

A Novel Theoretical Model for Molecular Recognition of Multiple-Site Interacting Systems Using Density Response Functions

K. R. S. Chandrakumar and Sourav Pal*

Physical Chemistry Division, National Chemical Laboratory, Pune 411 008, India

Received: December 31, 2000; In Final Form: March 30, 2001

We present the first theoretical study of molecular recognition of systems interacting via multiple sites, using the density response functions. The model, so-called, localized reactive model derived on the basis of these functions, has been employed for studying the interactions between Watson–Crick type DNA base pairs and the effect of substitution of various groups in the adenine–uracil base pairs. The estimated interaction energies for these base pairs agree remarkably with the corresponding literature and experimental values. It confirms the applicability of the present model to such complex multiple-site interaction cases and can be used as potential predictive tool for the understanding of such general molecular interaction or recognition processes.

Introduction

Molecular recognition is defined by the energy and the information involved in the binding and selection of substrates by a given receptor molecule. This amounts to a generalized principle extending over energetic as well as geometric features represented by the “lock–key” steric fit concept of Emil Fischer.¹ These complexes possess well-defined structural, conformational and various types of interactions that present different degrees of strength, directionality, and dependence on distance. It may be formed through the electrostatic forces, hydrogen bonding, van der Waals interaction, donor–acceptor interaction, and metal ion coordination. The interaction can also be accompanied by many simultaneous multibonding interactions of different kinds. These molecular interactions form the basis of the highly specific recognition processes that occur in biological systems, such as substrate binding to a receptor protein, enzymatic reactions, assembling of multiprotein complexes, etc.² In particular, the hydrogen bond (H-bond) determines the magnitude and the nature of the interactions of the biomolecules and is consequently responsible for the important unique properties of nucleic acids, DNA and RNA.³ The complementary Watson–Crick (WC) base pairs in DNA involve adenine (Ade) with thymine (Thy) or uracil (Ura) and guanine (Gua) with cytosine (Cyt).^{2,3} From a practical point of view, quantitative knowledge of the energies and geometry of these multiple molecular interactions is particularly important for the development and validation of the studies for the design of artificial receptor molecules. Owing to the importance of these molecular recognition processes, many theoretical efforts have been devoted to interpret the nature and the stabilization of the DNA base pair interactions using the semiempirical and ab initio quantum chemical methods. For a recent review, covering the current state of the field and discussing the extensive progress from semiempirical to ab initio methods, see ref 4. Since the small size and the existence of the experimental data, these base pairs (Gua–Cyt and Ade–Ura) have been chosen as prototypes of DNA structure in theoretical investigations. The quantum chemical treatment for these studies has provided some signifi-

cant and useful interesting information about their electronic structure and the stability of the weakly held molecular complexes.^{4–8} Although, there are several other theoretical models that have been proposed to explain the molecular recognition processes (for, e.g., electrostatic potential,⁹ molecular similarity indexes,¹⁰ electric field,¹¹ etc.), it still remains as a qualitative model. The quantitative assessment of structural and interaction features in terms of some other suitable alternative method will represent a challenging endeavor to the evaluation of the molecular complementarity or similarity recognition processes. Hence, it is natural to think about the formulation of some concrete and extremely appealing functions to describe the interaction between molecules. This focuses our principle aim of this Letter, and our approach is based on the density-based response functions, called local and global reactivity descriptors (LRD and GRD), that are derived from density functional theory (DFT).¹² DFT has provided the theoretical basis for the concepts like electronic chemical potential, electronegativity, and hardness, collectively known as global chemical reactivity descriptors¹³ and also for the local reactivity descriptors,¹⁴ like the Fukui function, local softness, etc. Using these response functions or descriptors, the interaction energy (IE) expression has been derived for the single site interaction, in the context of explaining the HSAB principle, by Gazquez and Mendez.¹⁵ In a recent study,¹⁶ we have made a critical study on these descriptors and subsequently proposed a quantitative approach to calculate the IE using the LRD and GRD for the weak interaction cases. Later, we have also proposed a more generalized IE expression for the single and multiple interaction cases and for the systems where the cooperative and noncooperative effects are important.¹⁷ From the viewpoint of the proposed model, we will present the description of the model and explain the basic questions in obtaining the quantitative aspects of the nature of molecular interaction, specifically for the case of DNA base pair interactions. We will also study the substitution effect on the Ade–Ura base pair interaction. This work represents the first such study in the literature, in which the DFT method based descriptors have been used to describe the interaction pattern in a quantitative way for the complex multiple interactions cases.

* Corresponding author. E-mail: pal@ems.ncl.res.in.

Theoretical Background

By using the energy perturbation method, as a function of the number of electrons and external potential, the local HSAB principle expresses the interaction energy between two systems A and B, interacting through a pair of sites of the two systems through the following equation.

$$(\Delta E_{\text{int}})_{\text{AK}} = \frac{-(\mu_{\text{A}} - \mu_{\text{B}})^2}{2} \left(\frac{S_{\text{A}} S_{\text{B}}}{S_{\text{A}} + S_{\text{B}}}_{\nu} \right) + \frac{-1}{2} \left(\frac{\lambda}{S_{\text{A}} + S_{\text{B}}}_{\mu} \right) \quad (1)$$

where μ_{A} and μ_{B} are chemical potentials of A and B, respectively, S_{A} and S_{B} are the local softnesses of A and B, respectively. We can further approximate eq 1 by replacing S_{A} or/and S_{B} by condensed local softnesses of the reacting sites of A and B. In our earlier work,¹⁶ we have defined the parameter λ as the change in the electron densities at the interacting site before and after the interaction process. This change will give the effective number of valence electrons that have participated in the interaction process. Thus, an expression for the term λ can be written as the difference of electron densities of the system A before and after the interaction,

$$\lambda_{\text{A}} = \sum_{i=1}^M \rho_{\text{Ai}}^{\text{eq}} - \sum_{i=1}^M \rho_{\text{Ai}}^0 \quad (2)$$

Alternately, the term λ can be defined as the difference of electron densities for the system B,

$$\lambda_{\text{B}} = \sum_{j=1}^N \rho_{\text{Bj}}^{\text{eq}} - \sum_{j=1}^N \rho_{\text{Bj}}^0 \quad (3)$$

where the first terms of the right-hand side of eqs 2 and 3 refer to the sum of the electron densities of each atom in A and B in the molecule AB at equilibrium, respectively, and the second terms in eqs 2 and 3 refer to the electron densities of each atom in the isolated systems A and B, respectively. The interaction between systems A and B is assumed to take place in two steps. In the first step, the interaction takes place at constant external potential through the equalization of chemical potential (ΔE_{ν}). This is represented by the first term of eq 1. In the second step, A and B evolve toward the equilibrium state through changes in the electron density of the global system produced by making changes in the external potential at constant chemical potential (ΔE_{μ}), as represented in the second term.

$$\Delta E_{\text{int}} = \Delta E_{\nu} + \Delta E_{\mu}$$

Very recently, we have proposed two models, namely localized reactive model (LRM) and smeared reactive model (SRM) to describe the interactions between molecular systems taking place on multiple sites, within the framework of DFT.¹⁷ For the details of the derivation for the models, LRM and SRM, one can refer to our recent work.¹⁷ In LRM, it is assumed that the molecules are interacting through pairs of reactive centers of the systems A and B, which are decoupled. Here one assumes that although the reaction proceeds simultaneously through multiple reaction centers, the IE can be calculated in a decoupled manner. Each interaction is assumed to be localized around the particular sites of reaction. Hence, the interaction energy for the complex AB, may be represented as the sum of interaction energies arising from many interaction modes of A and B (X–K, Y–L, Z–M, etc.). Thus, the net interaction is obtained as a logical extension of the single-site local HSAB principle to multiple sites by assuming the interaction occurs in a decoupled manner and by

assuming the additivity of energy. According to this model, the IE expression is given as

$$\Delta E_{\text{int}} \approx \frac{-(\mu_{\text{A}} - \mu_{\text{B}})^2}{2} \left(\sum_{\text{XK, YL, ZM, ...}} \frac{\text{CSA}_{\text{X}} \text{CSB}_{\text{K}}}{\text{CSA}_{\text{X}} + \text{CSB}_{\text{K}}} \right) + \frac{-\lambda}{2} \left(\sum_{\text{XK, YL, ZM, ...}} \frac{1}{\text{CSA}_{\text{X}} + \text{CSB}_{\text{K}}} \right)_{\mu} \quad (4)$$

where, CSA_{X} and CSB_{K} are the condensed local softness of the reactive atoms X and K, of the systems A and B, respectively. The first and second terms are the energy change at constant external potential (ΔE_{ν}) and chemical potential (ΔE_{μ}), respectively. In case a group of immediate neighboring atoms around the reacting sites contribute, the local softness of each atomic center in a site can be replaced by the corresponding group softness in eq 4. Another model, SRM, considers that the interaction between the molecular systems is assumed to take place only through a set of atoms that are directly involved in the interaction processes. The set of atoms includes only the reactive atoms of A and B and these reactive atoms should be adjacent or directly connected to each other. In the present study, the reactive atoms of the purine and pyrimidine bases are not directly connected to each other and hence SRM is not applicable to calculate the IE for the DNA base pairs. The computation of the terms that are present in the above equations can be found elsewhere.^{15–17}

Computational Details

The molecular geometries of the purine and pyrimidine bases (Gua, Cyt, Ade, and Ura and the uracil derivatives) were completely optimized using the ab initio Hartree–Fock (HF) quantum chemical calculations at the level of the 6-31G(d,p) basis set without any symmetry constraints. Systematic substitution effects on the stability of the Ade–Ura base pair interaction is studied, by substituting –H, –CH₃, –CF₃, –CN, and –NH₂ groups at the 5-position of uracil, as shown in Figure 1. The ab initio calculations were performed using the GAMESS¹⁸ system of programs on an IRIX-6.2 silicon graphics workstation. The parameter λ was calculated using eqs 2 and 3, through the Mulliken population scheme.¹⁹ The optimized geometry (HF/631G**) of the base pairs Gua–Cyt and Ade–Ura and its derivatives has been taken from the refs 5 and 7, respectively.

Results and Discussion

We will now briefly describe how the density response functions are closely related to the molecular recognition or the molecular interaction processes. To understand the subtle aspects of these molecular recognition processes, we consider the behavior of electron density and its variation with respect to the small perturbations due to the other molecular interactions. Generally, the interaction between the molecules and the prediction of changes in reactivity and selectivity of the atom centers of a given molecule can be described as the perturbation of the electron clouds of the interacting systems. Because of these electron cloud interactions, the electron density will be redistributed among themselves, which can lead to the specific strong, weak, or van der Waal bond type interactions. The knowledge of the perturbations of these electron clouds of the two interacting systems, such as chemical potential, softness, Fukui function, and local softness may give an insight about the nature of H-bond and its stability. Thus, it can be seen that the molecular recognition processes are closely related to these

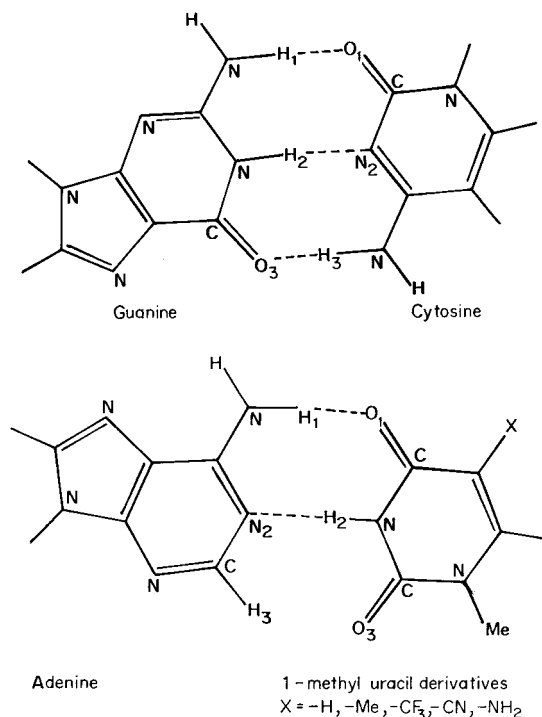


Figure 1. Multiple-site hydrogen bonding interactions between Watson–Crick type base pairs, guanine–cytosine and adenine–uracil. The bold and big letters are the reactive atoms. X = -H, -CH₃, -CF₃, -CN, -NH₂.

TABLE 1: Global Properties of the Systems, the Chemical Potential and Global Softness (Values in Atomic Units)

systems	chemical potential	global softness
Gua	-0.094	4.765
Cyt	-0.105	5.253
Ade	-0.106	4.564
H-Ura	-0.121	5.053
CH ₃ -Ura	-0.093	6.120
CF ₃ -Ura	-0.081	7.510
CN-Ura	-0.158	5.379
NH ₂ -Ura	-0.103	6.173

descriptors and hence these reactivity descriptors could be a better tool to describe the various types of interactions. In particular, the descriptors Fukui function and local softness are the important local descriptors that have a direct relationship with the change in reactivity or selectivity of molecules. The Fukui function has been found to be intimately related to the chemical reactivity of a system.^{16,20}

The global and local properties, chemical potential (CP), softness, and condensed local softness (CLS) for all systems are tabulated in Tables 1 and 2, respectively. The CPs of all molecules are not too different from each other, but their global and local softness values are substantially different from each other. The substitution of several groups on the uracil molecule has a significant effect in determining the value of the local and global descriptor values. In general, the GRD and LRD of Cyt and Ura are greater than that of Gua and Ade. On examining the value of λ from Table 3, it can be seen that the λ value for the base pair Gua-Cyt is considerably greater than that of other base pairs.

Let us now examine the IE values obtained through the model LRM (Table 3) along with the available experimental and theoretical IE values. The total IE of the base pairs ranges from -21.3 to -8.09 kcal/mol. The Gua-Cyt base pair is the most stable base pair, and its IE is -21.25 kcal/mol. The IE value is in excellent agreement with the experimental value of -21.0

TABLE 2: Condensed Local Softness of the Reactive Atoms (Values in Atomic Units)^a

system	condensed local softness		
	RA1	RA2	RA3
Gua	2.09	0.31	0.38
Cyt	0.30	0.18	2.50
Ade	0.29	0.27	0.48
H-Ura	0.42	0.30	2.30
CH ₃ -Ura	0.42	0.36	0.07
CF ₃ -Ura	0.60	0.45	2.93
CN-Ura	0.60	0.26	1.80
NH ₂ -Ura	0.64	0.36	0.58

^a Sx⁺ and Sx⁻ are calculated for H and for O, N, respectively, for different pairs of the system. RA1, RA2, and RA3 refer to the reactive atoms of the each system and the numbering is given in Figure 1.

TABLE 3: ΔE_{μ} and ΔE_{ν} and Total Interaction Energies for the Watson–Crick Type Base Pairs, Calculated through the Localized Reactive Model (Energy Values in kcal/mol)^a

system	λ	ΔE_{ν}	ΔE_{μ}	ΔE_{total}	ΔE^{ET}	
					HF ^b	MP2 ^b
Gua-Cyt	0.0243	-0.03	-21.22	-21.25	-25.5	-25.8
Ade-H-Ura	0.0106	-0.05	-11.73	-11.78	-9.4	-12.5
Ade-CH ₃ -Ura	0.0072	-0.02	-10.93	-10.95	-9.5	-12.5
Ade-CF ₃ -Ura	0.0090	-0.15	-7.94	-8.09	-9.9	-13.3
Ade-CN-Ura	0.0133	0.59	-14.49	-15.08	-10.3	-13.6
Ade-NH ₂ -Ura	0.0071	-0.01	-8.04	-8.05	-9.0	-12.1

^a The λ values are given in atomic units. ΔE^{ET} is the available theoretical IE values. ^b HF and MP2 are the available theoretical IE values, calculated by HF and MP2 methods using 6-31G* basis set. See refs 5–7.

kcal/mol⁸ and with the MP2 result -25.8 kcal/mol.⁵ Similarly, for the case Ade-Ura, the IE is -11.8 and this value is quite close to the experimental value⁸ -13.0 kcal/mol and with the MP2 value -12.5 kcal/mol.⁷ The IEs for the substituted Ade-Ura base pair are also in agreement with the available theoretical results, and the variation of IEs are found on the order of ± 2 kcal/mol. It could be due to the limited accuracy that is present in the calculation of the parameter λ as well as in the reactivity descriptors. Nevertheless, such a variation in the estimated interaction energies is very systematic and consistent with the available data. The accuracy of the calculation can be improved by choosing a larger size of the basis set and including the correlation effects. Although the GRD and LRD of substituted uracil derivatives are significantly different from each other, the IE values of these complexes that are calculated by LRM differ marginally in the range of a few kcal/mol. It is also in complete agreement with other theoretical results. This fact will allow us to permit the effect of changes in any of the atomic environment of a molecule and eventually can help in understanding the molecular interactions. We have also computed the IE for the above systems through the other model SRM, and the IE are considerably less than the actual values (the IEs of Gua-Cyt and Ade-Ura are -2.6 and -2.7 kcal/mol, respectively). This fact actually emphasizes the applicability and the validity of the LRM for such complex multiple-site interaction cases. A closer inspection on the contribution of the energy terms to the total interaction energy tabulated in Table 3 reveals that the most important component of the interaction energy arises from the ΔE_{μ} term. It also emphasizes that this term alone can explain the nature and stability of the complexes and it provides a driving force for the formation of the complexes. Hence, the charge redistribution process at constant chemical potential can be considered as a decisive modulating factor in determining the strength of the H-bonded and other

types of complexes that have been considered in this present study. The effect of the strong directional character and relative arrangement of atoms in the actual interaction is accomplished by the factor λ , as defined in eqs 2 and 3. The present model is sensitive to the basis set and the method that are employed for the computation of each quantity that present in eq 4. Nevertheless, despite the approximate nature of the model, the model can describe the essential features of the molecular interaction without any loss of generality. Hence, our methodology could be a main predictive tool for the general molecular interactions beyond the simple to the supramolecular chemistry. In our future work, we will be extending our methodology to address a wide range of some interesting problems such as the metal cation coordination with amino acids, nucleic acids, the effect of water molecular interaction in base pairs, cation- π interaction in proteins, the general hydrophilic and hydrophobic ligand-receptor complexes, etc.

Acknowledgment. K.R.S.C. acknowledges the Council of Scientific and Industrial Research, New Delhi, for a Senior Research Fellowship. We are thankful to Prof. J. Sponer, for providing some of the optimized coordinates of the DNA base pair.

References and Notes

- (1) (a) J. M. Lehn, *Supra Molecular chemistry: Concepts and perspectives*; VCH Publications: Weinheim, 1995. (b) Desiraju, G. R.; Steiner, T. *Weak hydrogen Bond*; In *Structural Chemistry and Biology*; Oxford University: Oxford, U.K., 1999.
- (2) (a) Saenger, W. *Principles and Nucleic Acid Structure*; Springer-Verlag: New York, 1984. (b) *Nucleic Acids In Chemistry and Biology*; Blackburn, G. M., Gait, M. J., Eds.; Oxford University Press: Oxford, U.K., 1990.
- (3) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 737.
- (4) (a) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, 99, 3247 and references therein. (b) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dynam.* **1996**, 14, 117.
- (5) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, 100, 1965.
- (6) (a) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed. Engl.* **1999**, 38, 2942. (b) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, 116, 2493. (c) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, 118, 7574. (d) Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, 101, 4851.
- (7) Kawahara, S.; Wada, T.; Kawauchi, S.; Uchimaru, T.; Sekino, M. *J. Phys. Chem. A* **1999**, 103, 8516.
- (8) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. R. *Biopolymers* **1979**, 18, 1149.
- (9) Pundlik, S. S.; Gadre, S. R. *J. Phys. Chem. B* **1997**, 101 (46), 9657.
- (10) *Ibid.* **1996**, 101, 3298.
- (11) Ponec, R. *Top. Curr. Chem.* **1995**, 174, 1. (b) Idem. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 554.
- (12) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press: New York, 1989.
- (13) (a) Pearson, R. G. *J. Am. Chem. Soc.* **1963**, 85, 3533. (b) *Chemical Hardness Struct. Bonding*; Sen, K. D., Ed.; Springer-Verlag: Berlin, 1993; Vol. 80. (c) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, 105, 7512. (d) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, 68, 3801. (e) Pearson, R. G. *J. Am. Chem. Soc.* **1985**, 107, 6801.
- (14) (a) Parr, R. G.; Yang, W. *J. Am. Chem. Soc.* **1984**, 106, 4049. (b) Yang, Y.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 821, 6723. (c) Yang, W.; Mortier, W. J. *J. Am. Chem. Soc.* **1986**, 108, 5708.
- (15) (a) Gazquez, J. L.; Mendez, F. *J. Phys. Chem.* **1994**, 98, 4591. (b) Gazquez, J. L.; Mendez, F. *J. Am. Chem. Soc.* **1994**, 116, 9298.
- (16) Pal, S.; Chandrakumar, K. R. S. *J. Am. Chem. Soc.* **2000**, 122, 4145 and references therein.
- (17) Chandrakumar, K. R. S.; Pal, S. Communicated
- (18) Schmidt, M. W.; et al. GAMESS, General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, 14, 1347.
- (19) Mulliken, R. S. *J. Chem. Phys.* **1955**, 23, 1833.
- (20) (a) Geerlings, P.; De Proft, F.; Langenaeker, W. *Adv. Quantum Chem.* **1999**, 33, 303. (b) Yang, W.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 6723.