C4	0.546	0.434	0.408	0.657	0.546	0.575
C5	-0.371	-0.246	-0.254	-0.400	-0.206	-0.295
C6	0.305	0.155	0.168	0.238	0.063	0.011
O2	-0.482	-0.394	-0.465	-0.474	-0.414	-0.464
N4	-0.524	-0.537	-0.558	-0.538	-0.522	-0.548
H1/3	0.257	0.260	0.269	0.251	0.246	0.263
H5	0.107	0.123	0.115	0.101	0.126	0.103
H6	0.130	0.128	0.121	0.122	0.113	0.144
H41	0.232	0.221	0.228	0.242	0.234	0.234
H42	0.248	0.237	0.227	0.227	0.222	0.214
CHELPG						
N1	-0.718	-0.608	-0.376	-0.928	-0.756	-0.147
C2	1.148	1.105	0.682	1.258	1.123	0.799
N3	-0.952	-0.866	-0.147	-0.904	-0.790	-0.537
C4	1.118	0.859	0.441	0.954	0.737	0.548
C5	-0.734	-0.430	-0.365	-0.905	-0.463	-0.500
C6	0.348	0.125	-0.167	0.712	0.314	-0.147
O2	-0.715	-0.616	-0.610	-0.731	-0.647	-0.637
N4	-1.047	-1.074	-1.033	-0.917	-0.897	-0.889
H1/3	0.371	0.366	0.335	0.418	0.392	0.366
H5	0.211	0.195	0.217	0.236	0.196	0.214
H6	0.116	0.094	0.175	0.024	0.030	0.166
H41	0.422	0.405	0.422	0.396	0.384	0.385
H42	0.432	0.445	0.426	0.387	0.377	0.379

^aGround and excited state parameters are at the HF/6-311G(d,p) and CIS/6-311G(d,p) levels, respectively.

The computed Mulliken and CHELPG charges of the keto-N1H and keto-N3H tautomers of cytosine in the ground and vertical singlet excited states are presented in Table 22. The charge distributions are modified in going from the ground state to vertical singlet excited states and such modifications are more pronounced in the singlet $n\pi^*$ excited states of both tautomers than those in the $\pi\pi^*$ excited states. In the vertical singlet $n\pi^*$ state of the keto-N1H tautomer, the N1 and N3 atoms lose and the C2 and C4 atoms gain significant amounts of electronic charge as compared to the corresponding ground state values (Table 22). In the keto-N3H tautomer, the N1, N3 and C5 atoms lose while the C2, C4 and C6 atoms gain significant amounts of electronic charge in the vertical singlet $n\pi^*$ excited state as compared to the corresponding ground state values (Table 22).

4.4.3 Phototautomerism²⁸⁹

The energy level diagrams comprising the lowest singlet $\pi\pi^*$ and $n\pi^*$ states of cytosine tautomers and their hydrated forms obtained at the CIS level are shown in Fig. 10. The excitation energies used in the energy level diagrams were computed applying the 6-311G(d,p) basis set using geometries optimized at different levels and basis sets as discussed earlier. The energy level diagrams show that the ordering of the $\pi\pi^*$ and $n\pi^*$ states of the keto-N1H and keto-N3H tautomers is changed as a result of the excited state geometry relaxation. Consequently, the optimized $n\pi^*$ excited states of the keto-N1H and keto-N3H tautomers lie lower than the corresponding optimized $\pi\pi^*$ excited states (Fig. 10). However, one can argue that such a change of the relative positions of these two states ($\pi\pi^*$ and $n\pi^*$) may be due to the different basis sets used in the geometry optimization and CIS calculations of excitation energies. However, the effect of using a different basis sets may not be important in this context since similar results were found in an earlier study also where the ground and excited state calculations were performed using the same basis set.⁵⁸ Such a crossing of the $\pi\pi^*$ and $n\pi^*$ states may be responsible for the non-radiative decay and thus weak fluorescence of the compound.²³ Further, the energy level diagrams show that among the vertical singlet $\pi\pi^*$ excited states, the state corresponding to the keto-N3H tautomer lies lowest both in gas phase and after hydration (Fig. 10). Therefore, energy transfer may take place from the lowest $\pi\pi^*$ state of the enol form (in the gas phase) or the keto-N1H tautomer (in an aqueous solution) to the corresponding state of the keto-N3H tautomer. Thus, under electronic excitation, cytosine may phototautomerize to the keto-N3H form. The computed energy differences

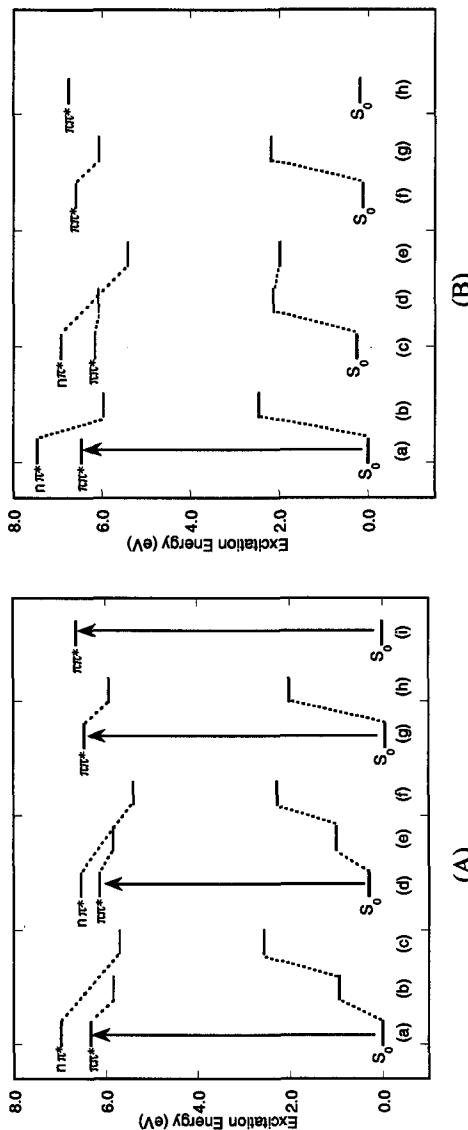


Figure 10. Energy level diagrams of cytosine tautomers: (A) isolated form, (B) hydrated form. In the case of isolated tautomers, (a), (d), (g), and (i) represent vertical states of the keto-N1H, keto-N3H, enol, and imino tautomers, respectively, (b), (e), and (h) represent optimized lowest singlet $n\pi^*$ excited state of the keto-N1H, keto-N3H and enol tautomers, respectively, while (c), and (f) represent the optimized lowest singlet $n\pi^*$ excited state of the keto-N1H and keto-N3H tautomers, respectively. In the case of hydrated tautomers, (a), (c), (f) and (h) represent the vertical states of the hydrated keto-N1H, keto-N3H, enol, and imino forms, respectively, (b), and (e) represent the optimized lowest singlet $n\pi^*$ excited state of the hydrated keto-N1H and keto-N3H tautomers, respectively, and (d), and (g) represent optimized lowest singlet $n\pi^*$ excited state of the hydrated keto-N3H and enol tautomer, respectively.

between the optimized $S_1(\pi\pi^*)$ state to the ground state lying vertically below to it of the keto-N1H and keto-N3H tautomers are 4.90 and 4.84 eV, and the corresponding scaled (scaling factor 0.72) values would be 3.53 and 3.48 eV, respectively. These values are closer to the experimental fluorescence peak observed near 3.85 eV (322 nm).²³ It appears that both of the keto-N1H and keto-N3H tautomers of cytosine would involve in fluorescence of the molecule.

4.5 AU, AT and GC Base Pairs

4.5.1 Electronic Transitions

Ground and singlet excited state properties including excited state geometries of the Watson-Crick AU, AT and GC base pairs were recently studied theoretically.^{113,114} Ground state geometries were optimized at the HF level, while the excited states were generated at the CIS level. The excited state geometries were optimized under the constraint of planarity. The 6-31++G(d,p) basis set was used in the study. The computed $\pi\pi^*$ and $n\pi^*$ transitions are shown in Table 23. As shown in Table 23, electronic excitations of the base pairs are localized on one or other monomeric unit. In an experimental study of the AT and GC polymers and natural DNA bases, electronic transitions were assigned to the corresponding monomer bases.²⁹¹

Among the vertical transitions of the AU base pair, the lowest singlet $\pi\pi^*$ excitation is characterized by an intense transition localized on the adenine moiety and is followed by a weak $\pi\pi^*$ excitation localized also at the same moiety (Table 23). The next $\pi\pi^*$ excitation is localized at the uracil moiety, especially at the C5'C6' bond, the transition intensity being slightly weaker than that of the first singlet $\pi\pi^*$ excitation. The last singlet $\pi\pi^*$ transition of the AU base pair shown in Table 23 is characterized by excitation from the π -type occupied orbitals of the adenine moiety to π^* -type virtual orbitals localized on the uracil moiety. Therefore, this excitation can be classified as one of charge transfer type. However, in this case a significant charge transfer was not found.¹¹³ The lowest singlet $n\pi^*$ excitation of the AU base pair is localized on the uracil moiety. It is assigned as excitation of a lone-pair electron of the C4'O4' group of uracil (Fig. 2). In the earlier study of uracil, the transition involving excitation of the O2 carbonyl oxygen lone pair was predicted to be higher in energy than transition involving the excitation out of the O4 carbonyl oxygen lone pair.^{128,129,275} The next singlet $n\pi^*$ transition is

localized at the adenine moiety and is characterized by excitation of nitrogen lone pairs of the purine ring.

Table 23. Vertical singlet $\pi\pi^*$ and $n\pi^*$ excitation energies (ΔE , eV), oscillator strengths (f) and assignments of the AU, AT and GC base pairs and their constituent bases^a^{113,114}.

ΔE	f	Assignment ^b	ΔE	f	ΔE	f
AU-base pair		A			U	
<i>$\pi\pi^*$ transitions</i>						
6.41 0.511 (A*)			6.43 0.412			
6.51 0.025 (A*)			6.56 0.021			
6.66 0.358 (U*)					6.66 0.463	
7.22 0.023 (A→U*)						
<i>$n\pi^*$ transitions</i>					6.42 0.000	
6.66 0.0001 (U*)						
7.29 0.0005 (A*)			7.13 0.0005			
AT-base pair		A			T	
<i>$\pi\pi^*$ transitions</i>						
6.39 0.518 (A*)			6.43 0.413			
6.49 0.183 (T*) ^c					6.49 0.441	
6.52 0.184 (A*)			6.56 0.021			
7.22 0.022 (A→T*) ^d			7.28 0.018			
<i>$n\pi^*$ transitions</i>					6.43 0.000	
6.67 0.0001 (T*)						
7.28 0.0006 (A*)			7.13 0.0005			
GC-base pair		G			C	
<i>$\pi\pi^*$ transitions</i>						
6.27 0.145 (G*)			6.23 0.291			
6.42 0.320 (C*)					6.19 0.151	
6.53 0.307 (G*)			6.62 0.317			
7.08 0.266 (G→C*) ^e			7.14 0.233			
<i>$n\pi^*$ transition</i>						
7.29 0.0010 (G*)			6.86 0.0003			

^aGeometries of the bases (A,G,U,T and C) are within the framework of the respective optimized base pair geometry, ^bA*, G*, U*, T* and C* indicate that the corresponding moiety of the respective base pair is excited. X→Y* indicates excitation from the purine (A, G) to the complementary pyrimidine (U, T, C), ^cSlightly contaminated by excitation of adenine unit,

^dSlightly contaminated by excitation to virtual orbitals of adenine, ^eSlightly contaminated by excitation to virtual orbitals of guanine.

A comparison of transition energies of the AU base pair with those of the isolated monomers computed within the framework of the optimized base pair geometry reveals that there is no significant change in the the $\pi\pi^*$ excitation energies in going from the isolated bases to the AU base pair, while the $n\pi^*$ transition energies are increased (blue-shift) upon base pair formation (Table 23). Such an increase in the $n\pi^*$ transition energies is in accordance with the established fact that hydrogen bonding causes blue-shift of $n\pi^*$ transitions.^{22,23,292} It has been suggested that the hydrogen bond strengths would be similar in the ground and $\pi\pi^*$ excited states while these would be decreased in $n\pi^*$ excited states.¹¹³ The oscillator strength of the first $\pi\pi^*$ transition of adenine is increased after the base pair formation, while for the second $\pi\pi^*$ transition, it has an insignificant effect (Table 23). In the case of uracil, after the base pair formation, the oscillator strength of the first $\pi\pi^*$ transition is decreased (Table 23). Further, while in uracil the first transition is predicted to be of the $n\pi^*$ type and the second transition is of $\pi\pi^*$ type, after base pair formation both the transition energies become nearly equal. This effect is similar to the blue shift in the $n\pi^*$ transition after the AU base pair formation.¹¹³ Experimentally, the lowest singlet excited state of uracil is predicted to be of the $n\pi^*$ type in an aprotic environment, while in a protic environment, the nature of the lowest singlet excited state is changed to $\pi\pi^*$ type.^{23,45,58,263} Thus a very weak coupling between the $\pi\pi^*$ spectral transitions of individual bases is revealed after the base pair formation (Table 23).

The lowest singlet $\pi\pi^*$ transition of the AT base pair is intense and is localized at the adenine moiety. The next higher energy transition is a comparatively weaker $\pi\pi^*$ transition localized at the C5'C6' bond of the thymine moiety (Table 23). The next singlet $\pi\pi^*$ transition originates from the excitation of the adenine moiety, the intensity being relatively weaker than that of the first $\pi\pi^*$ transition. The highest energy singlet $\pi\pi^*$ transition of the AT base pair shown in Table 23 is the charge transfer type¹¹⁴ and is characterized by excitation from π -type occupied orbitals of the adenine moiety to the π^* -type virtual orbitals localized at the thymine moiety. However, it is slightly contaminated by excitation to the virtual orbitals of the adenine moiety. The lowest singlet $n\pi^*$ transition of the AT base pair is localized at the thymine moiety and originates from excitation of the C4'O4' lone-pair (Fig. 2). The next singlet $n\pi^*$ excitation of the AT base pair is localized at the adenine moiety and is characterized by excitation of the nitrogen lone pairs of the purine ring.

A comparison of the excitation energies of the AT base pair with those of the isolated monomers computed within the framework of the optimized base pair geometry shows that there is a one-to-one correspondence between the $\pi\pi^*$ transition energies of the AT base pair and the corresponding transitions of the isolated bases (Table 23). The $n\pi^*$ transitions are blue-shifted due to the base pair (hydrogen bond) formation. The predicted blue-shift is in accordance with the well established fact that the $n\pi^*$ transitions are blue-shifted on hydrogen bonding.^{22,23,292} The oscillator strengths of $\pi\pi^*$ transitions of the AT base pair where excitations are localized at the adenine moiety are increased compared to the respective transitions of an isolated adenine (Table 23). On the contrary, the oscillator strength of the $\pi\pi^*$ transition of the AT base pair where the excitation is localized at the thymine moiety is decreased compared to the corresponding transition in isolated thymine (Table 23). The computed first transition of thymine is found to be of $n\pi^*$ type, and the second transition is of $\pi\pi^*$ -type. This sequence is reversed after the base pair formation (Table 23). The first $n\pi^*$ transition of the AT base pair where the excitation is localized at the thymine moiety has a higher energy than the first $\pi\pi^*$ -type transition, the excitation being localized again at the thymine moiety. These findings agree with the experimental observations that under the hydrogen bonding environment the nature of the lowest singlet excited state of thymine is changed from $n\pi^*$ to $\pi\pi^*$ -type.^{23,45,58,263}

The data presented in Table 23 show that the lowest singlet $\pi\pi^*$ excitation of the GC base pair is localized at the guanine moiety and is followed by a relatively more intense $\pi\pi^*$ transition localized at the cytosine moiety. The next transition is also of $\pi\pi^*$ -type and is localized at the guanine moiety. The highest singlet $\pi\pi^*$ transition of the GC base pair shown in Table 23 is of charge transfer type, involving mainly π -orbitals of guanine and π^* -orbitals of cytosine is being slightly contaminated by excitation to the unoccupied π^* -type orbitals of the guanine moiety also.¹¹⁴ The lowest singlet $n\pi^*$ transition of the GC base pair is localized at the guanine moiety and involves excitation of the carbonyl lone pair. The $\pi\pi^*$ excitation energies of the guanine moiety computed within the framework of the optimized geometry of the GC base pair are not significantly effected by the base pair formation, the maximum difference being within 0.1 eV. However, the $n\pi^*$ transition energy of the guanine moiety is appreciably increased upon base pair formation. This increase in the $n\pi^*$ transitions energies upon hydrogen bonding is, as discussed earlier, well known.^{22,23,292} The lowest singlet $\pi\pi^*$ excitation energy

of cytosine in the GC base pair is appreciably increased on the base pair formation (Table 23). This raise in excitation energy of the first $\pi-\pi^*$ transition may be due the influence of hydrogen bonds in the base pair. The differences between the transition energies of isolated guanine and cytosine forms to those of the GC base pair are more than those between the transition energies of isolated adenine and thymine or uracil and those of the AT or AU base pairs. These differences may be due to the presence of three hydrogen bonds in the GC base pair and comparatively larger deformation in the geometry of the isolated G and C bases.^{113,114}

Table 24. Hydrogen bond lengths (\AA), hydrogen bond angles ($^\circ$) and interaction energies (E_{int} , kcal/mol) of the AT, AU and GC base pairs in the ground and different singlet excited states.^a^{113,114}

AT base pair	S_0	$S_1(\pi-\pi^*)$	$S_2(\pi-\pi^*)$	$S_4(n-\pi^*)$	$S_6(n-\pi^*)$
N6...O4'	3.083	3.011	3.027	3.778	3.094
N1...N3'	3.023	3.071	3.025	3.090	3.152
H61(N6)...O4'	2.090	2.016	2.029	2.778	2.106
N1...H3'	2.010	2.063	2.014	2.086	2.149
O4'H61N6	172.6	171.1	174.5	161.8	170.2
N3'H3'N1	177.7	174.8	177.2	183.2	174.3
E_{int}	-9.9	-10.0	-10.6	-5.8	-7.1
AU base pair	S_0	$S_1(\pi\pi^*)$	$S_3(\pi\pi^*)$	$S_4(n\pi^*)$	$S_6(n\pi^*)$
N6...O4'	3.082	3.011	3.029	3.742	3.092
N1...N3'	3.019	3.066	3.022	3.087	3.146
H61(N6)...O4'	2.088	2.016	2.032	2.786	2.104
N1...H3'	2.007	2.059	2.010	2.083	2.143
O4'H61N6	172.6	171.0	174.2	161.9	170.2
N3'H3'N1	177.6	174.6	177.0	183.0	174.1
E_{int}	-10.1	-10.2	-10.7	-5.9	-7.3
GC base pair	S_0	$S_1(\pi-\pi^*)$	$S_2(\pi-\pi^*)$		
O6...N4'	2.932	2.937	2.958		
N1...N3'	3.053	3.023	3.066		
N2...O2'	3.028	3.037	3.098		
O6...H41'(N4')	1.926	1.929	1.954		
H1(N1)...N3'	2.048	2.021	2.059		
H21(N2)...O2'	2.028	2.034	2.100		
O6H41'N4'	175.7	177.7	177.3		
N1H1N3'	175.3	176.3	176.4		
N2H21O2'	176.9	177.4	177.7		
E_{int}	-24.8	-22.9	-20.4		

^a States are given in the ascending energy order (see Table 23).

4.5.2 Interaction Energies

The basis set superposition error (BSSE) corrected interaction energies were estimated using the Boys-Bernardi counterpoise correction scheme.²⁹³ The interaction energy (E_{int}) in the ground state was calculated using the formula:

$$E_{int} = E(XY) - E(X_{XY}) - E(Y_{XY}) \quad (6)$$

Where $E(XY)$ is the total energy of the XY base pair in the ground state; $E(X_{XY})$ and $E(Y_{XY})$ are the total energies of X (adenine in case of the AT (AU) base pair and guanine in the case of the GC base pair) and Y (thymine (uracil) in the case of the AT (AU) base pair and cytosine in the case of the GC base pair) monomeric moieties using the optimized XY base pair geometry and ghost atoms in place of the complementary base.

Boys-Bernardi counterpoise correction schemes for the ground state²⁹³ were utilized by us^{113,114} to compute BSSE corrected interaction energy for excited states. To the best of our knowledge there is no rigorous scheme to account for the BSSE correction of the interaction energy of molecules in excited states. We hope that our approach^{113,114} to compute BSSE corrected interaction energy for excited states would provide useful information. The interaction energy in the excited state ($E_{int}^{(n)}$) where the excitation is localized at the X monomeric moiety was estimated using the formula:

$$E_{int}^{(n)} = E^{(n)}(XY) - E^{(n)}(X_{XY}) - E^{(0)}(Y_{XY}) \quad (7)$$

While for the excited state where the excitation is localized at the Y monomeric moiety, the interaction energy was estimated using the formula:

$$E_{int}^{(n)} = E^{(n)}(XY) - E^{(0)}(X_{XY}) - E^{(n)}(Y_{XY}) \quad (8)$$

In equations ((7) and (8)), $E^{(n)}(XY)$ is the total energy of the XY base pair in the n^{th} excited state; $E^{(n)}(X_{XY})$ and $E^{(n)}(Y_{XY})$ are the total energies of the X and Y monomeric moieties, respectively, in the n^{th} excited state which corresponds to the n^{th} state of the XY base pair (since the n^{th} state of the XY base pair may not necessarily correspond to the n^{th} state of excited X or Y^{113,114}). $E^{(0)}(X_{XY})$ and $E^{(0)}(Y_{XY})$ are the ground state total energies of the X and Y monomeric moieties, respectively. In these calculations, the geometries of the X and Y monomeric moieties were those of the bases in the optimized geometry of the XY base pair in the n^{th} excited state while the ghost atoms were added in place of the complementary base.

The hydrogen bond parameters and interaction energies of the base pairs in the ground and excited states shown in Table 24 suggest that the hydrogen bond properties and interaction energies for the AT and AU base pairs are similar. The hydrogen bonds are more nonlinear in the singlet $n\pi^*$ states than those in the ground and singlet $\pi\pi^*$ states. It follows from our calculations that the base pairs are appreciably destabilized due to $n\pi^*$ excitations.

5. Concluding Remarks

Structures and properties of the nucleic acid bases and base pairs in the electronic excited states are usually appreciably different from those in the corresponding ground states. Excited state properties of such complex systems would not be revealed satisfactorily by one technique; instead, usually several techniques would be needed for this purpose. Quantum chemical calculations provide valuable information in this context. However, although a great progress has been made in theoretical techniques to study ground state properties of molecules, the situation with regard to excited states is not as good. Thus, applications of the comparatively more reliable MCSCF method to excited states are limited to smaller system. It is especially true for excited state geometries. The relatively more convenient and economical method in this regard is that of CIS and it also usually yields quite satisfactory results, but it is comparatively less accurate than the MCSCF method. Other single reference methods such as TD-DFT and EOM-CC are better than CIS but usually only vertical transition energies have been obtained using these methods so far. The theoretical results reviewed are mainly based on the CIS method. The information presented in this chapter can be summarized as follows.

All theoretical methods support the experimental findings that the 260 nm region in the absorption spectrum of adenine is composed of two transitions. Theoretical transition energies (CASPT2 and scaled CIS) are generally in agreement with the experimental ones. However, the agreement is not equally good with regard to the transition moment directions. It should be noted that the measured transition moment directions are dependent on environment. Precise information about $n\pi^*$ transitions is still lacking, although experimental and theoretical methods yield approximate data. Existence of an $n\pi^*$ transition close to the first $\pi\pi^*$ transition of adenine is indicated in some experiments, but theoretically it is only supported by the TD-DFT method.

Hydration induces red-shift in $\pi\pi^*$ transitions and significant blue-shifts in $n\pi^*$ transitions. For the N9H tautomer of adenine, the ring geometry is almost planar in the ground and lowest singlet $\pi\pi^*$ excited states, the amino group being nonplanar in both the states. However, in the singlet $n\pi^*$ excited state, the geometry is nonplanar including the amino group. For the N7H tautomer, the ring geometry is planar in the ground and lowest singlet $n\pi^*$ excited states while it is nonplanar in the lowest singlet $\pi\pi^*$ excited state. The amino group is pyramidal in the ground and the lowest singlet $\pi\pi^*$ excited states of the N7H tautomer. Further, the amino group of the N7H tautomer in the lowest singlet $n\pi^*$ excited state is strongly rotated about the CN (NH₂) bond. Hydration was found to induce planarity in the geometry of the tautomers. Electronic charge distribution as revealed by CHELPG charges was found to be similar in the ground and the lowest singlet $\pi\pi^*$ excited states but it was revealed to be appreciably different in the lowest singlet $n\pi^*$ excited state.

In the case of guanine, our computed transition energies are in agreement with the corresponding experimental data. However, this agreement is better in the lower energy region than that in the higher energy region. Theoretical calculations suggest that the 225 nm transition observed in some experiments^{226,227,250} could be due to a weak $\pi\pi^*$ transition. Our computed singlet $n\pi^*$ transitions support the finding of Clark²⁵¹ with regard to the existence of $n\pi^*$ transitions near 5.21 and 6.32 eV in guanine. The experimentally observed red-shift in the lowest singlet $\pi\pi^*$ transition of 7-methyl-guanine as compared to the corresponding transition of guanosine monophosphate is also revealed in CIS calculations of the keto-N9H and keto-N7H tautomers of guanine. Ground state ring geometries of both the keto-N9H and keto-N7H tautomers of guanine are planar, while the amino group is pyramidal. The lowest singlet $\pi\pi^*$ excited state geometries of both the tautomers are nonplanar including the amino group, the out-of-plane ring deformation being localized mainly at the six-membered ring. The singlet $n\pi^*$ excited state geometries of both the tautomers are characterized by a large increase in the C₆O₆ bond length and the rings also become slightly nonplanar. Hydration induces planarity in both the ground and lowest singlet $\pi\pi^*$ excited states. Ground and vertical singlet $\pi\pi^*$ excited state charge distributions are similar while that in the singlet $n\pi^*$ excited state is appreciably different. The main features of the difference in this context is a significant decrease and increase in the electronic charges at the O₆ and C₆ atomic sites of the guanine tautomers as compared to the corresponding ground and singlet $\pi\pi^*$ excited state values. The structural changes in guanine

tautomers in the $\pi\pi^*$ and $n\pi^*$ excited states are also accompanied by significant changes in electronic charges as compared to the corresponding vertical excited state charges. It has been found that the keto-N9H tautomer of guanine, under electronic excitation would phototautomerise to the keto-N7H form. The main 332 nm higher energy fluorescence would originate from the keto-N7H tautomer, while the weak fluorescence near 450 nm would originate from the normal keto-N9H tautomer.

The experimentally observed red-shift in the first $\pi\pi^*$ transition of thymine as compared to the corresponding transition of uracil is satisfactorily explained by the computed transitions using the CIS method. Hydration generally induces red-shift in $\pi\pi^*$ transition energies and blue-shift in $n\pi^*$ transition energies of the molecules. The computed scaled transition energies of the hydrated forms of uracil and thymine obtained using the CIS method are within about 0.3 eV of the corresponding observed experimental results. The computed transition moment directions of the first two transitions of uracil are also in good agreement with the corresponding experimental data observed in the crystal environment.²⁶⁹ The ground state geometries of uracil and thymine (except the hydrogen atoms of methyl group) are planar in the ground state and slightly nonplanar in the lowest singlet $n\pi^*$ excited state, where the C4O4 bond length is increased appreciably as compared to the ground state values. Thymine, in the lowest singlet $\pi\pi^*$ excited state adopts a boat type structure. Hydration of these molecules with three water molecules induces a slight planarity in geometry. Electronic charge distributions are appreciably modified in going from the ground states to excited states of the molecules. The possibility of phototautomerization of uracil (and possibly also of thymine) from the natural keto form to the mono-enol form, the fluorescence corresponding to which would be appreciably red-shifted as compared to the normal fluorescence peak of the molecule, is indicated by the CIS calculations.

The absorption transitions of cytosine have been found to be dominated by the normal keto-N1H form of the molecule. The computed transition energies of keto-N3H and imino tautomers of cytosine obtained at the CIS level after linear scaling were useful in explaining the experimental transitions of 3-methylcytosine and 3-methylcytidine. The first singlet $\pi\pi^*$ transition of the keto-N1H form was found to be blue-shifted after hydration. This blue-shift may be related to a large decrease in the dipole moment of the molecule in the vertical first singlet $\pi\pi^*$ excited state as compared to the ground state value. Ground state geometries of cytosine tautomers are found to be planar except the amino group which is pyramidal. However, excited state geometries are

found to be nonplanar both in the lowest singlet $\pi\pi^*$ and $n\pi^*$ excited states. The mode of hydration was found to be completely modified in the lowest singlet $n\pi^*$ excited state as compared to that in the corresponding ground and lowest singlet $\pi\pi^*$ excited states. Electronic charge distributions were also found to be modified in going from the ground state to vertical singlet excited states, these modifications being more pronounced in the lowest singlet $n\pi^*$ excited state than those in the singlet $\pi\pi^*$ excited state. Further, cytosine is expected to undergo phototautomerization to the keto-N3H form and contributions to fluorescence from both of the keto-N1H and keto-N3H tautomers would be expected.

Electronic excitations of the Watson-Crick AU, AT and GC base pairs are found to be localized at either of the monomeric units agreeing with the experimental observations. Some excited states of charge transfer type involving excitation from the occupied orbitals of one base to the virtual orbitals of the other complementary base of the base pair were also found with somewhat higher excitation energies. It has been shown that the base pair formation does not have a significant effect on the singlet $\pi\pi^*$ transitions of the constituent bases. However, $n\pi^*$ transition energies are appreciably blue-shifted on base pair formation. Changes in the excited state geometry of a base pair take place mainly at the monomer where the excitation is localized. Hydrogen bond parameters and interaction energies in the ground and excited states of the AT and AU base pairs were found to be almost the same. Hydrogen bonds in the base pairs are more nonlinear in the singlet $n\pi^*$ states as compared to those in the ground and singlet π^* states. Base pairs would be appreciably destabilized under the $n\pi^*$ excitation.

Thus the theoretical studies on the excited states of the nucleic acid bases and base pairs have been greatly helpful in unravelling several details regarding the nature and locations of transitions and excited state geometries and properties that were not understood earlier. However the knowledge obtained in this regard is expected to become richer in future as better computational technologies become available.

Acknowledgement

Authors are thankful to NIH-RCMI grant No. G1 2RR13459-21, NSF-CREST grant No. 9805465 & 9706268, and ONR grant No. N00034-03-1-0116 for financial assistance. The authors are also thankful to Prof. A. Sadlej,

Department of Chemistry, Nicolaus Copernicus University at Torun, Poland, Prof. A. Sokalski, Department of Chemistry, Wroclaw University of Technology, Wroclaw, Poland and Prof. P.C. Mishra, Department of Physics, Banaras Hindu University, Varanasi, India, for useful discussions and suggestions to the manuscript.

References

1. W. Saenger (1984) *Principles of Nucleic Structures*, Springer-Verlag, New York.
2. S. Scheiner (2000) *J. Phys. Chem. A* **104**, 5898.
3. A.L. Sobolewski and W. Domcke (1999) *Chem. Phys. Lett.* **300**, 533.
4. R.A. Kerr (1998) *Science* **280**, 202.
5. C. Baird, *Environmental Chemistry* (W.H. Freeman and Co., New York, 1995).
6. J. Darnell, H. Lodish and D. Baltimore, *Molecular Cell Biology* (Scientific American Books, New York, 1986).
7. J.-S. Taylor (1990) *Chem. Educat.* **67**, 835.
8. J. Peon and A.H. Zewail (2001) *Chem. Phys. Lett.* **348**, 255.
9. J.-M.L. Pecourt, J. Peon and B. Kohler (2000) *J. Am. Chem. Soc.* **122**, 9348, Errata (2001) *J. Am. Chem. Soc.* **123**, 5166.
10. J.-M.L. Pecourt, J. Peon and B. Kohler (2001) *J. Am. Chem. Soc.* **123**, 10370.
11. T. Gustavsson, A. Sharonov and D. Markovitsi (2002) *Chem. Phys. Lett.* **351**, 195.
12. A. Reuther, H. Iglev, R. Laenen and A. Laubereau (2000) *Chem. Phys. Lett.* **325**, 360.
13. A. Reuther, D.N. Nikogosyan and A. Laubereau (1996) *J. Phys. Chem.* **100**, 5570.
14. J. Cadet and P. Vigny (1990) in *The Photochemistry of Nucleic Acids*, H. Morrison (ed.), Vol. 1, Wiley, New York.
15. A.A. Oraevsky, A.V. Sharkov and D.N. Nikogosyan (1981) *Chem. Phys. Lett.* **83**, 276.
16. S. Kobayashi, M. Yamashita, T. Sato and S. Muramatsu (1984) *IEEE J. Quantum Electron.* **20**, 1383.
17. D.N. Nikogosyan, D. Angelov, B. Soep and L. Lindqvist (1996) *Chem. Phys. Lett.* **252**, 322.
18. T. Haupl, C. Windolph, T. Jochum, O. Brede and R. Hermann (1997) *Chem. Phys. Lett.* **280**, 520.
19. J. Stepanek and V. Baumruk (1990) *J. Mol. Struct.* **219**, 299.
20. T. Fujiwara, Y. Kamoshida, R. Morita and M. Yamashita (1997) *J. Photochem. Photobiol. B: Biol.* **41**, 114.
21. M. Daniels and W. Hauswirth (1971) *Science* **171**, 675.
22. J. Eisinger and A.A. Lamola (1971) in *Excited State of Proteins and Nucleic Acids*, R.F. Steiner and I. Weinryb (eds.), Plenum Press, New York-London.
23. P.R. Callis (1983) *Ann. Rev. Phys. Chem.* **34**, 329.

24. D. Duggan, R. Bowmann, B.B. Brodie, S. Udenfriend (1957) *Arch. Biochem. Biophys.* **68**, 1.
25. G. Ge, S. Zhu, T.D. Bradrick and S. Georghiou (1990) *Photochem. Photobiol.* **51**, 557.
26. S. Georghiou and A.M. Saim (1986) *Photochem. Photobiol.* **44**, 733.
27. J.D. Smith, R. Markham (1950) *Biochem. J.* **46**, 33.
28. L.S. Agroskin, N.V. Korolev, I.S. Kulaev, M.N. Mesel and N.A. Pomashchinkova (1960) *Dokl. Akad. Nauk SSSR* **131**, 1440.
29. R. Bersohn and I. Isenberg (1964) *J. Chem. Phys.* **40**, 3175.
30. S.V. Konev (1967) *Fluorescence and Phosphorescence of Proteins and Nucleic Acids*, Plenum Press, New York.
31. C. Helene (1966) *Biochem. Biophys. Res. Commun.* **22**, 237.
32. R.H. Steele and A. Szent-Gyorgyi (1957) *Proc. Natl. Acad. Sci.* **43**, 477.
33. J.W. Longworth (1962) *Biochem. J.* **84**, 104P.
34. E. Nir, Ch. Janzen, P. Imhof, K. Kleinermanns and M.S. de Vries (2001) *J. Chem. Phys.* **115**, 4604.
35. E. Nir, K. Kleinermanns, L. Grace and M.S. de Vries (2001) *J. Phys. Chem. A* **105**, 5106.
36. E. Nir, K. Kleinermanns and M.S. de Vries (2000) *Nature* **408**, 949.
37. E. Nir, P. Imhof, K. Kleinermanns and M.S. de Vries (2000) *J. Am. Chem. Soc.* **122**, 8091.
38. E. Nir, L. Grace, B. Brauer and M.S. de Vries (1999) *J. Am. Chem. Soc.* **121**, 4896.
- 39a. F. Piuzzi, M. Mons, I. Dimicoli, B. Tardivel and Q. Zhao (2001) *Chem. Phys.* **270**, 205.
- 39b. M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel and M. Elhanine (2002) *J. Phys. Chem. A* **106**, 5088.
40. D.C. Luhrs, J. Viallon and I. Fischer (2001) *Phys. Chem. Chem. Phys.* **3**, 1827.
41. N.J. Kim, G. Jeong, Y.S. Kim, J. Sung, S.K. Kim and Y.D. Park (2000) *J. Chem. Phys.* **113**, 10051.
42. N.J. Kim, H. Kang, G. Jeong, Y.S. Kim, K.T. Lee and S.K. Kim (2000) *J. Phys. Chem. A* **104**, 6552.
43. B.B. Brady, L.A. Peteanu and D.H. Levy (1988) *Chem. Phys. Lett.* **147**, 538.
44. Y. Tsuchiya, T. Tamura, M. Fujii and M. Ito (1988) *J. Phys. Chem.* **92**, 1760.
45. M. Fujii, T. Tamura, N. Mikami and M. Ito (1986) *Chem. Phys. Lett.* **126**, 583.
46. L.C. Sowers, G.V. Fazakerley, R. Eritja, B.E. Kaplan and M.F. Goodman (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5434.
47. P.A. Fagan, C. Fabrega, R. Eritja, M.F. Doogman and D. Wemmer (1996) *Biochemistry* **35**, 4026.
48. P.-O. Lycksel, A. Graslund, F. Claesens, L.W. McLaughlin, U. Larsson and R. Riegler (1987) *Nucleic Acids Res.* **15**, 9011.
49. T.M. Nordlund, S. Andersson, L. Nilsson, R. Riegler, A. Graslund and L.W. McLaughlin (1989) *Biochemistry* **28**, 9095.
50. A.N. Fletcher (1967) *J. Mol. Spectrosc.* **23**, 221.
51. T.M. Nordlund, D. Xu and K.O. Evans (1993) *Biochemistry* **32**, 12090.

52. C. Santhosh and P.C. Mishra (1991) *Spectrochim. Acta* **47A**, 1685.
53. R. Change (1971) *Basic Principles of Spectroscopy*, McGraw-Hill, New York.
54. J.R. Lakowicz (1983) *Principles of Fluorescence Spectroscopy*, Plenum Press, New York.
55. D. Voet and A. Rich (1970) *Prog. Nucl. Acid Res. Mol. Biol.* **10**, 183.
56. T. Cvitas, J.M. Hollas, G.H. Kirby (1970) *Mol. Phys.* **19**, 305.
57. L. Chinsky, A. Laigle, L. Peticolas and P.-Y. Turpin (1982) *J. Chem. Phys.* **76**, 1.
58. M.K. Shukla and P.C. Mishra (1999) *Chem. Phys.* **240**, 319.
59. J. Sponer, P. Hobza, and J. Leszczynski (2000) in *Computational Chemistry: Reviews of Current Trends*, Jerzy Leszczynski (ed.) Vol. 5, World Scientific.
60. J. Leszczynski (2000) in *Advances in Molecular Structure and Research*, M. Hargittai, I. Hargittai (Eds.) Vol. 6, JAI Press Inc., Stamford, Connecticut.
61. L. Gorb and J. Leszczynski (1999) in *Computational Molecular Biology, Theoretical and Computational Chemistry Book Series*, J. Leszczynski (ed.) Vol. 8, Elsevier.
62. J. Sponer, P. Hobza, and J. Leszczynski (1999) in *Computational Molecular Biology, Theoretical and Computational Chemistry Book Series*, J. Leszczynski (ed.) Vol. 8, Elsevier.
63. J. Leszczynski (1998) in *The Encyclopedia of Computational Chemistry*, Vol. V, John Wiley & Sons.
64. M. J. Nowak, L. Lapinski, J. S. Kwiatkowski, and J. Leszczynski (1997) in *Computational Chemistry: Reviews of Current Trends*, J. Leszczynski (ed.), Vol. 2, World Scientific.
65. J. Sponer, P. Hobza and J. Leszczynski (1996) in *Computational Chemistry: Reviews of Current Trends*, J. Leszczynski (ed.), Vol. 1, World Scientific.
66. P. Hobza and J. Sponer (1999) *Chem. Rev.* **99**, 3247.
67. M.K. Shukla and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 3021.
68. M.K. Shukla and J. Leszczynski (2000) *J. Mol. Struct. (THEOCHEM)* **529**, 99.
69. A. Pelmenschikov, D.M. Hovorun, O.V. Shishkin and J. Leszczynski (2000) *J. Chem. Phys.*, **113**, 5986.
70. A. Pelmenschikov, I. Zilberberg, J. Leszczynski, A. Famulari, M. Sironi and M. Raimondi (1999) *Chem. Phys. Lett.* **314**, 496.
71. D. Venkateswarlu and J. Leszczynski (1998) *J. Phys. Chem. A* **102**, 6161.
72. D.M. Hovorun, L. Gorb and J. Leszczynski (1999) *Int. J. Quant. Chem.* **75**, 245.
73. J. Gu and J. Leszczynski (2001) *Chem. Phys. Lett.* **335**, 465.
74. J. Gu and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 7353.
75. J. Gu and J. Leszczynski (2002) *J. Phys. Chem. A* **106**, 529.
76. J. Gu and J. Leszczynski (2001) *J. Phys. Chem. A* **105**, 10366.
77. J. Gu and J. Leszczynski (1999) *J. Phys. Chem. A* **103**, 2744.
78. J. Leszczynski (1998) *J. Phys. Chem. A* **102**, 2357.
79. J. Leszczynski (1992) *J. Phys. Chem.* **96**, 1649.
80. J. Sponer, J. Leszczynski and P. Hobza (2002) *Biopolymers (Nuc. Acid Sci.)* **61**, 3.
81. J. Sponer, J. Leszczynski and P. Hobza (2001) *J. Mol. Struct. (Theochem)* **573**, 43.
82. J. Sponer, J.E. Sponer and J. Leszczynski (2000) *J. Biomol. Struct. & Dynamics* **17**, 1087.

83. J. Sponer, M. Sabat, L. Gorb, J. Leszczynski, B. Lippert and P. Hobza (2000) *J. Phys. Chem. B* **104**, 7535.
84. J. V. Burda, J. Sponer and J. Leszczynski (2000) *J. Biol. Inorg. Chem.* **5**, 178.
85. J.E. Sponer, F. Glahe, J. Leszczynski, B. Lippert and J. Sponer (2001) *J. Phys. Chem. B* **105**, 12171.
86. J.E. Sponer, J. Leszczynski, F. Glahe, B. Lippert and J. Sponer (2001) *Inorg. Chem.* **40**, 3269.
87. J. Gu and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 1898.
88. J. Gu, J. Leszczynski and M. Bansal (1999) *Chem. Phys. Lett.* **311**, 209.
89. J.V. Burda, J. Sponer, and J. Leszczynski (2001) *Phys. Chem. Chem. Phys.* **3**, 4404.
90. L. Gorb and J. Leszczynski (1997) *Int. J. Quant. Chem.* **65**, 759.
91. L. Gorb and J. Leszczynski (1998) *J. Am. Chem. Soc.* **120**, 5024.
92. L. Gorb and J. Leszczynski (1998) *Int. J. Quant. Chem.* **70**, 855.
93. L. Gorb, Y. Podolyan, J. Leszczynski, W. Siebrand, A. Fernandez-Ramos and Z. Smedarchina (2002) *Biopolymers (Nuc. Acid Sci.)* **61**, 77.
94. L. Gorb, Y. Podolyan and J. Leszczynski (1999) *J. Mol. Struct. (Theochem)* **487**, 47.
95. O.V. Shishkin, A. Pelmenschikov, D.M. Hovorun and J. Leszczynski (2000) *Chem. Phys.* **260**, 317.
96. O.V. Shishkin, A. Pelmenschikov, D.M. Hovorun and J. Leszczynski (2000) *J. Mol. Struct.* **526**, 329.
97. T. Stepanenko, L. Lapinski, M.J. Nowak, J.S. Kwiatkowski and J. Leszczynski (2001) *Spectrochim. Acta* **57A**, 375.
98. Y. Podolyan, L. Gorb, A. Blue and J. Leszczynski (2001) *J. Mol. Struct. (Theochem)* **549**, 101.
99. Y. Podolyan, L. Gorb and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 7346.
100. Y. Podolyan, Y. Rubin and J. Leszczynski (2001) *Int. J. Quant. Chem.* **83**, 203.
101. Y. Podolyan, Y.V. Rubin and J. Leszczynski (2000) *J. Phys. Chem. A* **104**, 9964.
102. Y.V. Rubin and J. Leszczynski (1999) *Nucleosides & Nucleotides* **18**, 1119.
103. G. Fogarasi (2002) *J. Phys. Chem. A* **106**, 1381.
104. S.D. Wetmore, R.J. Boyd and L.A. Eriksson (2000) *Chem. Phys. Lett.* **322**, 129.
105. A.K. Chandra, M.T. Nguyen, T. Uchimaru and T. Zeegers-Huyskens (1999) *J. Phys. Chem. A* **103**, 8853.
106. T.V. Mourik (2001) *Phys. Chem. Chem. Phys.* **3**, 2886.
107. M.-P. Gaigeot and M. Ghomi (2001) *J. Phys. Chem. B* **105**, 5007
108. E.A. Syrstad, S. Vivekananda and F. Turecek (2001) *J. Phys. Chem. A* **105**, 8339.
109. J.K. Wolken and F. Turecek (2001) *J. Phys. Chem. A* **105**, 8352.
110. M. Brandl, M. Meyer and J. Suhnel (2000) *J. Phys. Chem. A* **104**, 11177.
111. D.B. Pedersen, M.Z. Zgierski, S. Denommee and B. Simard (2002) *J. Am. Chem. Soc.* **124**, 6686.
112. S.K. Mishra and P.C. Mishra (2001) *Spectrochim. Acta A* **57**, 2433.
113. M.K. Shukla and Jerzy Leszczynski (2002) *J. Phys. Chem. A* **106**, 1011.
114. M.K. Shukla and Jerzy Leszczynski (2002) *J. Phys. Chem. A* **106**, 4709.

115. M.K. Shukla and Jerzy Leszczynski (2000) *Int. J. Quantum Chem.* **77**, 240.
116. M.K. Shukla, Anil Kumar and P.C. Mishra (2001) *J. Mol. Struct. (THEOCHEM)* **535**, 269.
117. M.K. Shukla and P.C. Mishra (1998) *J. Chem. Infor. Computer Sci.* **38**, 678.
118. M.K. Shukla and P.C. Mishra (1998) *Spectrochim. Acta* **54A**, 937.
119. M.K. Shukla, S.K. Mishra, Anil Kumar and P.C. Mishra (2000) *J. Comput. Chem.* **21**, 826.
120. S.K. Mishra, M.K. Shukla and P.C. Mishra (2000) *Spectrochim. Acta* **56A**, 1355.
- 120a. M.K. Shukla and Jerzy Leszczynski, To be published.
121. L.M. Salter and G.M. Chaban (2002) *J. Phys. Chem.* **106**, 4251.
122. A. Broo (1998) *J. Phys. Chem. A* **102**, 526.
123. B. Mennucci, A. Toniolo and J. Tomasi (2001) *J. Phys. Chem. A* **105**, 4749.
124. A. Holmen, A. Broo, B. Albinsson and B. Norden (1997) *J. Am. Chem. Soc.* **119**, 12240.
125. B. Mennucci, A. Toniolo and J. Tomasi (2001) *J. Phys. Chem. A* **105**, 7126.
126. M.P. Fulscher, L. Serrano-Andres and B.O. Roos (1997) *J. Am. Chem. Soc.* **119**, 6168.
127. J.D. Petke, G.M. Maggiora and R.F. Christoffersen (1990) *J. Am. Chem. Soc.* **112**, 5452.
128. J. Lorentzon, M.P. Fulscher and B.O. Roos (1995) *J. Am. Chem. Soc.* **117**, 9265.
129. J.D. Petke, G.M. Maggiora and R.E. Christoffersen (1992) *J. Phys. Chem.* **96**, 6992.
130. M.P. Fulscher and B.O. Roos (1995) *J. Am. Chem. Soc.* **117**, 2089.
131. M.P. Fulscher, P.-A. Malmqvist and B.O. Roos, in *Time-Resolved Laser Spectroscopy in Biochemistry II*, Proceedings SPIEa, J. Lakowicz (ed.) SPIE: P.O. Box 10, Bellingham, WA 98227-0010: Vol. 1204, pg. 332.
132. A. Holmen, B. Norden and B. Albinsson (1997) *J. Am. Chem. Soc.* **119**, 3114.
133. E.L. Rachofsky, J.B.A. Roos, M. Krauss and R. Osman (2001) *J. Phys. Chem. A* **105**, 190.
134. A.C. Borin, L. Serrano-Andres, M.P. Fulscher and B.O. Roos (1999) *J. Phys. Chem. A* **103**, 1838.
135. A. Ohrn and O. Christiansen (2001) *Phys. Chem. Chem. Phys.* **3**, 730.
136. A. Broo and A. Holmen (1997) *J. Phys. Chem. A* **101**, 3589.
137. N. Ismail, L. Blancafort, M. Olivucci, B. Kohler and M.A. Robb (2002) *J. Am. Chem. Soc.* **124**, 6818.
138. J.B. Foresman, M. Head-Gordon, J.A. Pople and M.J. Frisch (1992) *J. Phys. Chem.* **96**, 135.
139. R.J. Bartlett and J.F. Stanton (1994) in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd (eds.), Vol. V, VCH Publishers, Inc. New York, p. 65.
140. R.J. Bartlett (1995) in *Modern Electronic Structure Theory, Part I*, D.R. Yarkony (ed.), World Scientific Publishing Com., Singapore, p. 1047.
141. J.E. Del Bene, J.D. Watts and R.J. Bartlett (1997) *J. Chem. Phys.* **106**, 6051.
142. M.E. Casida, C. Jamorski, K.C. Casida and D.R. Salahub (1998) *J. Chem. Phys.* **108**, 4439.

143. R.E. Stratmann, G.E. Scuseria and M.J. Frisch (1998) *J. Chem. Phys.* **109**, 8218.
144. K.B. Wiberg, R.E. Stratmann and M.J. Frisch (1998) *Chem. Phys. Lett.* **297**, 60.
145. S. Hirata and M. Head-Gordon (1999) *Chem. Phys. Lett.* **314**, 291.
146. S. Hirata, M. Head-Gordon and R.J. Bartlett (1999) *J. Chem. Phys.* **111**, 10774.
- 147a.M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian 98, Revision A.11.1, 2001, Gaussian, Inc., Pittsburgh PA.
- 147b.M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, J. A. Pople, Gaussian 94, Revision E.2, 1995, Gaussian, Inc., Pittsburgh PA.
148. M.W. Schmidt and M.S. Gordon (1998) *Annu. Rev. Phys. Chem.* **49**, 233.
149. B.O. Roos (1999) *Acc. Chem. Res.* **32**, 137.
150. K. Andersson and B.O. Roos (1995) in *Modern Electronic Structure Theory, Part I*, Vol. 2, D.R. Yarkony (ed.), World Scientific Publishing Com., Singapore, p. 55.
151. B.O. Roos, K. Andersson, M.P. Fulscher, P. Malmqvist and L. Serrano-Andres (1996) in *Advances in Chemical Physics*, Vol. XCIII, I. Prigogine and S.A. Rice (eds.), John Wiley & Sons, Inc., New York, p. 219.
152. M. Merchan, L. Serrano-Andres, M.P. Fulscher and B.O. Roos (1999) in *Recent Advances in Computational Chemistry*, Vol. 4, K. Hirao (ed.), World Scientific Publishing Com., Singapore, p. 161.
153. P.O. Lowdin (1963) *Rev. Mod. Phys.* **35**, 724.
154. P.O. Lowdin (1964) in *Electronic Aspects of Biochemistry*, B. Pullman (ed.), Academic Press, New York, pg. 167.
155. P.O. Lowdin (1965) *Mutation Res.* **2**, 218.
156. P.O. Lowdin (1965) *Adv. Quantum Chem.* **2**, 213.
157. G.G Sheina, S.G. Stepanian, E.D. Radchenko and Yu.P. Blagoi (1987) *J. Mol. Struct.* **158**, 275.
158. K. Szczepaniak, M. Szczesniak, W.B. Person (1988) *Chem.Phys.Lett.* **153**, 39.
159. K. Szczepaniak and M. Szczesniak (1987) *J. Mol. Struct.* **156**, 29.

160. E.D. Radchenko, A.M. Plohotnichenko, G.G. Sheina and Yu.P. Balgoi (1983) *Biophysics (Engl. Ed.)* **28**, 559.
161. K. Szczepaniak, M. Szczesniak, W. Szajda, W.B. Person and J. Leszczynski (1991) *Can. J. Chem.* **69**, 1705.
162. J. Schiedt, R. Weinkauf, D.M. Neumark and E.W. Schlag (1998) *Chem. Phys.* **239**, 511.
163. A.F. Bell, L. Hecht and L.D. Barron (1997) *J. Chem. Soc., Faraday Trans.* **93**, 553.
164. R.D. Brown, P.D. Godfrey, D.McNaughton and A.P. Pierlot (1989) *Chem. Phys. Lett.* **156**, 61.
165. R.D. Brown, P.D. Godfrey, D.McNaughton and A.P. Pierlot (1988) *J. Am. Chem. Soc.* **110**, 2329.
166. R.D. Brown, P.D. Godfrey, D.McNaughton and A.P. Pierlot (1989) *J. Chem. Soc. Chem. Commun.* 37.
167. Y. Ohta, K. Nishimoto, H. tanaka, Y. Baba and A. Kagemoto (1989) *Bull. Chem. Soc. Jpn.* **62**, 2441.
168. O.E. Kasende, K. Szczepaniak, W.B. Person and T. Zeegers-Huyskens (1997) *J. Mol. Struct.* **435**, 17.
169. P.A. Terpstra, C. Otto and J. Greve (1997) *J. Chem. Phys.* **106**, 846.
170. M.F. Goodman (1995) *Nature* **378**, 237.
171. A. Douhal, S.K. Kim and A.H. Zewail (1995) *Nature* **378**, 260.
172. K. Szczepaniak, W.B. Person, J. Leszczynski and J.S. Kwiatkowski (1998) *Polish J. Chem.* **72**, 402.
173. M. Chahinian, H.B. Seba and B. Ancian (1998) *Chem. Phys. Lett.* **285**, 337.
174. G. Ferenczy, L. Harsanyi, B. Rozsondai and I. Hargittai (1986) *J. Mol. Struct.* **140**, 71.
175. M.T. Chenon, R.J. Pugmire, D.M. Grant, R.P. Panzica and L.B. Townsend (1975) *J. Am. Chem. Soc.* **97**, 4636.
176. J.W. Eastman (1969) *Ber. Bunsen-Ges. Phys. Chem.* **73**, 407.
177. M. Dreyfus, G. Dodin, O. Bensaude and J.E. Dubois (1975) *J. Am. Chem. Soc.* **97**, 2369.
178. D. Shugar and A. Psoda (1990) in *Landolt-Bornstein-New Series: Biophysics of Nucleic Acids*, W. Saenger (ed.), Vol. VII/1d, Springer, Berlin, p. 308.
179. S.G. Stepanian, G.G. Sheina, E.D. Radchenko and Yu. Blagoi (1985) *J. Mol. Struct.* **131**, 333.
180. G.G. Sheina, E.D. Radchenko, S.G. Stepanian and Yu. Blagoi (1986) *Stud. Biophys.* **114**, 123.
181. M.J. Nowak, H. Rostkowska, L. Lapinski, J.S. Kwiatkowski and J. Leszczynski (1994) *J. Phys. Chem.* **98**, 2813.
182. M.J. Nowak, L. Lapinski, J.S. Kwiatkowski and J. Leszczynski (1991) *Spectrochim. Acta* **47A**, 87.
183. M.J. Nowak, H. Rostkowska, L. Lapinski, J.S. Kwiatkowski and J. Leszczynski (1994) *Spectrochim. Acta* **50A**, 10817.
184. M.J. Nowak, L. Lapinski, J.S. Kwiatkowski and J. Leszczynski (1996) *J. Phys. Chem.* **100**, 3527.

185. W.W. Hauswirth and M. Daniels (1971) *Photochem. Photobiol.* **13**, 157.
186. P. Vigny and M. Duquesne (1974) *Photochem. Photobiol.* **20**, 15.
187. A. Suwaiyan, M.A. Morsy and K.A. Odah (1995) *Chem. Phys. Lett.* **237**, 349.
188. M.A. Morsy, A.M. Al-Somali and A. Suwaiyan (1999) *J. Phys. Chem. B* **103**, 11205.
189. H. Ruterjans, E. Kaun, W.E. Hull and H.H. Limbach (1982) *Nucleic Acids Res.* **10**, 7027.
190. M.J. Nowak, K. Szczepaniak, A. Barski and D. Shugar (1978) *Z. Naturforsch. C* **33**, 876.
191. D. Shugar and K. Szczepaniak (1981) *Int. J. Quantum Chem.* **20**, 573.
192. M. Szczesniak, M.J. Nowak, K. Szczepaniak, W.B. Person and D. Shugar (1983) *J. Am. Chem. Soc.* **105**, 5969.
193. S. Chin, I. Scot, K. Szczepaniak and W.B. Person (1984) *J. Am. Chem. Soc.* **106**, 3415.
194. Y.D. Radchenko, G.G. Scheina, N.A. Smorygo and Y.P. Blagoi (1984) *J. Mol. Struct.* **116**, 387.
195. M. Szczesniak, K. Szczepaniak, J.S. Kwiatkowski, K. KuBulat and W.B. Person (1998) *J. Am. Chem. Soc.* **110**, 8319.
196. J.S. Kwiatkowski and J. Leszczynski (1996) *J. Phys. Chem.* **100**, 941.
197. J. Smets, L. Adamowicz and G. Maes (1996) *J. Phys. Chem.* **100**, 6434.
198. R.D. Brown, P.D. Godfrey, D. McNaughton and A.P. Pierlot (1989) *J. Am. Chem. Soc.* **111**, 2308.
199. M. Drefus, O. Bensaude, G. Dodin and J.E. Dubois (1976) *J. Am. Chem. Soc.* **98**, 6338.
200. R.J. McClure and B.M. Craven (1973) *Acta Crystallogr.* **29B**, 1234.
201. R. Kobayashi (1998) *J. Phys. Chem. A* **102**, 10813.
202. D.A. Estrin, L. Paglieri and G. Corongiu (1994) *J. Phys. Chem.* **98**, 5653.
203. A.K. Chandra, M.T. Nguyen and Th. Zeegers-Huyskens (2000) *J. Mol. Struct.* **519**, 1.
204. O.V. Shishkin (1995) *J. Chem. Soc. Chem. Commun.* 1538.
205. O.V. Shishkin (1998) *J. Mol. Struct.* **447**, 1.
206. O. V. Shishkin, L. Gorb and J. Leszczynski (2000) *Chem. Phys. Lett.* **330**, 603.
207. O.V. Shishkin, L. Gorb, P. Hobza and J. Leszczynski (2000) *Int. J. Quant. Chem.*, **80**, 1116.
208. D.M. Hovorun, L. Gorb and J. Leszczynski (1999) *Int. J. Quant. Chem.*, **75**, 245.
209. J. Sponer, J. Florian, P. Hobza and J. Leszczynski (1996) *J. Biomol. Struct. Dyn.* **13**, 827.
210. G.A. Leonard, McAuley-Hecht, T. Brown and W.N. Hunter (1995) *Acta Crystallogr.* **51D**, 136.
211. E.B. Starikov and T. Steiner (1997) *Acta Crystallogr.*, **53D**, 345.
212. P. Hobza, J. Sponer, E. Cubero, M. Orozco and F.J. Luque (2000) *J. Phys. Chem. B* **104**, 6286.
213. L.B. Clark, G.G. Peschel and I. Tinoco Jr. (1965) *J. Phys. Chem.* **69**, 3615.
214. D. Voet, W.B. Gratzer, R.A. Cox and P. Doty (1963) *Biopolymers* **1**, 193.
215. S.F. Mason (1954) *J. Chem. Soc.* 2071.

216. R.F. Stewart and J. Davidson (1963) *J. Chem. Phys.* **39**, 255.
217. T. Inagaki, A. Ito, K. Heida and T. Ho (1986) *Photochem. Photobiol.* **44**, 303.
218. A.F. Fucaloro and L.S. Forster (1971) *J. Am. Chem. Soc.* **93**, 6443.
219. Y. Matsuoks and B. Norden (1982) *J. Phys. Chem.* **86**, 1378.
220. W. Voelter, R. Records, E. Bunnenberg and C. Djerassi (1968) *J. Am. Chem. Soc.* **90**, 6163.
221. J.C. Sutherland and K. Griffin (1984) *Biopolymers* **23**, 2715.
222. R.F. Stewart and L.H. Jensen (1964) *J. Chem. Phys.* **40**, 2071.
223. P.R. Callis, E.J. Rosa and W.T. Simpson (1964) *J. Am. Chem. Soc.* **86**, 2292.
224. L.B. Clark and I. Tinoco (1965) *J. Am. Chem. Soc.* **87**, 11.
225. W.C. Brunner and M.F. Maestre (1975) *Biopolymers* **14**, 555.
226. C.A. Sprecher and W.C. Johnson Jr. (1977) *Biopolymers* **16**, 2243.
227. D.W. Miles, S.J. Hann, R.K. Robins, H. Eyring (1968) *J. Phys. Chem.* **72**, 1483.
228. H.H. Chen and L.B. Clark (1973) *J. Chem. Phys.* **58**, 2593.
229. L.B. Clark (1989) *J. Phys. Chem.* **93**, 5345.
230. L.B. Clark (1990) *J. Phys. Chem.* **94**, 2873.
231. L.B. Clark (1995) *J. Phys. Chem.* **99**, 4466.
232. C. Santhosh and P.C. Mishra (1990) *J. Mol. Struct.* **220**, 25.
233. C. Santhosh and P.C. Mishra (1992) *Proc. Ind. Acad. Sci. (Chem. Sci.)* **104**, 649.
234. P.C. Mishra and K. Jug (1994) *J. Mol. Struct. (Theochem)* **305**, 139.
235. P.-T. Chou, M.L. Martinez, W.C. Cooper, D. McMorrow, S.T. Collins and M. Kasha (1992) *J. Phys. Chem.* **96**, 5203.
236. M.K. Shukla and P.C. Mishra (1998) *Chem. Phys.* **230**, 187.
237. P.R. Callis (1986) *Photochem. Photobiol.* **44**, 315.
238. W. Hug and I. Tinoco (1973) *J. Am. Chem. Soc.* **95**, 2803.
239. H. Ito and Y.J. I'Haya (1976) *Bull. Chem. Soc. Jpn* **49**, 3466.
240. S.K. Srivastava and P.C. Mishra (1980) *J. Mol. Struct.* **65**, 199.
241. A.L. Sobolewski and W. Domcke (2000) *Chem. Phys.* **259**, 181.
242. A.L. Sobolewski and W. Domcke (1999) *Chem. Phys. Lett.* **315**, 293.
243. I.N. Levine, *Quantum Chemistry*, 1991, Prentice-Hall Inc.
244. C.M. Breneman and K.B. Wiberg (1990) *J. Comput. Chem.* **11**, 361.
245. J. Andreasson, A. Holmen and Bo Albinsson (1999) *J. Phys. Chem. B* **103**, 9782.
246. B. Mennucci, A. Toniolo and J. Tomasi (2000) *J. Am. Chem. Soc.* **122**, 10621.
247. W. Rettig and B. Zietz (2000) *Chem. Phys. Lett.* **317**, 187.
248. W. Rettig (1986) *Angew. Chem. Int. Ed. Engl.* **25**, 971.
249. K. Bhattacharyya and M. Chowdhury (1993) *Chem. Rev.* **93**, 507.
250. L.B. Clark (1977) *J. Am. Chem. Soc.* **99**, 3934.
251. L.B. Clark (1994) *J. Am. Chem. Soc.* **116**, 5265.
252. P.C. Mishra (1989) *J. Mol. Struct.* **195**, 201.
253. C. Santhosh and P.C. Mishra (1989) *J. Mol. Struct.* **198**, 327.
254. C. Santhosh and P.C. Mishra (1991) *J. Photochem. Photobiol. A* **61**, 281.
255. D. Theiste, P.R. Callis and R.W. Woody (1991) *J. Am. Chem. Soc.* **113**, 3260.
256. P.R. Callis, B. Fanconi and W.T. Simpson (1971) *J. Am. Chem. Soc.* **93**, 6679.
257. R.W. Wilson and P.R. Callis (1980) *Photochem. Photobiol.* **31**, 323.

258. P. Vigny and J.P. Ballini (1977) in *Excited States in Organic Chemistry and Biochemistry*, B. Pullman, N. Goldblum (eds.), D. Reidel, Boston, pg. 1.
259. R.W. Wilson, J.P. Morgan and P.R. Callis (1975) *Chem. Phys. Lett.* **36**, 618.
260. R. Cammi and J. Tomasi (1995) *J. Comput. Chem.* **16**, 1449.
261. E. Cancès, B. Mennucci and J. Tomasi (1997) *J. Chem. Phys.* **107**, 3032.
262. M. Isaacon (1972) *J. Chem. Phys.* **56**, 1803.
263. R.S. Becker and G. Kogan (1980) *Photochem. Photobiol.* **31**, 5.
264. R.S. Becker, R.V. Bensasson and C. Satel (1981) *Photochem. Photobiol.* **33**, 115.
265. D. Murray and R. Becker (1983) *J. Phys. Chem.* **87**, 625.
266. C. Salet, R. Bensasson and R.S. Becker (1979) *Photochem. Photobiol.* **30**, 325.
267. W.A. Eaton and T.P. Lewis (1970) *J. Chem. Phys.* **53**, 2164.
268. P.R. Callis and W.T. Simpson (1971) *J. Am. Chem. Soc.* **92**, 3593.
269. J.S. Novros and L.B. Clark (1986) *J. Phys. Chem.* **90**, 5666.
270. B.G. Anex, A.F. Fucaloro and J. Durra-Ahmed (1975) *J. Phys. Chem.* **79**, 2636.
271. A. Holmen, A. Broo and B. Albinsson (1994) *J. Phys. Chem.* **98**, 4998.
272. R.F. Stewart and L.H. Jensen (1967) *Acta Crystallogr.* **23**, 1102.
273. R. Taylor and O. Kennard (1982) *J. Am. Chem. Soc.* **104**, 3209.
274. J.S. Novros and L.B. Clark (1986) *J. Phys. Chem.* **90**, 5666.
275. M.K. Shukla and J. Leszczynski (2002) *J. Phys. Chem. A* **106**, 8642.
276. T.Yamada and H.Fukutome (1968) *Biopolymers* **6**, 43.
277. F. Zaloudek, J.S. Novros and L.B. Clark (1985) *J. Am. Chem. Soc.* **107**, 7344.
278. W.C. Johnson Jr., P.M. Vipond and J.C. Girod (1971) *Biopolymers* **10**, 923.
279. A. Kaito, M. Hatano, T. Ueda, S. Shibuya (1980) *Bull.Chem. Soc. Jpn.* **53**, 3073.
280. K. Raksany and I. Foldvary (1978) *Biopolymers* **17**, 887.
281. D.W. Miles, M.J. Robins, R.K. Robins, M.W. Winkley and H. Eyring (1969) *J. Am. Chem. Soc.* **91**, 831.
282. W.H. Inskeep, D.W. Miles and H. Eyring (1970) *J. Am. Chem. Soc.* **92**, 3866.
283. P.R. Callis and W.T. Simpson (1970) *J. Am. Chem. Soc.* **92**, 3593.
284. T.P. Lewis and W.A. Eaton (1971) *J. Am. Chem. Soc.* **93**, 2054.
285. C.C. Bott and T. Kurucsev (1977) *Spectrosc. Lett.* **10**, 495.
286. R.W. Wilson and P.R. Callis (1976) *J. Phys. Chem.* **80**, 2280.
287. A.F. Fucaloro and L.S. Forster (1971) *J. Am. Chem. Soc.* **93**, 6443.
288. D.W. Miles, R.K. Robins and H. Eyring (1967) *Proc. Natl. Acad. Sci.* **57**, 1138.
289. M.K. Shukla and J. Leszczynski (2002) *J. Phys. Chem. A* **106**, 11338.
290. T.M. Korter, D.W. Pratt and J. Kupper (1998) *J. Phys. Chem. A* **102**, 7211.
291. P.-J. Chou and W.C. Johnson Jr. (1993) *J. Am. Chem. Soc.* **115**, 1205.
292. G.J. Brealey and M. Kasha (1955) *J. Am. Chem. Soc.* **77**, 4462.
293. S.F. Boys and F. Bernardi (1970) *Mol. Phys.* **19**, 553.

INDEX

- 1-MeC 212
- 1-MeC:5FU 213
- 1-MeC reduction product 214
- 2-aminopyridine 190
- 6-thioguanine 190
- 6-thiopurine 190
- 8-aminoguanine 19

- A-tetrad 188
- absorption 62
- AcesII program 142
- acid-base pairs 152
- adenine 162, 186
- adenine oxidation products 235
- adenine reduction products 233
- alanine 24
- AM1 81, 92, 93, 95, 98, 101, 102, 104, 108, 111-114, 121-123, 125, 127
- ammonia complexes 145, 146, 151
- apparent surface charge (ASC) 7
- atomic axial tensors (AAT) 58
- atomic polar tensor (APT) 58

- basis set requirements for the coupling constants 143, 144
- basis set superposition error 330
- benzene complexes 155
- bond length alternation (BLA) 28

- C-tetrad 173, 188
- C=O stretching 31
- CASSCF method 254
- cavity surface 11
- charge transfer 87, 104
- charges
 - adenine 262–264
 - cytosine 322, 323
 - guanine 285–288
 - uracil and thymine 301, 302, 304
- chiral solutes 55

- CIS method 254
Complete Active Space (CAS) 141
configuration interaction singles (CIS) 89, 93–95
coupled cluster (CC) theory 140, 142
coupled perturbed Hartree-Fock (CPHF) 19
coupled perturbed Kohn-Sham (CPKS) 19
CPCM 9
cytosine 162, 184, 186, 191, 200
cytosine oxidation products 218
cytosine reduction products 216
- DALTON program 141–143
deMon program 142
density functional theory (DFT) 140, 142, 143
derivatives of solvation free energy 18
DFT on free radicals 215
diamagnetic spin-orbit (DSO) term 138–140
diazine 50
dipolar nuclear spin-spin coupling constant 135
dipole moment 86–88, 90–92, 98, 99, 101–104, 110, 115–117, 119, 122, 125–127
dispersion interactions 14
DMABN 65
DNA polymerase 164
donor-acceptor solutes 26, 44
DPCM 9
dressing effect 88, 100, 103, 110
dual fluorescence 65
- effective Hamiltonian 3
effective NMR spin Hamiltonian 135, 136
electric (hyper)polarizability 34
electronic Hamiltonian in magnetic field 138
electronic transitions
 adenine 259–267
 base pairs 325–329
 cytosine 310–316
 guanine 274–281
 uracil and thymine 290–297
electrostatic interactions 6
emission 62
energy properties 22, 23
EOMCC method 254
EPR/ENDOR 213
escaped charge 12

excitation 61

Fermi contact (FC) term 138, 139, 145, 152

first derivatives with respect to nuclear coordinates 20

first hyperpolarizability 89, 92, 93, 98, 111, 112, 117, 120, 121, 123, 125,
126

focused models 2

formic acid dimer 146

free energy functional 3

frequency-dependent response 36

G-tetrad 164, 168, 173, 175, 178, 180, 186–188, 190, 191

G:C base pair 212

generalized Fock 17

geometries

adenine 267–271

cytosine 316–322

guanine 281–285

uracil and thymine 297–304

GEPOL 12

GIAO 49

glycine 24

ground state properties 255–258

guanine 162, 173, 178, 186, 191

guanine oxidation products 231

guanine reduction products 228

guanosine 168, 169

harmonic approximation 29

helium dimer 154–155

heptads 178, 182, 190

hexads 178, 190

hydrogen bonded clusters 51

hydrogen cyanide complexes 147, 149, 150, 151

hydrogen fluoride complexes 145, 148, 155

hyperpolarizabilities 38

hypoxanthine 186

IEF 10

indirect nuclear spin-spin coupling constant as energy derivative 137

indirect nuclear spin-spin coupling constant, definition of 136

interaction energy 329–331

intermolecular coupling constants

carbon-nitrogen 133, 151, 153

fluorine-fluorine 145, 148, 155

fluorine-proton 145, 148, 155

- indirect spin-spin 132
- nitrogen-chlorine 150, 151
- nitrogen-nitrogen 133, 147, 150–154
- oxygen-oxygen 144, 147, 149
- phosphorus-nitrogen 133, 154
- phosphorus-phosphorus 150
- phosphorus-proton 133, 154
- proton-carbon 133, 145, 156
- proton-nitrogen 133, 134, 145, 151–154
- proton-oxygen 144, 145
- proton-proton 147, 152, 156
- IR intensities 29
- IR spectra 29
- ketones 31
- linear response theory 37, 137, 138
- local field effect 87, 103, 104, 110, 125
- magnetic susceptibility 47
- molecular dynamics simulation 189, 192
- multiconfigurational self consistent field (MCSCF) 140–142
- multiplicative scheme 81, 99, 101–103, 105–114, 119–121, 125–127
- NMR shielding constant as energy derivative 137
- NMR shielding constant, definition of 135
- non-planar thymine products 240
- nuclear contribution to (hyper)polarizabilities 43
- nuclear shielding 47
- nucleic acid database 162
- nucleic acids 133, 151–154
- octads 178, 180, 190
- optical rotation (OR) 54
- paramagnetic spin-orbit (PSO) term 138, 139, 145
- PCM 3
- pentads 163, 178, 190, 192
- phototautomerism
 - cytosine 313–315
 - guanine 288–290
 - uracil 305–310
- planar radicals 215
- polarizability 87, 92, 98, 102, 103, 107, 108, 109, 125, 126
- protein data bank 162, 200
- proteins 133, 151, 154

- push-pull molecules 26, 44
- radiation chemistry of DNA 242
- ratings of DFT results on free radical structures 238
- reaction potential 3
- repulsion interactions 14
- response properties 22
- Restricted Active Space (RAS) 141
- RPA 37
- second harmonic generation (SHG) 85, 87, 110, 112, 114, 121–123
- second order polarization propagator approach (SOPPA) 140
- self-consistent field (SCF) 140–142
- solute-solvent non-equilibrium 30, 62
- solvent accessible surface 12
- solvent effects on isomers 24
- solvent excluding surface 12
- solvent interaction operators 16
- spin-dipole (SD) term 138, 139
- spin densities 210
- sum-over-states expression 137
- summation-over-states (SOS) 88–90
- T-tetrad 175, 186, 188
- TDDFT method 254
- TDHF (time-dependent Hartree-Fock) 91–93, 98, 104–107, 109–114, 120, 122, 123, 125–127
- tesserae 12
- tetrads 163, 164, 168, 169, 173, 175, 176, 183, 188, 190, 192
- through-space mechanism 132
- thymine 162, 200
- TICT 65
- transition properties 22
- triads 163, 169, 170–172, 174, 182, 183, 185, 186, 190, 192, 195, 199
- thymine oxidation products 225
- thymine reduction products 221
- U-tetrad 173, 187, 188
- uracil 162
- urea 38
- UV-induced DNA damage 250
- van der Waals surface 12
- vector potential of magnetic field 138, 139
- Vibrational Circular Dichroism (VCD) 58
- vibrational frequencies 29

- vibrational polarizability 44
- water complexes 144–146, 149
- xenon dimer 155
- ZINDO 81, 93, 95, 114–116, 118–121, 126, 127