



NATO Science for Peace and Security Series - B:
Physics and Biophysics

Biophysics and the Challenges of Emerging Threats

Edited by
Joseph D. Puglisi



Springer



*This publication
is supported by:*

The NATO Science for Peace
and Security Programme

Biophysics and the Challenges of Emerging Threats

NATO Science for Peace and Security Series

This Series presents the results of scientific meetings supported under the NATO Programme: Science for Peace and Security (SPS).

The NATO SPS Programme supports meetings in the following Key Priority areas: (1) Defence Against Terrorism; (2) Countering other Threats to Security and (3) NATO, Partner and Mediterranean Dialogue Country Priorities. The types of meeting supported are generally "Advanced Study Institutes" and "Advanced Research Workshops". The NATO SPS Series collects together the results of these meetings. The meetings are coorganized by scientists from NATO countries and scientists from NATO's "Partner" or "Mediterranean Dialogue" countries. The observations and recommendations made at the meetings, as well as the contents of the volumes in the Series, reflect those of participants and contributors only; they should not necessarily be regarded as reflecting NATO views or policy.

Advanced Study Institutes (ASI) are high-level tutorial courses intended to convey the latest developments in a subject to an advanced-level audience

Advanced Research Workshops (ARW) are expert meetings where an intense but informal exchange of views at the frontiers of a subject aims at identifying directions for future action

Following a transformation of the programme in 2006 the Series has been re-named and re-organised. Recent volumes on topics not related to security, which result from meetings supported under the programme earlier, may be found in the NATO Science Series.

The Series is published by IOS Press, Amsterdam, and Springer, Dordrecht, in conjunction with the NATO Public Diplomacy Division.

Sub-Series

| | | |
|----|--|-----------|
| A. | Chemistry and Biology | Springer |
| B. | Physics and Biophysics | Springer |
| C. | Environmental Security | Springer |
| D. | Information and Communication Security | IOS Press |
| E. | Human and Societal Dynamics | IOS Press |

<http://www.nato.int/science>

<http://www.springer.com>

<http://www.iospress.nl>



Series B: Physics and Biophysics

Biophysics and the Challenges of Emerging Threats

Edited by

Joseph D. Puglisi

Stanford University
SMRL & Dept. of Structural Biology
D105A Fairchild Science Building
Stanford CA
USA



Springer

Published in Cooperation with NATO Public Diplomacy Division

Proceedings of the NATO Advanced Study Institute on
Biophysics and the Challenges of Emerging Threats
Erice, Sicily, Italy
19–30 June 2007

Library of Congress Control Number: PCN applied for

ISBN 978-90-481-2367-4 (PB)
ISBN 978-90-481-2366-7 (HB)
ISBN 978-90-481-2368-1 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved

© 2009 Springer Science+Business Media B.V.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

TABLE OF CONTENTS

| | |
|--|-----|
| List of Contributors | vii |
| A Simple Model for Protein Folding | 1 |
| <i>Eric R. Henry and William A. Eaton</i> | |
| Complementarity of Hydrophobic/Hydrophilic Properties in Protein–Ligand Complexes: A New Tool to Improve Docking Results..... | 21 |
| <i>Timothy V. Pyrkov, Anton O. Chugunov, Nikolay A. Krylov, Dimitry E. Nolde, and Roman G. Efremov</i> | |
| Structures of CVNH Family Lectins | 43 |
| <i>Angela M. Gronenborn</i> | |
| Biophysical Approaches to Study DNA Base Flipping..... | 51 |
| <i>Saulius Klimašauskas, Zita Liutkevičiūtė, and Dalia Daujotytė</i> | |
| The Diversity of Nuclear Magnetic Resonance Spectroscopy | 65 |
| <i>Corey W. Liu, Viktor Y. Alekseyev, Jeffrey R. Allwardt, Alexander J. Bankovich, Barbara J. Cade-Menun, Ronald W. Davis, Lin-Shu Du, K. Christopher Garcia, Daniel Herschlag, Chaitan Khosla, Daniel A. Kraut, Qing Li, Brian Null, Joseph D. Puglisi, Paul A. Sigala, Jonathan F. Stebbins, and Luca Varani</i> | |
| Improved Dye Stability In Single-Molecule Fluorescence Experiments | 83 |
| <i>Colin Echeverría Aitken, R. Andrew Marshall, and Joseph D. Puglisi</i> | |
| The Evaluation of Isotope Editing and Filtering for Protein–Ligand Interaction Elucidation by NMR | 101 |
| <i>Ian M. Robertson, Leo Spyropoulos, and Brian D. Sykes</i> | |

| | |
|--|-----|
| Ribosome: An Ancient Cellular Nano-Machine For Genetic Code Translation | 121 |
| <i>Ada Yonath</i> | |
| Course Abstracts and Posters | 157 |
| Author Index | 179 |

LIST OF CONTRIBUTORS

William A. Eaton National Institutes of Health, Laboratory of Chemical Physics Chief, NIDDK, NIH, Bldg 5, Room 104, Bethesda, MD 20892-0520 USA, E-mail: eaton@helix.nih.gov

Roman G. Efremov Russian Academy of Sciences, Institute of Bioorganic Chemistry, Molecular Modeling Group, 16/10 Ul. Miklukho-Maklaya, GSP 117997, Moscow V-437, Russia, E-mail: efremov@nmr.ru

Angela M. Gronenborn University of Pittsburgh, Department of Structural Biology, 3501 Fifth Avenue, 1051 BST3, Pittsburgh PA 15213 USA, E-mail: amg100@pitt.edu

Saulius Klimasauskas Institute of Biotechnology, V. Graiciuno 8, 2028, Vilnius, Lithuania, E-mail: klimasau@ibt.lt

Corey W. Liu Stanford University, Stanford Magnetic Resonance Laboratory, D105 Fairchild Science Building, Stanford, CA 94305-5126 USA, E-mail: liuc@stanford.edu

Joseph D. Puglisi Stanford University, SMRL & Dept. of Structural Biology, D105A Fairchild Science Building, Stanford, CA 94305-5126 USA, E-mail: puglisi@stanford.edu

Brian D. Sykes University of Alberta, Department of Biochemistry, Edmonton, Alberta, Canada T6G 2H7, E-mail: brian.sykes@ualberta.ca

Ada Yonath The Weizmann Institute Of Science, Department Of Structural Biology, Helen & Milton A. Kimmelman Building, Rehovot, Israel. 76100, E-mail: ada.yonath@weizmann.ac.il

A SIMPLE MODEL FOR PROTEIN FOLDING

ERIC R. HENRY AND WILLIAM A. EATON*

*Laboratory of Chemical Physics, National Institute of Diabetes
and Digestive and Kidney Diseases, National Institutes of Health,
Bethesda, MD 20892-0520, USA*

**Corresponding author: e-mail: eaton@helix.nih.gov*

Abstract We describe a simple Ising-like statistical mechanical model for folding proteins based on the α -carbon contact map of the native structure. In this model residues can adopt two microscopic states corresponding to the native and non-native conformations. In order to exactly enumerate the large number of possible configurations, structure is considered to grow as continuous sequences of native residues, with no more than two sequences in each molecule. Inter-residue contacts can only form within each sequence and between residues of the two native sequences. As structure grows there is a tradeoff between the stabilizing effect of inter-residue contacts and the entropy losses from ordering residues in their native conformation and from forming a disordered loop to connect two continuous sequences. Folding kinetics are calculated from the dynamics on the free energy profile, as in Kramers' reaction rate theory. Although non-native interactions responsible for roughness in the energy landscape are not explicitly considered in the model, they are implicitly included by determining the absolute rates for motion on the free energy profile. With the exception of α -helical proteins, the kinetic progress curves exhibit single exponential time courses, consistent with two state behavior, as observed experimentally. The calculated folding rates are in remarkably good agreement with the measured values for the 25 two-state proteins investigated, with a correlation coefficient of 0.8. With its coarse-grained description of both the energy and entropy, and only three independently adjustable parameters, the model may be regarded as the simplest possible analytical model of protein folding capable of predicting experimental properties of specific proteins.

Introduction

A large number of experimental studies over the past decade have established that for most single domain proteins containing ~ 100 residues or less only two populations of molecules are observed in both equilibrium and kinetic studies. These populations correspond to the folded state and an ensemble

of widely varying unfolded structures that collectively make up the unfolded state. As a result of this two-state behavior, there is simply an exchange between folded and unfolded populations as the folding reaction proceeds, so that all measured properties can be described as a simple linear combination of folded and unfolded state properties. Provided the inter-conversion among the conformations of the unfolded state is fast compared to the folding rate, any observed property that distinguishes the two populations will exhibit a perfectly exponential time course. The surprise from recent theoretical work is that highly simplified statistical mechanical models are remarkably successful in predicting both two-state behavior and protein folding rates [1–8]. In this paper, we describe one such model and use it to predict the number of thermodynamic states and relative rates of protein folding.

The basic assumptions of the model were motivated by three major developments. The first was the verification by Socci, Onuchic and Wolynes [9] of the energy landscape ideas of Bryngelson and Wolynes that the kinetics of folding could be described by diffusion on a low-dimensional free energy surface and that global order parameters, such as the number of residues ordered in the native conformation or the number of native contacts, could serve as reaction coordinates [10, 11]. The second was the observation by Plaxco et al. of a significant correlation between the folding time and a simple measure of the topology of the native structure – the mean separation in sequence of contacting residues, which they called the contact order [12, 13]. This finding had important implications for theoretical models. It suggested that the contact map of the native protein is sufficient structural information for describing folding kinetics, enormously simplifying the problem by making unnecessary the explicit consideration of non-native interactions [14]. It also suggested that a simple energy function would suffice.

The third key result was the development of Ising-like models, beginning with the work of Zwanzig, Szabo and Bagchi [15, 16], and the extension by Muñoz et al. to explain equilibrium and kinetic data on the formation of a 16-residue β -hairpin from protein GB1 [3, 4, 17]. To simplify the partition function they assumed that a residue could be in the native state only if it is contained in a continuous stretch of native residues, that molecules never contain more than one such stretch, and that inter-residue interactions only occur if all intervening residues in the sequence are in a native conformation. The importance of this so-called single-sequence approximation [18] is that it dramatically reduces the number of configurations that must be enumerated, from 2^N to $N(N+1)/2 + 1$ for an N -residue polypeptide. The kinetic equations could therefore be solved by standard matrix methods and both the equilibrium and kinetic parameters could be readily optimized to fit experimental data. The model produced a free energy profile with only two minima, explaining the experimentally observed two-state behavior and single-exponential time course for the β -hairpin. According to the model, the energy of the transition

state ensemble is lower than the energy of the unfolded state, accounting for the observed negative activation energy for folding.

An interesting feature of the β -hairpin analysis was that it used the same physical picture that would be necessary to describe folding of a single domain protein – namely the tradeoff between stabilization from inter-residue interactions and the loss of conformational entropy associated with ordering the residues in their native conformation. This is shown schematically in Fig. 1. Motion along the reaction coordinate is first uphill in free energy due to the loss of entropy without sufficient compensation by inter-residue contacts. At the top of the hill (the transition state) the stabilizing interactions from adding residues are roughly equal to the entropy loss of ordering additional residues. Past this point stabilization from inter-residue interactions more than compensates for the conformational entropy loss and motion on the reaction coordinate is downhill in free energy in a rapid run to the native conformation. If contacts are mainly local in sequence, stabilizing interactions are realized early along the reaction coordinate, and the top of the free energy hill is reached with a relatively small entropy loss. In this case the barrier height is low and the folding rate fast. If the contacts are more distant in sequence, then more residues must be ordered to realize the stabilizing effect of these interactions, hence a larger barrier and a slower rate. This is the physical basis of the contact order correlation [5, 12, 13].

These three key results – the accuracy of the energy landscape description of the lattice simulations, the contact-order correlation, and the success of the β -hairpin model – suggested that the β -hairpin model might also work for proteins. To this end Muñoz and Eaton formulated a more realistic model by allowing native structure to grow in as many as two (double-sequence approximation) or three (triple-sequence approximation) continuous stretches of native

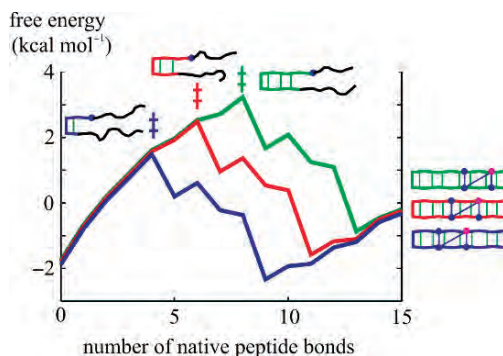


Fig 1. Effect of contact order on folding rate of β hairpin. Three profiles are shown for three different locations of the hydrophobic cluster. (From Muñoz et al. [4].) As the hydrophobic cluster, indicated by the four circles connected by bars, moves away from the turn, the contact order decreases and the free energy barrier increases. See text for discussion.

structure simultaneously, while still allowing for enumeration of all configurations [5]. As in the β -hairpin model [3, 4], interactions between two residues were only allowed if all intervening residues are in a native conformation, thereby excluding the possibility of connecting native regions by disordered loops. Muñoz and Eaton examined 18 single-domain proteins exhibiting two-state behavior. Free-energy profiles were calculated using the number of native residues (or more precisely, the number of native peptide bonds) as the reaction coordinate, with the interaction energy for each protein adjusted to yield the experimental equilibrium constant. With one or two exceptions these profiles showed only two deep minima, consistent with the experimental observation of two-state behavior. The heights of the barrier separating the minima, as well as the folding rates calculated from the profiles, correlated surprisingly well with the folding rates for the 18 proteins measured in the absence of chemical denaturants (correlation coefficient > 0.8). Muñoz and Eaton also compared calculated and experimental values on the effects of mutations in the most thoroughly studied two-state protein, chymotrypsin inhibitor 2 (CI 2), but obtained much less good agreement with the experimental results (i.e. ϕ values). Subsequently, Bruscolini and Pelizzola developed methods for calculating the partition function for all 2^N configurations, and applied it to the kinetics of folding the GB1 β -hairpin and CI2 [19].

Models similar to the β -hairpin model but allowing for the possibility of connecting native regions by disordered loops were simultaneously developed by Alm and Baker [1] and by Galzitskaya and Finkelstein [7]. These authors did not compare free-energy barrier heights or calculate folding rates, but instead focused on identifying the ensemble of structures that make up the transition state for comparison with mutation experiments. This was done using algorithms to find the free-energy maxima along lowest free-energy pathways connecting the folded and unfolded states. Subsequently, Ivankov and Finkelstein [8] calculated folding rates at mid-transition, i.e. at concentrations of chemical denaturants where the folded and unfolded states are equally populated. To make the computation possible, they represented the chain as a series of links, each containing several residues, and solved the kinetic equations for the network of configurations using a steady-state approximation. They found a correlation coefficient for 20 two-state proteins of 0.59. Most recently, Alm et al. [20] have compared measured folding rates at mid-transition for 37 two-state proteins with barrier heights calculated from the transition-state ensemble, and found a correlation coefficient of 0.67.

In the present study we extend the model of Muñoz and Eaton to include the possibility of forming a disordered loop between stretches of native residues in the double sequence approximation. We also simplify their model by using an α -carbon instead of an atomistic contact map and by using a less detailed treatment of both the entropy and the stabilization energies. After describing the model we calculate the free energy profiles and rates for

25 proteins that are known from a combination of equilibrium and kinetic measurements to exhibit two-state behavior.**

Description of Model and Partition Function

The basic ingredient of the model is that each residue of the polypeptide chain can exist in one of two possible states – native (n) or non-native (c), as in an Ising model [15, 16]. To greatly reduce the number of possible configurations, we employ the double sequence approximation in which no more than two continuous stretches of native residues are allowed in each molecule (e.g., ...*cnnnccnnnnccc*...). The free energy and thermodynamic weight of a stretch of native residues of length j beginning at residue i are, respectively,

$$G_{ji} = \sum_{\text{contacts}} \varepsilon - T \sum_{k=i}^{i+j-1} \Delta s_k, \quad w_{ji} = \exp(-G_{ji} / RT) \quad (1)$$

where the first term is the total energy of native contacts among residues in the stretch, and Δs_k is the entropy cost of fixing residue k in the native conformation. The number of configurations containing exactly 2 contiguous stretches of native residues, without disordered loops connecting the stretches, is given by the binomial coefficient:

$$\binom{N+1}{2m} = \frac{(N+1)!}{(2m)!(N+1-2m)!}. \text{ For } N = 100, \text{ this approximation } (m = 2)$$

requires enumeration of $\sim 4 \times 10^6$ configurations. The model allows contacts between residues in a native segment and between residues in two different native segments [7, 20]. Connecting native segments with a disordered loop increases the number of configurations by factor of ~ 2 . The partition function for this model is:

$$\begin{aligned} Z = & 1 + \sum_{j(1)=1}^n \sum_{i(1)=1}^{n-j(1)+1} w_{j(1)i(1)} + \sum_{j(1)=1}^n \sum_{i(1)=1}^{n-j(1)+1} \sum_{j(2)=1}^{n-(i(1)+j(1))} \sum_{i(2)=i(1)+j(1)+1}^{n-j(2)+1} w_{j(1)i(1)} w_{j(2)i(2)} \\ & + \sum_{j(1)=1}^n \sum_{i(1)=1}^{n-j(1)+1} \sum_{j(2)=1}^{n-(i(1)+j(1))} \sum_{i(2)=i(1)+j(1)+1}^{n-j(2)+1} w_{j(1)i(1)} w_{j(2)i(2)} v_{j(1)i(1);j(2)i(2)} \end{aligned} \quad (2)$$

where $v_{j(1)i(1);j(2)i(2)}$ is the weighting factor reflecting the change in free energy arising from the coupling between the native segments $j(1)i(1)$ and $j(2)i(2)$, given by

** This article is a shortened version of the more detailed study published elsewhere that considers several different partition functions and alternative assumptions in a combinatorial analysis 21. Henry, E.R. and W.A. Eaton, *Combinatorial modeling of protein folding kinetics: free energy profiles and rates*. Chemical Physics, 2004. **307**(2–3): p. 163–185.

$$v_{j(1)i(1);j(2)i(2)} = \exp\left(-\Delta G_{j(1)i(1);j(2)i(2)} / RT\right) \quad (3)$$

$$\Delta G_{j(1)i(1);j(2)i(2)} = \left(\sum_{\substack{\text{contacts} \\ j(1)i(1) \leftrightarrow j(2)i(2)}} \varepsilon \right) - T\Delta s_{\text{loop}}$$

Here Δs_{loop} is the entropy loss due to constraining the endpoints of the disordered loop incorporating residues $i(1) + j(1)$ to $i(2) - 1$, inclusive. A schematic picture of the model is shown in Fig. 2.

To evaluate the partition function the simplest description of interactions is employed, in which both non-native interactions and amino-acid type are ignored, with identical energies for all contacts in the known three-dimensional structure. A contact between residues is defined by a $C_{\alpha} - C_{\alpha}$ distance less than or equal to 0.7 nm. Contacts between adjacent residues and residues separated by one amino acid are ignored because of the high probability that their side chains are in contact in the unfolded state. For a given protein the same energy is assigned to each inter-residue contact. This energy is individually adjusted for each protein, in order to obtain the experimentally measured equilibrium constant from the free-energy profile. In the same spirit, simple entropy functions are used. The same entropy decrease (Δs_{conf}) for the non-native to native transition is assigned to every residue of every protein.

Our treatment of the entropy loss on forming a non-native loop draws from recent studies of loop formation in disordered polypeptides [22–29]. For a chain with a Gaussian distance distribution between two interior residues, the entropy change is simply obtained from the ratio of the volume occupied at contact to the effective volume explored by the residues, i.e.

$$\Delta s_{\text{loop}} = R \ln \left(\frac{2\pi a^3 / 3}{(2\pi \langle r^2 \rangle / 3)^{3/2}} \right) \quad (4)$$

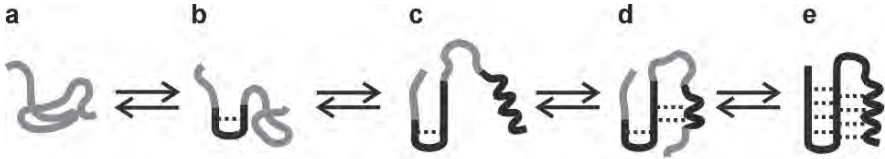


Fig 2. Schematic of a microscopic folding pathway. Region of chain in which all residues are in native conformation colored in black (unfolded in grey). (a) The unfolded state. (b) β -hairpin begins to form. (c) Second stretch of native residues becomes ordered into a helix (d) Contact forms between hairpin and helix region with disordered loop in between. (e) Native structure.

where a is the center-to-center distance at contact, and r is the distance between the two residues at the ends of the disordered loop. The contact distance a for a specific loop is identified with the $C_\alpha - C_\alpha$ distance (computed from the native structure of the molecule) between the two native residues that anchor the ends of the loop). The volume at contact (the numerator in (Equation (4)) has been reduced by a factor of 2 from the full spherical volume to account for the steric interference by adjacent residues in making a contact to form an interior loop [30]. To obtain the length dependence we employ the Flory relation for a random-walk chain: $\langle r^2 \rangle = C_n n l^2$ where $l = 0.36 \text{ nm}$ is the projection of the $C_\alpha - C_\alpha$ vector on the direction of an extended strand, $n + 1$ is the number of residues in the loop, and C_n (the Flory characteristic ratio) is an empirically-determined parameter that yields the experimentally measured value of $\langle r^2 \rangle$, and may be regarded as the number of residues of a polypeptide that make up a segment of an equivalent random-walk chain [31]. To estimate the range of possible values for C_n , we use the experimental results on end-to-end contact formation in small disordered peptides. These experiments also show that the rate for peptides of more than ten residues varies very nearly as $n^{-3/2}$, as expected for a random-walk chain, and that at less than ten residues the dependence is much less because of chain stiffness [32, 33]. The values of C_n that yield $\langle r^2 \rangle$ consistent with the experimentally measured rates for peptides containing about 20 residues range from 3 (for the repeating sequence ala-gly-gln) to 9 (for a threonine-based peptide [26]). We therefore use C_n as a parameter to be optimized with the constraint that $3 < C_n < 9$. For all loops with $n < 10$, the entropy change (strictly speaking, a free energy change because of chain stiffness) computed for $n = 10$ is used.

Calculations of the equilibrium constants and rates are performed with either the number of ordered residues (P) or the number of native contacts (Q) as a reaction coordinate – the only possible choices in this model. For each reaction coordinate, the equilibrium constant is calculated from the free energy versus reaction coordinate profile by placing a dividing line at the half-way point along the profile. The population to the left of the dividing line is regarded as unfolded and that to the right as folded. The equilibrium constant is given by the ratio of these populations.

Calculation of Folding Rates

The kinetics of the folding process are modeled as motion along a one-dimensional free-energy profile, with the number of ordered residues P or number of native contacts Q as the reaction coordinate. Instead of solving a diffusion

equation to describe this motion, a simpler approach is used of solving a system of differential equations describing reversible hopping between adjacent discrete values of the reaction coordinate. In this picture the evolution of the vector of state populations $[p_i(t)]$ is described by the rate equation

$$\frac{d}{dt} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{N-1} \\ p_N \end{pmatrix} = \begin{pmatrix} -k_{0,1} & k_{1,0} & 0 & \cdots & \cdots & 0 \\ k_{0,1} & -(k_{1,0} + k_{1,2}) & k_{2,1} & \cdots & \cdots & 0 \\ 0 & k_{1,2} & -(k_{2,1} + k_{2,3}) & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & k_{N-1,N-2} & 0 \\ \vdots & \vdots & \vdots & \vdots & -(k_{N-1,N} + k_{N-1,N-2}) & k_{N,N-1} \\ 0 & \cdots & \cdots & 0 & k_{N-1,N} & -k_{N,N-1} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{N-1} \\ p_N \end{pmatrix} \quad (5)$$

where k_{ij} is the rate coefficient for hopping from reaction coordinate i to the adjacent reaction coordinate j .

In addition to the equilibrium probabilities $p_{eq}(j)$ computed from the free-energy profile, the prediction of the folding kinetics using Equation (5) also requires values for the hopping rates $k_{jj \pm 1}$. An obvious condition to be

satisfied by these rates is that of detailed balance $\frac{k_{j,j+1}}{k_{j+1,j}} = \frac{p_{eq}(j+1)}{p_{eq}(j)}$, which still

leaves the free choice of one of the rates. The simplest assumption that mimics diffusion on the free energy profile invokes a linear free-energy relation between the activation free energy for hopping between adjacent points on the reaction surface and the relative free energy at the two points:

$$\Delta G_{j,j+1}^\ddagger = \frac{1}{2} \Delta G_{j,j+1} \quad (6)$$

which yields the rate coefficients in the form

$$\begin{aligned} k_{j,j+1} &= \gamma \left(\frac{p_{eq}(j+1)}{p_{eq}(j)} \right)^{1/2} \\ k_{j+1,j} &= \gamma \left(\frac{p_{eq}(j+1)}{p_{eq}(j)} \right)^{-1/2} \end{aligned} \quad (7)$$

Free Energy Profiles and Characterization of Two-State Folding

The analysis of the model was conducted as a least-squares fit of predicted to experimental folding rates. The adjustable parameters are the interaction energy ε (adjusted individually for each protein to reproduce the experimental equilibrium constants), the rate parameter γ that determines absolute folding

rates, the conformational entropy associated with ordering a residue, and the Flory characteristic ratio, C_n , for describing the entropy loss upon forming a disordered loop connecting two stretches of native residues. A Marquardt-Levenberg algorithm was used to minimize the residuals between the logarithms of a set of experimental folding rates and the predicted rates. The structures of the 25 proteins considered are shown in Fig. 3 with their α -carbon contact maps. Their basic equilibrium and kinetic folding data are given in Table 1. All of these proteins have been reported to exhibit two-state behavior in the absence of denaturant from a combination of equilibrium and kinetic experiments.

Figure 4 shows free energy profiles obtained from the least-squares fits for each of the two reaction coordinates – the number of ordered residues P (Fig. 4a) and the number of native contacts Q (Fig. 4b). The most populated conformations of the denatured state are at $Q = 0$, i.e. conformations with no native contacts. This is an unphysical aspect of the model, since some native contacts should be present in the denatured state. Simulations using simplified representations of proteins in lattice [9, 57] and off-lattice [58, 59] models, for example, show a free energy minimum at fractional Q values between 0.2 and 0.3. The free energy minimum at $Q = 0$ results from the neglect by the model of non-native interactions which would otherwise provide compensating stabilization of denatured conformations to produce more compact structures and the possibility of native interactions.

Figure 5 shows that the free energy barrier is the result of almost complete cancellation of large entropy and energy terms and that the entropy function is curved (curvature in either the entropy or energy function is required to produce a barrier). Also shown is the “combinatorial entropy”, defined simply as $R \ln \Omega(Q)$, where $\Omega(Q)$ is the total number of possible states having a value Q . ($\Omega(Q)$ states are only equally populated at infinite temperature where the thermodynamic weight is the same for all conformations; at room temperature an infinitesimally small fraction are populated because of the differences in free energy among the conformations, explaining why there is in fact no Levinthal paradox [11, 15]).

Using P as the reaction coordinate, the profiles are similar. The increase in combinatorial entropy associated with placing two native residues anywhere in the sequence produces the initial sharp dip in the free energy. The free energy is then uphill because there is little change in the combinatorial entropy and the entropy cost of ordering residues is not fully compensated by the stabilization associated with forming contacts. There are no barriers more than 1 kcal/mol for the α -helical proteins using either P or Q as the reaction coordinate. These proteins are “downhill folders” in the folding scenario classification of Bryngelson et al. [61]. In contrast, with just a few

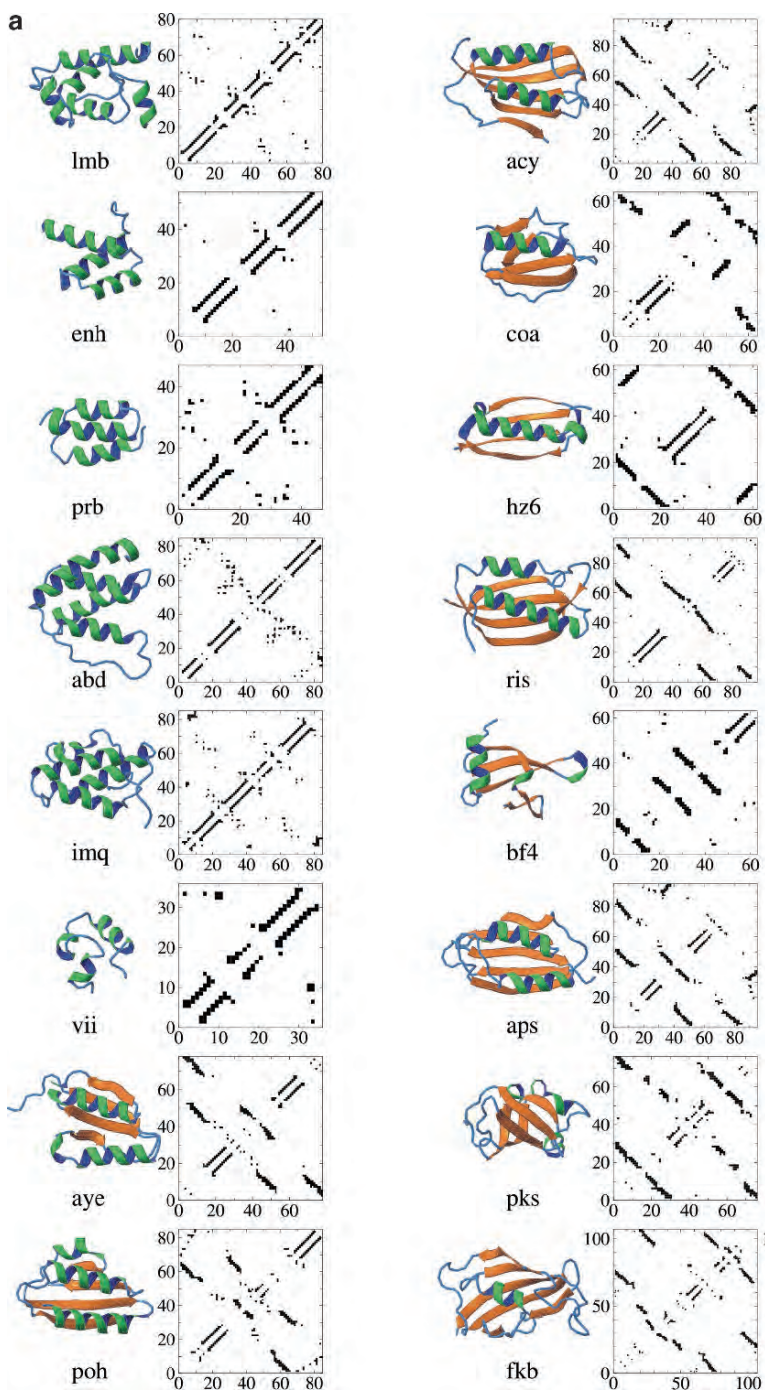


Fig 3. Structures and α -carbon contact maps. The structures were drawn with MOLMOL [60]. The contact map for each protein shows all contacts between residues defined by a distance between α carbons of 0.7 nm or less.

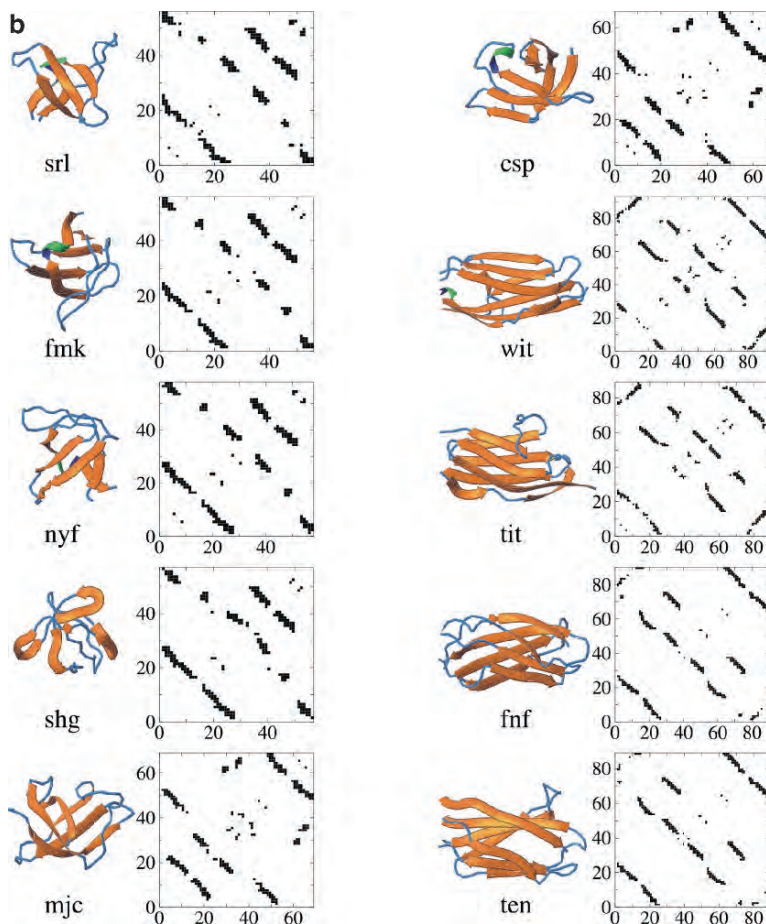


Fig 3. (continued)

TABLE 1. Proteins included in analysis.[†]

| Name [data source] | PDB designation (method) ^a | Chain length | % α | % β | ΔG kcal/ mol | T °C | k_f s ⁻¹ |
|------------------------------------|---|-----------------|------------|-----------|----------------------------|------|-----------------------|
| Monomeric lambda repressor [34] | 1lmb(X-ray) ^b | 80 | 73 | 0 | 3.0 | 37 | 4900 |
| Engrailed homeodomain [35] | 1enh(X-ray) | 54 | 70 | 0 | 2.1 | 25 | 37500 |
| Albumin-binding domain PRB [36] | 1prb(NMR) ^c | 47 | 68 | 0 | 2.8 | 74 | 1.0×10^6 |
| ACBP [37] | 2abd(NMR) | 85 | 61 | 0 | 7.1 | 5 | 279 |
| Bacteriocin Im9 [2, 38] | 1imq(NMR) ^d | 85 | 53 | 0 | 6.6 | 10 | 1450 |
| Villin 14T headpiece [39] | 1vii(X-ray) | 36 | 53 | 0 | 3.6 | 27 | 2.3×10^5 |

(continued)

TABLE 1. (continued)

| | | | | | | | |
|--|--------------------------|-----|----|----|-----|----|------|
| Procarboxypeptidase A2 [40]* | 1aye(X-ray) ^c | 78 | 36 | 24 | 4.4 | 25 | 760 |
| HIS-containing phosphocarrier [41]* | 1poh(X-ray) | 85 | 34 | 27 | 4.6 | 20 | 14.9 |
| Acylphosphatase (common type) [42]* | 2acy(X-ray) | 98 | 25 | 42 | 4.4 | 28 | 2.3 |
| Chymotrypsin inhibitor 2 [43]* | 1coa(X-ray) | 64 | 22 | 22 | 7.6 | 25 | 56 |
| IgG-binding domain streptococcus protein L [44]* | 1hz6(X-ray) ^f | 62 | 34 | 45 | 4.6 | 22 | 60 |
| Ribosomal protein S6 [45]* | 1ris(X-ray) | 97 | 29 | 48 | 8.0 | 25 | 332 |
| Sso7d protein [46] | 1bf4(X-ray) | 63 | 24 | 52 | 5.9 | 20 | 1040 |
| Acylphosphatase (muscle) [42] | 1aps(NMR) | 95 | 19 | 38 | 5.4 | 28 | 0.23 |
| SH3 domain (PI3 kinase) [47]* | 1pks(NMR) ^g | 76 | 17 | 34 | 3.4 | 20 | 0.35 |
| FK506-binding protein [48]* | 1fkb(X-ray) | 107 | 12 | 38 | 5.5 | 25 | 4.3 |
| | 1a7x(X-ray) ^h | | | | | | |
| SH3 domain (src) [49]* | 1fmk(X-ray) ⁱ | 56 | 5 | 32 | 4.1 | 22 | 57 |
| | 1srl(NMR) ^j | | | | | | |
| SH3 domain (fyn) [50]* | 1nyf(NMR) ^k | 58 | 5 | 41 | 6.0 | 20 | 94 |
| SH3 domain (α spectrin) [51]* | 1shg(X-ray) ^l | 57 | 5 | 46 | 4.1 | 25 | 4.1 |
| Cold-shock protein A [52]* | 1mjc(X-ray) | 69 | 4 | 48 | 2.9 | 10 | 200 |
| Cold-shock protein B [53]* | 1csp(X-ray) | 67 | 5 | 54 | 3.0 | 25 | 690 |
| Twitchin IgSF 18 [54]* | 1wit(NMR) | 93 | 3 | 52 | 4.0 | 20 | 1.5 |
| Titin [54]* | 1tit(NMR) ^m | 89 | 0 | 36 | 7.5 | 25 | 32 |
| Fibronectin type III module 9 [55]* | 1fnf(X-ray) ⁿ | 90 | 0 | 54 | 1.2 | 25 | 0.4 |
| Tenascin (short form) [56]* | 1ten(X-ray) | 89 | 0 | 54 | 4.8 | 20 | 2.9 |

† Proteins marked with an asterisk were included in the parameter optimization procedure. The reference to equilibrium and kinetic data is given next to the protein name.

^a All NMR structures are the energy-minimized average of multiple NMR structures, with the exception of 2abd (29 structures reported) and 1aps (five structures reported).

^b Two molecules from asymmetric unit: 1mba include residues 6–85 of segment 3, 1mbb includes residues 6–85 of segment 4.

^c Residues 7–53.

^d Residues 2–86.

^e Residues 1–78.

^f Residues 11–72.

^g Residues 4–79.

^h Two molecules from asymmetric unit: a7xa includes residues 1–107 of segment A, a7xb includes residues 1–107 of segment B.

ⁱ Residues 4–59.

^j Residues 9–64.

^k Residues 3–60.

^l Residues 6–62.

^m Residues 10–98.

ⁿ Residues 185–274.

exceptions, there are much larger barriers and only two deep minima for almost all of the $\alpha\beta$ and β proteins. It is necessary, however, to have some quantitative criterion to establish whether or not a profile predicts two-state behavior.

An invariant property of a two-state system is that the relaxation to equilibrium shows an exponential time course. Prediction of the experimental observation of an exponential time course requires that essentially all of the amplitude be associated with a single eigenvalue of the rate matrix that describes hopping back and forth between adjacent discrete values of the reaction coordinate, P or Q . We therefore calculated the fractional amplitude of the dominant relaxation for every profile. These are given in Fig. 4 for each protein, together with the dividing line that yields the maximum amplitude. Allowing for noise and uncertainties, a reasonable experimental criterion for a single exponential would be that more than 90% of the amplitude be contained in a single exponential relaxation. A finding of less than $\sim 90\%$ indicates that the profile is not two-state, either because there is no free-energy barrier, as in the case of the all- α proteins, or because there is an intermediate that is significantly populated during folding. Using this criterion two-state behavior is predicted for the vast majority of β and $\alpha\beta$ proteins. Interestingly, among the α -helical proteins the villin subdomain (vii) and the albumin binding domain (prb) have steep downhill free-energy profiles, and have been shown in recent experiments [36, 39] not only to be among the fastest-folding proteins, but also to exhibit rates closest to those expected for barrierless folding [62].

Calculated Folding Rates and Parameters of the Model

A comparison of the observed and the calculated rates, determined as the smallest non-zero eigenvalue of the rate matrix (Equation (5)), is shown in Fig. 6. There is a good correlation between the mean of the calculated rates and the observed rates using either P or Q as reaction coordinates. The lack of a difference between the P and Q reaction coordinates is consistent with the notion that several different global order parameters may serve as reaction coordinates [61, 63].

The agreement between calculated and observed folding rates is remarkable, considering the simplicity of the model. Only three adjustable parameters of the model were optimized to give the best agreement of the calculated and experimental rates: Δs_{conf} , the scale factor γ for the hopping rate along the free energy profile, and the Flory characteristic ratio C_n for the disordered loop entropy. Another parameter, the interaction energy ϵ , was separately varied for each protein to obtain agreement between the calculated and

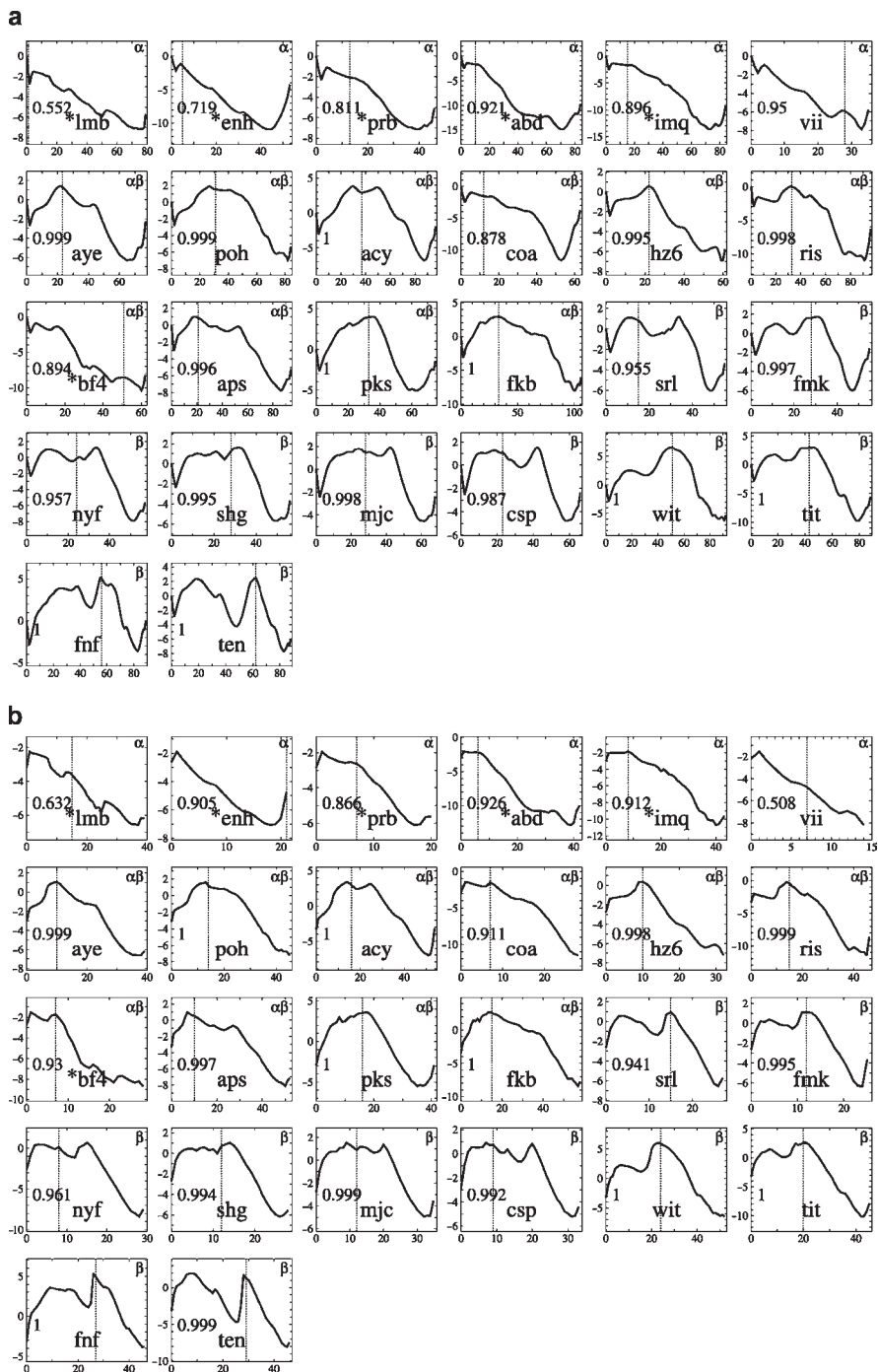


Fig. 4. Free energy profiles. The free energy in kcal/mol is plotted as a function of either the number of ordered residues (P) (panel *a*) or the number of native contacts (Q) (panel *b*). The profiles are calculated using parameters that gave best agreement between calculated and observed

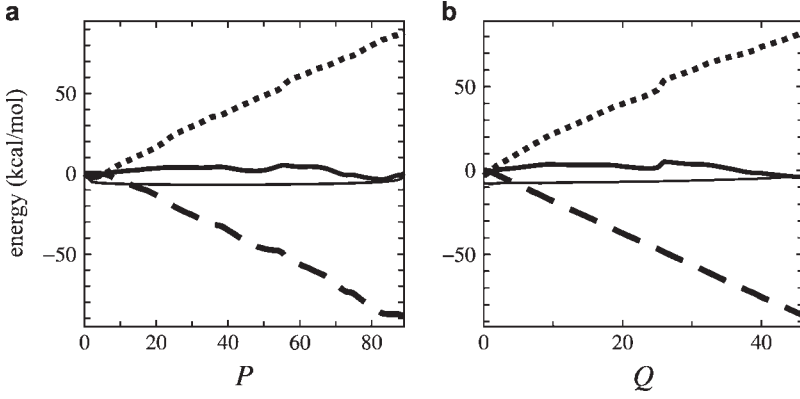


Fig 5. Free energy profiles for a single protein (fnf) decomposed into energy and entropy contributions. (a) P reaction coordinate, free energy (thick continuous line), energy (dashed line), total entropy (multiplied by temperature) (dotted line), and combinatorial entropy (multiplied by temperature) (thin continuous line) defined by $R \ln \Omega(P)$, where $\Omega(P)$ is the total number of conformations having P ordered residues. The energy and entropy were calculated from the relations:

$$\bar{E}(P) = \frac{\sum_k \varepsilon_k(P) w_k(P)}{\sum_k w_k(P)}, \quad TS(P) = \bar{E}(P) + RT \ln \sum_k w_k(P)$$

(b) For the Q reaction coordinate, the energy and entropy were calculated from the relations:

$$E(Q) = Q\varepsilon, \quad TS(P) = E(Q) + RT \ln \sum_k w_k(Q)$$

observed equilibrium constant. However, given the value of the conformational entropy the value of ε for a given protein is uniquely determined by the requirement that the free energy profile yield the experimentally determined equilibrium constants. Thus, the separate values of ε for each protein do not contribute to the parametric complexity of the model.

← rates using the least squares criterion. The free energy versus Q profiles were smoothed by re-binning the state populations after adjusting the number of bins to approximate one half the number of residues in the protein. In each panel the fractional amplitude associated with the smallest non-zero eigenvalue is given. The dividing (dotted) line that produced the maximum amplitude is shown. An asterisk preceding the PDB file name indicates that there is no clear demarcation between folded and unfolded populations. In these so-called “downhill scenarios”, the equilibrium constant was calculated as the ratio of the populations at $P = 2$ and the maximum P (panel a) and $Q = 0$ and the maximum Q (panel b). (a) Profiles calculated using P as reaction coordinate ($\Delta s_{\text{conf}} = -3.32 \text{ cal mol}^{-1} \text{ K}^{-1}$, $\gamma = 2.1 \times 10^6 \text{ s}^{-1}$, $C_n = 8.8$). (b) Profiles calculated using Q as reaction coordinate ($\Delta s_{\text{conf}} = -3.27 \text{ cal mol}^{-1} \text{ K}^{-1}$ and $\gamma = 5.6 \times 10^5 \text{ s}^{-1}$, $C_n = 8.9$).

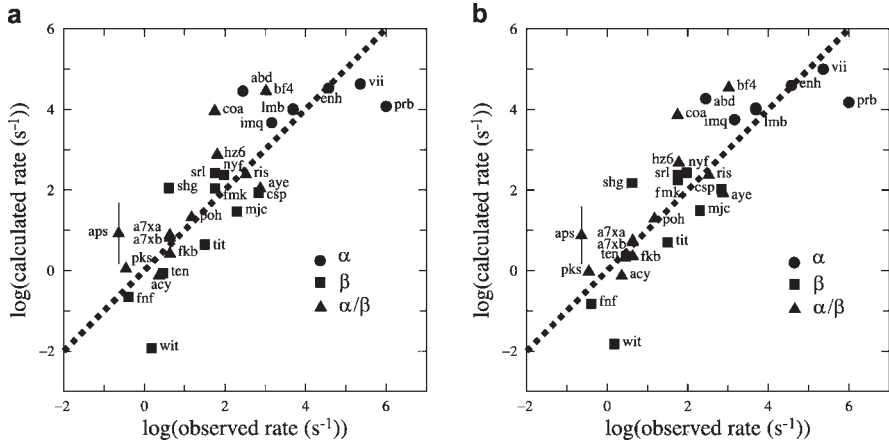


Fig 6. Comparison of observed and calculated rates. (a) P as reaction coordinate ($r = 0.81$, $p = 1 \times 10^{-6}$). (b) Q as reaction coordinate ($r = 0.83$, $p = 3 \times 10^{-7}$).

In a Kramers description the pre-exponential factor is proportional to the diffusion coefficient along the reaction coordinate, or in this case the hopping rate, γ . The present results suggest that the pre-exponential factor is relatively constant, implying a similar degree of energetic roughness for the landscapes of all of the proteins.

Conclusions

A simple statistical mechanical, Ising-like model, which only considers the trade-off between the loss of conformational entropy and stabilization from native inter-residue contacts, is a surprisingly accurate predictor of two-state-ness and relative folding rates of small single-domain proteins. With a coarse-grained description of both the energy and entropy, the model described here may be regarded as the simplest possible analytical theory of protein folding. It should therefore be a useful starting point in determining what additional physics must be included to develop a more complete and realistic theory that makes more accurate predictions.

The model must also be capable of explaining the relative effect of mutations on folding equilibrium constants and rates, i.e. ϕ values. Similar models have already been used by Baker, Finkelstein, Serrano and coworkers [7, 8, 20, 46], with variable success in explaining ϕ values. Although more realistic energy functions produce some improvement [20, 46], there has been no attempt in these studies to optimize the parameters of the model to obtain the best agreement with experiments. This is the next obvious step.

Once the model has been optimized to simultaneously explain two-stateness, relative rates, and ϕ values, it will be worthwhile to exploit the enormous amount of information contained in the partition function. The partition function and associated kinetic scheme not only describe the nature of the transition state ensemble, but also contain the distribution of microscopic pathways that connect the folded and unfolded states. This distribution is in principle contained in all-atom molecular dynamics simulations of folding [39, 64], and is the goal of single molecule protein folding experiments [65]. If the experiments are successful and the simulations produce reliable distributions, they will provide yet another important test of theoretical models.

Acknowledgement We are indebted to Victor Muñoz for developing the original model that is both the basis and the motivation for the present investigation. We also thank Attila Szabo for many insightful discussions. This research was supported by the Intramural Research Program of NIDDK, NIH.

References

1. Alm, E. and D. Baker, *Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures*. Proc. Natl. Acad. Sci. U. S. A., 1999. 96(20): 11305–11310.
2. Baker, D., *A surprising simplicity to protein folding*. Nature, 2001. 405: 39–42.
3. Munoz, V., P.A. Thompson, J. Hofrichter, and W.A. Eaton, *Folding dynamics and mechanism of beta-hairpin formation*. Nature, 1997. 390(6656): 196–199.
4. Munoz, V., E.R. Henry, J. Hofrichter, and W.A. Eaton, *A statistical mechanical model for beta-hairpin kinetics*. Proc. Natl. Acad. Sci. U. S. A., 1998. 95(11): 5872–5879.
5. Munoz, V. and W.A. Eaton, *A simple model for calculating the kinetics of protein folding from three-dimensional structures*. Proc. Natl. Acad. Sci. U. S. A., 1999. 96(20): 11311–11316.
6. Munoz, V., *What can we learn about protein folding from Ising-like models?* Curr. Opin. Struct. Biol., 2001. 11(2): 212–216.
7. Galzitskaya, O.V. and A.V. Finkelstein, *A theoretical search for folding/unfolding nuclei in three-dimensional protein structures*. Proc. Natl. Acad. Sci. U. S. A., 1999. 96(20): 11299–11304.
8. Ivankov, D.N. and A.V. Finkelstein, *Theoretical study of a landscape of protein folding-unfolding pathways. Folding rates at midtransition*. Biochemistry, 2001. 40(33): 9957–9961.
9. Socci, N.D., J.N. Onuchic, and P.G. Wolynes, *Diffusive dynamics of the reaction coordinate for protein folding funnels*. J. Chem. Phys., 1996. 104(15): 5860–5868.
10. Bryngelson, J.D. and P.G. Wolynes, *Spin-glasses and the statistical-mechanics of protein folding*. Proc. Natl. Acad. Sci. U. S. A., 1987. 84(21): 7524–7528.
11. Bryngelson, J.D. and P.G. Wolynes, *Intermediates and barrier crossing in a random energy model (with applications to protein folding)*. J. Phys. Chem., 1989. 93(19): 6902–6915.
12. Plaxco, K.W., K.T. Simons, and D. Baker, *Contact order, transition state placement and the refolding rates of single domain proteins*. J. Mol. Biol., 1998. 277(4): 985–994.
13. Plaxco, K.W., K.T. Simons, I. Ruczinski, and B. David, *Topology, stability, sequence, and length: Defining the determinants of two-state protein folding kinetics*. Biochemistry, 2000. 39(37): 11177–11183.
14. Go, N., *Theoretical studies of protein folding*. Ann. Rev. Biophys. Bioeng., 1983. 12: 183–210.
15. Zwanzig, R., A. Szabo, and B. Bagchi, *Levinthal's paradox*. Proc. Natl. Acad. Sci. U. S. A., 1992. 89(1): 20–22.

16. Zwanzig, R., *Simple model of protein-folding kinetics*. Proc. Natl. Acad. Sci. U. S. A., 1995. 92(21): 9801–9804.
17. Thompson, P.A., V. Munoz, G.S. Jas, E.R. Henry, W.A. Eaton, and J. Hofrichter, *The helix-coil kinetics of a heteropeptide*. J. Phys. Chem. B, 2000. 104(2): 378–389.
18. Schellman, J.A., *The factors affecting the stability of hydrogen-bonded polypeptide structures in solution*. J. Phys. Chem., 1958. 62: 1485–1494.
19. Bruscolini, P. and A. Pelizzola, *Exact solution of the Munoz-Eaton model for protein folding*. Phys. Rev. Lett., 2002. 88(25): art. no. 258101.
20. Alm, E., A.V. Morozov, T. Kortemme, and D. Baker, *Simple physical models connect theory and experiment in protein folding kinetics*. J. Mol. Biol., 2002. 322(2): 463–476.
21. Henry, E.R. and W.A. Eaton, *Combinatorial modeling of protein folding kinetics: free energy profiles and rates*. Chemical Physics, 2004. 307(2–3): 163–185.
22. Bieri, O., J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, *The speed limit for protein folding measured by triplet-triplet energy transfer*. Proc. Natl. Acad. Sci. U. S. A., 1999. 96(17): 9597–9601.
23. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Measuring the rate of intramolecular contact formation in polypeptides*. Proc. Natl. Acad. Sci. U. S. A., 2000. 97(13): 7220–7225.
24. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Dynamics of intramolecular contact formation in polypeptides: Distance dependence of quenching rates in a room-temperature glass*. Phys. Rev. Lett., 2001. 87(25): art. no. 258101.
25. Lapidus, L.J., P.J. Steinbach, W.A. Eaton, A. Szabo, and J. Hofrichter, *Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics*. J. Phys. Chem. B, 2002. 106(44): 11628–11640.
26. Lapidus, L.J., W.A. Eaton, and J. Hofrichter, *Measuring dynamic flexibility of the coil state of a helix-forming peptide*. J. Mol. Biol., 2002. 319(1): 19–25.
27. Buscaglia, M., B. Schuler, L.J. Lapidus, W.A. Eaton, and J. Hofrichter, *Kinetics of intramolecular contact formation in a denatured protein*. J. Mol. Biol., 2003. 332(1): 9–12.
28. Huang, F. and W.M. Nau, *A conformational flexibility scale for amino acids in peptides*. Angew. Chem.-Int. Edit., 2003. 42(20): 2269–2272.
29. Krieger, F., B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber, *Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding*. J. Mol. Biol., 2003. 332(1): 265–274.
30. Buscaglia, M., L.J. Lapidus, W.A. Eaton, and J. Hofrichter, *Effect of denaturants on loop dynamics in polypeptides*. Biophys. J., 2006. 91(1): 276–288.
31. Flory, P.J., *Statistical Mechanics of Chain Molecules*. 1969, New York: Wiley.
32. Camacho, C.J. and D. Thirumalai, *Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin-inhibitor*. Proc. Natl. Acad. Sci. U. S. A., 1995. 92(5): 1277–1281.
33. Guo, Z.Y. and D. Thirumalai, *Kinetics of protein-folding - nucleation mechanism, time scales, and pathways*. Biopolymers, 1995. 36(1): 83–102.
34. Myers, J.K. and T.G. Oas, *Contribution of a buried hydrogen bond to lambda repressor folding kinetics*. Biochemistry, 1999. 38(21): 6761–6768.
35. Mayor, U., C.M. Johnson, V. Daggett, and A.R. Fersht, *Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation*. Proc. Natl. Acad. Sci. U. S. A., 2000. 97(25): 13518–13522.
36. Wang, T., Y.J. Zhu, and F. Gai, *Folding of a three-helix bundle at the folding speed limit*. J. Phys. Chem. B, 2004. 108(12): 3694–3697.
37. Kragelund, B.B., P. Hojrup, M.S. Jensen, C.K. Schjerling, E. Juul, J. Knudsen, and F.M. Poulsen, *Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family*. J. Mol. Biol., 1996. 256(1): 187–200.
38. Ferguson, N., A.P. Capaldi, R. James, C. Kleanthous, and S.E. Radford, *Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9*. J. Mol. Biol., 1999. 286(5): 1597–1608.

39. Kubelka, J., W.A. Eaton, and J. Hofrichter, *Experimental tests of villin subdomain folding simulations*. J. Mol. Biol., 2003. 329: 625–630.
40. Villegas, V., A. Azuaga, L. Catasus, D. Reverter, P.L. Mateo, F.X. Aviles, and L. Serrano, *Evidence for a 2-state transition in the folding process of the activation domain of human procarboxypeptidase-A2*. Biochemistry, 1995. 34(46): 15105–15110.
41. Van Nuland, N.A.J., W. Meijberg, J. Warner, V. Forge, R.M. Scheek, G.T. Robillard, and C.M. Dobson, *Slow cooperative folding of a small globular protein HPr*. Biochemistry, 1998. 37(2): 622–637.
42. Taddei, N., F. Chiti, P. Paoli, T. Fiaschi, M. Bucciattini, M. Stefani, C.M. Dobson, and G. Ramponi, *Thermodynamics and kinetics of folding of common-type acylphosphatase: Comparison to the highly homologous muscle isoenzyme*. Biochemistry, 1999. 38(7): 2135–2142.
43. Jackson, S.E. and A.R. Fersht, *Folding of chymotrypsin inhibitor-2.1. Evidence for a 2-state transition*. Biochemistry, 1991. 30(43): 10428–10435.
44. Kim, D.E., C. Fisher, and D. Baker, *A breakdown of symmetry in the folding transition state of protein L*. J. Mol. Biol., 2000. 298(5): 971–984.
45. Otzen, D.E., O. Kristensen, M. Proctor, and M. Oliveberg, *Structural changes in the transition state of protein folding: Alternative interpretations of curved chevron plots*. Biochemistry, 1999. 38(20): 6499–6511.
46. Guerois, R. and L. Serrano, *The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways*. J. Mol. Biol., 2000. 304(5): 967–982.
47. Guijarro, J.I., C.J. Morton, K.W. Plaxco, I.D. Campbell, and C.M. Dobson, *Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy*. J. Mol. Biol., 1998. 276(3): 657–667.
48. Main, E.R.G., K.F. Fulton, and S.E. Jackson, *Folding pathway of FKBP12 and characterisation of the transition state*. J. Mol. Biol., 1999. 291(2): 429–444.
49. Grantcharova, V.P. and D. Baker, *Folding dynamics of the src SH3 domain*. Biochemistry, 1997. 36(50): 15685–15692.
50. Plaxco, K.W., J.I. Guijarro, C.J. Morton, M. Pitkeathly, I.D. Campbell, and C.M. Dobson, *The folding kinetics and thermodynamics of the Fyn-SH3 domain*. Biochemistry, 1998. 37(8): 2529–2537.
51. Viguera, A.R., J.C. Martinez, V.V. Filimonov, P.L. Mateo, and L. Serrano, *Thermodynamic and kinetic-analysis of the Sh3 domain of spectrin shows a 2-state folding transition*. Biochemistry, 1994. 33(8): 2142–2150.
52. Reid, K.L., H.M. Rodriguez, B.J. Hillier, and L.M. Gregoret, *Stability and folding properties of a model beta-sheet protein, Escherichia coli CspA*. Protein Sci., 1998. 7(2): 470–479.
53. Schindler, T., M. Herrler, M.A. Marahiel, and F.X. Schmid, *Extremely rapid protein-folding in the absence of intermediates*. Nat. Struct. Biol., 1995. 2(8): 663–673.
54. Clarke, J., E. Cota, S.B. Fowler, and S.J. Hamill, *Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway*. Struct. Fold. Des., 1999. 7(9): 1145–1153.
55. Plaxco, K.W., C. Spitzfaden, I.D. Campbell, and C.M. Dobson, *A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules*. J. Mol. Biol., 1997. 270(5): 763–770.
56. Clarke, J., S.J. Hamill, and C.M. Johnson, *Folding and stability of a fibronectin type III domain of human tenascin*. J. Mol. Biol., 1997. 270(5): 771–778.
57. Sali, A., E. Shakhnovich, and M. Karplus, *How does a protein fold?* Nature, 1994. 369(6477): 248–251.
58. Klimov, D.K. and D. Thirumalai, *Viscosity dependence of the folding rates of proteins*. Phys. Rev. Lett., 1997. 79(2): 317–320.
59. Shea, J.E. and C.L. Brooks, *From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding*. Annu. Rev. Phys. Chem., 2001. 52: 499–535.

60. Koradi, R., M. Billeter, and K. Wuthrich, *MOLMOL: a program for display and analysis of macromolecular structures*. J. Mol. Graphics, 1996. 14: 51–55.
61. Bryngelson, J.D., J.N. Onuchic, N.D. Socci, and P.G. Wolynes, *Funnels, pathways, and the energy landscape of protein-folding - a synthesis*. Proteins, 1995. 21(3): 167–195.
62. Kubelka, J., J. Hofrichter, and W.A. Eaton, *The protein folding “speed limit”*. Curr. Opin. Struct. Biol., 2004. 14: 76–78.
63. Berezhkovskii, A. and A. Szabo, *One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions*. J. Chem. Phys., 2005. 122(1): art. no. 014503.
64. Zagrovic, B., C.D. Snow, M.R. Shirts, and V.S. Pande, *Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing*. J. Mol. Biol., 2002. 323(5): 927–937.
65. Schuler, B., E.A. Lipman, and W.A. Eaton, *Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy*. Nature, 2002. 419(6908): 743–747.

COMPLEMENTARITY OF HYDROPHOBIC/HYDROPHILIC PROPERTIES IN PROTEIN–LIGAND COMPLEXES: A NEW TOOL TO IMPROVE DOCKING RESULTS

TIMOTHY V. PYRKOV^{1,2*}, ANTON O. CHUGUNOV¹,
NIKOLAY A. KRYLOV^{1,3}, DIMITRY E. NOLDE¹, AND
ROMAN G. EFREMOV¹

¹*M.M. Shemyakin & Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya, 16/10, 117997 GSP, Moscow V-437, Russia*

²*Moscow Institute of Physics and Technology (State University), Institutskii per., 9, 141700, Dolgoprudny, Moscow region, Russia*

³*Joint Supercomputer Center, Russian Academy of Sciences, Leninskii prospect, 32a, 119991, Moscow, Russia*

*Corresponding author: e-mail: pyrkov@nmr.ru

Abstract Computational techniques designed to predict the spatial structure of ligand–receptor complexes (molecular docking) are widely used in investigations of molecular details of protein functioning and in drug design. Here, a brief review of docking methods is given and recent advances in improvement of their accuracy and efficiency are discussed. Two acute problems of standard docking algorithms are considered: proper ranking of putative docking solutions and simulation of receptor flexibility. The recent trends to overcome these problems are demonstrated with the results obtained in our Laboratory. Particular attention is paid to protein–ligand hydrophobic and stacking interactions, which are not always adequately represented in scoring criteria of docking applications.

Introduction

Nowadays, computational modeling approaches have become an integral part of both the fundamental studies aimed at understanding of molecular mechanisms underlying the protein functioning and the applied drug design programmes [1, 2].

Molecular modeling approaches aimed at prediction of the spatial structure of ligand–receptor complex, given the 3D-model of the latter, are

referred to as molecular docking. The benefits of these methods, should they indeed be reliable, are numerous: detection of key amino acids in the receptor active site, unveiling the structure-dynamic basis of enzymatic reactions at atomic scale, rational design of ligands and/or receptor sites with predefined kinetic properties, etc. Computational docking is used for estimation of affinity of novel compounds in virtual high throughput screening (HTS) projects that greatly reduces costs of projects aimed at finding novel prodrugs [3–7]. However, the precision of the docking algorithms appears to be satisfactory only in individual cases. And each time a study of a new biomolecular system is being undertaken, the researchers have to test the available computational techniques to find those that fit best the object under investigation. That is why, the work on developing and optimization of docking algorithms remains now an area of active research.

In recent years, a number of studies have appeared aimed at comparison of different docking/scoring schemes [8–15]. These may provide some guidelines to which of the numerous docking software is more universally precise or even give recommendations about algorithms which are more suitable for biomolecular systems with particular properties.

Here, we will consider the problems of modern docking algorithms, and the approaches that have been utilized to improve the predictive power of molecular docking. To better understand the underlying effects it is wholesome to consider the process of docking in more detail. Generally, it can be divided into two independent (though interconnected) parts – the conformational search algorithm and the scoring function (Fig. 1).

The conformational search algorithm is a method that starting from a random orientation of two molecules leads to the final 3D structure of

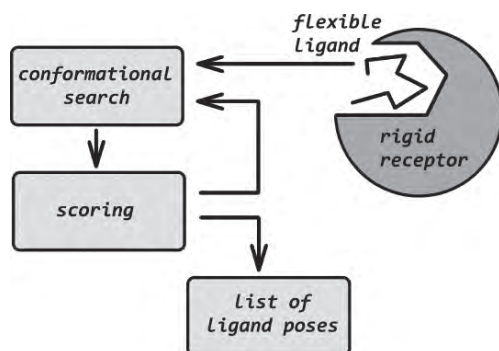


Fig. 1. Principal scheme of a molecular docking procedure. The cycle of generating a set of ligand poses and scoring them to select those that fit the active site best to further optimize their conformation continues until the termination conditions are met. For detailed description see text.

biomolecular complex through a chain of stepwise conformational changes. Usually, these are internal rotations around chemical bonds in molecules and translation and rotation of one molecule (the ligand) relative to another (the receptor) as a rigid body. Here, the number of torsion angles that are allowed to change is a bottleneck of the algorithm. And considering the great number of such bonds in a protein receptor it becomes clear why its structure is usually kept rigid in standard docking approaches. Most of the modern docking methods have achieved a remarkable efficiency in reproducing the orientation and conformation of a ligand in the active site, where it has preliminary been removed from. Unfortunately, the conformation of the binding site is not mandatory fitted to the ligand molecule and the ability of docking to place the ligand correctly starts to decrease dramatically when one uses a 3D structure of the receptor either determined experimentally without a ligand or theoretically modeled. This happens because the induced-fit effect – the fitting of receptor local conformation to the structure of the ligand molecule – is not taken into account. A number of approaches have been proposed to overcome this problem, and we will discuss them in more details below.

Another problem comes from the second part of the docking procedure – the scoring function. This function is used to score (or rank) various putative ligand conformations placed in the binding site at each step of the conformational search procedure according to some predefined criteria. The most primitive such criteria are bump checks which are applied at early stages of docking. Finally the docking process arrives at a number of ligand poses – the putative solutions of the docking problem – highly ranked by their affinity to the receptor. However, due to inaccuracies of scoring functions the native-like docking solutions may be not ranked at the top. In order not to lose it, the final list of ligand poses can be re-scored by a more precise or more suitable criterion – this approach is usually referred to as the consensus docking/scoring.

We will also discuss the results of studies aimed at improving the efficiency of docking methods achieved in our laboratory. Our main interest lies in investigation of the role of hydrophobic/hydrophilic contacts in ligand recognition by proteins. The applicability of methods for quantitative estimation of the hydrophobic interactions at the ligand–receptor interface in a complex to improve the scoring of docking poses will be demonstrated.

Scoring Functions

A scoring function is supposed to tell us which of the ligand orientations fits the binding site better or (if different ligands are compared) – which of them binds with greater affinity. Thus, it should more or less precisely describe intermolecular interactions between a ligand and a receptor and provide an

estimation of the free energy of binding. Usually, the interaction energy is decomposed into separate terms describing physical interactions. A weighted linear combination of such terms represents a scoring function.

It is common to divide all scoring approaches used in modern docking algorithms into three types: (a) force field based, (b) empirical, and (c) statistical. They differ in the scope of physical interaction they take into account and the way these interactions are quantitatively estimated. Detailed description of such scoring criteria can be found in recent reviews and references therein [1, 2]. Here we will only briefly consider their main features.

The force field based scoring functions are straightforwardly transferred from the molecular mechanics terms that describe non-covalent interactions. These are the Coulomb electrostatic and van der Waals components. The latter is usually described by the 12-6 Lennard-Jones potential. However, the so called “soft” 4-8 equation can also be used, as e.g. in the “goldscore” function [16]. This potential is more suitable in a case of moderate clashes between a ligand and a receptor, which may arise from non-perfect fit (Fig. 2). The main limitation of the force field based scores is that they are designed to optimize molecular structure according to the enthalpic part of the free energy in vacuum. Meanwhile binding of a ligand to a receptor is accompanied by the change in entropy and desolvation effects [2].

Knowledge-based scoring functions take the advantage of analysis of known X-ray structures of receptor-ligand complexes converted into statistical potentials. Such potentials are designed for predicting the orientation of a ligand in the active site rather than the binding energy. Nevertheless, they can be converted to energy units. One of the most popular such

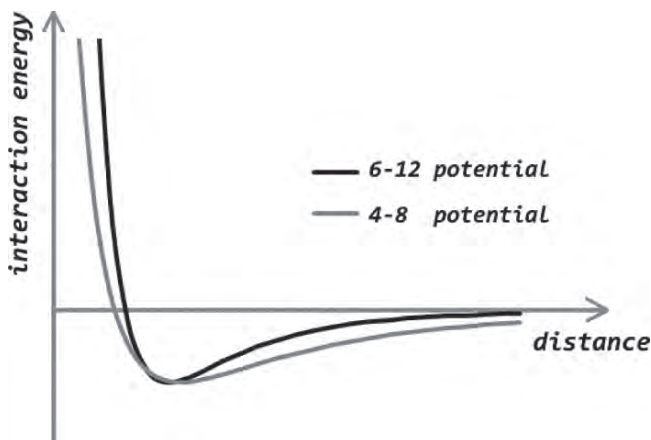


Fig. 2. Lennard-Jones potential used to describe van-der-Waals interactions (and sometimes other, e.g. hydrogen-bonds) in scoring functions. Soft 4-8 potential is often used to forgive moderate clashes between atoms of a ligand and a receptor.

function is the PMF-score [17] (Equation 1). Here, the pairwise potentials are parameterized for each pair of atom types and determine the average bulk density ρ_{bulk}^{ij} of atom type j and its radial distribution function $\rho_{seg}^{ij}(r)$ relative to atom type i , where r is the distance between them. The correction factor $f_{Vol_correct}^j(r)$ describes the volume occupied by the ligand molecule, and the factor $k_B T$ reduces it to the units of free energy of binding.

$$PMFscore = -k_B T \times \ln \left(\sum_i \left[f_{Vol_correct}^i(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right] \right) \quad (1)$$

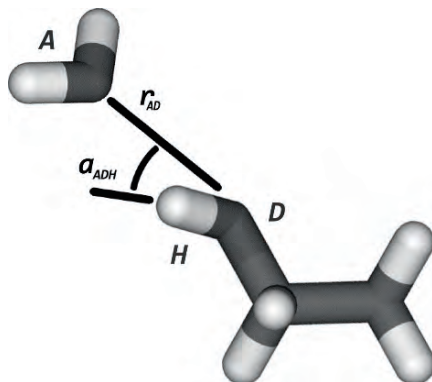
Such scoring functions implicitly account for the special features of ligand–receptor interactions that could be otherwise difficult to be described explicitly. However, this approach is highly dependent on the atom typing procedure and may be difficult to be interpreted in terms of commonly recognized interactions such as electrostatic, hydrophobic, etc.

More often, docking algorithms are based on empirical scores, which usually represent a combination of force field based terms with explicit empirical terms of hydrogen bonds, lipophilic contact, metal coordination interactions, etc. One of the most popular is the “chemscore” developed by Eldridge et al. in 1997 [18], see Equation 2. Here, the intermolecular interactions are represented by a linear combination of terms: hydrogen bonds (C_{hbond}), metal coordination bonds (C_{metal}), lipophilic contacts (C_{lipo}), and the number of ligand torsion angles “frozen” upon binding as a measure of the entropy change. The distinctive feature of empirical scores is their easy parameterization. Development of such scoring functions is based on the regression analysis that fits the weighting coefficients so that the final equation reproduces experimental data such as the measured affinity or geometry of the receptor–ligand complex.

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{lr}) + C_{hbond} \sum g(r_{AD}) h(\alpha_{ADH}) + C_{metal} \sum f(r_{lm}) + C_{roth} H_{roth} \quad (2)$$

Another distinctive feature of such scoring functions is the way how the interactions terms are represented. Here, the value of the score is substantially determined by the weighting coefficients, while the particular form of the interaction term plays less important role. Therefore, it is possible to simplify the equations and save the calculation time. Thus, the presence of hydrogen bonds can be identified through simple geometric criteria (Fig. 3) – the distance between donor and acceptor $g(r_{AD})$ and the angle between them and the hydrogen atom $h(\alpha_{ADH})$. Similarly to that, the terms of metal coordination bonds $f(r_{lm})$ or hydrophobic contact $f(r_{lr})$

Fig. 3. Geometrical parameters of a hydrogen bond that are used in scoring functions are the distance between the donor and the acceptor r_{AD} and the angle between them and the hydrogen atom α_{ADH} .



can be described by simple dependencies like the Fermi-type function [19]. Here, r_{lr} is the distance between the hydrophobic atoms of a ligand and a receptor, and r_{lm} is the distance between the acceptor atoms of a ligand and metal ions in the receptor active site.

Of course, the efficiency of empirical scores depends on the quality of the training sets used to parameterize the weighting coefficients. They are not likely to be applicable to protein-targets of ligand molecules that differ greatly from those in the training set. For example, if the score was trained to work better with complexes where hydrophilic contacts (hydrogen bonds, salt bridges, etc.) dominate, then it cannot yield reliable results for those where hydrophobic contact plays the determinative role in receptor–ligand recognition.

Empirical scoring function may be the best choice when a representative training set can be constructed. In this case, the experimental information about native receptor–ligand contacts can be implicitly encoded in the weighting coefficients. This effect has been used in some studies to develop the system-specific or customizable scores for docking [4, 20–22]. Two distinct approaches can be applied here. The first variant is to use experimental data on binding of different ligands to a certain protein-target. Then the target-specific score is applied to search for other ligands from the libraries of low molecular weight compounds to find those that are able to bind the target with greater affinity. Another approach is to design ligand-specific criteria and is less wide-spread since it requires a considerable amount of experimental data on interactions of a particular class of ligands with proteins. However, there have been presented successful studies aimed at development of such scores for peptides [23], saccharides [24], ATP [25]. Taking into account the ever increasing amount of experimental data on biomolecular complexes it is likely that the use of customizable algorithms will gain more and more popularity in molecular docking investigations.

Hydrophobic Interactions

Apolar substances tend to avoid contacts with polar solvent (water) by grouping with other apolar molecules. This results in effective hydrophobic (lipophilic) interaction which is recognized to play an important role in interactions and specific recognition in biomolecular systems. The manifestation of the hydrophobic effect is ubiquitous: interactions of lipids forming a bilayer cell membrane with each other or with transmembrane proteins is to much extent driven by hydrophobicity [26]; in particular cases it plays the main role in specific recognition of ligands by protein-receptors [27, 28].

Despite this, the nature of the physical forces lying in the basis of hydrophobic effect remains uncertain. Many different intermolecular interactions contribute to this phenomenon: electrostatic, van-der-Waals, hydrogen-bonding, entropic effect, etc. That is why, there still has not been proposed a universal quantitative approach to estimate the energetics of hydrophobic contact.

The importance of this effect in molecular biology has lead to development of empirical methods. One of the most primitive ways, used in the “chemscore” function [18], is to sort all atoms into hydrophobic (aliphatic and aromatic carbons, sulfur, halogens) and hydrophilic (all other atoms). The number of close contacts between hydrophobic atoms can serve as a rough measure of the magnitude of the receptor–ligand hydrophobic interaction. However, this approach has serious limitations. Thus, considering the important case of heterocycle-containing ligands (i.g. adenosinetriphosphate, ATP) one can see that the whole molecule will be characterized as hydrophilic. Meanwhile, it has been convincingly demonstrated that the adenine moiety of ATP tends to bind in hydrophobic pockets of protein active sites [29–32] and thus can be considered as hydrophobic, too (Fig. 4).

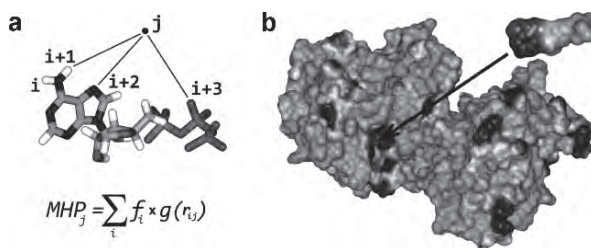


Fig. 4. The concept of MHP (Molecular Hydrophobicity Potential) is based on atomic hydrophobicity constants that contribute to the MHP value at any spatial point *j* (a). (b) Complementarity of hydrophilic and hydrophobic properties in ATP–Ca-ATPase complex (only the nucleotide-binding domain of Ca-ATPase is shown). Favorable hydrophobic match between the adenine ring of the ligand and its binding pocket is clearly seen (colored in dark).

More elaborated methods are based on experimental data on partitioning of organic compounds between polar and apolar environments (e.g. water and octanol/hexane/chloroform). The logarithm of the ratio of concentrations of a molecule in two solvents $\log P$ (Equation 3) serves as a quantitative measure of its hydrophobicity. The empirical approaches to describe molecular hydrophobic/hydrophilic properties are based on the concept of linear interaction energy, i.e. the molecular $\log P$ can be decomposed into a linear combination of atomic/fragmental hydrophobicity constants. One of the most widely used parameterization is based on partition data for ~900 compounds in water/octanol system [33], which was later updated on a larger training set [34]. Here, each atom is attributed to one of the 120 topological atom types, including explicitly given hydrogens. Other parameterizations are known – without explicit hydrogens [35], fragmental [36, 37], and so on.

Empirical atomic hydrophobicity constants can be used to view the spatial distribution of molecular hydrophobic/hydrophilic properties – e.g. on the molecular surface. This can be done using the formalism of the Molecular Hydrophobicity Potential (MHP) also known as Molecular Lipophilicity Potential (MLP). The hydrophobic properties are calculated on molecular surfaces or at any point in space using a distance-dependent decay function (see Fig. 3):

$$MHP_j = \sum_i f_i \times g(r_{ij}) \quad (3)$$

where MHP_j is the value of MHP at spatial point j , f_i is the atomic hydrophobicity constant, $g(r_{ij})$ is the distance-decay function. It is noteworthy that there is no exact expression for the “hydrophobic field” and thus all such functions are of intuitive and empirical nature. Usually, exponential decay is used $g(r) = e^{-r}$ [38] or $g(r) = e^{-r/2}$ [27], however hyperbolic $g(r) = (1 + r)^{-1}$ [39] or Fermi-like cut-off $g(r) = (1 + e^{-(r - r^0)})^{-1}$ [19] functions are also applied in some studies. Some of these functions are supposed to behave better when applied to low molecular weight compounds, others – to macromolecules such as proteins [19].

Different quantitative criteria have been proposed to characterize the contribution of hydrophobic interactions into protein–ligand binding, estimating the complementarity between hydrophobic/hydrophilic properties of both molecules. Some approaches are based on a score depending on the sign and magnitude of the hydrophobicity of compared atoms [40–42] or surface points [28]. Other methods simply estimate the hydrophobic complementarity by the number of contacts between hydrophobic atoms or groups [18], by the fraction of buried hydrophobic surface [43], or by matching of surface regions [44] of molecules.

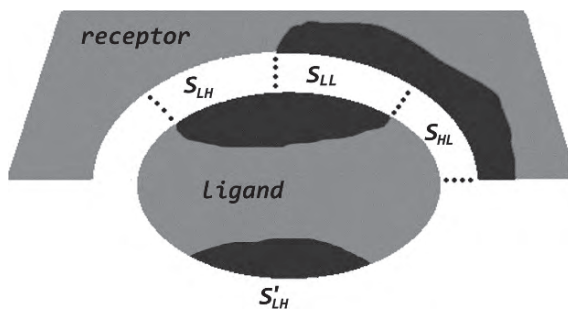


Fig. 5. The scheme illustrating the types of contacts between lipophilic (L) and hydrophilic (H) areas of a ligand in the binding site. S_{LL} and S_{HL} are the contacts of lipophilic areas of the receptor with lipophilic and hydrophilic areas of the ligand, respectively. S_{LH} and S'_{LH} are the contacts of lipophilic areas of the ligand with hydrophilic area of the binding site and surrounding water, respectively.

$$MHP_{match} = \frac{2S_{LL}}{2S_{LL} + S_{HL} + S_{LH} + S'_{LH}} \quad (4)$$

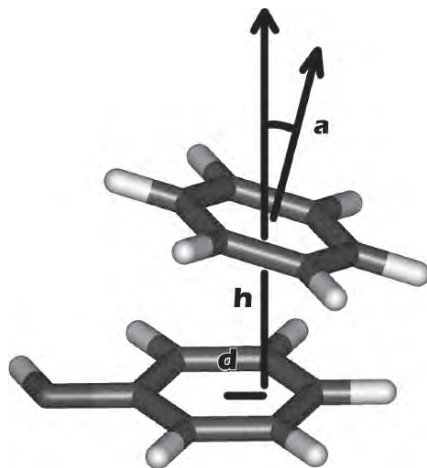
Testing these different approaches we found that in case of ATP it is better to use the fraction of buried hydrophobic surface area of ligand MHP_{match} as a quantitative criterion to score the ATP docking poses [25] (Fig. 5; Equation 4). Besides, we have found that a moderate shift to the more hydrophobic range (+0.03 atomic $\log P$ units) improves the distribution of properties for low-molecular-weight compounds, particularly nucleosides and nucleotides. Meanwhile, such moderate shift does not influence protein MHP greatly.

Stacking

Aromatic stacking is another type of intermolecular contacts which plays a distinctive role in biological recognition that deserves to talk about. To our view, stacking is not always paid due attention in many scoring functions, however it is often observed in receptor–ligand complexes since most of drug-like compounds contain aromatic fragments.

Stacking interaction can occur between two aromatic groups and results in their specific arrangement in space. The most typical example is DNA, where nitrogen bases form parallel stacking contacts [45, 46]. Other variants are also possible – a so-called T-shaped arrangement is observed for such compounds as benzene [47]. Furthermore, aromatic substances tend to form cation–pi interactions where positively charged groups interact with

Fig. 6. Scheme of geometrical parameters used to describe a stacking contact between two aromatic rings. Displacement (d) and height (h) are calculated for the center of one aromatic ring relative to another ring's plane. Angle α is calculated as the angle between the normal vectors of both rings.



aromatic electron clouds [48–50]. Like the hydrogen bonds, the aromatic contact can be described by geometrical criteria. In our studies of ATP – protein interactions we found that stacking of adenine with aromatic side-chains of protein amino acids is quite typical [25], which is also confirmed by numerous other studies [29–32, 51–53]. To identify a stacking contact we analyzed the mutual orientation of two aromatic fragments in terms of the height h and displacement d of one cycle with respect to the other, and the angle α between them (Fig. 6).

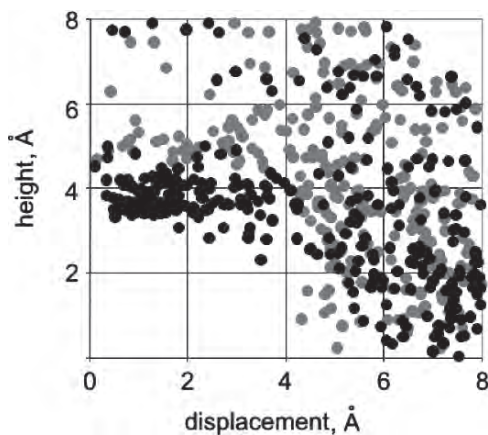
Initially, it was unclear what values of these parameters correspond to the presence of a stacking contact – different cut-offs have been used [54, 55]. Additional ambiguity is related to the fact that many aromatic groups tend to form both parallel and T-shaped contacts as it has been demonstrated for amino acids in proteins [55, 56] and in model systems consisting of simple aromatic hydrocarbons – benzene, naphthalene, etc. [4, 57–60]. Our analysis of spatial structures of complexes of adenine-containing ligands with proteins retrieved from the Brookhaven Protein Data Bank (PDB) [61] has revealed that the adenine base prefers to make parallel stacking with $h = 3 - 4 \text{ \AA}$, and sometimes T-shaped contact with $h = 4.5 - 5.5 \text{ \AA}$ as is shown in Fig. 7 for Tyr. Interestingly, the guanidine group of Arg often forms parallel contacts with adenine similar to aromatic amino acids.

So, we proposed to use the following criterion to describe the interaction of two aromatic fragments (including flat guanidine group):

$$S = S_d(d) \times S_h(h) \times S_\alpha(\alpha) \quad (5)$$

where d and h are, respectively, displacement and height of the center of one aromatic ring relative to each other, and α is the angle between their planes (Fig. 6); the weighting functions in the product are of the following form:

Fig. 7. The distribution of aromatic rings of Tyr around the adenine moiety of various ligands in complexes with protein receptors. The height, the displacement, and the angle α between the two interacting cycles are defined as in Fig. 6. Dark and grey circles correspond to parallel ($\cos^2\alpha = 0.6-1.0$) and T-shaped ($\cos^2\alpha = 0.0-0.4$) geometry, respectively.



$$S_{\alpha}(\alpha) = \max(\cos^2 \alpha; \sin^2 \alpha) \quad (6)$$

where the particular form of $S_{\alpha}(\alpha)$ defines whether we observe parallel ($\cos^2\alpha \geq 0.5$) or “T-shaped” edge-to-face ($\sin^2\alpha > 0.5$) stacking;

$$S_h(h) = \begin{cases} 1.0; & h \leq h_1 \\ (h_0 - h) / (h_0 - h_1); & h_1 < h < h_0 \\ 0.0; & h_0 \leq h \end{cases} \quad (7)$$

where $h_1 = 4.0 \text{ Å}$ and $h_0 = 5.0 \text{ Å}$ for parallel stacking and $h_1 = 5.0 \text{ Å}$ and $h_0 = 6.0 \text{ Å}$ for edge-to-face stacking;

$$S_d(d) = \begin{cases} 1.0; & d \leq d_1 \\ (d_0 - d) / (d_0 - d_1); & d_1 < d < d_0 \\ 0.0; & d_0 \leq d \end{cases} \quad (8)$$

where $d_1 = 2.0 \text{ Å}$ and $d_0 = 3.0 \text{ Å}$ for any type of stacking arrangement. The parameters of stacking were derived from the statistical analysis of contacts of nucleobases with aromatic amino acids in proteins [25].

The presented method of identifying aromatic contacts in receptor–ligand complexes along with calculations of molecular hydrophobic/hydrophilic properties has been implemented in the PLATINUM web service (Protein–Ligand ATtractions Investigation NUMerically) on the web-site of our Laboratory <http://model.nmr.ru/platinum/>. This web service provides tools to analyze and re-score the results of protein – small molecule docking obtained with a standard docking software. According to our studies, this software can be quite useful in particular cases [25, 62] – see the detailed discussion in CONSENSUS APPROACHES section below.

Comparing Docking/Scoring Algorithms

It would be appropriate here to say a few words about criteria used to compare the efficiency of different docking/scoring schemes. The general idea is that the computational protocol under consideration can be compared to known structural/physicochemical data and should reproduce the native binding modes and binding free energies of ligands for a set of receptor–ligand complexes.

Root-mean-square deviation (RMSD) is the most common way to compare docking poses. RMSD between poses a and b is calculated over 3D-coordinates of all N heavy atoms of ligand by the following Equation 11:

$$RMSD = \sqrt{\frac{\sum_i^N ((x_i^a - x_i^b)^2 + (y_i^a - y_i^b)^2 + (z_i^a - z_i^b)^2)}{N}} \quad (9)$$

For two completely similar positions of a ligand in 3D-space $RMSD = 0$. In docking applications $RMSD = 2 \text{ \AA}$ is usually considered acceptable, i.e. if docking yields ligand poses within this cutoff of RMSD from crystallographic (or NMR) structure for most of the reference receptor-ligand complexes, it is likely to produce reliable predictions for others.

Alternatively the Tanimoto coefficient of interaction fingerprints (TcIFP) can be used. IFP represents a more complex notion and has not yet achieved wide use. In some cases even low RMSD may correspond to distinctly different intermolecular contact pattern, and vice versa. One way to solve this problem may be using some quantitative measure of similarity of contacts with receptor, such as hydrogen bonds, salt-bridges, and so forth. IFP provides such a measure.

Introduced by Deng et al. (2004) [54], IFP represents a conversion of 3D pattern of contacts between ligand and receptor into 1D bit string. Here, a set of bits, corresponding to various intermolecular contacts is assigned to each protein residue in the vicinity of the ligand-binding site (Fig. 8) When a contact with a particular residue is observed, e.g. a hydrogen-bond is formed, the corresponding bit is raised. And two bit-strings coding two docking poses can be compared now indicating whether they are similar or not. To compare the strings A and B, coding two ligand poses a Tanimoto coefficient is used:

$$TcIFP = \frac{|A \cap B|}{|A \cup B|} \quad (10)$$

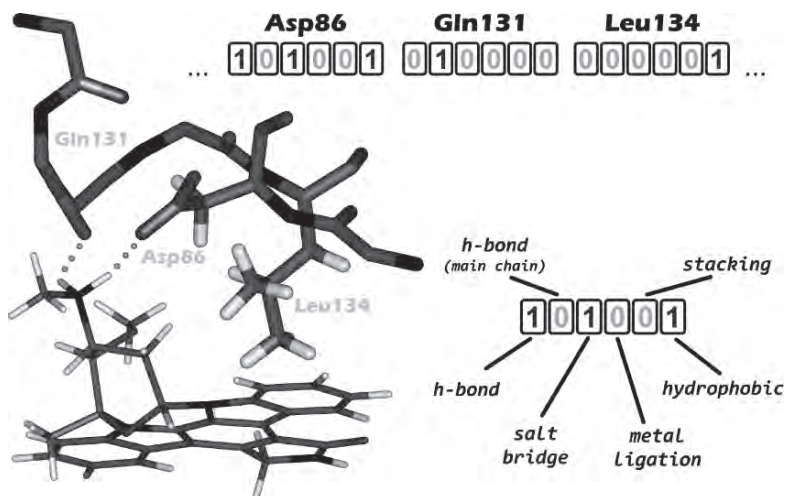


Fig. 8. An example of IFP bit string used in our program PLATINUM. Inhibitor staurosporin is shown in Cyclin-Dependent Kinase 2 active site (PDB code 1AQ1). Interactions of selected residues are shown as they are encoded into an IFP bit string.

Here $A \cap B$ is the number of raised bits common to both poses, and $A \cup B$ is the number of bits raised for either ligand A or B . Consequently the value of TcIFP can vary between 0 (different poses) and 1 (identical poses).

RMSD and TcIFP can tell whether a docking method predicts correct (native-like) orientation of a ligand in a receptor binding site. However, often more than one docking solution is considered. And even if the top-ranked pose is not correct it is interesting, to what rank in the list it had been assigned – i.e. how efficient the scoring method is. This question also arises in the virtual HTS projects, where enrichment of the top of the list of docking solutions with high-affinity (true binding) compounds is an important measure, which is usually expressed in terms of the enrichment factor and the receiver operating characteristic (ROC) curve [63].

The ROC curve demonstrates the percentage of the true binders found in a top ranked subset of the docking solutions (Fig. 9a). The area under the ROC curve can serve as a measure of the efficiency of a scoring function. If a score is completely unable to distinguish between true ligands and junk compounds then the area will be $\frac{1}{2}$ of the whole graph, which corresponds to random distribution. However, the area of the perfect ranking will also be less than 1. The enrichment factor is calculated as a ratio between the percentage of true binders in a top ranked subset and the percentage in the whole set (Fig. 9b). Both these values characterize the efficiency of a scoring function and may vary for the same set of docking poses ranked by different scoring criteria.

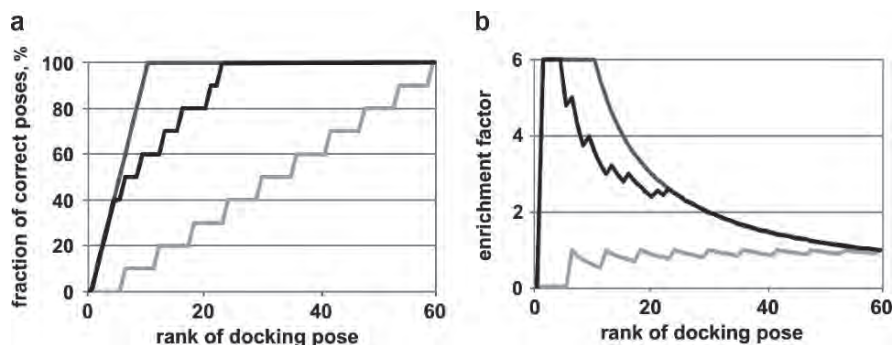


Fig. 9. The ROC curve (a) and enrichment factor (b) graphs for the results of ATP docking to Ca-ATPase (PDB code 1WPE) scored by the ATP-specific hydrophobicity-based criterion (black) [25]. For comparison, the graphs for perfect (dark grey) and random (light grey) scorings are given.

Receptor Flexibility

All docking methods imply that the atomic 3D-structure of the receptor is known and in standard docking algorithms it is kept rigid. Only the small molecule of ligand is allowed full torsion flexibility. The rationale for keeping the receptor rigid is the enormous number of its degrees of freedom which could lead to the combinatorial explosion, were they not frozen. Meanwhile, the necessity of simulating the flexibility of the receptor has been demonstrated in structural studies of protein–protein or ligand–protein associations showing that these processes are often accompanied by conformational changes in protein structure [64–67]. This effect is referred to as induced fit.

Nevertheless, starting from the earliest docking tools, the best known of which is DOCK [68], which were based on the rigid key-and-lock mechanism of ligand–receptor recognition, the conformational search algorithms become able to simulate flexibility of larger and larger molecules. This is due to two reasons. Firstly, the computational power of modern computers is increasing rapidly. Secondly, the accumulated structural data for proteins (which usually are the receptors in docking) provide guidelines to restrict the conformational space of the receptor to a reasonable number of principal degrees of freedom.

Many approaches to take the receptor flexibility into account have emerged in the leading scientific groups throughout the world, while they have not yet become a routine of the standard docking protocols. It would be appropriate to mention here the most popular of them. The “relaxed complex” method [69] is the most obvious solution to the problem and implies docking of a ligand to a set of representative conformers of the receptor.

The latter could be retrieved from X-ray or NMR [70–73] or generated by molecular dynamics (MD) or other simulation techniques [69, 74–83]. The reliability of the docking results in this case depends on the representative quality of the ensemble of receptor conformers. Another popular method is based on the rotamer libraries [84–92] of amino acids side-chains, which are sampled for residues in the vicinity of the binding site. Many other approaches to simulate the receptor flexibility in docking can be found in excellent recent reviews [93–97].

Our focus is the efficiency of scoring functions when docking is applied to the “flexible receptor” model. We studied the complex of ATP with the ATP-binding domain of Sarcoplasmic Ca-ATPase. This integral transmembrane protein pumps Ca^{2+} ions across the membrane against the electrochemical gradient in exchange for ATP hydrolysis. The structure of the enzyme has been resolved at different stages of the enzymatic cycle [98]. It has been identified that the transmembrane TM-domain, that houses the Ca^{2+} -binding site, consists of ten transmembrane alpha-helical segments and it is linked to three cytoplasmic domains (A, actuator or anchor; N, nucleotide-binding; and P, phosphorylation). The N-domain is inserted into the P-domain and these domains together are also referred to as the ATP-binding domain. Interestingly, N- and P-domains are well apart in the absence of the substrate (the open form) and come close to each other upon binding of ATP (the closed form) so that it could interact with active sites on both these domains simultaneously (Fig. 10).

Prediction of interactions of the ATP molecule with the ATP-binding domain would be a challenge for docking methods if the substrate-free open form of the receptor is used. Some approaches to model the global domain motions of the receptor in docking have been proposed. In the earliest attempt to take into account such receptor rearrangements, hinge-bending motions of protein domains were considered, while the domains themselves were held rigid [99]. Normal mode analysis has also been used to sample large-scale loop

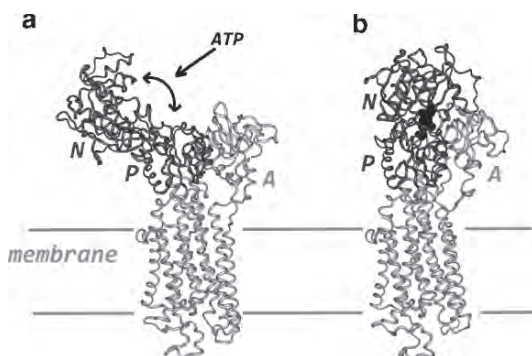


Fig. 10. Ca-ATPase in the open (*left*) and closed (*right*) forms of the ATP-binding domain illustrated with the structures from PDB entries 1SU4 and 1VFP, respectively. The ATP-binding domain which in turn consists of the N- and P-domains is shown in dark grey. ATP is shown in black in the right structure. The anchor domain is indicated with the symbol “A”.

motions of proteins [95, 100]. Recently, a hybrid method combining MD and harmonic dynamics was applied to incorporate domain motions into the docking procedure [101]. We used MD simulations of the ligand-free ATP-binding domain to generate an ensemble of protein conformers and demonstrated that it was possible to find a closed form of the receptor in MD trajectory starting from either closed or open form structure of Ca-ATPase [102].

When the problem of sampling the receptor conformational space had been solved, another one arose – how to score the docking solutions properly. It becomes clearly seen that standard scoring functions may not be applicable in this case when docking of ATP to an X-ray structure (PDB code 1WPE) and to MD-conformers (starting structure 1T5S) are compared (Fig. 11). The standard function “goldscore” (based on van der Waals and hydrogen-bond terms) of the docking package GOLD [16] though not perfect, performed satisfactory for the X-ray structure, but for the ensemble of Ca-ATPase MD-conformers its efficiency dropped to nearly random. At the same time the performance of the ATP-specific score [25], which is based on more delocalized hydrophobic interactions, remains almost the same. These results indicate that in the “flexible receptor” model incorporating hydrogen bonding terms into the scoring function may be inadequate, since even moderate conformational changes of side-chains may seriously alter the value of such terms. Similar conclusions were drawn from a study of docking ligands to high- and low-resolution protein–ligand complexes with different scoring functions [103]. So, we may conclude that additional optimization of scoring functions may be required for docking ligands to binding sites that have a loosely defined 3D-structure. We were lucky that the term of hydrophobic match played important role in ATP – Ca-ATPase recognition, however this is not always the case and further investigations are required.

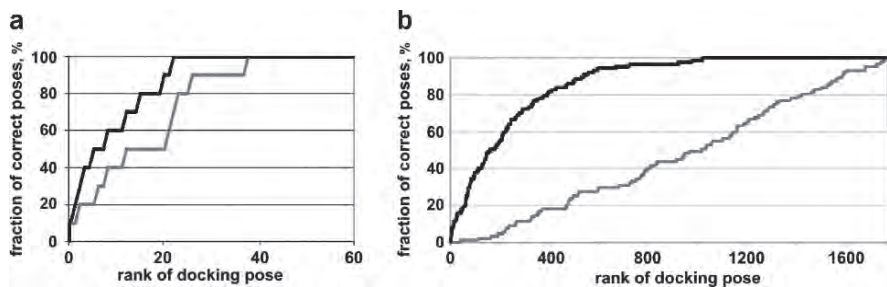


Fig. 11. Ranking of ATP docking posed obtained (a) for the X-ray structure of Ca-ATPase (PDB code 1WPE) and for (b) MD-conformers of its ATP-binding domain (starting structure 1T5S) using the docking program GOLD. The ROC graphs for ATP-specific score (black) and “goldscore” function (grey) are shown. The total number of correct docking poses was 10 for the X-ray structure and ~400 for the ensemble of MD-conformers.

Consensus Approaches

Combination of readily available scoring methods instead of developing them from the beginning can also improve the predictive power of docking and is commonly referred to as consensus approach. If a native-like docking solution had been generated but was not scored properly, one can use a more efficient scoring method to re-rank the list of docking poses and put the correct pose at top [104–109]. One example of such “consensus docking” was demonstrated with successive implementation of two GOLD scoring functions. One of them, “goldscore” [16], is a force-field-based scoring function that considers both van der Waals interactions and protein–ligand hydrogen-bonds (H-bonds) and the other one, “chemscore” [18], is an empirical scoring function comprising H-bonding, acceptor-metal and lipophilic terms. Considering different interaction terms, these two scoring criteria supplemented each other in a successive filter when docking poses obtained with “goldscore” were then re-scored by the “chemscore” values resulting in a considerable improvement of the success rate [110]. Similar results were obtained for rescoring the “chemscore” docking results with the “goldscore” function.

The term “consensus scoring” refers to slightly different technique. Here, the list of docking solutions is re-ranked by a criterion that combines a number of different scoring functions into a weighted combination [8–15]. In recent years, the increasing amount of experimental data has evoked interest in development of system-specific (customizable) scoring functions that in turn can be also used in a consensus docking. The weighting coefficients of interaction terms in such scoring functions are optimized for a particular protein-target or a class of ligand molecules [23–25].

Regarding our success in development of ATP-specific score, we undertook a study to find out whether similar ligand-specific approach can be extended over a broader class of ligands. Using a simple model of two-term score (hydrogen bonds and hydrophobic match), we have identified some relationships indicating that the hydrophobic match term should have higher weight in scores for non-peptide, medium-sized (10–25 heavy atoms) or amphiphilic compounds [62].

To make this approach available to a broader community, we have designed the PLATINUM web interface, where one can upload the results of docking and retrieve the quantitative estimation of hydrophobic/hydrophilic, hydrogen bond and stacking interactions for each of the docking poses. These data can then be used in a consensus re-scoring (Fig. 12). Besides protein–ligand complexes, PLATINUM can estimate hydrophobic match in other systems, such as e.g. peptide–lipid bilayer, etc., which can also be easily visualized online.

Results

Receptor name: 1VFP_prt.mol2

Reference ligand: 1VFP_atp.pdb

| Ligand name | r.m.s.d. | IFP | # H-bonds | $S_{L/L}$ | $S_{H/H}$ | S_{buried} | S_{total} | Match ¹ | Match ² | # Stack. | # Stack. Gua- π |
|-------------------|----------|-------|-----------|-----------|-----------|--------------|-------------|--------------------|--------------------|----------|---------------------|
| 1VFP_atp | 0.00 | 1 | 5.36 | 0.95 | 248.23 | 290.06 | 319.68 | 0.7794 | 0.0424 | 1.933 | 0 |
| gold_soh_atp_m1_2 | 1.94 | 0.714 | 5.55 | 5.25 | 248.80 | 289.11 | 322.97 | 0.7866 | 0.208 | 0.804 | 0 |
| gold_soh_atp_m1_5 | 8.16 | 0.208 | 2.04 | 0.00 | 234.87 | 287.97 | 335.25 | 0.7006 | 0 | 0 | 0.121 |
| gold_soh_atp_m1_4 | 1.19 | 0.632 | 4.27 | 0.25 | 244.18 | 280.63 | 313.54 | 0.7796 | 0.0128 | 1.891 | 0 |
| gold_soh_atp_m1_1 | 1.21 | 0.684 | 4.16 | 0.06 | 242.53 | 279.18 | 309.37 | 0.7842 | 0.0034 | 1.742 | 0 |
| gold_soh_atp_m1_3 | 3.89 | 0.55 | 1.87 | 1.77 | 229.30 | 252.59 | 311.52 | 0.7418 | 0.1083 | 1.659 | 0 |

Fig. 12. Screenshot of interaction terms for docking poses of ATP for Ca-ATPase structure (PDB code 1VFP) yielded by the PLATINUM web server. All docking poses can be compared to the reference ligand structure (1VFP_atp) in terms of RMSD and IFP. The interaction terms calculated by PLATINUM are the following: $S_{L/L}$, $S_{H/H}$, S_{buried} are the surface areas of lipophilic, hydrophilic, and whole contacts with the receptor molecule, respectively (the whole ligand surface area S_{total} is given for reference); Match¹ and Match² are the fractions of favorable lipophilic and hydrophilic/solely lipophilic surface areas; #H-bonds, #Stack, #Stack(Gua- π) are the counts of hydrogen bonds, stacking and stacking with guanidine group, respectively (the numbers are fractional due to the weighting functions of geometrical parameters of these contacts – see STACKING section).

Summary

The challenges and advances in modern scoring/docking techniques were reviewed here with particular attention paid to the role of hydrophobic interactions in receptor–ligand recognition. Among these were discussed the issues of system-specific scoring functions and methods to model receptor flexibility in docking. While some scoring functions do not account for them explicitly, we have demonstrated in our studies that using the MHP approach to measure the complementarity of hydrophobic/hydrophilic properties at the interface between ligand and receptor can efficiently complement standard scoring criteria. The PLATINUM web service designed in our Laboratory provides flexible tools for calculation and visualization of molecular hydrophobic/hydrophilic properties in receptor–ligand complexes.

However, further considerable investigations in the field of molecular docking will be required since there are known large classes of receptor–ligand systems, whose structure is predominantly determined by other intermolecular interactions. Overestimation of the role of hydrophobic match in such cases can only spoil the results. In future we hope to develop the system-specific approach to construction of scoring functions and to incorporate into the PLATINUM interface some predefined scoring criteria for particular receptor/ligand classes as e.g. for ATP binding proteins.

Acknowledgements This work was supported by the grants of Russian Foundation for Basic Research (No 06-04-49194-a, No 07-04-01514-a), by the Programme RAS MCB, and by the grant of the President of Russian Federation No MK-125.2008.4.

References

1. Kitchen D.B., Decornez H., Furr J.R., Bajorath J. *Nat. Rev. Drug. Discov.* (2004) **3**, 935–949.
2. Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C.R. *Br. J. Pharmacol.* (2008) **153**, S7–S26.
3. Abagyan R., Totrov M. *Curr. Opin. Struct. Biol.* (2001) **5**, 375–382.
4. Jansen J.M., Martin E.J. *Curr. Opin. Chem. Biol.* (2004) **8**, 359–364.
5. Cavasotto C.N., Orry A.J.W., Abagyan R.A. *Curr. Comput. Aided Drug Des.* (2005) **1**, 423–440.
6. Cavasotto C.N., Orry A.J. *Curr. Top. Med. Chem.* (2007) **7**, 1006–1014.
7. Rester U. *Curr. Opin. Drug Discov. Devel.* (2008) **11**, 559–568.
8. Bursulaya B.D., Totrov M., Abagyan R., Brooks C.L. *J. Comput. Aided Mol. Des.* (2003) **17**, 755–763.
9. Schulz-Gasch T., Stahl M. *J. Mol. Model* (2003) **9**, 47–57.
10. Ferrara P., Gohlke H., Price D.J., Klebe G., Brooks C.L. *J. Med. Chem.* (2004) **47**, 3032–3047.
11. Kellenberger E., Rodrigo J., Muller P., Rognan D. *PROTEINS* (2004) **57**, 225–242.
12. Kontoyianni M., McClellan L.M., Sokol G.S. *J. Med. Chem.* (2004) **47**, 558–565.
13. Perola E., Walters W.P., Charifson P.S. *PROTEINS* (2004) **56**, 235–249.
14. Wang R., Lu Y., Fang X., Wang S. *J. Chem. Inf. Comput. Sci.* (2004) **44**, 2114–2125.
15. Zhou Z., Felts A.K., Friesner R.A., Levy R.M. *J. Chem. Inf. Model* (2007) **47**, 1599–1608.
16. Jones G., Willett P., Glen R.C., Leach A.R., Taylor R.D. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* (1997) **267**, 727–748.
17. Muegge I. *J. Med. Chem.* (2006) **49**, 5895–5902.
18. Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P. *J. Comput. Aided. Mol. Des.* (1997) **11**, 425–445.
19. Heiden W., Moeckel G., Brickmann J. *J. Comput. Aided. Mol. Des.* (1993) **7**, 503–514.
20. Ferrara P., Curioni A., Vangrevelinghe E., Meyer T., Mordasini T., Andreoni W., Acklin P., Jacoby E. *J. Chem. Inf. Model* (2006) **46**, 254–263.
21. Amini A., Shrimpton P.J., Muggleton S.H., Sternberg M.J. *PROTEINS* (2007) **69**, 823–831.
22. Catana C., Stouten P.F.W. *J. Chem. Inf. Model.* (2007) **47**, 85–91.
23. Rognan D., Lauemoller S.L., Holm A., Buus S., Tschinke V. *J. Med. Chem.* (1999) **42**, 4650–4658.
24. Laederach A., Reilly P.J. *PROTEINS* (2005) **60**, 591–597.
25. Pyrkov T.V., Kosinsky Y.A., Arseniev A.S., Priestle J.P., Jacoby E., Efremov R.G. Complementarity of hydrophobic properties in ATP-protein binding: a new criterion to rank docking solutions. *PROTEINS* (2007) **66**, 388–398.
26. Efremov R.G., Chugunov A.O., Pyrkov T.V., Priestle J.P., Arseniev A.S., Jacoby E. *Curr. Med. Chem.* (2007) **14**, 393–415.
27. Gaillard P., Carrupt P.A., Testa B., Boudon A. *J. Comput. Aided. Mol. Des.* (1994) **8**, 83–96.
28. Testa B., Carrupt P.A., Gaillard P., Billois F., Weber P. *Pharm. Res.* (1996) **13**, 335–343.
29. Denissiouk K.A., Rantanen V.V., Johnson M.S. *PROTEINS* (2001) **44**, 282–291.
30. Cappello V., Tramontano A., Kock U. *PROTEINS* (2002) **47**, 106–115.
31. Kuttner Y.Y., Sobolev V., Raskind A., Edelman M. *PROTEINS* (2003) **52**, 400–411.
32. Mao L., Wang Y., Liu Y., Hu X. *J. Mol. Biol.* (2004) **336**, 787–807.
33. Viswanadhan V.N., Ghose A.K., Revankar G.R., Robins R.K. *J. Chem. Inf. Comput. Sci.* (1989) **29**, 163–172.
34. Ghose A.K., Viswanadhan V.N., Wendoloski J.J. *J. Phys. Chem.* (1998) **102**, 3762–3772.
35. Wang R., Gao I., Lay L. *Perspect. Drug Discov. Des.* (2000) **19**, 47–66.
36. Klopman G., Li J.-Y., Wang S., Dimayuga M. *J. Chem. Inf. Comput. Sci.* (1994) **34**, 752–781.

37. Meylan W.M., Howard P.H. *J. Pharm. Sci* (1995) **84**, 83–92.
38. Fauchere J.L., Quaredon P., Kaetterer L. *J. Mol. Graphics* (1998) **6**, 203–206.
39. Audry E., Dubost J.P., Colleter J.C., Dallet P. *Eur. J. Med. Chem.* (1986) **21**, 71–72.
40. Wang R., Luhua L., Shaomeng W. *J. Comp. Aided. Mol. Des.* (2002) **16**, 11–26.
41. Meng E.C., Kuntz I.D., Abraham D.J., Kellogg G.E. *J. Comput. Aided. Mol. Des.* (1994) **8**, 299–306.
42. Efremov R.G., Alix A.J.P. *J. Biomol. Struct. Dyn.* (1993) **11**, 483–507.
43. Bohm H.J. *J. Comput. Aided Mol. Des.* (1994) **8**, 243–256.
44. Exner T.E., Keil M., Brickmann J. *J. Comput. Chem.* (2002) **23**, 1176–1187.
45. Gotoh O. *Adv. Biophys.* (1983) **16**, 1–52.
46. Spomer J., Leszczynski J., Hobza P. *J. Biomol. Struct. Dynam.* (1996) **14**, 117–135.
47. Jorgensen W.L., Severance D.L. *J. Am. Chem. Soc.* (1990) **112**, 4768–4774.
48. Waters M.L. *Curr. Opin. Chem. Biol.* (2002) **6**, 736–741.
49. Meyer E.A., Castellano R.K., Diederich F. *Angew. Chem. Int. Ed.* (2003) **42**, 1210–1250.
50. Tewari A.K., Dubey R. *Bioorg. Med. Chem.* (2008) **16**, 126–143.
51. Traxler P., Furet P. *Pharmacol. Ther.* (1999) **82**, 195–206.
52. Saito M., Go M., Shirai T. *Protein Eng. Des. Sel.* (2006) **19**, 67–75.
53. Sigel H., Griesser R. *Chem. Soc. Rev.* (2005) **34**, 875–900.
54. Deng Z., Chuaqui C., Singh J. *J. Med. Chem.* (2004) **47**, 337–344.
55. Chakrabarti P., Bhattacharyya R. *Prog. Biophys. Mol. Biol.* (2007) **95**, 83–137.
56. Chelli R., Gervasio F.L., Procacci P., Schettino V. *J. Am. Chem. Soc.* (2002) **124**, 6133–6143.
57. Tsuzuki S., Honda K., Uchimaru T., Mikami M., Tanabe K. *J. Am. Chem. Soc.* (2001) **124**, 104–112.
58. Small D., Zaitsev V., Jung Y., Rosokha S.V., Head-Gordon M., Kochi J.K. *J. Am. Chem. Soc.* (2004) **126**, 13850–13858.
59. Sato T., Tsuneda T., Hirao K. *J. Chem. Phys.* (2005) **123**, 104307.
60. Malathy Sony S.M., Ponnuswamy M.N. *Crystal Growth & Design* (2006) **6**, 736–742.
61. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. *Nucleic. Acid. Res.* (2000) **28**, 235–242.
62. Pyrkov T.V., Priestle J.P., Jacoby E., Efremov R.G. *SAR QSAR Environ. Res.* (2008) **19**, 91–99.
63. Witten I.H., Frank E. Morgan Kaufman, New York (1999).
64. Betts M.J., Sternberg M.J. *Protein Eng.* (1999) **12**, 271–283.
65. Najmanovich R., Kuttner J., Sobolev V., Edelman M. *PROTEINS* (2000) **39**, 261–268.
66. McGovern S.L., Shoichet B.K. *J. Med. Chem.* (2003) **46**, 2895–2907.
67. Ferrara P., Jacoby E. *J. Mol. Model.* (2007) **13**, 897–905.
68. Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E. *J. Mol. Biol.* (1982) **161**, 269–288.
69. Lin J.H., Perryman A.L., Schames J.R., McCammon J.A. *J. Am. Chem. Soc.* (2002) **124**, 5632–5633.
70. Bouzida D., Rejto P.A., Arthurs S., Colson A.B., Freer S.T., Gehlhaar D.K., Larson V., Luty B.A., Rose P.W., Verkhivker G.M. *Int. J. Quantum Chem.* (1999) **72**, 73–84.
71. Erickson J.A., Jalaie M., Robertson D.H., Lewis R.A., Vieth M. *J. Med. Chem.* (2004) **47**, 45–55.
72. Murcia M., Ortiz A.R. *J. Med. Chem.* (2004) **47**, 805–820.
73. Barril X., Morley S.D. *J. Med. Chem.* (2005) **48**, 4432–4443.
74. Pang Y.P., Kozikowsky A.P. *J. Comput. Aided Mol. Des.* (1994) **8**, 669–681.
75. Kua J., Zhang Y., Eslami A.C., Butler J.R., McCammon J.A. *Protein Sci.* (2003) **12**, 2675–2684.
76. Kua J., Zhang Y., McCammon J.A. *J. Am. Chem. Soc.* (2002) **124**, 8260–8267.
77. Efremov R.G., Kosinsky Y.A., Nolde D.E., Tsivkovskii R., Arseniev A.S., Lutsenko S. *Biochem. J.* (2004) **382**, 293–305.
78. Schames J.R., Henchman R.H., Siegel J.S., Sottriffer C.A., Ni H., McCammon J.A. *J. Med. Chem.* (2004) **47**, 1879–1881.

79. Wong C.F., Kua J., Zhang Y., Straatsma T.P., McCammon J.A. *PROTEINS* (2005) **61**, 850–858.
80. Sivanesan D., Rajnarayanan R.V., Doherty J., Pattabiraman N. *J. Comput. Aided Mol. Des.* (2005) **19**, 213–228.
81. Mancinelli F., Caraglia M., Budillon A., Abbruzzese A., Bismuto E. *J. Cell Biochem.* (2006) **99**, 305–318.
82. Totrov M., Abagyan R. *Curr. Opin. Struct. Biol.* (2008) **18**, 178–184.
83. Zavodszky M.I., Kuhn L.A. *Protein Sci.* (2004) **14**, 1104–1114.
84. Tuffery P., Etchebest C., Hazout S., Lavery R. *J. Biomol. Struct. Dyn.* (1991) **8**, 1267–1289.
85. Dunbrack R.L. Jr., Karplus M. *J. Mol. Biol.* (1993), **230**, 543–574.
86. Leach A.R. *J. Mol. Biol.* (1994) **235**, 245–356.
87. Leach A.R., Lemon A.P. *PROTEINS* (1998) **33**, 227–239.
88. Lovell S.C., Word J.M., Richardson D.C. *PROTEINS* (2000) **40**, 389–408.
89. Frimmurer T.M., Peters G.H., Iversen L.F., Andersen H.S., Moller N.P.H., Olsen O.H. *Biophys. J.* (2003) **84**, 2273–2281.
90. Kallblad P., Dean P.M. *J. Mol. Biol.* (2003) **326**, 1651–1665.
91. Yang A.Y.C., Kallblad P., Mancera R. *J. Comput. Aided Mol. Des.* (2004) **18**, 235–250.
92. Hartmann C., Antes I., Lengauer T. *PROTEINS* (2009) **74**, 712–726.
93. Carlson H.A., McCammon J.A. *Mol. Pharmacol.* (2000) **57**, 213–218.
94. Carlson H.A. *Curr. Opin. Chem. Biol.* (2002) **6**, 447–452.
95. Cavasotto C.N., Kovacs J.A., Abagyan R.A. *J. Am. Chem. Soc.* (2005) **127**, 9632–9640.
96. May A., Zacharias M. *Biochim. Biophys. Acta* (2005) **1754**, 225–231.
97. Alonso H., Bliznyuk A.A., Gready J.E. *Med. Res. Rev.* (2006) **26**, 531–568.
98. Toyoshima C. *Arch. Biochem. Biophys.* (2008) **476**, 3–11.
99. Sandak B., Wolfson H.J., Nussinov R. *PROTEINS* (1998) **32**, 159–174.
100. Keseru G.M., Kolossvary I. *J. Am. Chem. Soc.* (2001) **123**, 12708–12709.
101. Tatsumi R., Fukunishi Y., Nakamura H. *J. Comput. Chem.* (2004) **25**, 1995–2005.
102. Pyrkov T.V., Kosinsky Y.A., Arseniev A.S., Priestle J.P., Jacoby E., Efremov R.G. *J. Chem. Inf. Model.* (2007) **47**, 1171–1181.
103. Bindewald E., Skolnick J. *J. Comput. Chem.* (2005) **26**, 374–383.
104. Charifson P.S., Corkery J.J., Murcko M.A., Walters W.P. *J. Med. Chem.* (1999) **42**, 5100–5109.
105. Wang R., Wang S. *J. Chem. Inf. Comput. Sci.* (2001) **41**, 1422–1426.
106. Clark R.D., Strizhev A., Leonard J.M., Blake J.F., Matthew J.B. *J. Mol. Graph. Mod.* (2002) **20**, 281–295.
107. Paul N., Rognan D. *PROTEINS* (2002) **47**, 521–533.
108. Xing L., Hodgkin E., Liu Q., Sedlock D. *J. Comput. Aided Mol. Des.* (2004) **18**, 333–344.
109. Mozziconacci J.C., Arnoult E., Bernard P., Do Q.T., Marot C., Morin-Allory L. *J. Med. Chem.* (2005) **48**, 1055–1068.
110. Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D. *PROTEINS* (2003) **52**, 609–623.

STRUCTURES OF CVNH FAMILY LECTINS

ANGELA M. GRONENBORN

*Department of Structural Biology, University of Pittsburgh
School of Medicine, Pittsburgh PA 15261, USA
E-mail: amg100@pitt.edu;*

Abstract Members of the CVNH family are found in a restricted range of eukaryotic organisms as diverse as filamentous ascomycetes and seedless plants. All CVNH proteins so far exhibit a fold that matches the unique fold of the cyanobacterial protein. The CVNH domain is a versatile protein module, and, with some exceptions, comprises 101–150 aa with two sequential repeats of 50 amino acids. We determined high resolution structures of CVNHs from *Tuber borchii*, *Ceratopteris richardii*, *Neurospora crassa*, and *Gibberella zeae*, representing different phylogenetic groups. All proteins exhibit the same fold and the overall structures resemble that of the founding member of the family, CVN, albeit with noteworthy differences in loop conformation and detailed local structure.

Introduction

The CVNH sequence and fold was first discovered in a protein from *Nostoc ellipsosporum*, where it was detected as the active, HIV-inactivating species in an aqueous extract. This founding member of the family, CVN, is an 11-kDa protein that blocks virus entry into target cells by binding to high-mannoses on viral surface glycoproteins. In this manner CVN interferes with pivotal interactions between viral and target cell receptors, preventing virus-cell fusion. The initial biochemical characterization was performed on protein isolated from the cyanobacterial extract, yielding the primary amino acid sequence and disulfide bonding pattern and all subsequent work was carried out on protein expressed from a synthetic gene over-expressed in *E. coli*. The *N. ellipsosporum* CVN gene was only recently identified and characterized (Gronenborn, unpublished data). Structurally, several high-resolution three-dimensional structures of CVN with and without bound carbohydrate have

been determined by NMR and X-ray crystallography since its discovery. Two different quaternary structures of CVN are observed. In dilute solution, the major form of the protein is monomeric and its structure was solved by heteronuclear, multidimensional NMR. It comprises a compact, bilobal fold with pseudo-symmetry that had no structural similarities to other proteins in the PDB. Interestingly, the amino acid sequence repeats of CVN do not constitute the repeated structural domains. Rather, the two symmetrically related domains, A and B, are formed by strand exchange across the two sequence repeats. In the monomer, the pseudo-symmetric domains A and B consist of residues [1–38/90–101] and [39–89], respectively, each comprising a triple-stranded β -sheet with a β -hairpin packed on top. At the pseudo-twofold axis both domains are held together by a cluster of hydrophobic residues comprising V39, H90, W49 and Y100, residing in two helical turns and the tips of the two β -hairpins. In addition, the interaction between W49 and D89, involving both hydrophobic contacts and a hydrogen bond between the N ϵ H atom of W49 and the O δ atom of D89, contributes to this interface. In the middle of the sequence (residues 51–56) a helical linker is found in domain B whose counterpart in domain A consists of the N- and C-termini. The subsequently determined X-ray structure consisted of a domain-swapped dimer in which the monomers open up, and the two halves of the dimer contain essentially the same interactions as the monomer. The only significant structural difference between dimer and monomer is found in domain B in the 51–56 linker region that is helical in the monomer whereas it is essentially extended in the dimer. Depending on experimental conditions, either form predominates in solution and can be isolated for biophysical, structural and functional studies.

Recently, a family of CVN homologous proteins, collectively designated as CVNHs, has been discovered in multicellular ascomycetous fungi and in the fern *Ceratopteris richardii* [7]. Interestingly, a number of distinct features are present in these related sequences: (i) in fungal but not fern homologs, Cys residues corresponding to the two S–S bonds of CVN are absent. (ii) the sugar binding residues in CVN are not conserved in CVNH proteins, rendering predictions about carbohydrate ligand specificity fraught with difficulty. (iii) multiple CVNH sequences can be present in ferns and fungi (up to five in *Aspergillus oryzae*), and the CVNH module may represent solely a domain in a multi-domain protein. Therefore, these homologs can potentially act in a multi-functional manner. Possible activities for these proteins could be as diverse as cell–cell recognition and attachment or nutrient state responsiveness.

In order to address questions related to structure/function of CVNH proteins, we have determined the structures of four representative members of the CVNH family. Solution structures of *T. borchii*, *C. richardii*, *N. crassa*, and

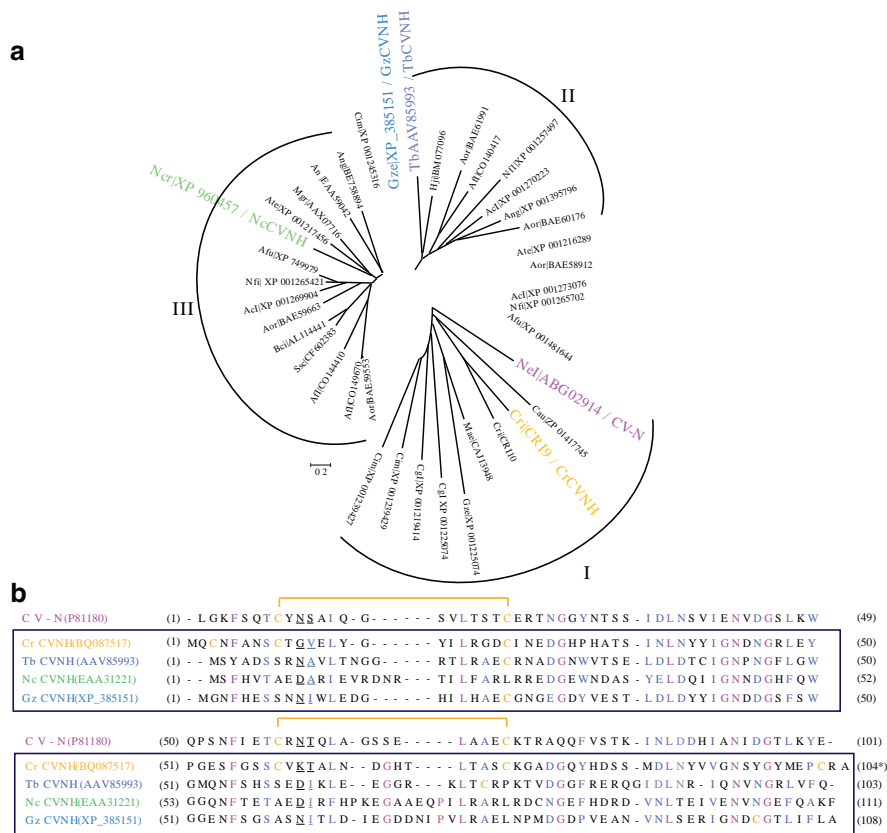


Fig. 1. Phylogenetic tree of the CVNH family and sequence alignment of CVN and four CVNH members.

G. zeae CVNHs were solved and we derived the structural basis of sequence conservation for these proteins. Using carbohydrate glycoarrays and direct NMR titrations we also ascertained their sugar binding specificities. The results of this work provide the basis for understanding protein–carbohydrate interaction in general and suggest testable hypotheses for further elucidation of the physiological roles played by CVNHs in their respective organisms.

Structures

Sequence analysis of an extended set of CVNHs revealed three distinct groups in the phylogenetic tree (Fig. 1a). Group I includes CVN, two additional prokaryotic cyanovirin-like polypeptides (MVN and CaVN), five

Fig. 2. 600 MHz ^{15}N - $^1\text{H}_\text{N}$ HSQC spectrum of GzCVNH.

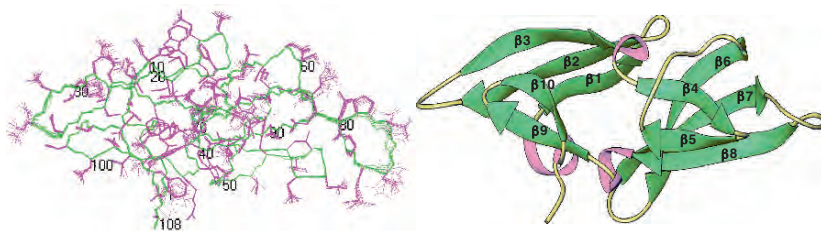


Fig. 3. Ensemble of 20 conformers and ribbon representation of the solution structure of GzCVNH.

of all resonances is observed. Spectra of similar quality were obtained for the other members as well.

Like CVN, the GzCVNH structure exhibits a pseudosymmetric, bilobal architecture. The same fold is observed for the other three new members. Structural comparison between CVNH members and CVN yields backbone r.m.s.d. values of 0.9–1.2 Å, with CrCVNH being the closest to CVN (Table 1). Similar values (0.1–1.5 Å) were also observed for all pairwise comparisons between the three new members (Table 1).

Using the same domain designation as described previously for CVN [3], domain A of GzCVNH comprises residues 1 to 39 and 95 to 108 (53 residues), residues 1 to 40 and 92 to 104 (53 residues) in CrCVNH, residues 1 to 40 and 93 to 103 (51 residues) in TbCVNH, and residues 1 to 42 and 100 to 111 (56 residues) in NcCVNH. Domain B, on the other hand, extends from residues 40 to 94 (55 residues) in GzCVNH, 41 to 91 (51 residues) in CrCVNH, residues 41 to 92 (52 residues) in TbCVNH, and 43 to 99 (57 residues) in NcCVNH. As for CVN, the tandem sequence repeats do not constitute the individual domains. Instead, each domain is formed by strand exchange between the two repeats. Several residues are completely conserved between the two-tandem repeats, with significant sequence identities noted within each CVNH sequence and between member sequences, ranging from 11% to 40%. Of these residues, the most conserved are F4, L18, G27, L36, G41, N42, G45, F54, L69, G78, L87, N93, and G96 (numbering according to the CVN sequence). All of these residues are located within the hydrophobic core and several loop regions, suggesting a critical structural role or their involvement in carbohydrate binding. The secondary structure elements in domain A are formed by a triple-stranded β -sheet ($\beta 1$, $\beta 2$, and $\beta 3$), a β -hairpin ($\beta 9$ and $\beta 10$) and two 3_{10} -helical turns ($\alpha 1$ and $\alpha 2$), again similar in the parent molecule CVN. Likewise, domain B is composed of a triple-stranded β -sheet ($\beta 6$, $\beta 7$, and $\beta 8$), a β -hairpin ($\beta 4$ and $\beta 5$) and two 3_{10} -helical turns ($\alpha 3$ and $\alpha 4$) (Fig. 4).

TABLE 1. Sequence identity (%) and r.m.s.* deviations (Å) for different CVNHs.

| | CVN | CrCVNH | TbCVNH | NcCVNH | GzCVNH |
|--------|-----|--------|--------|--------|--------|
| CVN | — | 0.94 | 1.01 | 1.09 | 1.19 |
| CrCVNH | 25 | — | 1.10 | 1.10 | 1.42 |
| TbCVNH | 24 | 22 | — | 1.22 | 1.51 |
| NcCVNH | 22 | 19 | 28 | — | 1.54 |
| GzCVNH | 30 | 28 | 37 | 33 | — |

Sequence identity scores were obtained using ClustalW for alignment of the entire sequence of the individual CVNHs.

*r.m.s. deviations of the lowest energy structure in each CVNH ensemble were calculated for residues 11–14, 17–23, 29–35, 40–42, 46–48, 60–63, 69–74, 81–86, 91–94, and 97–100 for CVN, 12–15, 18–24, 30–36, 41–43, 47–49, 61–64, 70–75, 82–87, 92–95, and 98–101 for CrCVNH, 10–13, 18–24, 30–36, 41–43, 47–49, 61–64, 71–76, 84–89, 93–96, and 99–102 for TbCVNH, 10–13, 20–26, 32–38, 43–45, 49–51, 63–66, 78–83, 90–95, 100–103, and 106–109 for NcCVNH, and 11–14, 18–24, 30–36, 41–43, 47–49, 61–64, 75–80, 87–92, 97–100, 103–106 for GzCVNH.

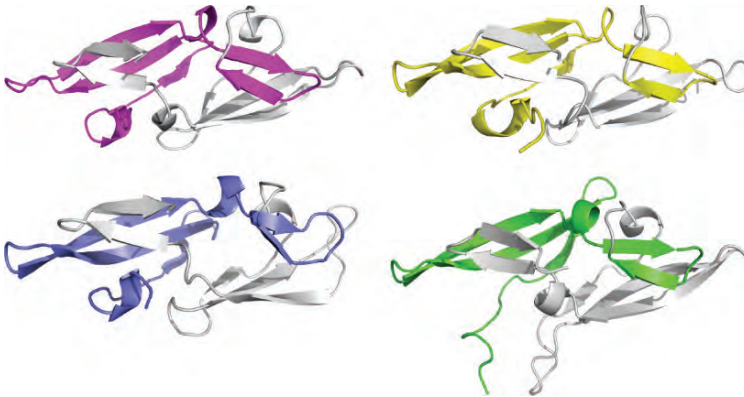


Fig. 4. Ribbon diagrams of the structures of CVN, CrCVNH, TbCVNH, and NcCVNH. The first sequence repeat in each protein is color coded [CVN (magenta), CrCVNH (yellow), Tb CVNH (blue), and Nc CVNH (green)] to illustrate that the domain structure does not coincide with the sequence repeat.

The close similarity in fold for all four proteins is caused by significant amino acid conservation throughout the hydrophobic core and at the domain interface of individual CVNH structures.

Despite the significant degree of sequence conservation and high structural similarity among CVNH members, several clear differences emerged. The first involves the presence or absence of disulphide bonds. Based on the original CVN structure and initial biochemical data it appeared that the disulphides were critical for maintaining the fold and antiviral activity of the molecule [6].

Interestingly, equivalent S–S bonds to those in CVN (between C8–C22 and C58–C73) are only found in CrCVNH (between C9–C23 and C59–C74) (Fig. 1b). In CrCVNH, an additional disulphide bond is present, namely between C3 and C102. For the other three proteins, TbCVNH, NcCVNH and GzCVNH, no S–S bridges are seen. In TbCVNH and GzCVNH, only one Cys residue corresponding to C22 of CVN is conserved. The other Cys residues are replaced by S7 (CVN C8), S/A59 (CVN C58) and P/L75 (CVN C73). One Cys is present at position 72 in TbCVNH and position 101 in GzCVNH. However, given that no possible interacting Cys are found spatially close to position 58 or 101, no disulphide bond can be formed. In NcCVNH, all corresponding Cys are replaced by hydrophobic residues (Ala at positions 7/61 and Leu at positions 25/82). These residues are buried and hydrophobic interactions substitute for the covalent S–S bonds present in CVN. A second area of significant differences concerns the loop regions between β strands, especially those connecting strands 1 and 2 and strands 6 and 7 (Fig. 4). The number of residues within the loops varies considerably and their orientation is also different. The loops between strands 1 and 2 and strands 6 and 7 in TbCVNH contain one more residue than CrCVNH or CVN. These loops in NcCVNH are at least four residues longer than the equivalent loops in TbCVNH or CrCVNH. An increase in the loop length is clearly noted for all members of the fungi family and MV-N, the closest homolog to CVN. This is particularly true for the loop connecting strands 6 and 7. In addition, the β 8 strand of TbCVNH which is shorter than the corresponding strands in the other CVNHs. A third difference is an altered relative orientation of the two domains for all CVNH members. This different orientation is vividly apparent upon superposition of domains A and inspecting the relative orientation of domains B. Upon superposition of only domain A, notable differences are clearly observed in the second 3_{10} -helical turn (α 2), the β -hairpin composed of strands 4 and 5, the hinge preceding α 3, the fourth 3_{10} -helical turn (α 4) and the loop connecting strands 7 and 8 (Fig. 5). The major determinant for these orientational variations is a different residue composition around the pseudo twofold axis.

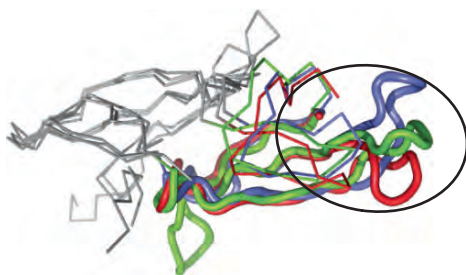


Fig. 5. Variation in loop conformation between β -strands β 7 and β 8.

It is interesting to note that, unlike CVN that can exist both in monomeric and domain-swapped dimeric forms [1], so far we have not observed a domain-swapped structure for any of the present CVNH members.

In addition, each protein exhibits different carbohydrate binding sites, with TbCVNH possessing a single binding site on domain A, NcCVNH one site only on domain B and CrCVNH two sites [5]. Carbohydrate binding specificities are distinct as well, and no potent HIV-inactivation was observed with any of these proteins. This contrasts the findings for the founding member CVN whose anti-HIV activity has been linked to the multivalent and multi-site interaction with high-mannose sugars on gp120 [2].

Acknowledgements I am indebted to all members of my research group, past and present, who contributed to the entire 'CVN' story. In particular, Laura Barrientos, Leo Koharudin, Elena Matei and John Louis played important roles. In addition, fruitful collaborations with Corine Sandström and Simone Ottonello are gratefully acknowledged. Work in the author's laboratory was in part funded by a National Institutes of Health grant (R01GM080642).

References

1. Barrientos, L.G., Louis, J.M., Botos, I., Mori, T., Han, Z., O'Keefe, B.R., Boyd, M.R., Wlodawer, A., and Gronenborn, A.M. (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure*, 10, 673–686.
2. Barrientos, L.G., Matei, E., Lasala, F., Delaglio, F., and Gronenborn, A.M. (2006) Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implication for viral entry inhibition. *Protein Eng Des Sel*, 19, 525–535.
3. Bewley, C.A., Gustafson, K.R., Boyd, M.R., Covell, D.G., Bax, A., Clore, G.M., and Gronenborn, A.M. (1998) Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Biol*, 5, 571–578.
4. Clore, G.M., and Gronenborn, A.M. (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol*, 16, 22–34.
5. Koharudin, L.M.I., Viscomi, A.R., Jee, J., Ottonello, S., and Gronenborn, A.M. (2008) The evolutionary conserved family of Cyanovirin-N homologs (CVNHs): Structures and carbohydrate specificity. *Structure*, 16, 570–584.
6. Mori, T., Shoemaker, R.H., Gulakowski, R.J., Krepps, B.L., McMahon, J.B., Gustafson, K.R., Pannell, L.K., and Boyd, M.R. (1997) Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochem Biophys Res Commun*, 238, 218–222.
7. Percudani, R., Montanini, B., and Ottonello, S. (2005) The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins*, 60, 670–678.

BIOPHYSICAL APPROACHES TO STUDY DNA BASE FLIPPING

SAULIUS KLIMAŠAUSKAS*, ZITA LIUTKEVIČIŪTĖ,
AND DALIA DAUJOTYTĖ#

*Laboratory of Biological DNA Modification
Institute of Biotechnology
LT-02241 Vilnius, Lithuania*

**Correspondence to: Prof. Saulius Klimašauskas
Institute of Biotechnology, Graičiūno 8
LT-02241 Vilnius, Lithuania;*

Email: klimasau@ibt.lt

*#Present address: MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 0QH, United Kingdom.*

Abstract The most dramatic yet localized enzyme-induced conformational deformation of the helical structure of DNA is base flipping, in which a nucleobase is unpaired, removed from the stack and further rotated out to assume a fully extrahelical position. Since its first demonstration in crystal structures of cytosine methyltransferase-DNA complexes numerous studies revealed that base flipping is a fundamental mechanism in DNA modification and repair, mediates sequence-specific recognition by restriction endonucleases, and is part of replication, transcription and recombination events. Here we discuss experimental and theoretical approaches used to study DNA base flipping in different systems.

Introduction

Normally, DNA exists as the B-form double-stranded helix in which partner bases on the two complementary strands make Watson-Crick pairs. The base pairs are stacked face-to-face to form the inner core of the double helix with the sugar-phosphate backbone wrapping around the outer edge of the structure. Import inherent features of the helical DNA are conformational plasticity and flexibility. Although the double helix is thermodynamically stable at physiological conditions, it undergoes dynamic conformational fluctuations including spontaneous transient disruptions of base pairing interactions (a phenomenon called DNA breathing). The helical structure is often perturbed

by interactions with proteins and other cellular components. The most common distortions of the DNA helix include bending/kinking and strand separation, which at the nucleotide level constitute base unstacking and base pair opening, respectively. The most dramatic and yet highly localized non-covalent distortion of the DNA helix is base flipping, in which a nucleobase is unpaired, removed from the stack and further rotated out to assume a fully extrahelical conformation. Such conformations are very unstable in free DNA but can be stabilized upon interaction with other biomolecules. The first example of DNA base flipping appeared with a high-resolution co-crystal structure of the HhaI methyltransferase reaction complex¹, and subsequently was found in most systems where an enzyme needs to gain access to a DNA base. Base flipping is now recognized as a fundamental mechanism in DNA modification and repair², and is also used for opening the DNA during replication, transcription and recombination^{3, 4}. Recently, sequence-specific DNA recognition by restriction endonucleases⁵ and the recognition of hemimethylated CpG sites by the UHRF1⁶⁻⁸ protein have been shown to employ expulsion of a complementary central base pair, or a single 5-methylcytosine residue, respectively, to achieve specific recognition of their targets. These examples suggest that many other enzymes or DNA-binding proteins may employ this mechanism in their interactions with DNA.

Here we attempt to discuss the variety of experimental approaches that were developed to study the occurrence and the mechanisms of base flipping in double helical nucleic acids.

X-Ray Crystallography

X-ray crystallography of protein-DNA complexes is the sole experimental method that can provide the ultimate proof of base flipping. Examples of crystallographically proven base-flipping systems include DNA methyltransferases, DNA glycosylases, apurinic/apyrimidinic endonucleases, glucosyltransferases, restriction endonucleases and some other systems (see Table 1). Crystallographic studies showed that DNA base flipping occurs in various modes (see Figure 1) such as sole flipping of the target base itself^{1, 9}, flipping of a base located on the opposite DNA strand to the target base¹⁰ or flipping of both nucleosides of a target base pair^{5, 11, 12}. Crystal structures reveal many structural details at atomic resolution however they yield static snapshots, usually at the end of a flipping pathway. Therefore, crystallization of mutant proteins or DNA substrates containing rotationally restricted nucleotide analogs to trap intermediate base-flipping conformers can give further insights into the flipping process and expands the applicability of X-ray crystallography^{13, 14}.

TABLE 1. Base-flipping systems proven by crystal structures of protein-DNA complexes.

| Specific protein | Catalytic reaction | Reference |
|---|---|-----------|
| DNA MTases | | |
| M.HhaI | Forms 5-methylC in DNA | 1 |
| M.HaeIII | Forms 5-methylC in DNA | 15 |
| M.TaqI | Forms N6-methylA in DNA | 78 |
| M.T4Dam | Forms N6-methylA in DNA | 79 |
| M.EcoDam | Forms N6-methylA in DNA | 11 |
| DNA glycosylases | | |
| T4-Pdg (formerly known as T4 endonuclease V) | Removes pyrimidine dimers from DNA | 80 |
| Human UDG | Removes uracil from DNA | 81 |
| <i>E.coli</i> MUG | Removes uracil or thymine from DNA containing G:T or G:U | 82 |
| Human AAG | Removes 3-methylA from DNA | 83 |
| <i>E.coli</i> AlkA | Removes 3-methylA from DNA | 84 |
| hOGG1 | Removes 8-oxoG from DNA | 85 |
| <i>B. stearothermophilus</i> EndoIII | Removes oxidized pyrimidine from DNA | 86 |
| <i>E. coli</i> MutY | Removes adenines from mismatch base pair | 87 |
| Apurinic/apyrimidinic endonucleases | | |
| <i>E.coli</i> endonuclease IV | Cleaves the DNA backbone 5' of apurinic/apyrimidinic sites | 12 |
| Human apurinic/apyrimidinic endonuclease (HAP1 or APE1) | Cleaves 5' to apurinic/apyrimidinic sites | 88 |
| Other DNA repair proteins | | |
| <i>S. cerevisiae</i> Rad4 | Binds to the lesion and recruits the multi-subunit transcription factor TFIIH | 10 |
| <i>E. coli</i> AlkB | Oxidizes N-alkylated base lesions to restore standard bases in single-stranded DNA and RNA | 17 |
| Human ABH2 | Oxidizes 1-methylA damage to restore A in double-stranded DNA | |
| <i>Anacystis nidulans</i> DNA photolyase | Repairs pyrimidine dimers by catalyzing the cleavage of the cyclobutane ring | 59 |
| Glucosyltransferases | | |
| T4 bacteriophage BGT | Transfers the glucose moiety of UDP-glucose to the 5-hydroxymethylcytosine bases making β -glucosidic bond | 9 |
| T4 bacteriophage AGT | Transfers the glucose moiety of UDP-glucose to the 5-hydroxymethylcytosine bases making α -glucosidic bond | 89 |
| Restriction endonucleases | | |
| R.HinP1I | Cleaves the DNA backbone at recognition site | 90 |
| R.Ecl18kI | Cleaves the DNA backbone at recognition site | 5 |
| R.PspGI | Cleaves the DNA backbone at a recognition site | 91 |
| Other | | |
| SRA domain of UHRF1 (also known as ICBP90, Np95) | Directs Dnmt1 methyltransferase to hemi-methylated CpG sites | 6-8 |

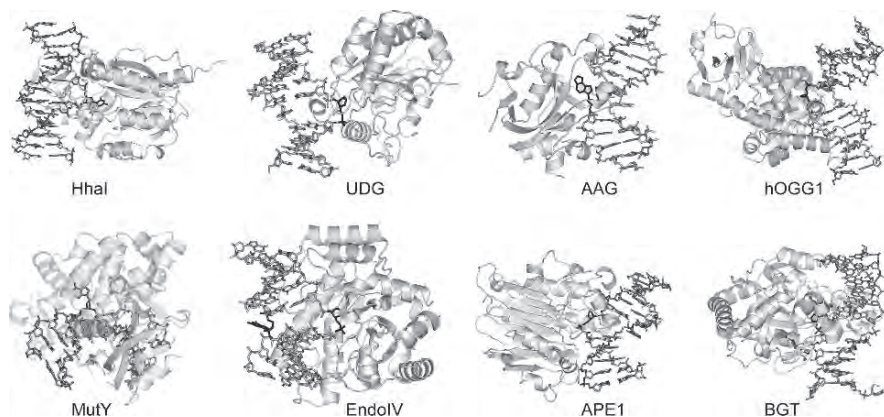


Fig. 1. Examples of base flipping from crystal structures of protein-DNA complexes (see Table 1 for further details).

However, a major limitation of the method is that co-crystallization of proteins with their DNA substrates is often tedious or even impossible. Covalent cross-linking with catalysis-based analogs^{1, 15-17} or alkyldisulfide tethers¹⁸ can be used to obtain protein-DNA complexes amenable to crystallization. In lack of co-crystals, topological considerations are often used to predict base flipping in cases when catalytic residues are located in a concave pocket of a protein and thus cannot interact with the target base in B-DNA without a substantial conformational rearrangement of the DNA.

NMR Spectroscopy And Computational Methods

NMR spectroscopy is a powerful technique that is well established to reveal various aspects of nucleic acids structure¹⁹. In contrast to X-ray crystallography, NMR can give direct insights into dynamic aspects of base flipping. On the other hand, analysis of large protein-DNA complexes is often a challenge due to slow molecular tumbling or insufficient solubility. Therefore, computational approaches are also used for base flipping analysis, mostly focusing on dynamic aspect of this process and studying the same problems as solution (NMR) methods.

The first attempt to study enzyme-induced base flipping by NMR in solution was performed for the M.HhaI DNA methyltransferase²⁰. ¹⁹F chemical shift analysis of the free DNA duplex and the M.HhaI-DNA complexes revealed the existence of multiple conformers of the target 5-fluorocytosine along the

base flipping pathway that were not seen in the previous crystal structures. Comparison of the T_1 , T_2 and $T_{1\rho}$ spin relaxation times of ^{19}F for the free DNA duplex and the enzyme-DNA binary complex showed no dramatic acceleration of the internal motional processes in the DNA duplex upon binding of M.HhaI. More recently, NMR analysis of interactions between cyclobutane pyrimidine dimer photolyase and its DNA substrates was performed employing ^{13}C or ^{15}N labeled DNA substrates²¹. Intimate contacts between the DNA lesion and a Trp residue in a cavity in the enzyme were inferred based on chemical shift differences of ^1H - ^{13}C HSQC resonances of the cyclobutane pyrimidine moiety upon binding of the deuterated protein. Base pairing in the rest of the DNA duplex was assessed from analysis of the imino region of a ^1H - ^{15}N HSQC spectrum, indicating a very localized but dramatic conformational change (i.e. base-flipping) at the damaged dinucleotide.

A series of NMR experiments have been devoted to study the dynamics of base pairing in free DNA duplexes in solution²². As mentioned above, double helical nucleic acids undergo spontaneous conformational fluctuations at physiological conditions including transient disruptions of base pairing. The imino protons of guanine and thymine are largely inaccessible to bulk solvent in the closed base pairs, but can be exchanged with those of water in an open state. The lifetimes of the closed and open state for individual base pairs can be derived from analysis of spin inversion recovery or spin saturation transfer from water. Base pair lifetimes have been found to be in the range of 1–5 ms for A:T and 10–50 ms for G:C, which can vary by a substantial margin depending on the sequence context²².

Analogous comparative experiments were performed on DNA-protein complexes and corresponding free DNA duplexes in order to assess the roles of enzymes in base flipping^{20, 23}. A lack of or a small increase of the breathing rate upon binding of enzyme is typically observed and interpreted as a passive mechanism by which an enzyme merely catches the spontaneously flipped out base^{14, 23, 24}. It should be noted that, computational estimates of the minimum rotation of a base that is required to allow proton exchange are in the range of 35°–40°. This is only 20–25% of the full 180° rotation observed in many biological systems²³; the estimated activation energy barrier of 9 kcal/mol²² accounts for roughly a half of the total 15–20 kcal/mol required for a complete expulsion of the nucleotide^{25, 26}. Since the majority of stacking interactions may still be preserved in such open intermediates, the nucleobases remain largely obscured within the DNA stack in such open base pairs. Therefore, it appears that such bases would be insufficiently exposed to be simply captured in a concave active site of an enzyme in a passive manner^{14, 23} and the conformational motions observed in such NMR experiments reflect very early events along the flipping pathway. A more plausible model for passive base flipping comes from the capture of extrahelical guanine bases by macrocyclic

glycans such as β -cyclodextrin. β -Cyclodextrin traps guanines that appear upon spontaneous flipping of G nucleotides in DNA leading to denaturation (low temperature melting) of DNA at a rate of 0.003s^{-1} (0.2min^{-1})²⁷. HhaI methyltransferase actively flips (pushes) the base out of the DNA helix²⁸, and, remarkably, the passive base-flipping rate exactly matches an estimated rate of target cytosine flipping ($k_{\text{flip}} \sim k_{\text{chem}} = 0.2\text{min}^{-1}$) in the M.HhaI Q237G mutant, which is deficient in promoting active base flipping²⁸. Clearly, the events of spontaneous flipping of nucleobases into extended extrahelical positions in DNA occur at rates several orders of magnitude lower than the NMR-detectable imino proton exchange. Therefore, the NMR derived exchange rates are of little relevance for assigning an active or passive role for an enzyme in base flipping. Most DNA modification and repair enzymes operate at turnover rates faster than 1min^{-1} and thus cannot depend to a significant extent on spontaneous DNA base flipping.

Theoretical methods based on empirical force fields offer an alternative to study base flipping dynamics. An “umbrella” sampling was used to calculate the individual potentials of mean force (PMF) for flipping of the Watson-Crick paired C and G bases in free DNA duplexes²⁵. The calculated data showed that the minor and major groove barriers to cytosine flipping in the absence of enzyme are similar (15 and 17 kcal/mol, respectively). Thus, the minor groove pathway is not more expensive energetically than the major groove pathway and cytosine flipping may occur *via* either groove²⁵. Free-energy profiles for base flipping suggested that the HhaI methyltransferase is able to lower the free-energy barrier to flipping by 17 kcal/mol and stabilizes the fully flipped state. The helix is destabilized by interactions with the enzyme facilitating flipping *via* the major groove of DNA. It was suggested that S87 plays an important role in the flipping process by forming initial H-bonds both with the target cytosine and the “orphan” guanine bases, thus distorting the helix²⁹. However certain details of the proposed model appear to be in conflict with experimental evidence²⁸, which can be resolved via further concerted efforts using computation and experimental approaches.

Biochemical Studies

Biochemical studies of DNA cytosine-5 methyltransferases revealed that binding affinity of DNA duplexes containing mismatched base pairs at the target site is increased as compared to cognate duplexes^{30, 31}. This unexpected feature was observed for DNA methyltransferases (reviewed in²) as well as DNA repair enzymes³². Improved binding is even more prominent if the target base or entire nucleotide is removed³³. An inverse correlation between

the target base pair stability and the DNA binding affinity indicates that the binding energy is partially used to destabilize the target base pair and advance the flipping (substrate strain). DNA straining by pushing with a protruding residue²⁸, backbone compression with a serine pinch^{34, 35} or bending of the DNA helix³⁶⁻³⁸ are examples of strategies used by enzymes to accelerate the base flipping process. These studies also showed that the target base is not recognized by MTases during initial flipping, and thus the target base can be substituted with fluorescent probes (see below) or structural nucleotide analogs³⁹⁻⁴¹.

Fluorescence Spectroscopy

DNA nucleobases are good chromophores, however, the absorbance is very weakly dependent on their conformational state. A higher sensitivity is generally achieved using fluorescence spectroscopy, but the DNA nucleobases are weak fluorophores, and their chemical analogs need to be employed to achieve the required sensitivity and enhanced selectivity.

The most widely used fluorescent probe for analysis of base flipping is 2-aminopurine (2AP). This structural analog of adenine has a high quantum yield in aqueous solution but is highly quenched when incorporated into DNA⁴². The major decay pathway of the excited state occurs *via* charge transfer to nearby guanine bases⁴³⁻⁴⁵. Since the fluorescence intensity of 2AP is highly dependent on base stacking in the DNA helix, removal of both stacking partners due to base flipping should be easily detectable by monitoring emission intensity. Indeed, substitution of the target base with 2AP usually leads to a strong increase (several ten-fold) in fluorescence upon binding of a base-flipping enzyme^{46, 47}. A smaller enhancement (several-fold) in fluorescence is observed if the flipped out base appears in close proximity with non-polar or aromatic residues in the flipped out state^{46, 48, 49}. A moderate increase cannot serve as a reliable signature of base flipping, since complete flipping of 2AP from the DNA into a quenched environment and partial unstacking of the 2AP base in the helix would give similar fluorescence intensity changes^{37, 50, 51}. On the other hand, kinetic stopped-flow experiments can provide useful information on particular aspects of the base flipping mechanisms as exemplified by DNA MTases^{28, 52} and uracil-glycosylases^{53, 54}.

More detailed information on the environment of the fluorophore can be obtained using time-resolved fluorescence spectroscopy. First time-resolved fluorescence measurements of MTase-DNA complexes in single crystals and in solution established that the fluorescence decay function of 2AP shows a pronounced, characteristic response to base flipping: the loss of the very

short (~100 ps) decay component and the large increase in the amplitude of the long (~10 ns) component^{49, 55}. This approach is though more technically demanding and cannot be routinely used in a standard laboratory.

A series of new fluorescent analogues of adenine and cytosine have recently been developed and are now available as components for synthetic DNA oligonucleotides. Although many of these compounds are capable of Watson-Crick pairing, they may potentially perturb interactions with enzymes in the major groove due to substantial increase in steric bulk. Two such analogs were used to study base flipping by DNA photolyase. 6MAP, a derivative of pteridone, has a high quantum yield (excitation and emission maxima are at 330 nm and 430 nm, respectively) and the fluorescence decreases when the base is incorporated into DNA⁵⁶. Similarly, the cytosine derivative pyrrolo-dC was employed to probe how far base flipping propagates along the duplex⁵⁷⁻⁵⁹.

Photochemical Approaches

Base flipping events can be directly sensed by monitoring nucleobase-mediated charge transfer along the DNA helix. Such long-range charge effects can be studied by analyzing oxidation of guanines at one end of the DNA duplex that is induced by a light-harvesting intercalating rhodium photo-oxidant tethered to the other end of the DNA. It was shown that yields of the long-range guanine damage correlate with protein-induced changes in base stacking in the center of the duplex. For example, M.HhaI effectively abolishes the charge transfer by inserting its protruding Gln237 residue to replace the target cytosine in the DNA⁶⁰. Alternatively, light-sensitive nucleobase analogs can be incorporated into DNA to probe protein contacts via photo-induced covalent cross-linking reactions. When incorporated at the target positions, 5-iodouracil and 2-thiouracil formed covalent cross-links with the EcoRV⁶¹ and HhaI⁶² DNA methyltransferases, respectively, upon exposure to UV light. Such experiments can identify the region of a protein proximal to the target base, but do not *per se* indicate base flipping. However, base flipping can be implied when the modified residues map to a distal location in the cross-linked protein⁶¹.

Chemical Probing

Detection of extrahelical conformations in DNA may rely on differential chemical reactivity of individual residues (nucleobase, sugar or phosphodiester moiety). Hyper-reactivity of a nucleotide is typically a sign of its enhanced accessibility, which may come from (partial) unpairing or unstacking events. Protein-induced base flipping is thus expected to show hyper-reactivity at or

around the extrahelical residue itself. Many chemicals have been developed to probe such structural variations in DNA⁶³⁻⁶⁵. The most widely used approach for subsequent visualization of the chemically modified bases is piperidine treatment. In cases when the chemical modification blocks base pairing, strand cleavage with nuclease S1⁶⁶ or polymerase extension can be used⁶⁷. Three examples of chemical probing are discussed below.

Potassium permanganate leads to oxidation of thymine to *cis*-thymine glycol (5,6-dihydroxy-5,6-dihydrothymine)⁶⁸, which undergoes further degradation leading to cleavage of the DNA strand upon piperidine treatment. The permanganate attack on the C6-C5 edge takes place from a face of the pyrimidine ring, which is thus strongly dependent upon solvent accessibility of the thymine residue. The suitability of potassium permanganate for "positive display" of flipped-out thymines in DNA was demonstrated for cytosine and adenine DNA methyltransferases^{50, 69} and for a sequence-specific transposase⁷⁰.

Haloacetaldehydes are known to primarily react with unpaired adenine and cytosine bases in DNA yielding 1,*N*6-ethenoadenine and 3,*N*4-ethenocytosine derivatives, respectively⁷¹. Such haloacetaldehyde-modified residues can be detected by piperidine-induced strand cleavage⁷². Recently, the chloroacetaldehyde reaction has been successfully used for mapping flipped-out cytosine bases in a series of bacterial DNA MTases, and two restriction endonucleases⁷³.

Hydroxyl radicals (generated in the Fenton reaction) attack DNA by abstracting H atoms in the deoxyribose which eventually leads to strand breakage at the modified residues. The Fenton reaction shows no base or sequence selectivity and is thus a proven probe for footprinting of protein-DNA complexes⁷⁴. However, footprints of type II bacterial cytosine-5 MTases showed no regions of enhanced deoxyribose accessibility^{75, 76}. In contrast, enhanced susceptibility to hydroxyl radicals at and around the target nucleotides was observed with the EcoR124I DNA methyltransferase⁷⁷ suggesting that a different base flipping strategy is employed by type I restriction-modification enzymes.

Chemical probing is a simple assay that does not require specialized equipment. On the other hand, since flipping of DNA bases is observed upon interaction with proteins, only those chemical reactions that proceed at near physiological conditions can be employed.

Conclusion

X-ray crystallography of protein-DNA complexes tops experimental methods in providing the ultimate proof of base flipping. NMR spectroscopy can give further insight into dynamic aspects of base flipping, but its potential has not yet been fully exploited. An attractive "low resolution" method to study

base flipping in solution is fluorescent spectroscopy. However, it requires modified bases such as 2-aminopurine to be introduced in DNA, and does not always yield unequivocal results. Chemical probing as well as biochemical/kinetic studies involving protein mutants and DNA substrates containing nucleotide analogs can provide valuable insights into protein-DNA interactions leading to base flipping.

In spite of extensive studies, many mechanistic aspects of base flipping still remain obscure. In light of new base flipping enzymes and proteins that emerged recently, further efforts combining various experimental and computational approaches are required to unravel this exciting enigma of macromolecular interactions.

Acknowledgements The authors thank all current and former members of the Laboratory of Biological DNA Modification and our numerous collaborators for their important contributions to the studies of DNA base flipping by the M.HhaI methyltransferase. Work in the laboratory was supported by grants from Volkswagen Stiftung, the Howard Hughes Medical Institute, the Royal Society, NATO Science Program and the Lithuanian State Science and Study Foundation.

References

1. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell*. 1994;76:357-369.
2. Cheng X, Roberts RJ. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res*. 2001;29(18):3784-3795.
3. Duval-Valentin G, Ehrlich R. Dynamic and structural characterization of multiple steps during complex formation between E. coli RNR polymerase and the tetR promoter from pSC101. *Nucleic Acids Res*. 1987;15(2):575-595.
4. Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science*. 2005;309:2219-2222.
5. Bochtler M, Szczepanowski RH, Tamulaitis G, et al. Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J*. 2006;25(10):2219-2229.
6. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*. 2008.
7. Avvakumov GV, Walker JR, Xue S, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*. 2008.
8. Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*. 2008.
9. Lariviere L, Morera S. A base-flipping mechanism for the T4 phage β -glucosyltransferase and identification of a transition state analog. *J. Mol. Biol*. 2002;324:483-490.
10. Min JH, Pavletich NP. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*. 2007;449(7162):570-575.
11. Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X. Structure and Substrate Recognition of the Escherichia coli DNA Adenine Methyltransferase. *J Mol Biol*. 2006;358:1-12.
12. Hosfield DJ, Guan Y, Haas BJ, Cunningham RP, Tainer JA. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*. 1999;98:397-408.

13. Horton JR, Ratner G, Banavali NK, et al. Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase. *Nucleic Acids Res.* 2004;32:3877-3886.
14. Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature.* 2007;449:433-438.
15. Reinisch KM, Chen L, Verdine GL, Lipscomb WN. The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell.* 1995;82:143-153.
16. Gilboa R, Zharkov DO, Golan G, et al. Structure of Formamidopyrimidine-DNA Glycosylase Covalently Complexed to DNA. *J Biol Chem.* 2002;277(22):19811-19816.
17. Yang CG, Yi C, Duguid EM, et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature.* 2008;452(7190):961-965.
18. Banerjee A, Santos WL, Verdine GL. Structure of a DNA glycosylase searching for lesions. *Science.* 2006;311(5764):1153-1157.
19. Palmer AG, Kroenke C, Loria P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 2001;339:204-239.
20. Klimasauskas S, Szyperski T, Serva S, Wuethrich K. Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution. *EMBO J.* 1998;17(1):317-324.
21. Torizawa T, Ueda T, Kuramitsu S, et al. Investigation of the cyclobutane pyrimidine dimer (CPD) photolyase DNA recognition mechanism by NMR analyses. *J Biol Chem.* 2004;279(31):32950-32956.
22. Dornberger U, Leijon M, Fritzsche H. High Base Pair Opening Rates in Tracts of GC Base Pairs. *J Biol Chem.* 1999;274:6957-6962.
23. Cao C, Jiang YL, Stivers JT, Song F. Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat Struct Mol Biol.* 2004;11(12):1230-1236.
24. Roberts RJ, Cheng X. Base flipping. *Annu Rev Biochem.* 1998;67:181-198.
25. Banavali NK, MacKerell AD. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J Mol Biol.* 2002;319:141-160.
26. Giudice E, Varnai P, Lavery R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* 2003;31(5):1434-1443.
27. Spies MA, Schowen RL. The trapping of a spontaneously "flipped-out" base from double helical nucleic acids by host-guest complexation with β -cyclodextrin: the intrinsic base-flipping rate constant for DNA and RNA. *J Am Chem Soc.* 2002;124:14049-14053.
28. Daujotyte D, Serva S, Vilkaitis G, Merkiene E, Venclovas C, Klimasauskas S. HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. *Structure.* 2004;12:1047-1055.
29. Huang N, Banavali NK, MacKerell AD, Jr. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Nat. Acad. Sci. USA.* Jan 7 2003;100(1):68-73.
30. Klimasauskas S, Roberts RJ. M.HhaI binds tightly to substrates containing mismatches at the target base. *Nucleic Acids Res.* 1995;23:1388-1395.
31. Yang AS, Shen JC, Zingg JM, Mi S, Jones PA. HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res.* 1995;23:1380-1387.
32. Krosky DJ, Song F, Stivers JT. The origins of high-affinity enzyme binding to an extrahelical DNA base. *Biochemistry.* Apr 26 2005;44(16):5949-5959.
33. Klimasauskas S, Roberts RJ. Disruption of the target G-C base-pair by the HhaI methyltransferase. *Gene.* 1995;157:163-164.
34. Parikh SS, Mol CD, Slupphaugh G, Bharati S, Krokan HE, Tainer JA. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* 1998;17(17):5214-5226.
35. Stivers JT, Jiang YL. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem Rev.* 2003;103:2729-2759.

36. Hopkins BB, Reich N. Simultaneous DNA binding, bending, and base flipping: evidence for a novel M.EcoRI methyltransferase:DNA complex. *J. Biol. Chem.* 2004;279:37049-37060.
37. Gowher H, Jeltsch A. Molecular Enzymology of the EcoRV DNA-(Adenine-N6)-Methyltransferase: Kinetics of DNA Binding and Bending, Kinetic Mechanism and Linear Diffusion of the Enzyme on DNA. *J Mol Biol.* 2000;303:93-110.
38. Hosfield DJ, Guan Y, Haas BJ et al. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at a basic sites and three-metal-ion catalysis. *Cell* 1999;98:397-408.
39. Jiang YL, Stivers JT. Base-Flipping Mutations of Uracil DNA Glycosylase: Substrate Rescue Using a Pyrene Nucleotide Wedge. *Biochemistry.* 2002;41:11248-11254.
40. Beuck C, Singh I, Bhattacharya A, et al. Polycyclic aromatic DNA-base surrogates: high-affinity binding to an adenine-specific base-flipping DNA methyltransferase. *Angew Chem Int Ed Engl.* 2003;42(33):3958-3960.
41. Wang P, Nicklaus MC, Marquez VE, et al. Use of oligodeoxyribonucleotides with conformationally constrained abasic sugar targets to probe the mechanism of base flipping by HhaI DNA (cytosine C5)-methyltransferase. *J Am Chem Soc.* 2000;122(50):12422-12434.
42. Ward DC, Reich E, Stryer L. Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J Biol Chem.* 1969;244(5):1228-1237.
43. Kelley SO, Barton JK. Electron transfer between bases in double helical DNA. *Science.* 1999;283(5400):375-381.
44. Jean JM, Hall KB. 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc Natl Acad Sci U S A.* 2001;98(1):37-41.
45. Rachofsky EL, Osman R, Ross JBA. Probing structure and dynamics of DNA with 2-Aminopurine: effects of local environment on fluorescence. *Biochemistry.* 2001;40:946-956.
46. Allan BW, Reich N. Targeted base stacking disruption by the EcoRI DNA methyltransferase. *Biochemistry.* 1996;35(47):14757-14762.
47. Holz B, Klimasauskas S, Serva S, Weinhold E. 2-Amino purine as a fluorescence probe for DNA base flipping by methyltransferases. *Nucleic Acids Res.* 1998;26(4):1076-1083.
48. Tamulaitis G, Zaremba M, Szczepanowski RH, Bochtler M, Siksnys V. Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res.* 2007;35(14):4792-4799.
49. Lenz T, Bonnist EY, Pljevaljcic G, et al. 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.TaqI: crystal structures and time-resolved fluorescence. *J Am Chem Soc.* 2007;129(19):6240-6248.
50. Reddy YV, Rao DN. Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. *J Mol Biol.* 2000;298(4):597-610.
51. Christine KS, MacFarlane AWt, Yang K, Stanley RJ. Cyclobutylpyrimidine dimer base flipping by DNA photolyase. *J Biol Chem.* 2002;277(41):38339-38344.
52. Vilkaitis G, Dong A, Weinhold E, Cheng X, Klimasauskas S. Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. *J Biol Chem.* 2000;275(49):38722-38730.
53. Stivers JT, Pankiewicz KW, Watanabe KA. Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry.* 1999;38(3):952-963.
54. Wong I, Lundquist AJ, Bernards AS, Mosbaugh DW. Presteady-state analysis of a single catalytic turnover by Escherichia coli uracil-DNR glycosylase reveals a "pinch-pull-push" mechanism. *J Biol Chem.* 2002;277(22):19424-19432.
55. Neely RK, Daujotyte D, Grazulis S, et al. Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Res.* 2005; 33(22):6953-6960.
56. Hawkins ME, Pfeleiderer W, Jungmann O, Balis FM. Synthesis and fluorescence characterization of pteridine adenosine nucleoside analogs for DNA incorporation. *Anal Biochem.* 2001;298(2):231-240.

57. Yang K, Matsika S, Stanley RJ. 6MAP, a fluorescent adenine analogue, is a probe of base flipping by DNA photolyase. *J Phys Chem B*. 2007;111(35):10615-10625.
58. Yang K, Stanley RJ. The extent of DNA deformation in DNA photolyase- substrate complexes: a solution state fluorescence study. *Photochem Photobiol*. 2008;84(3):741-749.
59. Mees A, Klar T, Gnau P, et al. Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science*. 2004;306(5702):1789-1793.
60. Rajski SR, Barton JK. How different DNA-binding proteins affect long-range oxidative damage to DNA. *Biochemistry*. 2001;40:5556-5564.
61. Jeltsch A, Roth M, Friedrich T. Mutational analysis of target base flipping by the EcoRV adenine-N6 DNA methyltransferase. *J Mol Biol*. 1999;285(3):1121-1130.
62. Daujotyte D, Klimasauskas S. Affinity photo-crosslinking study of the DNA base flipping pathway by HhaI methyltransferase. *Nucleic Acids Symp Ser*. 2000(44):271-272.
63. Nielsen PE. Chemical and photochemical probing of DNA complexes. *J. Mol Recognit*. 1990;3(1):1-25.
64. Lilley DMJ. Probes for DNA structure. *Methods Enzymol*. 1992;212:133-139.
65. Rokita SE. Chemical and enzymatic probes for nucleic acids structure. In: Beaucage SL, Bergstrom DE, Glick GD, Jones RA, eds. *Current Protocols in Nucleic Acids Chemistry*. Vol 6.6: John Wiley & sons; 2001:1-16.
66. McLean MJ, Larson JE, Wohlrab F, Wells RD. Reaction conditions affect the specificity of bromoacetaldehyde as a probe for DNA cruciforms and B-Z junctions. *Nucleic Acids Res*. 1987;15(17):6917-6935.
67. Guerin M, Leng M, Rahmouni AR. High resolution mapping of E.coli transcription elongation complex in situ reveals protein interactions with the non-transcribed strand. *EMBO J*. 1996;15(19):5397-5407.
68. Sasse-Dwight S, Gralla JD. KMnO₄ as a probe for lac promoter DNA melting and mechanism in vivo. *J Biol Chem*. 1989;264(14):8074-8081.
69. Serva S, Weinhold E, Roberts RJ, Klimasauskas S. Chemical display of thymine residues flipped out by DNA methyltransferases. *Nucleic Acids Res*. 1998;26(15):3473-3479.
70. Bischerour J, Chalmers R. Base-flipping dynamics in a DNA hairpin processing reaction. *Nucleic Acids Res*. 2007;35(8):2584-2595.
71. Kusmirek JT, Singer B. Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. *Biochemistry*. 1982;21(22):5717-5722.
72. Kohwi-Shigematsu T, Kohwi Y. Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *Methods Enzymol*. 1992;212:155-180.
73. Daujotyte D, Liutkeviciute Z, Tamulaitis G, Klimasauskas S. Chemical mapping of cytosines enzymatically flipped out of the DNA helix. *Nucleic Acids Res*. 2008.
74. Price MA, Tullius TD. Using hydroxyl radical to probe DNA structure. *Methods Enzymol*. 1992;212:194-219.
75. Renbaum P, Razin A. Footprint Analysis of M.SssI and M.HhaI Methyltransferases Reveals Extensive Interactions with the Substrate DNA Backbone. *J Mol Biol*. 1995;248:19-26.
76. Finta C, Kiss A. Footprint analysis of the BspRI DNA methyltransferase-DNA interaction. *Nucleic Acids Res*. 1997;25(14):2841-2846.
77. Mernagh DR, Kneale GG. High resolution footprinting of a type I methyltransferase reveals a large structural distortion within the DNA recognition site. *Nucleic Acids Res*. 1996;24(24):4853-4858.
78. Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nature Struct. Biol*. 2001;8(2):121-125.
79. Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. *Cell*. 2005;121:349-361.

80. Vassilyev DG, Kashiwagi T, Mikami Y, et al. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell*. 1995;83:773-782.
81. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*. 1996;384:87-92.
82. Barrett TE, Savva R, Panayotou G, et al. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: Mismatch recognition by complementary-strand interactions. *Cell*. 1998;92:117-129.
83. Lau AY, Schaerer OD, Samson L, Verdine GL, Ellenberger T. Crystal structure of a human alkylbase-DNA repair enzymes complexed to DNA: Mechanisms for nucleotide flipping and base excision. *Cell*. 1998;95:249-258.
84. Hollis T, Ichikawa Y, Ellenberger T. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J*. 2000;19(4):758-766.
85. Bruner SD, Norman DP, Verdine GL. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*. 2000;403:859-866.
86. Fromme JC, Verdine GL. Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J*. 2003;22:3461-3471.
87. Fromme JC, Banerjee A, Huang N, Verdine GL. Structural Basis for Removal of Adenine Mispaiored with 8-Oxoguanine by Muty Adenine DNA Glycosylase. *Nature*. 2004;427:652-656.
88. Mol CD, Izumi T, Mitra S, Tainer JA. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*. 2000;403:451-456.
89. Lariviere L, Sommer N, Morera S. Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. *J Mol Biol*. 2005;352:139-150.
90. Horton JR, Zhang X, Maunus R, et al. DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res*. 2006;34:939-948.
91. Szczepanowski RH, Carpenter MA, H. C, et al. Central base pair flipping and discrimination by PspGI. *Nucleic Acids Research*. 2008.

THE DIVERSITY OF NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

COREY W. LIU^{1,*}, VIKTOR Y. ALEKSEYEV², JEFFREY R. ALLWARDT³, ALEXANDER J. BANKOVICH⁴, BARBARA J. CADE-MENUN⁵, RONALD W. DAVIS⁶, LIN-SHU DU⁷, K. CHRISTOPHER GARCIA⁸, DANIEL HERSCHLAG⁹, CHAITAN KHOSLA¹⁰, DANIEL A. KRAUT¹¹, QING LI¹², BRIAN NULL¹³, JOSEPH D. PUGLISI¹⁴, PAUL A. SIGALA⁹, JONATHAN F. STEBBINS¹⁵, AND LUCA VARANI¹⁶

¹*Stanford Magnetic Resonance Laboratory, Stanford University, Stanford, CA, USA*

²*Department of Chemistry, Stanford University, Stanford, CA, USA and Currently: Genencor, Palo Alto, CA, USA*

³*Department of Geological & Environmental Sciences, Stanford University, Stanford, CA, USA and Currently: ConocoPhillips Company, Houston, Texas, USA*

⁴*Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA; Department of Structural Biology, Stanford University, Stanford, CA, USA and Currently: Department of Microbiology & Immunology, University of California, San Francisco, CA, USA*

⁵*Department of Geological & Environmental Sciences, Stanford University, Stanford, CA, USA; Currently: Agriculture and Agri-Food Canada, Swift Current, Canada*

⁶*Stanford Genome Technology Center, Stanford University, Stanford, CA, USA; Department of Biochemistry, Stanford University, Stanford, CA, USA and Bio-X Program, Stanford University, Stanford, CA, USA*

⁷*Department of Geological & Environmental Sciences, Stanford University, Stanford, CA, USA and Currently: Air Products and Chemicals, Inc. Allentown, Pennsylvania, USA*

⁸*Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA; Department of Structural Biology, Stanford University, Stanford, CA, USA and Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA*

⁹*Department of Biochemistry, Stanford University, Stanford, CA, USA*

¹⁰*Department of Chemistry, Stanford University, Stanford, CA, USA; Department of Biochemistry, Stanford University,*

Stanford, CA, USA and Department of Chemical Engineering, Stanford University, Stanford, CA, USA

¹¹*Department of Biochemistry, Stanford University, Stanford, CA, USA and Currently: Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois, USA*

¹²*Department of Chemistry, Stanford University, Stanford, CA, USA and Institute for Research in Biomedicine, Bellinzona, Switzerland*

¹³*Stanford Genome Technology Center, Stanford University, Stanford, CA, USA; Department of Biochemistry, Stanford University, Stanford, CA, USA and Bio-X Program, Stanford University, Stanford, CA, USA*

¹⁴*Stanford Magnetic Resonance Laboratory, Stanford University, Stanford, CA, USA and Department of Structural Biology, Stanford University, Stanford, CA, USA*

¹⁵*Department of Geological & Environmental Sciences, Stanford University, Stanford, CA, USA*

¹⁶*Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA; Department of Structural Biology, Stanford University, Stanford, CA, USA and Institute for Research in Biomedicine, Bellinzona, Switzerland*

**Corresponding author: e-mail: liuc@stanford.edu*

Introduction

The discovery of the physical phenomenon of Nuclear Magnetic Resonance (NMR) in 1946 gave rise to the spectroscopic technique that has become a remarkably versatile research tool. One could oversimplify NMR spectroscopy by categorizing it into the two broad applications of structure elucidation of molecules (associated with chemistry and biology) and imaging (associated with medicine). But, this certainly does not do NMR spectroscopy justice in demonstrating its general acceptance and utilization across the sciences. This manuscript is not an effort to present an exhaustive, or even partial review of NMR spectroscopy applications, but rather to provide a glimpse at the wide-ranging uses of NMR spectroscopy found within the confines of a single magnetic resonance research facility, the Stanford Magnetic Resonance Laboratory. Included here are summaries of projects involving protein structure determination, mapping of intermolecular interactions, exploring fundamental biological mechanisms, following compound cycling in the environmental, analysis of synthetic solid compounds, and microimaging of a model organism.

Protein Structure Determination

One of the most commonly recognized uses of NMR spectroscopy is that of structure determination of small molecules in the world of chemistry and biomolecules (protein, RNA, DNA) in the world of biology, and certainly this remains a preeminent application of the technique.

ACYL CARRIER PROTEINS OF POLYKETIDE SYNTHASES [1, 2]

Polyketide synthases (PKSs) catalyze the biosynthesis of a large and diverse group of pharmacologically important natural products. Two major types of polyketide synthase systems exist in nature. Type I PKSs are gigantic multifunctional proteins with active sites organized into modules. Type II PKSs consist of several monofunctional proteins that are iteratively used in the biosynthetic pathway. The acyl carrier protein (ACP) is a key component to both pathways. During polyketide biosynthesis, ACPs perform the central role of transferring polyketide intermediates between active sites of the polyketide synthase. ACPs tether the growing acyl chains by means of a 4'-phosphopantetheine prosthetic group (a unique feature of all ACPs) as they migrate between the active sites of PKSs during the course of chain elongation and modification. The prosthetic group is covalently attached to a highly conserved serine residue via a phosphodiester linkage and provides a terminal sulfhydryl group for the attachment of acyl groups through a thioester bond. The 4'-phosphopantetheine moiety is often described as a long and flexible arm that can reach into the active sites of polyketide synthase enzymes during chain elongation steps, although direct biophysical evidence for this hypothesis is lacking.

TYPE I PKS

In modular polyketide synthases (schematic of 6-deoxyerythronolide B synthase (DEBS) shown in Fig. 1), specific intermodular transfer of polyketide intermediates between two different polypeptides (for example, between the ACP domain of module 2 [ACP2] and the KS domain of module 3 [KS3] in DEBS) is mediated by intermodular docking domains or linker sequences located at the N and C termini of the acceptor and donor modules of adjacent PKS polypeptides. The ACP domain from module 2 of DEBS was chosen as a target for structural characterization by NMR spectroscopy because it is involved in a number of well-characterized domain-domain interactions and because its size (~10 kDa) is amenable to structure determination by NMR (Fig. 2). The solution structure of ACP2 from DEBS serves as a prototype for this class of domains from modular polyketide synthase assemblies.

To demonstrate this potential, homology models were constructed for the other ACP domains in DEBS utilizing the ACP2 structure. A comparison

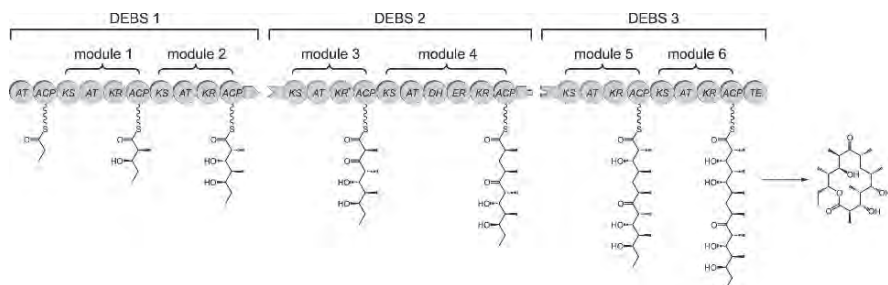


Fig. 1. Schematic diagram of the modular organization of 6-deoxyerythronolide B synthase (DEBS). Three polypeptides (DEBS1, DEBS2, and DEBS3) each contain two modules that are, in turn, composed of distinct catalytic domains. KS indicates ketosynthase; AT, acyl transferase; KR, ketoreductase; ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase; and TE, thioesterase. DEBS1 also contains a loading didomain, consisting of an AT and an ACP. KR⁰ denotes an inactive KR domain. Inter-protein linker (docking) domains are shown as matching tabs.

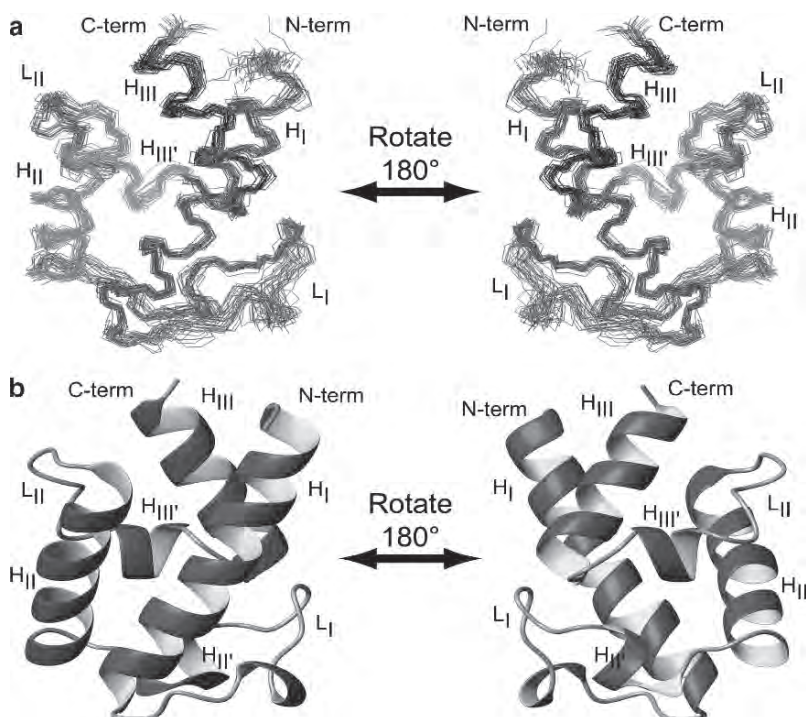


Fig. 2. NMR solution structure of DEBS ACP2. (a) Superposition of residues 14–95 for the 30 conformers from torsion angle dynamics calculation. (b) Ribbon diagram of the minimized mean structure of ACP2. Secondary structure elements are labeled as follows: helices – H_I, H_{II}, H_{III}, H_{III}', H_{III}'', H_{III}'''; loops – L_I, L_{II}.

of steric and electrostatic surface properties of these domains revealed plausible mechanistic explanations for several earlier observations with regard to ACP domain specificity in the context of polyketide chain transfer and chain elongation. In addition, three single-site mutants of the ACP2 domain were constructed and evaluated for their ability to support polyketide chain elongation in reactions with KS-AT didomain fragments from module 3 and from module 6 of DEBS. Two of the three mutated residues appeared to have an effect on the specificity of ACP interactions.

TYPE II PKS

In the type II polyketide synthase systems acyl carrier proteins are small acidic proteins of ~9kDa. The NMR solution structure was determined of the holo-form (with 4'-phosphopantetheine prosthetic group) of the ACP from the frenolycin producing PKS (Fig. 3).

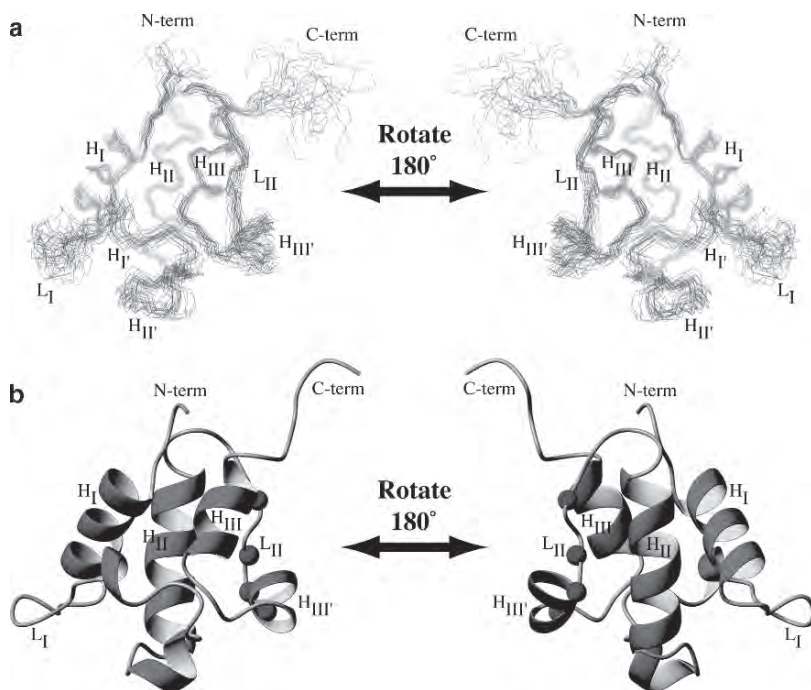


Fig. 3. NMR structure of frenolycin holo-ACP. **(a)** Ensemble of 24 torsion angle dynamics conformers, using the backbone atoms of residues 5–16, 39–52, and 69–77 for superposition. **(b)** Ribbon diagram of the minimized mean structure of frenolycin holo-ACP. Secondary structure elements are labeled as described in the text [helices (H_I, H_{II}, H_{III}, H_I', H_{II}', H_{III}') and loop (L_I)]. Residues identified as undergoing slow conformational exchange (alanine 56, threonine 58, aspartate 59, and glycine 63) are shown as spheres mapped onto the ribbon diagrams in **(b)**.

^{15}N backbone dynamics determined for holo-ACP found S^2 order parameters to be consistent with the tertiary fold. Folded, secondary structure containing regions of the protein displayed high order parameter values (approaching 1) and disordered loop and termini regions displaying lower order parameter values (Fig. 4). Interestingly, heteronuclear exchange experiments performed on holo-ACP found residues in the loop II region between helix II and helix III' to be exchanging between two conformations (Fig. 5). Analysis of this finding in light of other ACP results suggests a functional pur-

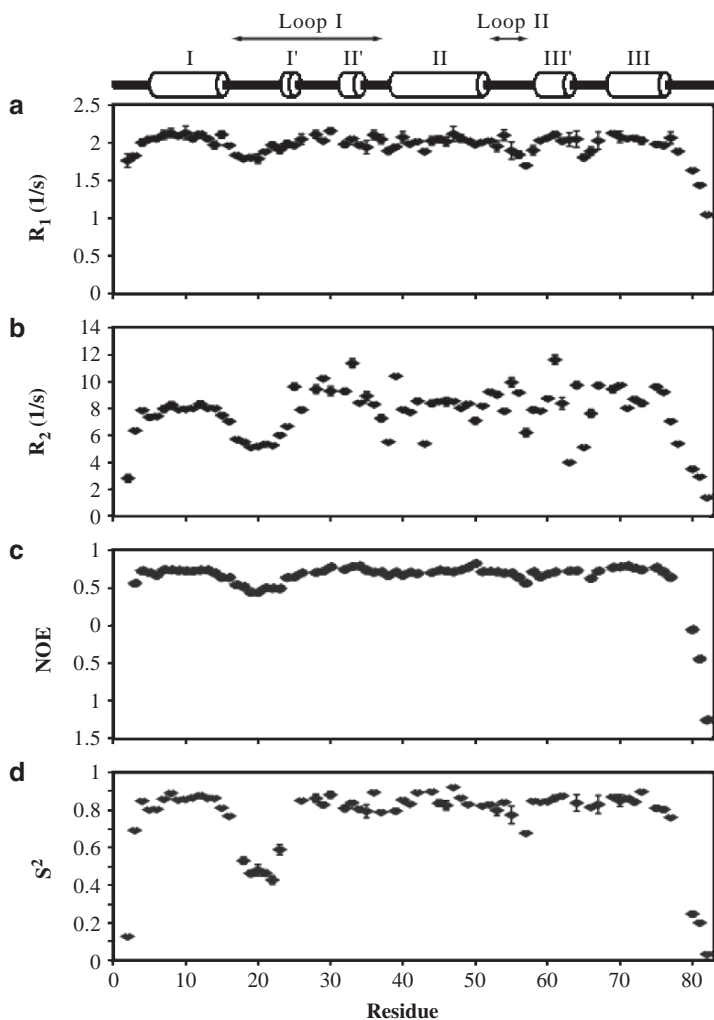
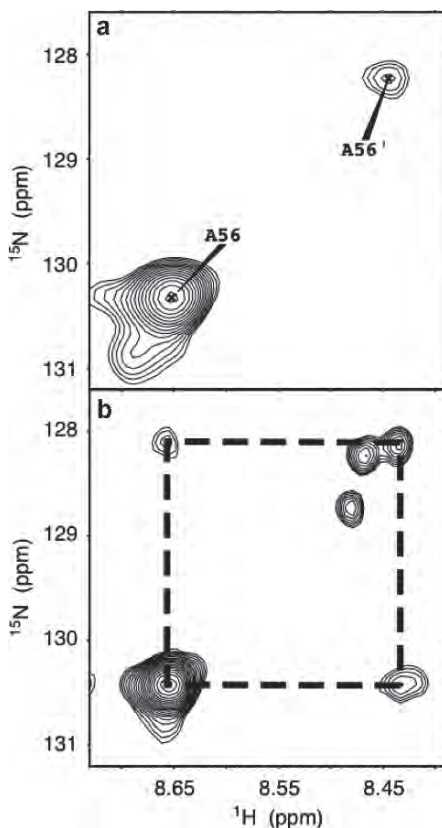


Fig. 4. ^{15}N relaxation parameters for backbone dynamics analysis of fren holo-ACP. (a) R_1 ($1/T_1$), (b) R_2 ($1/T_2$), (c) NOE, and (d) S^2 . Secondary structure elements are indicated above the plots. Uncertainties in R_1 , R_2 , and S^2 were standard deviations resulting from fits of the data to relaxation models. NOE uncertainties were conservatively estimated to be (10%).

Fig. 5. Conformational exchange of loop II and helix III'. **(a)** Region of the ^1H - ^{15}N HSQC spectrum showing the two amide resonances of alanine 56: the major form, A56, and the minor form, A56'. **(b)** Same region from a two-dimensional ^{15}N - ^1H exchange spectrum (100 ms mixing time). The exchange peaks between the major and minor forms of alanine 56 are connected by the dashed box. The remaining two peaks are spurious resonances from sample degradation.



pose for this conformational exchange with respect to the phosphopantetheine arm and insertion of acyl group substrates during polyketide formation.

Mapping of Intermolecular Interactions

A structure alone does not necessarily expose functional aspects of the molecules in question. More often than not other studies are necessary to give such insight; as demonstrated by the dynamics and exchange experiments presented above. In cases where protein-ligand or protein-protein interactions occur, information such as NMR chemical shift mapping, coupled with structures (whether NMR or X-ray derived), can be quite informative in discerning regions of interest.

T CELL RECEPTOR DOCKING FOOTPRINTS ON PEPTIDE-MAJOR HISTOCOMPATIBILITY COMPLEX [3]

T cell receptor (TCR) recognition of peptide-MHC (pMHC) is central to the cellular immune response. A large database of TCR-pMHC structures is needed

to reveal general structural principles, such as whether the repertoire of TCR/MHC docking modes is dictated by a “recognition code” between conserved elements of the TCR and MHC genes. Although ~17 co-crystal structures of unique TCR–pMHC complexes have been determined, co-crystallization of soluble TCR and pMHC remains a major technical obstacle in the field.

Given a large set of TCR that binds to a single MHC, would an unrestricted array of docking orientations be found or would there be subsets depending on the V α and V β chain usage by the TCR? For this question, high-resolution structures are not necessary; instead, a method that could paint the pMHC surface with the TCR footprint would suffice. In this way, footprint mapping could serve as a filtering step to characterize a collection of TCRs specific for a given pMHC.

Solution NMR spectroscopy is well suited for experimental characterization of the TCR–pMHC footprint in a relatively short amount of time. After assignment of the protein backbone NMR resonances, structural information can be obtained through chemical shift perturbation mapping. This approach is most effective when the structures of one or both of the components are known, which increases the information contained within resonance assignments. The protein that has a known structure is isotopically labeled, ligand is added, and the resonances that correspond to residues involved in intermolecular contacts are identified by changes in chemical shift. These perturbations are then mapped on the known structure to define a binding site. In this fashion, many ligands can be mapped on a single assigned and isotopically labeled protein.

This work mapped the 2C TCR binding interaction with its allogeneic ligand H–2L^d–QL9 (Fig. 6) and identified a group of NMR-shifted residues that delineated a clear surface (Fig. 7) of the MHC that was defined as the TCR footprint. The docking footprint was found to be highly accurate compared with a crystal structure of the same complex. And a subsequent NMR mapping analysis performed on a high-affinity mutant of the TCR was highly consistent and correlated to the increased affinity.

The results established a foundation to explore the molecular dynamics of pMHC complexes and to rapidly determine the footprints of many L^d-specific TCRs.

Exploring Fundamental Biological Mechanisms

Now having looked at cases of macromolecular structure and interactions capable of being addressed by NMR spectroscopy, it is also important to realize that NMR can be used to probe minute details of fundamental processes such as enzyme catalysis.

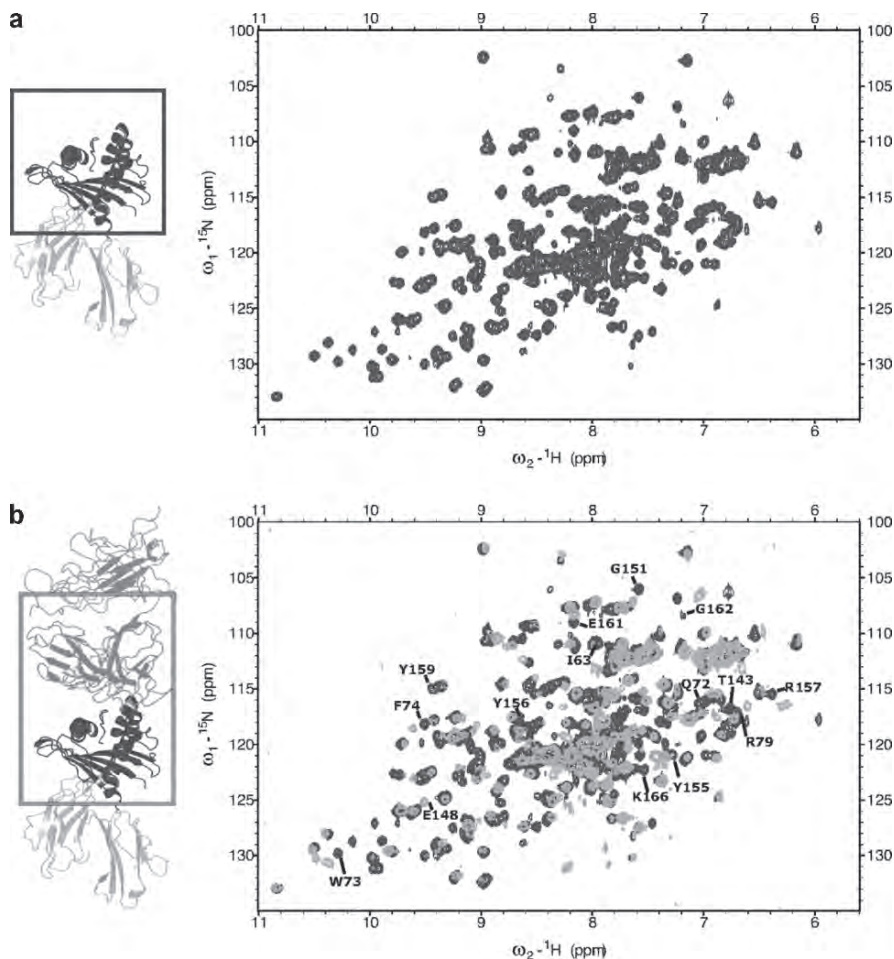


Fig. 6. 800 MHz ^1H - ^{15}N HSQC spectra collected on the platform MHC. (a) Free MHC. (b) Superposition of the free MHC (dark) and its complex with 2C TCR (light). Only the region of the MHC highlighted in the box is present in our NMR sample; the TCR and peptide are not isotopically labeled and not visible in the spectra. Some of the peaks that shift upon complex formation are noted.

HYDROGEN BONDING IN THE KETOSTEROID ISOMERASE OXYANION HOLE [4]

A longstanding proposal in enzymology is that enzymes are electrostatically and geometrically complementary to the transition states of the reactions they catalyze and that this complementarity contributes to catalysis. Experimental evaluation of this contribution, however, has been difficult. This study systematically dissected the potential contribution to catalysis from electrostatic complementarity in ketosteroid isomerase (KSI).

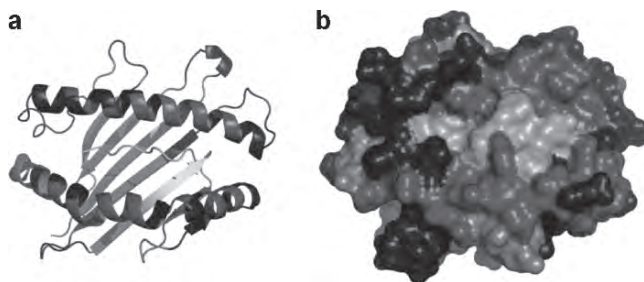


Fig. 7. Residues that shift upon complex formation mapped on crystal structure. Residues with ambiguous or missing assignments are a light shade (primarily the β -sheet scaffold to the rear), residues which shift are shown in a medium shade, and residues which do not shift are shown in a dark shade; the unlabelled peptide running between the two helices, for which we have no information, is also a light shade. (a) Cartoon representation of the H-2Ld-peptide structure with NMR shifts colored. (b) Surface representation of H-2Ld-peptide with NMR shifts colored.

Phenolates, analogs of the transition state and reaction intermediate, bind and accept two hydrogen bonds in an active site oxyanion hole (Fig. 8). The binding of substituted phenolates of constant molecular shape but increasing pK_a models the charge accumulation in the oxyanion hole during the enzymatic reaction. As charge localization increases, the NMR chemical shifts of protons involved in oxyanion hole hydrogen bonds increase by 0.50–0.76 ppm/ pK_a unit, suggesting a bond shortening of ~ 0.02 Å/ pK_a unit (Fig. 9).

Nevertheless, there is little change in binding affinity across a series of substituted phenolates ($\Delta\Delta G = -0.2$ kcal/mol/ pK_a unit). The small effect of increased charge localization on affinity occurs despite the shortening of the hydrogen bonds and a large favorable change in binding enthalpy ($\Delta\Delta H = -2.0$ kcal/mol/ pK_a unit). This shallow dependence of binding affinity suggests that electrostatic complementarity in the oxyanion hole makes at most a modest contribution to catalysis of ~ 300 -fold. The proposal is that geometrical complementarity between the oxyanion hole hydrogen-bond donors and the transition state oxyanion provides a significant catalytic contribution, and suggests that KSI, like other enzymes, achieves its catalytic prowess through a combination of modest contributions from several mechanisms rather than from a single dominant contribution.

Compound Cycling in the Environment

Going away from the molecular biology leaning NMR applications presented thus far, are a couple of uses for environmental and physical sciences. The first being the ability to utilize NMR to screen compounds cycling through the environment.

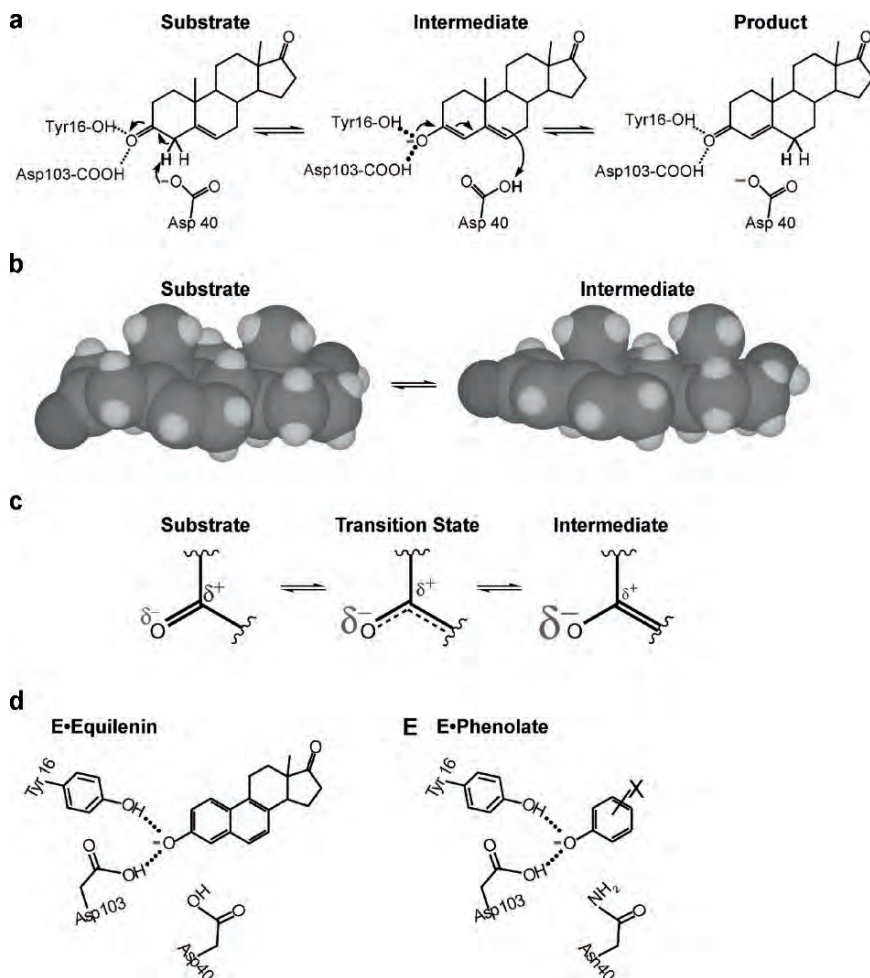


Fig. 8. Geometric and Electrostatic Changes in the KSI Reaction. (a) Mechanism of KSI catalyzed isomerization of 5-androstene-3,17-dione (substrate) to 4-androstene-3,17-dione (product). In the first step a general base, Asp40, removes a proton from the steroid to form a dienolate intermediate (via a dienolate-like transition state), which receives hydrogen bonds from an oxyanion hole consisting of Tyr16 and protonated Asp103. In the second step of the reaction the steroid is reprotonated at a different position to give the product. (b) Geometric changes accompanying the first half of the KSI reaction. Oxygen is shown in dark, carbon in grey, hydrogen in light. The ring geometry changes in the transition state and in the intermediate, becoming more planar in the intermediate. The sp^2 -hybridized carbonyl oxygen in the substrate becomes a predominantly sp^3 -hybridized oxyanion. (c) Electrostatic changes at the carbonyl group accompanying the first half of the KSI reaction. The larger “ δ^- ” refers to increased negative charge on the oxygen atom as the reaction proceeds. The dienolate-like transition state is expected to be between the substrate and dienolate intermediate in charge arrangement, but closer to the high-energy intermediate. (d) Schematic depiction of the steroid equilenin bound at the KSI active site. Equilenin geometrically and electrostatically resembles the dienolate reaction intermediate and transition state. (e) Schematic depiction of a single-ringed phenolate bound at the active site of KSID40N, the mutant enzyme used for this work. The Asp40Asn mutation mimics the protonated aspartate found in the intermediate and equilenin complexes, see (a) and (d), and leads to tighter binding of phenolate and other intermediate analogs.

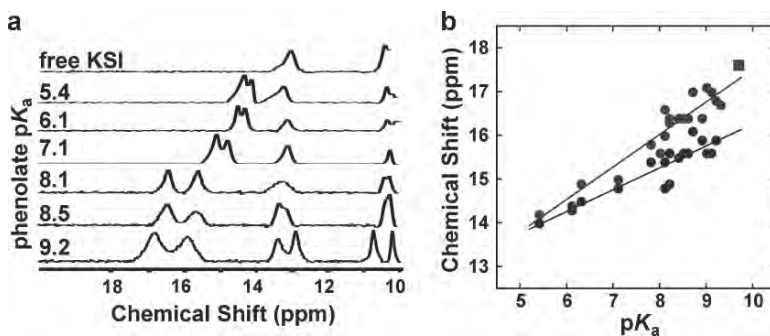


Fig. 9. ¹H NMR Downfield Chemical Shifts for Substituted Phenolates Bound to tKSID40N. (a) Representative spectra, with phenol pK_a shown on the left. From top to bottom: free enzyme, 3,4-dinitrophenol, 3-fluoro-4-nitrophenol, 4-nitrophenol, 3-fluoro-5-trifluoromethylphenol, 3,4-dichlorophenol, and 3-iodophenol. (b) Correlation between increasing phenolate pK_a and increasing chemical shift of observed downfield peaks. Circles are the two downfield peaks observed for phenolate binding. A linear fit gives slopes of 0.76 ± 0.06 and 0.50 ± 0.06 ppm/pK_a unit for the most downfield (upper line) and the second-most downfield (lower line) peak, respectively. The square is the main downfield peak observed with the intermediate analog equilenin (pK_a = 9.7), for comparison.

CHARACTERIZING PHOSPHOROUS IN THE ENVIRONMENT AND AGRICULTURE [5]

³¹P-NMR spectroscopy has substantially advanced the knowledge of phosphorus (P) compound content in soil, water and other environmental samples. Compounds include phosphonates orthophosphate; orthophosphate monoesters; orthophosphate diesters; pyrophosphate, and polyphosphate (Fig. 10).

Figure 11 shows solution ³¹P-NMR spectra for several extracted environmental samples. These include: forest floor material from grass in an oak-grass savannah in California; an alkaline agricultural soil from Idaho; marine sediment trap material from Monterey Bay, California; a sewage sludge sample from England; and a humic acid extracted from an Australian soil. These give a good idea of the variation in P forms, and their relative proportions, in environmental samples.

Because ³¹P is the only naturally occurring P isotope (100% natural abundance), all P species within a sample can potentially be detected by NMR spectroscopy. Thus, if the NMR experiments were carefully run, the area under each peak should be proportional to the number of that particular type of P nucleus, allowing NMR spectroscopy to quantitatively identify different P forms in a sample.

³¹P-NMR is a powerful approach that has been applied to a wide variety of studies including: phosphorous transformation; agricultural, forest, and natural

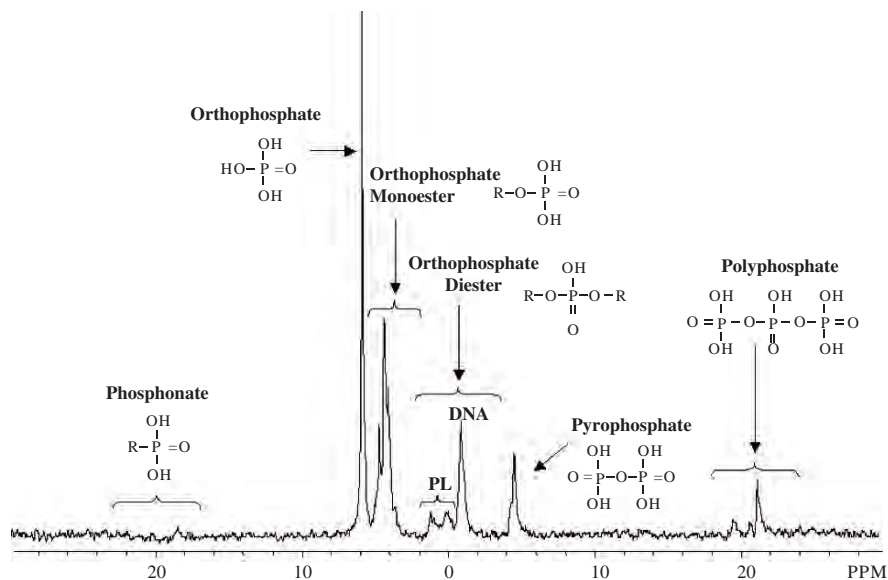


Fig. 10. A solution ^{31}P -NMR spectrum of a forest floor sample extracted with NaOH-EDTA. This spectrum shows the diversity of P species in natural samples, including phosphonates, orthophosphate, orthophosphate monoesters, orthophosphate diesters such as phospholipids (PL) and deoxyribonucleic acids (DNA), pyrophosphate and polyphosphate.

ecosystems; particle size separations and humic substances; manure, compost, sewage sludge and fertilizers; and freshwater, estuary and marine sediments.

Analysis of Synthetic Solid Compounds

The next non-molecular biology application of NMR spectroscopy involves its use for the analysis of man-made solid materials, particularly for structure and dynamics in crystalline, glassy, and liquid silicates and oxides, and on glass-forming liquids; The goal being to relate experimental measurements of atomic-scale processes to macroscopic properties of interest to the earth sciences (including mineralogy, igneous and metamorphic petrology, volcanology and geochemistry), materials sciences, and physical chemistry.

PROPERTIES OF HIGH-PRESSURE MELTS [6]

Silicate melts have played a major role in many events in the early histories of the Earth and moon and continue to shape planetary evolution. The densities and viscosities of silicate melts depend strongly on pressure, in part because of potentially measurable structural rearrangements. In an attempt to further

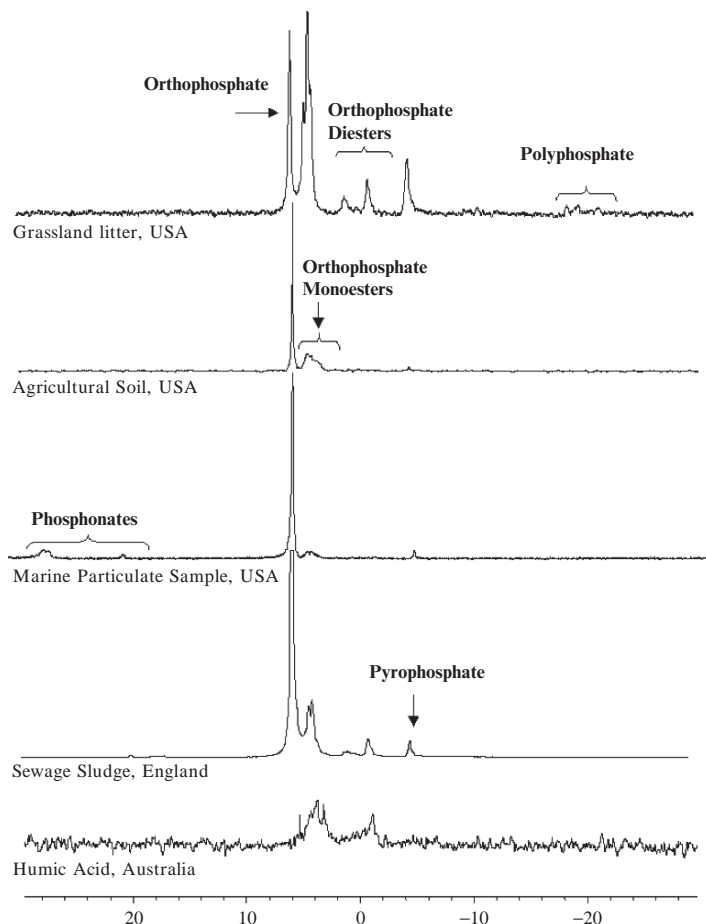


Fig. 11. ^{31}P -NMR spectra of a range of environmental samples. These include: forest floor material from under grass in an oak-grass savannah in California; an alkaline agricultural soil from Idaho; a marine particulate sample from Monterey Bay, California; a sewage sludge sample from England; and a humic acid extracted from Australian soil.

understand these changes and how they affect macroscopic properties, ^{27}Al magic angle spinning (MAS) NMR was used to determine the coordination of the Al cations in a series of aluminosilicate glasses quenched from melts at pressures of 2 to 8 GPa (Fig. 12).

The ^{27}Al MAS spectra in Fig. 12 indicate that the structures of all four- and five-component high-pressure glasses studied contain $^{\text{IV}}\text{Al}$ and $^{\text{VI}}\text{Al}$ and the abundances of these species increase with the quench pressure during formation. Integration of the fitted areas for each peak determined the relative abundances of Al coordinated by four, five, and six O atoms, which could be used to calculate the average Al coordination. The NMR results, along with viscosity and density measurements, support the concept that

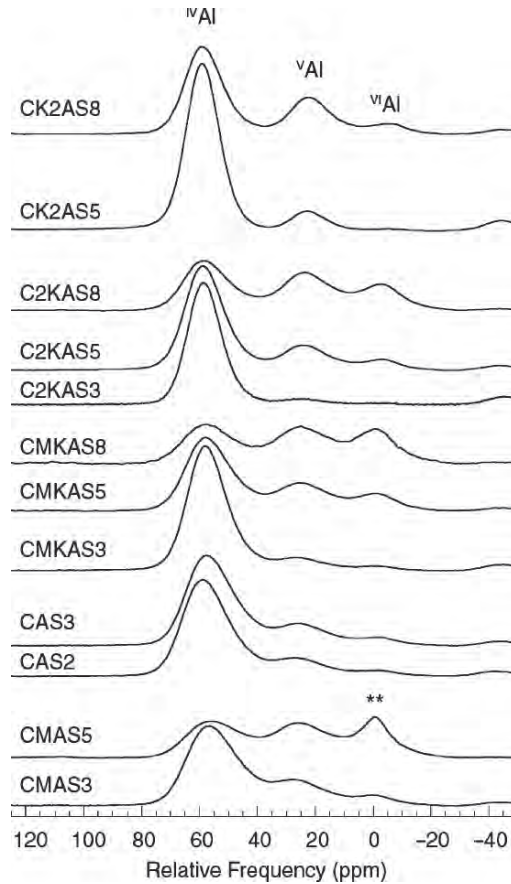


Fig. 12. Background-subtracted ^{27}Al MAS NMR spectra (18.8 T) of the high-pressure glasses scaled by equal area. The “**” symbols indicate signal resulting from a small amount of a crystalline impurity in CMAS5 only.

pressure-induced network structural transitions have direct implications for the macroscopic properties of high-pressure melts.

Microimaging of a Model Organism

Finally we return to a biological focus but not of NMR spectroscopy as broadly covered above, but into the realm of magnetic resonance imaging (MRI) as is now so prevalent in the medical community.

MICROIMAGING OF FRUIT FLIES [7]

The study of tiny, highly tractable model organisms is a powerful paradigm for understanding genetic and biochemical physiology, knowledge readily

carried through to vertebrate models and ultimately human medicine. *In vivo* methods for measuring small signaling molecules, metabolites, and neurotransmitters in model organisms, as well as dynamic *qualia* such as electrical potential, pH, fluid flow and molecular turnover, are highly desirable but very challenging. To this end Magnetic Resonance (MR), best known for anatomical imaging in widespread clinical applications, is an intriguing method to consider for *Drosophila* research.

This study demonstrated the feasibility of MR methods for imaging the fruit fly *Drosophila* with an NMR spectrometer, at a resolution relevant for undertaking future studies of the *Drosophila* brain and other organs. *Drosophila* adults and pupae from a range of developmental stages were imaged at high resolution by several distinct modes of imaging (Figs. 13 and 14). The fruit fly

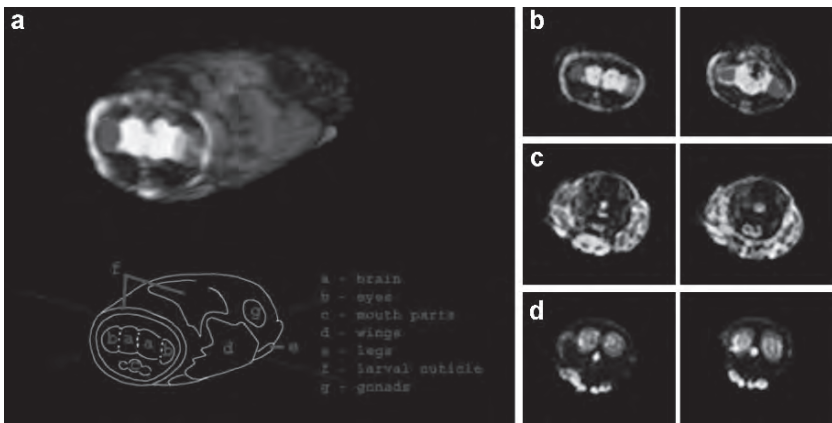


Fig. 13. Late *D. melanogaster* pupa. (a) Exterior view and labeled schematic. At right, selected example slices from the (b) head (c) thorax and (d) abdomen.

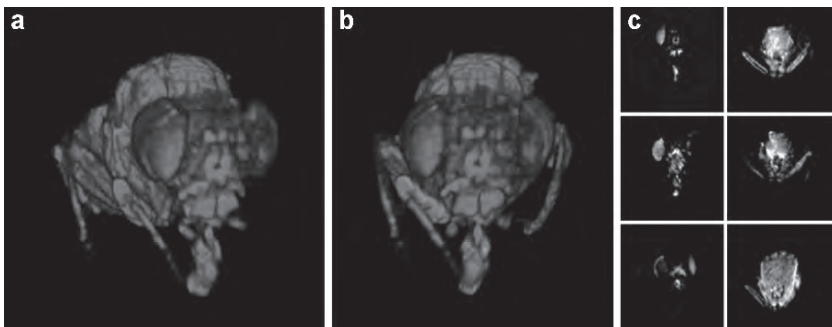


Fig. 14. Adult *D. bifurca*. (a, b) Two exterior views of 3D rendering. (c) Example 2D slices from the image set, three from the head (left) and three from the thorax (right).

has long been a principal model organism for elucidating biology and disease, but lacking capabilities like those of MRI. This feasibility marks progress toward the development of new *in vivo* research approaches in *Drosophila* without the requirement for light transparency or destructive assays.

Closing

Even with this small sampling of projects, the diversity of NMR spectroscopy applications cannot be disputed. Its pervasiveness throughout the sciences means that for any given research project, it is reasonable to think that some aspect of NMR spectroscopy could provide some benefit.

References

1. Li Q, Khosla C, Puglisi JD, Liu CW. "Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein." *Biochemistry*. 2003 Apr 29;42(16):4648–4657.
2. Alekseyev VY, Liu CW, Cane DE, Puglisi JD, Khosla C. "Solution structure and proposed domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase." *Protein Sci*. 2007 Oct;16(10):2093–2107.
3. Varani L, Bankovich AJ, Liu CW, Colf LA, Jones LL, Kranz DM, Puglisi JD, Garcia KC. "Solution mapping of T cell receptor docking footprints on peptide-MHC." *Proc Natl Acad Sci U S A*. 2007 Aug 7;104(32):13080–13085.
4. Kraut DA, Sigala PA, Pybus B, Liu CW, Ringe D, Petsko GA, Herschlag D. "Testing electrostatic complementarity in enzyme catalysis: hydrogen bonding in the ketosteroid isomerase oxyanion hole." *PLoS Biol*. 2006 Apr;4(4):e99.
5. Cade-Menun BJ. "Characterizing phosphorus in environmental and agricultural samples by ^{31}P nuclear magnetic resonance spectroscopy." *Talanta*. 2005 Apr 15;66(2):359–371.
6. Allwardt JR, Stebbins JF, Terasaki H, Du L-S, Frost DJ, Withers AC, Hirschmann MM, Suzuki A, Ohtani E. "Effect of structural transition on properties of high-pressure silicate melts: ^{27}Al NMR, glass densities, and melt viscosities." *Am Mineral*. 2007; 92, 1093–1104.
7. Null B, Liu CW, Hedehus M, Conolly S, Davis RW. "High-resolution, *in vivo* magnetic resonance imaging of *Drosophila* at 18.8 Tesla." *PLoS ONE*. 2008 Jul 30;3(7):e2817.

IMPROVED DYE STABILITY IN SINGLE-MOLECULE FLUORESCENCE EXPERIMENTS

COLIN ECHEVERRÍA AITKEN^{1,+}, R. ANDREW MARSHALL^{2,+}, AND JOSEPH D. PUGLISI^{3,4*}

¹ *Biophysics Program, Stanford University School of Medicine, Stanford, CA 94305-5126, USA*

² *Department of Chemistry, Stanford University, Stanford, CA 94305-5080, USA*

³ *Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305-5126, USA*

⁴ *Stanford Magnetic Resonance Laboratory, Stanford University School of Medicine, Stanford, CA 94305-5126, USA*

*Corresponding author: e-mail: puglisi@stanford.edu

Abstract Complex biological systems challenge existing single-molecule methods. In particular, dye stability limits observation time in single-molecule fluorescence applications. Current approaches to improving dye performance involve the addition of enzymatic oxygen scavenging systems and small molecule additives. We present an enzymatic oxygen scavenging system that improves dye stability in single-molecule experiments. Compared to the currently-employed glucose-oxidase/catalase system, the protocatechuate-3,4-dioxygenase system achieves lower dissolved oxygen concentration and stabilizes single Cy3, Cy5, and Alexa488 fluorophores. Moreover, this system possesses none of the limitations associated with the glucose oxidase/catalase system. We also tested the effects of small molecule additives in this system. Biological reducing agents significantly destabilize the Cy5 fluorophore as a function of reducing potential. In contrast, antioxidants stabilize the Cy3 and Alexa488 fluorophores. We recommend use of the protocatechuate-3,4,-dioxygenase system with antioxidant additives, and in the absence of biological reducing agents. This system should have wide application to single-molecule fluorescence experiments.

⁺ These authors contributed equally

Introduction

Single-molecule techniques eliminate ensemble averaging, thus revealing transient or rare species in heterogeneous systems [1–3]. These approaches have been employed to probe myriad biological phenomena, including protein and RNA folding [4–6], enzyme kinetics [7, 8], and even protein biosynthesis [1, 9, 10]. In particular, immobilization-based fluorescence techniques such as total internal reflection fluorescence microscopy (TIRF-M) have recently allowed for the observation of multiple events on the milliseconds to seconds timescale [11–13].

Single-molecule fluorescence methods are challenged by the instability of single fluorophores. The organic fluorophores commonly employed in single-molecule studies of biological systems display fast photobleaching, intensity fluctuations on the millisecond timescale (blinking), or both. These phenomena limit observation time and complicate the interpretation of fluorescence fluctuations [14, 15].

Molecular oxygen (O_2) modulates dye stability. Triplet O_2 efficiently quenches dye triplet states responsible for blinking. This results in the formation of singlet oxygen [16–18]. Singlet O_2 reacts efficiently with organic dyes, amino acids, and nucleobases [19, 20]. Oxidized dyes are no longer fluorescent; oxidative damage impairs the folding and function of biomolecules. In the presence of saturating dissolved O_2 , blinking of fluorescent dyes is suppressed, but oxidative damage to dyes and biomolecules is rapid. Enzymatic O_2 -scavenging systems are commonly employed to ameliorate dye instability. Small molecules are often employed to suppress blinking at low O_2 levels.

The most commonly used system employs coupled glucose oxidase and catalase (GODCAT) in a glucose-containing buffer [9–13, 21]. Blinking is controlled by addition of β -mercaptoethanol (BME) [22]. This system allows for observation on the tens of seconds timescale and has been employed in numerous studies. Nonetheless, the utility of this system is limited. Both enzymes employ highly colored cofactors; glucose oxidase employs flavin adenine dinucleotide, while catalase uses a heme group. Catalase is inhibited by a variety of molecules, including BME, DTT, and ascorbic acid [23–25]. Certain *in vitro* conditions, such as experiments in cell extract, are incompatible with the enzymatic system. Moreover, mismatched glucose oxidase and catalase activities can result in a build-up of hydrogen peroxide, a reactive oxygen species that may damage fluorophores and biomolecules (Fig. 1a) [26]. While it suppresses blinking on the millisecond timescale, BME induces transitions to long-lived (seconds) dark states, or slow blinking, in certain fluorophores. Rasnik et al. recently characterized this effect and that of Trolox, a water-soluble vitamin E analog that suppresses slow blinking, as well as photobleaching, as compared to BME [27]. Other chemical additives – anti-fading agents, anti-oxidants,

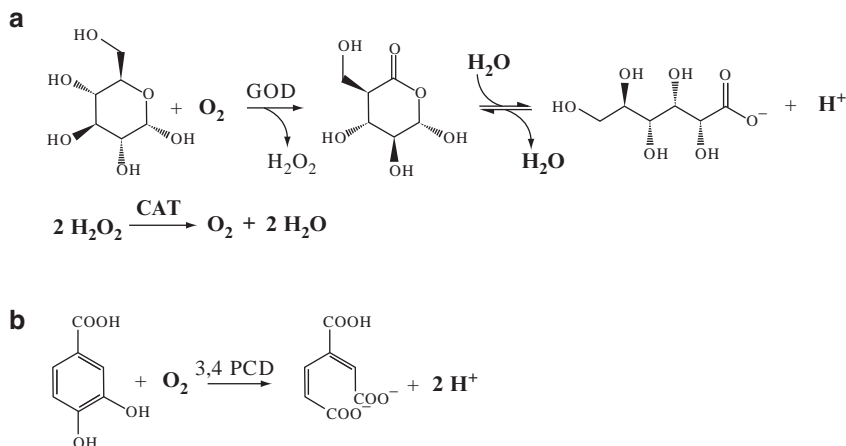


Fig. 1. Schematics of chemical reactions in GODCAT (a) and PCD systems (b).

triplet state quenchers, singlet oxygen quenchers – have been employed to improve dye stability in other systems [27–36].

We applied the protocatechuic acid (PCA)/protocatechuate-3,4-dioxygenase (PCD) O_2 -scavenging system to single-molecule TIRF-M experiments. PCD is a well-characterized enzyme that has been applied to biological systems requiring anaerobic conditions [37–40]. PCD is multimeric enzyme that employs a nonheme iron center to catalyze the conversion of PCA to β -Carboxy-*cis,cis*-muconic acid in one step, resulting in the consumption of one mole of O_2 and the production of two protons per mole of PCA converted (Fig. 1b). We characterized the effect of PCD catalyzed O_2 -scavenging on the stability of three dyes commonly used in single-molecule TIRF-M experiments, Cy3, Cy5, and Alexa488. We further characterized the effects of commonly used biological reducing agents and other chemical additives. Our system significantly improves the stability of single fluorophores and should have broad application to single-molecule experiments.

Improved Dye Stability With PCD

OXYGEN-SCAVENGING MEASUREMENTS

We performed simple kinetic assays to compare the GODCAT and PCD systems. A dissolved oxygen probe submerged in glucose containing buffer under constant stirring was employed to determine initial reaction rates and steady-state O_2 concentration; reaction progress was followed upon addition of enzyme. Under conditions currently employed in the literature (~ 100 nM glucose oxidase, ~ 1.5 μ M catalase, 56 mM glucose) the GODCAT system

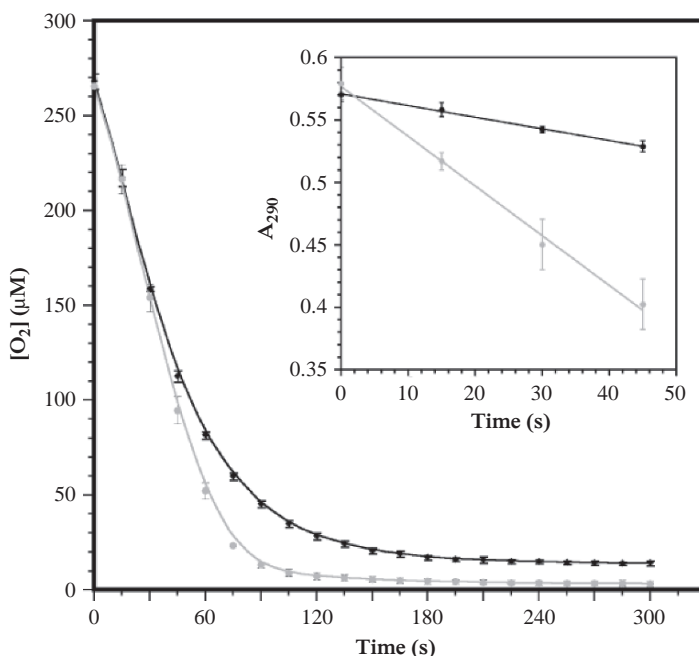


Fig. 2. Dissolved oxygen timecourses for GODCAT (black) and standardized PCD (gray) systems. Conditions for each reaction are as follows: GODCAT ~ 100 nM glucose oxidase, ~ 1.5 μM catalase, 56 mM glucose; PCD ~ 50 nM PCD, 2.5 mM PCA. (inset) Timecourses of PCA consumption as monitored by absorbance at 290-nm. Each reaction contains 200 μM PCA and either 50 nM (black) or 10 nM (gray) PCD.

consumes O_2 at an initial rate of $4 \mu M s^{-1}$ and achieves a steady-state O_2 concentration of $14 \pm 1 \mu M$ within approximately 3 min (Fig. 2).

To enable direct comparison of dye stability in the GODCAT and PCD systems, we standardized O_2 -scavenging by PCD to that of the GODCAT system according to our dissolved oxygen assay. At ~ 50 nM PCD and 2.5 mM PCA, the PCD system reproduces the initial rate of O_2 -scavenging by the GODCAT (Fig. 2). However, PCD achieves a steady-state O_2 concentration of $3 \pm 1.5 \mu M$, five-fold lower than GODCAT; this oxygen concentration is maintained for several hours with no change in pH (data not shown). We observed no appreciable loss in activity for enzymes stored either at $-80^\circ C$ or $4^\circ C$ for periods exceeding 1 month, similar to the observations of Patil et al. (data not shown) [40]. We also monitored reaction progress with a second assay; PCA absorbs at 290 nm, and its disappearance can be monitored with a UV/VIS spectrophotometer. Initial rates determined using this method agreed with dissolved oxygen measurements (Fig. 2, inset). This complementary assay allows for quick verification of PCD activity with equipment common to most labs.

SINGLE-MOLECULE MEASUREMENTS

We next compared the photophysical stability of single Cy3, Cy5, and Alexa488 fluorophores in the standardized GODCAT and PCD systems. These fluorophores were immobilized via covalent linkage to short, biotinylated oligonucleotides. Intensity trajectories for at least 700 molecules were collected for each condition. Trajectories were subsequently parsed into three component events: initial fluorescence, blinked-off fluorescence, and blinked-on fluorescence as (Fig. 3). Events limited by the completion of observation were discarded. Mean lifetimes for each component event were obtained by fitting to single-exponential probability functions (Fig. 3). Mean intensities and signal-to-noise ratios were also determined for each condition. The relative frequency, duration, and intensity of these component events can be used to define the utility of a fluorophore in single-molecule experiments; an ideal fluorophore should exhibit prolonged initial fluorescence, minimal blinking, and high signal-to-noise.

Using this analysis, we determined photo-stability of Cy3, Cy5, and Alexa488 in the GODCAT O_2 -scavenging system. For all three fluorophores, initial fluorescence is observed from approximately 90% of immobilized molecules. The two dyes most commonly employed in single-molecule TIRF-M studies – Cy3 and Cy5 – display long initial lifetimes, infrequent blinking, and average signal to noise values of approximately 4.0 (Fig. 4). The rhodamine-derived Alexa488 gives significantly shorter initial lifetimes, increased blinking, and average signal to noise values of approximately 2.5 (Fig. 4).

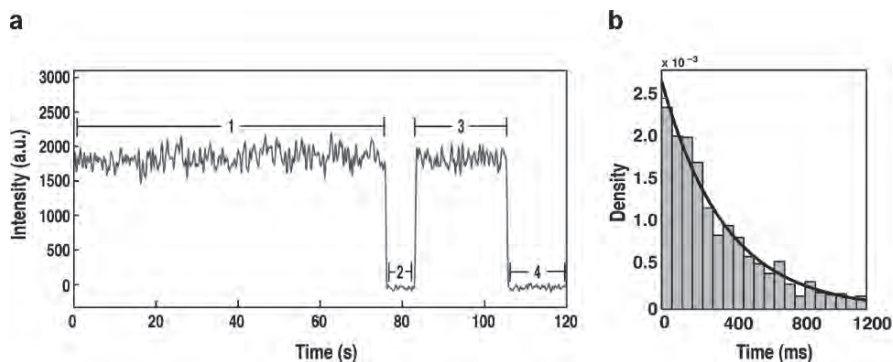


Fig. 3. Analysis of single fluorophores. (a) Representative fluorescence vs. time trace for a single dye. Initial events (1), off-blink (2), and on-blink events (3), were segregated prior to lifetime analysis; events limited by acquisition length (4) were discarded. (b) Initial, off-blink, and on-blink event distributions were fit, as described, to an exponential probability function to determine mean lifetimes, here Cy5 initial lifetimes with PCD.

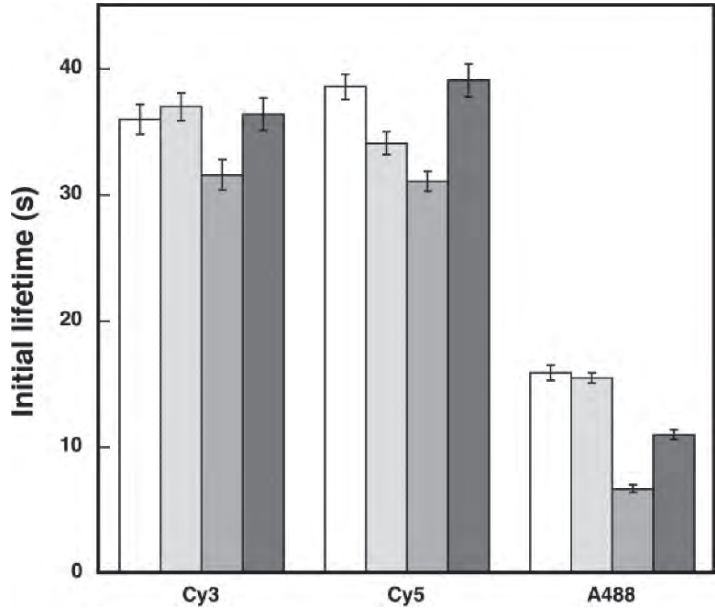


Fig. 4. Comparison of initial dye lifetimes in the standardized PCD (white), dilute PCD (light gray), GODCAT (gray), and GODCAT+PCA (dark gray) systems.

TABLE 1. Mean lifetimes and event frequencies (per molecule) in the GODCAT (a) and PCD (b) systems.

| | Cy3 | | | Cy5 | | | A488 | | |
|---------------------|---------|----------|-----------|---------|----------|-----------|---------|----------|-----------|
| | Initial | On-blink | Off-blink | Initial | On-blink | Off-blink | Initial | On-blink | Off-blink |
| (a) | | | | | | | | | |
| Mean Lifetime | 31.6 | 10.3 | 2.0 | 31.1 | 20.2 | 24.5 | 6.7 | 3.3 | 13.4 |
| (s) Events per mol. | | | | | | | | | |
| | ±1.6 | ±0.6 | ±0.2 | ±0.8 | ±1.8 | ±2.1 | ±0.3 | ±0.1 | ±0.3 |
| | 0.88 | 0.36 | 0.26 | 0.97 | 0.08 | 0.09 | 0.71 | 2.14 | 2.17 |
| (b) | | | | | | | | | |
| Mean Lifetime | 36.0 | 10.5 | 21.5 | 38.6 | 18.9 | 22.4 | 15.9 | 4.4 | 6.8 |
| (s) Events per mol. | | | | | | | | | |
| | ±1.2 | ±0.6 | ±1.2 | ±1.0 | ±1.7 | ±2.0 | ±0.6 | ±0.1 | ±0.2 |
| | 0.92 | 0.38 | 0.31 | 0.95 | 0.08 | 0.12 | 0.87 | 2.07 | 2.10 |

PCD stabilizes Cy3, Cy5, and Alexa488, as compared to GODCAT. PCD extends initial lifetimes up to 140% with no significant effect on blinking frequency or lifetimes (Table 1). In fact, similar results can be obtained using a PCD system with a five-fold lower enzyme concentration, provided the

system is allowed to achieve steady-state (Fig. 4). UV/VIS activity measurements confirmed a five-fold decrease in the initial rate of O_2 -scavenging with this dilute system (inset, Fig. 2). This suggests that the rate of O_2 -scavenging currently employed in the literature exceeds the minimum requirements for dye stability. All subsequent PCD measurements were obtained with this dilute system (10 nM PCD, 2.5 mM PCA, 1 mM Trolox).

To test the origin of dye stabilization in the PCD system, we added PCA to the GODCAT system. Interestingly, addition of PCA to the GODCAT system recapitulates the improvements in Cy3 and Cy5 stability observed with PCD (Fig. 4). PCA appears to stabilize the Cy3 and Cy5 fluorophores. However, addition of PCA to the GODCAT system does not replicate Alexa488 initial lifetimes observed with PCD. PCA extends Alexa488 initial lifetimes to 11.0 ± 0.4 s, whereas lifetimes longer than 15 s are observed in both the standardized and dilute PCD systems (15.9 ± 0.6 s and 15.4 ± 0.4 s, respectively). Increased Alexa488 stability in the PCD system is due not only to PCA, but to other factors as well, potentially lower steady-state O_2 concentrations. These results highlight the unique environmental responses of distinct dyes.

The Effect of Chemical Additives On Dye Stability

Many small molecule additives that purportedly stabilize emitting fluorophores under various conditions have been identified in the literature[27–36]. We characterized the effects of four previously described chemical additives in the PCD system: mercaptoethylamine (MEA), diazabicyclo[2.2.2]octane (DABCO), ascorbic acid, and n-propyl gallate (nPG). MEA, DABCO, and ascorbic acid were used at 10 mM. Due to limited solubility in aqueous solutions, nPG was used at 100 μ M. These molecules possess a variety of chemical functionalities (Fig. 5), and haven been reported to act as either singlet oxygen quenchers, triplet state quenchers, or antioxidants. We observed no effect on PCD O_2 -scavenging in the presence of each additive, as determined by dissolved oxygen measurements (data not shown). However, we observed a variety of effects on dye stability (Fig. 5).

MEA

MEA has been described as an effective triplet-state quencher [28]. However, we observed deleterious effects on Cy5 stability in the presence of MEA (Fig. 5). Most notably, MEA shortens initial and on-blink lifetimes (5.3 ± 0.2 s and 4.5 ± 0.1 s, respectively), lengthens off-blink lifetimes to 30.6 ± 0.5 s, increases the frequency of blinking events 20-fold (2.4 events per molecule), and decreases signal to noise by 25%. MEA improves the initial lifetimes of Cy3

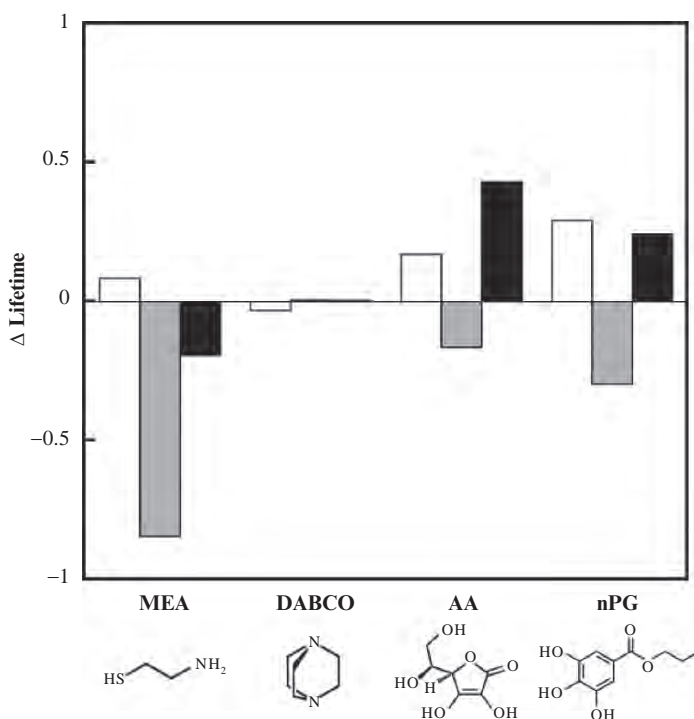


Fig. 5. Bar plot of the fractional change in initial lifetimes of Cy3 (white), Cy5 (gray), and Alexa488 (black) in the PCD system with respective additives.

fluorophores (40.1 ± 1.6 s) but decreases signal to noise by 20%. No significant changes in Alexa488 fluorescence were observed in the presence of MEA.

DABCO

DABCO is a purported anti-fading agent, presumably through triplet state and singlet oxygen quenching (Fig. 5) [33, 35, 36, 41]. However, we observed little to no effects on dye stability upon addition of DABCO. Initial, on-blink, and off-blink lifetimes of all three fluorophores are largely unaffected, though a small increase in the frequency of blinking events was observed in all cases. Likewise, we observed no significant effects on signal-to-noise.

ASCORBIC ACID

Ascorbic acid is an antioxidant and is thought to prevent photobleaching by quenching radical species [28, 31]. Addition of ascorbic acid increased initial dye lifetimes for both Cy3 (43.3 ± 1.2 s) and Alexa 488 (22.2 ± 0.7 s) (Fig. 5). In contrast, ascorbic acid slightly shortens Cy5 initial lifetimes (28.5 ± 0.7 s). We observed no effect on blinking or signal-to-noise.

NPG

nPG is also an antioxidant and has similarly been employed to counteract photobleaching [30, 32, 34, 42]. In the presence of nPG, we observed a profile of effects similar to those of ascorbic acid (Fig. 5). nPG extends both Cy3 and Alexa488 initial lifetimes (47.8 ± 1.3 s, and 19.3 ± 0.7 s, respectively); Cy5 initial lifetimes decreased (24.1 ± 0.5 s). Blinking and signal-to-noise were unaffected.

Reducing Agents Destabilize Single Fluorophores

Biological reducing agents are commonly used to reduce protein disulfide groups *in vitro* [43–46]. β -mercaptoethanol (BME) has been shown to suppress fast (millisecond) blinking of Cy5; most single-molecule TIRF-M experiments are performed in the presence of millimolar concentrations of BME. However, Rasnik et al. recently demonstrated that BME induces slow blinking of Cy5 [27]. L-glutathione, another thiol-containing compound, was similarly observed to suppress fast blinking, probably by quenching triplet states, while inducing slow blinking. The authors suggest that thiol-containing compounds adversely affect dye photophysics.

To test the effect biological reducing agents, we characterized the stability of Cy3, Cy5, and Alexa488 with PCD in the presence of 10mM BME, dithiothreitol (DTT), and tris(2-carboxyethyl)phosphine (TCEP). The redox potentials of BME and DTT at pH 7 are -0.26 V and -0.33 V, respectively [46, 47]; TCEP reduces DTT in solution [44]. Dissolved oxygen measurements showed no effect on PCD O_2 -scavenging in the presence of BME, DTT, or TCEP (data not shown). In agreement with prior studies, we observed drastic reduction of Cy5 stability in the presence of thiol-containing compounds Table 2).

TABLE 2. Cy5 event frequencies and lifetimes for initial, on-blink, and off-blink events (per molecule) in the absence or presence of 10mM BME, DTT, and TCEP.

| | No reductant | | BME | | DTT | | TCEP | |
|-----------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|
| | Events per mol. | Lifetime (s) | Events per mol. | Lifetime (s) | Events per mol. | Lifetime (s) | Events per mol. | Lifetime (s) |
| Initial | 0.96 | 34.1 ± 0.9 | 0.91 | 10.6 ± 0.4 | 0.88 | 4.3 ± 0.1 | 0.51 | 0.4 ± 0.1 |
| On-blink | 0.11 | 15.1 ± 1.2 | 1.02 | 9.8 ± 0.3 | 0.79 | 4.3 ± 0.1 | 1.02 | 0.4 ± 0.1 |
| Off-blink | 0.12 | 21.4 ± 1.6 | 1.04 | 39.2 ± 1.2 | 0.80 | 64.8 ± 2.1 | 1.02 | 48.2 ± 1.9 |

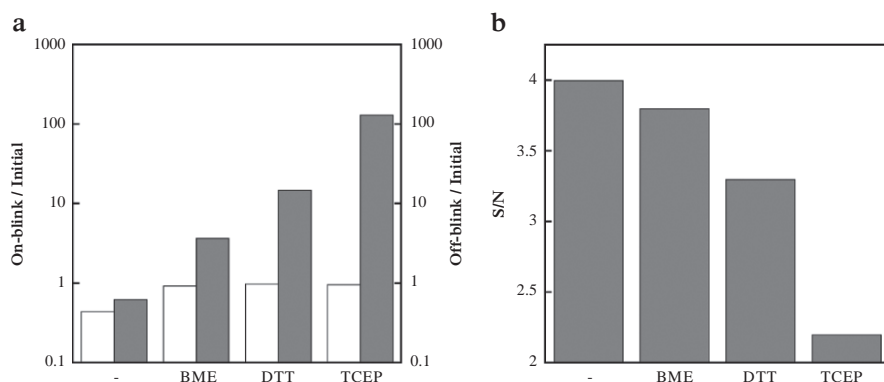


Fig. 6. Cy5 dye stability in the presence of biological reducing agents. **(a)** Bar plot of the ratios of on-blink-to-initial lifetimes (white) and off-blink-to-initial lifetimes (gray) for Cy5 in the presence of reducing agents. **(b)** Bar plot of signal-to-noise ratio (S/N) for Cy5, determined on a per molecule basis.

Both BME and DTT severely shorten Cy5 initial lifetimes, and increase blinking frequency approximately ten-fold. In fact, in the presence of reducing agents, all observed fluorescence on-states – initial and blinking events – are of equal duration. Off-blink events are drastically lengthened, to 64.8 ± 2.1 s, in the presence of DTT. Both BME and DTT reduce Cy5 signal-to-noise. Interestingly, TCEP, which contains no thiol groups, has even more striking effects (Table 2). Biological reducing agents as a whole, not only those containing thiol groups, destabilize emitting fluorophores.

Notably, the magnitude of this destabilization scales with reducing potential. Cy5 initial lifetimes are shortened approximately 4-fold, 9-fold, and 100-fold by BME, DTT, and TCEP, respectively (Fig. 6). On-blink lifetimes are similarly shortened as a function of reducing potential; off-blink lifetimes increase in the presence of all reducing agents. However, reducing agents do not affect the overall contribution of on-blink events as compared to initial events, while that of off-blink events is drastically increased as a function of reducing power; normalization to initial lifetimes reveals these trends (Fig. 6b). Signal-to-noise are decreased to 3.8, 3.3, and 2.2 in the presence of BME, DTT, and TCEP, respectively (Fig. 6).

Biological reducing agents show mixed effects on the stability of Cy3 and Alexa488 (data not shown). BME and DTT extend Cy3 initial lifetimes, whereas TCEP decreases initial lifetimes. As with Cy5, all three reducing agents decrease Cy3 signal-to-noise as a function of reducing power. BME effect only affects Alexa488 blinking, increasing the frequency of these events 50%. DTT, however, drastically reduces Alexa488 signal-to-noise, to the

point that intensity trajectories are virtually uninterpretable. TCEP hinders almost all aspects of Alexa488 fluorescence: initial lifetimes are shortened, blinking events are more frequent, and signal to noise is decreased.

As a group, biological reducing agents – not only those containing thiol groups – severely impair dye stability. These effects are most drastic for Cy5, and correlate with reducing power, not chemical functionality.

Discussion

As described above, the GODCAT system suffers from several limitations. Recently, Patil et al. presented PCD as an alternative O_2 -scavenging system; this system has since been employed in numerous published studies to control or minimize solution O_2 concentrations [48]. PCD is a single enzyme system possessing none of the aforementioned limitations. Catalysis is achieved by a nonheme iron center requiring no additional co-factors. The catalytic mechanism of oxidation is well understood, and does not involve production of reactive species. PCD from various *Pseudomonas* species is commercially available, stable for months, and active in a variety of buffers of pH 6–9.

We matched the initial O_2 -scavenging rate of PCD to that of the GODCAT system. This standardized PCD system achieves lower steady-state O_2 concentrations than GODCAT. Five-fold dilution of the PCD system results in a corresponding decrease in the initial rate of O_2 -scavenging, but still results in similar steady-state O_2 concentrations.

PCD significantly improves the stability of Cy3, Cy5, and Alexa488 in single-molecule TIRF-M experiments. PCD extends initial dye lifetimes up to 140% with no deleterious effects. This improvement is conserved with five-fold dilute PCD, provided the system is permitted to achieve equilibrium. Interestingly, the cofactor PCA appears to stabilize single fluorophores. Inclusion of PCA in the GODCAT system recapitulates the improvements in Cy3 and Cy5 dye stability observed with PCD. The antioxidant properties of PCA and its ethyl ester, ethyl protocatechuate have been characterized previously[49, 50]. In fact, PCA shares chemical similarities with Trolox and nPG, both of which improve dye stability in fluorescence applications. However, removal of Trolox from the PCD system results in increased blinking as described by Rasnik et al. (data not shown). Moreover, PCA is not solely responsible for improved Alexa488 stability in the PCD system. Others factors, whether reaction products of PCD or lower steady-state O_2 levels, likely contribute to improved Alexa488 stability. This underscores the particular environmental sensitivities of distinct fluorophores.

Both Trolox and PCA stabilize single fluorophores. However, these effects are achieved in the absence of biological reducing agents. Most *in vitro* studies of biological systems are performed in buffers containing reducing agents. Thiol-containing reducing agents have been observed to cause slow blinking of Cy5 on a timescale that limits the observation of biological systems. To identify reducing agents compatible with single-molecule fluorescence experiments, we compared the effects of DTT, TCEP, and BME on the stability of Cy5. All three reducing agents increased blinking frequency in the PCD system. The effect of biological reducing agents such as BME, then, is not limited to those containing thiol moieties. We also observe decreased initial lifetimes, increased off-blink lifetimes, and decreased signal to noise in the presence of reducing agents. These effects scale strongly with reducing power. Cy5 photochemistry is populated by various emitting and dark states [27, 51–58] that might interact with strong reducing agents. We suggest removal of biological reducing agents from single-molecule TIRF-M experiments with Cy5; *in vitro* measurements performed on the minutes timescale can successfully be performed in the absence of reducing agents.

We also tested the effects on dye stability of four small molecules employed in the literature to stabilize fluorophores: MEA, DABCO, nPG, and ascorbic acid. Surprisingly, the effect of MEA was to destabilize single fluorophores, similar to BME and DTT. DABCO had little effect on dye stability. In fact, the efficacy of DABCO has been questioned elsewhere [33, 34]. In contrast to MEA and DABCO, nPG and ascorbic acid showed significantly stabilized emitting fluorophores, specifically Cy3 and Alexa488. nPG and ascorbic acid are well known antioxidants, and may improve dye stability through interactions with dark radical ion species proposed to precede photobleaching. However, neither ascorbic acid nor nPG improve the stability of Cy5, potentially because their redox potentials do not permit efficient interaction with Cy5 radical species. Our results highlight the role of redox interactions in fluorophore stability. More detailed investigation of this complex physical landscape might permit for the rational pairing of specific fluorophores and antioxidants.

Materials and Methods

LABELED OLIGONUCLEOTIDES

5'-amino, 3'-biotin functionalized oligonucleotides were synthesized by IDT (Coralville, IA) and labeled with NHS esters of Cy3, Cy5 (GE Healthcare, Piscataway, NJ), and Alexa Fluor 488 (Alexa488) (Molecular Probes, Carlsbad, CA). Labeling reactions were performed under the following conditions:

250 μ M oligo, 2.5 mM dye NHS-ester, 500 mM NaCl, and 100 mM KBO₃ pH 9. Reactions were allowed to proceed at room temperature for 1 h. Labeled oligos were subsequently PAGE-purified, eluted by crush and soak, ethanol precipitated, and stored at -20°C .

ENZYMES AND OTHER REAGENTS

All reagents are from Sigma (St. Louis, MO) unless otherwise noted. TP50 is 50 mM KCl, 50 mM Tris-OAc pH 7.5. Glucose oxidase from *Aspergillus niger* and catalase from bovine liver were obtained as lyophilized powders and stored as 50% glycerol stocks in TP50 at -20°C . The lyophilized powder of protocatechuate 3,4-dioxygenase from *Pseudomonas* species was resuspended and stored as a 50% glycerol stock in 50 mM KCl, 1 mM EDTA, 100 mM Tris-HCl pH 8. The concentration of each enzyme was estimated using molecular weight information provided by Sigma (glucose oxidase = 160 kDa, catalase = 250 kDa, PCD = 700 kDa). Purified PCDs from *Burkholderia cepacia* and *Pseudomonas putida* were obtained as generous gifts from D. Ballou and D. Ohlendorf, respectively, and used as supplied. 3,4-dihydroxybenzoic acid was resuspended to 100 mM in deionized water and adjusted to pH 9 using NaOH. 100 mM Trolox solution was prepared by dissolving Trolox in methanol followed by the addition of de-ionized water and 1 M NaOH to the appropriate volume (final concentration methanol = 10% at pH \sim 9.5). Ascorbic acid was dissolved to 100 mM in water and brought to pH 7 with 1 M NaOH. n-Propyl gallate was dissolved to 10 mM in 1% ethanol/water (v/v). 1,4-diazabicyclo [2.2.2]octane dissolved in water and brought to pH 7.5 with 1 M HCl, final concentration of 1 M. Mercaptoethylamine was dissolved in water and brought to pH 7.5 with glacial acetic acid, final concentration 1 M. Dithiothreitol was dissolved to 1 M in deionized water. Tris(2-carboxyethyl)phosphine hydrochloride solution (0.5 M in water) and β -mercaptoethanol were used as supplied.

DISSOLVED OXYGEN KINETICS

Dissolved O₂ (DO) measurements were performed using an Orion 835A meter (Thermo, Waltham, MA). Probe calibration was performed according to the manufacturer's protocol prior to each set of experiments, using ZERO OXYGEN SOLUTION (Oakton, Vernon Hills, IL) and water-saturated air. The DO probe was inserted into a 20 mL glass scintillation vial containing 4 mL TP50 (TP50 + 1.25% glucose for glucose oxidase/catalase measurements) under constant stirring. Stock solutions of glucose oxidase, catalase, PCD, and PCA were diluted to 500 μ L each and delivered by pipetting to the reaction vial. Measurements were recorded manually.

PCA KINETICS

Conversion of PCA to β -carboxy-cis,cis-muconic acid was monitored by absorbance at 290-nm using a Shimadzu UV1201 UV/VIS spectrophotometer (Kyoto, Japan). Two microliters of PCA stock solution was rapidly mixed with 1 mL of dilute PCD in buffer to a final concentration of 200 μ M PCA. Measurements for the determination of initial rates were recorded manually.

SINGLE-MOLECULE TIRF

Flow-cells for single-molecule TIRF microscopy were prepared following a protocol derived from Ha et al. [59]. Fluorescently-labeled oligos were immobilized at \sim 10 pM, yielding hundreds of well resolved single dyes per viewing field. Cy3 and Cy5 functionalized oligos were immobilized in the same flow cell and simultaneously imaged. Alexa488 functionalized oligos were immobilized and imaged in a separate flow cell. Unless otherwise mentioned, measurements were performed in the presence of 1 mM Trolox. All measurements were performed after the selected O_2 -scavenging system had achieved steady-state as determined by previous dissolved oxygen measurements. Measurements were performed using a prism-based total internal reflection instrument built in-house [9]. Image acquisition and analysis was controlled with the MetaMorph software package (Molecular Devices, Sunnyvale, CA). Spot picking and fluorescence intensity trajectories were recorded post-acquisition using MetaMorph. Analysis of fluorescence intensity trajectories was performed using MATLAB (Mathworks, Natick, MA).

Acknowledgements We thank D. Ohlendorf and D. Ballou for gifts of purified PCD enzymes, and the Puglisi lab members – notably Manolia Margaris and Sean “Jaguar” McKenna – for feedback and advice. This work was supported by NIH grant GM51266. C.E. Aitken is supported by an NIH Molecular Biophysics training grant (T32 GM008294).

References

1. Bai, L., T. J. Santangelo, and M. D. Wang. 2006. Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct* 35:343–360.
2. Cornish, P. V., and T. Ha. 2007. A survey of single-molecule techniques in chemical biology. *Acc Chem Biol* 2:53–61.
3. Weiss, S. 1999. Fluorescence spectroscopy of single biomolecules. *Science* 283:1676–1683.
4. Zhuang, X., and M. Rief. 2003. Single-molecule folding. *Curr Opin Struct Biol* 13:88–97.
5. Schuler, B. 2005. Single-molecule fluorescence spectroscopy of protein folding. *Chemphyschem* 6:1206–1220.
6. Lipman, E. A., B. Schuler, O. Bakajin, and W. A. Eaton. 2003. Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.
7. Min, W., B. P. English, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie. 2005. Fluctuating enzymes: lessons from single-molecule studies. *Acc Chem Res* 38:923–931.
8. Xie, X. S., and H. P. Lu. 1999. Single-molecule enzymology. *J Biol Chem* 274:15967–15970.

9. Blanchard, S. C., R. L. Gonzalez, H. D. Kim, S. Chu, and J. D. Puglisi. 2004. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* 11:1008–1014.
10. Blanchard, S. C., H. D. Kim, R. L. Gonzalez, Jr., J. D. Puglisi, and S. Chu. 2004. tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A* 101:12893–12898.
11. Munro, J. B., R. B. Altman, N. O'Connor, and S. C. Blanchard. 2007. Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell* 25:505–517.
12. Lee, T. H., L. J. Lapidus, W. Zhao, K. J. Travers, D. Herschlag, and S. Chu. 2007. Measuring the folding transition time of single RNA molecules. *Biophys J* 92:3275–3283.
13. Joo, C., S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong, and T. Ha. 2006. Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* 126:515–527.
14. Bagshaw, C. R., and D. Cherny. 2006. Blinking fluorophores: what do they tell us about protein dynamics? *Biochem Soc Trans* 34:979–982.
15. Ha, T. J., A. Y. Ting, J. Liang, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss. 1999. Temporal fluctuations of fluorescence resonance energy transfer between two dyes conjugated to a single protein. *Chem Phys* 247:107–118.
16. Hubner, C. G., A. Renn, I. Renge, and U. P. Wild. 2001. Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen. *J Chem Phys* 115:9619–9622.
17. Renn, a., J. Seelig, and V. Sandoghdar. 2006. Oxygen-dependent photochemistry of fluorescent dyes studied at the single molecule level. *Mol Phys* 104:409–414.
18. Piwonski, H., R. Kolos, A. Meixner, and J. Sepiol. 2005. Optimal oxygen concentration for the detection of single indocarbocyanine molecules in a polymeric matrix. *Chem Phys Lett* 405:352–356.
19. Sies, H., and C. F. Menck. 1992. Singlet oxygen induced DNA damage. *Mutat Res* 275:367–375.
20. Davies, M. J. 2004. Reactive species formed on proteins exposed to singlet oxygen. *Photochem Photobiol Sci* 3:17–25.
21. Harada, Y., K. Sakurada, T. Aoki, D. D. Thomas, and T. Yanagida. 1990. Mechanochemical coupling in actomyosin energy transduction studied by in vitro movement assay. *J Mol Biol* 216:49–68.
22. Kishino, A., and T. Yanagida. 1988. Force measurements by micromanipulation of a single actin filament by glass needles. *Nature* 334:74–76.
23. Orr, C. W. M. 1967. Studies on ascorbic acid. 2. Physical changes in catalase following incubation with ascorbate or ascorbate and copper (2). *Biochemistry* 6:3000–3006.
24. Orr, C. W. M. 1967. Studies on Ascorbic Acid. I. Factors influencing ascorbate-mediated inhibition of catalase. *Biochemistry* 6:2995–3000.
25. 1994. In *The Enzyme Handbook*. D. Schomburg, M. Salzmänn, and D. Stephan, editors. Springer Handbook of Enzymes, Vol. 25. Eds. Dietmar Schomburg and Ida Schomburg. Berlin: Springer Berlin Heidelberg, 2006. 194–210.
26. Englander, S. W., D. B. Calhoun, and J. J. Englander. 1987. Biochemistry without Oxygen. *Anal Biochem* 161:300–306.
27. Rasnik, I., S. A. McKinney, and T. Ha. 2006. Nonblinking and longlasting single-molecule fluorescence imaging. *Nat Methods* 3:891–893.
28. Widengren, J., A. Chmyrov, C. Eggeling, P. A. Lofdahl, and C. A. Seidel. 2007. Strategies to improve photostabilities in ultrasensitive fluorescence spectroscopy. *J Phys Chem A* 111:429–440.
29. van Dijk, M. A., L. C. Kapitein, J. van Mameren, C. F. Schmidt, and E. J. G. Peterman. 2004. Combining optical trapping and single-molecule fluorescence spectroscopy: enhanced photobleaching of fluorophores. *J Phys Chem B* 108:6479–6484.
30. Gaigalas, A. K., L. Wang, K. D. Cole, and E. Humphries. 2004. Photodegradation of fluorescein in solutions containing n-propyl gallate. *J Phys Chem A* 108:4378–4384.
31. Dittrich, P. S., and P. Schwill. 2001. Photobleaching and stabilization of fluorophores used for single-molecule analysis with one- and two-photon excitation. *Appl Phys B – Lasers Optics* 73:829–837.
32. Berrios, M., K. A. Conlon, and D. E. Colflesh. 1999. Antifading agents for confocal fluorescence microscopy. *Methods Enzymol* 307:55–79.

33. Florijn, R. J., J. Slats, H. J. Tanke, and A. K. Raap. 1995. Analysis of antifading reagents for fluorescence microscopy. *Cytometry* 19:177–182.
34. Longin, A., C. Souchier, M. Ffrench, and P. A. Bryon. 1993. Comparison of anti-fading agents used in fluorescence microscopy: image analysis and laser confocal microscopy study. *J Histochem Cytochem* 41:1833–1840.
35. Von Trebra, R., and T. H. Koch. 1982. DABCO stabilization of coumarin dye lasers. *Chemical Physics Letters* 93:315–317.
36. Johnson, G. D., R. S. Davidson, K. C. McNamee, G. Russell, D. Goodwin, and E. J. Holborow. 1982. Fading of immunofluorescence during microscopy: a study of the phenomenon and its remedy. *J Immunol Methods* 55:231–242.
37. Brown, C. K., M. W. Vetting, C. A. Earhart, and D. H. Ohlendorf. 2004. Biophysical analyses of designed and selected mutants of protocatechuate 3,4-dioxygenase. *Annu Rev Microbiol* 58:555–585.
38. Bull, C., and D. P. Ballou. 1981. Purification and Properties of Protocatechuate 3,4-Dioxygenase from *Pseudomonas-Putida* – a New Iron to Subunit Stoichiometry. *J Biol Chem* 256:2673–2680.
39. Ballou, D. P., C. Bull, and T. Walsh. 1981. Nonheme iron intradiol catechol dioxygenases – structural and mechanistic studies. *Fed Proc* 40:1554–1554.
40. Patil, P. V., and D. P. Ballou. 2000. The use of protocatechuate dioxygenase for maintaining anaerobic conditions in biochemical experiments. *Anal Biochem* 286:187–192.
41. Ahmad, M., M. D. Rahn, and T. A. King. 1999. Singlet oxygen and dye-triplet-state quenching in solid-state dye lasers consisting of Pyrromethene 567-doped poly(methyl methacrylate). *Appl Optics* 38:6337–6342.
42. Giloh, H., and J. W. Sedat. 1982. Fluorescence microscopy – reduced photobleaching of rhodamine and fluorescein protein conjugates by normal-propyl gallate. *Science* 217:1252–1255.
43. Getz, E. B., M. Xiao, T. Chakrabarty, R. Cooke, and P. R. Selvin. 1999. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem* 273:73–80.
44. Burns, J. a., J. C. Butler, J. Moran, and G. M. Whitesides. 1991. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *J Org Chem* 56:2648–2650.
45. Jocelyn, P. C. 1987. Chemical-reduction of disulfides. *Methods Enzymol* 143:246–256.
46. Cleland, W. W. 1964. Dithiothreitol new protective reagent for Sh groups. *Biochemistry* 3:480–482.
47. Mickey, B., and J. Howard. 1995. Rigidity of microtubules is increased by stabilizing agents. *J Cell Biol* 130:909–917.
48. Vandegriff, K. D., R. J. Rohlf, M. D. Magde, and R. M. Winslow. 1998. Hemoglobin-oxygen equilibrium curves measured during enzymatic oxygen consumption. *Anal Biochem* 256:107–116.
49. Jovanovic, S. V., S. Steenzen, M. Tomic, B. Marjanovic, and M. G. Simic. 1994. Flavonoids as antioxidants. *J Am Chem Soc* 116:4846–4851.
50. Huang, S. C., G. C. Yen, L. W. Chang, W. J. Yen, and P. D. Duh. 2003. Identification of an antioxidant, ethyl protocatechuate, in peanut seed testa. *J Agric Food Chem* 51:2380–2383.
51. Buschmann, V., K. D. Weston, and M. Sauer. 2003. Spectroscopic study and evaluation of red-absorbing fluorescent dyes. *Bioconjugate Chem* 14:195–204.
52. Eggeling, C., J. Widengren, L. Brand, J. Schaffer, S. Felekyan, and C. A. M. Seidel. 2006. Analysis of photobleaching in single-molecule multicolor excitation and forster resonance energy transfer measurement. *J Phys Chem A* 110:2979–2995.
53. Ha, T., and J. Xu. 2003. Photodestruction intermediates probed by an adjacent reporter molecule. *Phys Rev Lett* 90:223002.
54. Heilemann, M., E. Margeat, R. Kasper, M. Sauer, and P. Tinnefeld. 2005. Carbocyanine dyes as efficient reversible single-molecule optical switch. *J Am Chem Soc* 127:3801–3806.
55. Huang, Z. X., D. M. Ji, S. F. Wang, A. D. Xia, F. Koberling, M. Patting, and R. Erdmann. 2006. Spectral identification of specific photophysics of Cy5 by means of ensemble and single molecule measurements. *J Phys Chem A* 110:45–50.

56. Huang, Z. X., D. M. Ji, A. D. Xia, F. Koberling, M. Patting, and R. Erdmann. 2005. Direct observation of delayed fluorescence from a remarkable back-isomerization in Cy5. *J Am Chem Soc* 127:8064–8066.
57. Widengren, J., and P. Schwille. 2000. Characterization of photoinduced isomerization and back-isomerization of the cyanine dye Cy5 by fluorescence correlation spectroscopy. *J Phys Chem A* 104:6416–6428.
58. Ying, L. M., H. T. Li, J. Green, J. Huppert, S. Balasubramanian, and D. Klenerman. 2003. Single-molecule screening of red dyes. *Biophys J* 84:472A.
59. Ha, T., I. Rasnik, W. Cheng, H. P. Babcock, G. H. Gauss, T. M. Lohman, and S. Chu. 2002. Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* 419:638–641.

THE EVALUATION OF ISOTOPE EDITING AND FILTERING FOR PROTEIN–LIGAND INTERACTION ELUCIDATION BY NMR

IAN M. ROBERTSON, LEO SPYRACOPOULOS, AND BRIAN D. SYKES*

Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

**Corresponding author: e-mail: Brian.Sykes@ualberta.ca*

Abstract A series of experiments that aid in the structural characterization of protein–ligand complexes by NMR have been assessed. Methods have been established to identify intermolecular NOEs between labeled proteins and unlabeled peptides and/or drug ligands, while omitting signal from intramolecular NOEs within both labeled and unlabeled constituents. The protein–peptide complex chosen to illustrate the value of such techniques is the C-terminal domain of the cardiac muscle regulatory protein Troponin C (cCTnC_{91–161}) bound to the N-terminal domain of the inhibitory protein Troponin I (cTnI_{35–72}). The measurement of intermolecular NOE contacts between cCTnC and cTnI_{35–72} was accomplished by the ¹³C-edited/filtered NOESY-HSQC [19] and ¹³C-edited/filtered HMQC-NOESY [9] experiments. The assignment of the bound peptide was facilitated by the ¹³C, ¹⁵N-filtered TOCSY, and ¹³C, ¹⁵N-filtered NOESY experiments [4, 6, 15].

Introduction

The determination of structures of protein–ligand complexes is an important role of structural biology, but is also one of the most experimentally challenging aspects of solution NMR. Of particular interest is the structural assessment of unlabeled ligands such as peptides and/or drugs associated with uniformly labeled proteins by isotope filtering [16]. Characterizing the binding interface can be difficult, especially when trying to elucidate intermolecular contacts between an unlabeled ligand and a labeled protein. Several methods have been designed in which the signals from the labeled protein are filtered, leaving only signals from the unlabeled ligand to be detected, or the reverse. The wide variation of ¹H–¹³C scalar coupling constants (120–220 Hz) make isotope-based NMR spectral purging experiments difficult to design.

In this review, isotope filtering strategies will be introduced, experimentally tested, and compared.

The ^{13}C -edited/filtered HMQC-NOESY [9] and ^{13}C -edited/filtered NOESY-HSQC [19] are variations of edited/filtered NOESY schemes that isolate NOEs between a labeled protein and an unlabeled ligand, while filtering out intramolecular NOE contacts either within the labeled protein or unlabeled ligand. The other series of experiments that will be described, the two-dimensional (2D) $^{13}\text{C},^{15}\text{N}$ -filtered TOCSY and 2D $^{13}\text{C},^{15}\text{N}$ -filtered NOESY [4, 6, 15] display spectra of the unlabeled ligand with no intramolecular or intermolecular NOEs from the labeled protein. The ^{13}C -edited/filtered HMQC-NOESY, and two 2D $^{13}\text{C},^{15}\text{N}$ -filtered experiments use a similar approach to filter signals from the labeled protein. The experiments use nominal one-bond ^1J -couplings between a labeled isotope (^{13}C , or ^{15}N) and a ^1H to tune delay periods. The tuned delays generate multiple quantum coherences (MQCs) via the ^1J -coupling Hamiltonian (\hat{H}), which are unobservable and also purged using pulsed field gradients (PFGs). The ^{13}C -edited/filtered NOESY-HSQC uses an elegant approach; delays are used that allow the one-bond ^1H - ^{13}C coupling \hat{H} to mimic the effect of a broad-band composite 90° pulse, such that a wide range of $^1\text{J}_{\text{HC}}$ values can be filtered.

The two edited/filtered experiments; the ^{13}C -edited/filtered HMQC-NOESY and ^{13}C -edited/filtered NOESY-HSQC, are examples of experiments that attain similar results via different routes. The ^{13}C -edited/filtered HMQC-NOESY first selects for ^1H resonance signals coupled to ^{13}C (F_1 -*edited*) and then after the mixing period filters out the same ^1H - ^{13}C signals (F_3 -*filtered*) detection of the unlabeled ligand ^1H signals occurs during t_3 . In the ^{13}C -edited/filtered NOESY-HSQC the opposite path of magnetization transfer is followed; first the ^1H - ^{12}C nuclei is selected for (F_1 -*filtered*) and following the NOE transfer the ^1H - ^{13}C is selected (F_3 -*edited*). In the ^{13}C -edited/filtered NOESY-HSQC, the final signal detected during acquisition is the ^1H - ^{13}C coupled nuclei. Figure 1 presents a schematic explanation of the two different experiments described above.

Other methods have been developed to improve the efficiency of the edited/filtered experiments. Zwahlen et al. [22] and Kupče et al. [8] used the approximate correlation between $^1\text{J}_{\text{HC}}$ and ^{13}C chemical shift to design an isotope-filter. An adiabatic pulse is applied, such that the frequency is swept according to the ^{13}C chemical shift range, and the signal from the labeled protein is inverted according to $^1\text{J}_{\text{HC}}$. More recently, Valentine et al. [21] developed a new procedure in which they combined the composite-rotation and adiabatic-sweeping techniques to improve signal filtering and the removal of artifacts. In practice we found that the simpler pulse sequences can be easily implemented on modern NMR spectrometers and effectively filter protein signals.

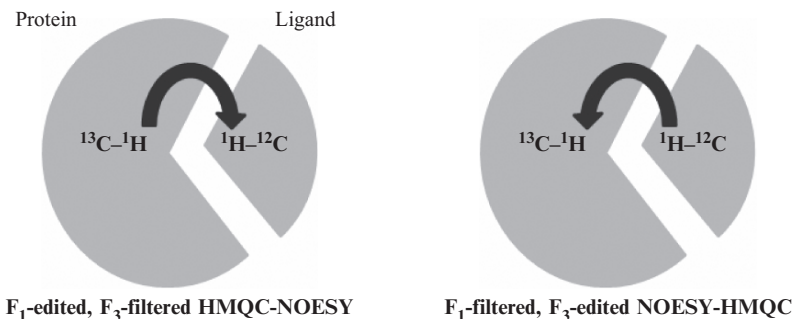


Fig. 1. An introduction into the edited/filtered experiment nomenclature. The F_1 -edited, F_3 -filtered HMQC-NOESY is an experiment that first detects the signals from a proton attached on the ^{13}C nuclei, and then after the NOE transfer, filters out the ^{13}C attached protons prior to detection [9]. F_1 -filtered, F_3 -edited NOESY-HSQC initially suppresses the signal from an attached proton to the heteronuclei, and following the NOE mixing period detects the ^1H - ^{13}C signal [19].

The biological system chosen to examine the efficiency of the isotope edited/filtered experiments is the cardiac contractile network. The C-terminal domain of the contractile regulatory protein Troponin C (cTnC) and the N-terminal region of the inhibitory protein Troponin I (cTnI₃₅₋₇₂) associate to form a tight complex [11]. Contraction in cardiac muscle is stimulated by the binding of Ca^{2+} to the N-terminal domain of cTnC (cNTnC), subsequently cNTnC opens and binds the C-terminal region of cTnI (otherwise associated with actin) [12]. The interaction between cTnC and cTnI₃₅₋₇₂ is maintained during the contraction and relaxation cycle of muscle. The complex has a dissociation constant of $\sim 1 \mu\text{M}$ [11], making it ideal for studying the efficacy of the edited/filtered experiments. The previously determined structure of cTnC-cTnI₄₀₋₇₁ [20] will also help gauge the efficiency of the filter, and accuracy of the NOE assignments. Of course, the experiments described here may be successfully applied to observe contacts made between a labeled protein and other unlabeled ligands, such as a drug. The object of this report is to use the cTnC-cTnI₃₅₋₇₂ complex as a model to compare the advantages of the two edited/filtered experiments, as well as describe a technique for generating a structure of a protein associated with a ligand.

Experimental Procedures

The expression vector for cardiac Troponin C₉₁₋₁₆₁ (cTnC) was designed [2] and uniformly labeled ^{13}C , ^{15}N cTnC was isolated from *E. coli* as previously described [10]. The unlabeled peptide cTnI₃₅₋₇₂, Ac-AKKKSKISASRKLQLKT-LLLQIAKQELEREAEERRGEK-NH₂ was synthetically prepared by GL

Biochem Ltd. All NMR experiments described in the article were run at 30 °C on Varian Inova 600 MHz and 800 MHz spectrometers. Details for the particular experiments are given in the corresponding figure legends. Figure 2 shows the ^{13}C F_1 -edited, F_3 -filtered HMQC-NOESY, and the ^{13}C F_1 -filtered, F_3 -edited NOESY-HSQC as well as the ^{13}C , ^{15}N -filtered NOESY. Detailed discussions of the pulse schemes may be found in references [4, 6, 9, 15, 19]. All acquired spectra were processed with NMRPipe [3], and analyzed with NMRView [7]. The previously determined structures used to illustrate the topics presented in this paper are the Ca^{2+} saturated chicken cCTnC [18] (pdb: 3CTN) and the troponin complex core structure [20] (pdb: 1J1E).

are indicative of 180° pulses. All the pulse phases are along x , unless otherwise indicated. For simplicity, the phase cycling from the pulse sequences has been omitted. The delay τ_m represents the NOE mixing time present in all three of the experiments. (a) The ^{13}C -edited/filtered NOESY-HSQC pulse scheme as presented by Stuart et al. [19]. The unfilled pulses represent composite 180° pulses that function to invert over the wide range of ^{13}C resonances. The pulse sequence also purges ^{15}N coupled protons, and not show here is the ^{13}CO decoupling applied during t_2 . The delays selected for the selection of the coupling constant $J_{\text{HC}} = 140 \text{ Hz}$ are $\tau = 3.57 \text{ ms}$, $\tau_a = 5.355 \text{ ms} + 0.5\tau_{c0} - n\Delta\tau$, $\tau_b = 0.5(3.57 \text{ ms} - \tau_{c0}) + n\Delta\tau$, $\tau_c = \tau_{c0} + n(1/\text{SW1} - 2\Delta\tau)$, $\Delta\tau = (5.425 \text{ ms} - \tau_{a0})/(N - 1)$, $\tau_{a0} = 0.9 \text{ ms}$, $\tau_{c0} = 0.65 \text{ ms}$, $\tau_d = 1.785 \text{ ms}$ and $\tau_e = 1.785 \text{ ms}$. In order to optimize the filter efficiency for $^1J_{\text{HN}}$, the delays should be set to $\tau_a - \tau_b + 2\zeta = 1/(2J_{\text{HN}})$. For more detailed descriptions on the selection of the different delays refer to Stuart et al. [19]. The decoupling of the ^{13}C nuclei during acquisition is accomplished by the GARP method [17]. The gradients applied are $g1 = (1.5 \text{ ms}, 17 \text{ G/cm})$; $g2 = (450 \mu\text{s}, 10.4 \text{ G/cm})$; $g3 = (450 \mu\text{s}, 24 \text{ G/cm})$; $g4 = (450 \mu\text{s}, -13.8 \text{ G/cm})$; $g5 = (450 \mu\text{s}, 27.6 \text{ G/cm})$; $g6 = (1 \text{ ms}, 7 \text{ G/cm})$; $g7 = (500 \mu\text{s}, 7 \text{ G/cm})$; $g8 = (500 \mu\text{s}, 10.4 \text{ G/cm})$; $g9 = (1.2 \text{ ms}, -17.6 \text{ G/cm})$; $g10 = (300 \mu\text{s}, 20.7 \text{ G/cm})$. (b) The ^{13}C -edited/filtered HMQC-NOESY pulse scheme. This pulse scheme has been simplified from the original [9] including the removal of the ^{15}N purging and ^{13}CO decoupling. This simplified version is optimized for methyl–methyl NOEs, and is best run in D_2O . The delay $\tau_a = 3.57 \text{ ms}$ during the HMQC transfer was chosen to select for aliphatic ^{13}C coupled protons ($J_{\text{HC}} = 140 \text{ Hz}$). The delays in the filtering period after the NOE mixing time; $\tau_b = 4.17 \text{ ms}$, $\tau_c = 3.33 \text{ ms}$ and $\tau_d = 0.84 \text{ ms}$ were chosen to cover a range of aliphatic ^{13}C coupled ^1H signals (120–150 Hz). The gradient strengths and lengths are: $g1 = (0.1 \text{ ms}, 10 \text{ G/cm})$; $g2 = (1 \text{ ms}, 5 \text{ G/cm})$; $g3 = (0.4 \text{ ms}, 8 \text{ G/cm})$. (c) The pulse sequence applied for observing NOEs within the unlabeled ligand, the ^{13}C , ^{15}N filtered NOESY experiment [4, 5, 15]. The pulse sequence was written by L.E. Kay 1996, and incorporated into the Varian BioPack pulse sequence suite. The delays chosen to filter the aliphatic and aromatic ^{13}C – ^1H , as well as the ^1H scalar coupled to ^{15}N were; $\tau_a = 4.0 \text{ ms}$, $\tau_b = 0.43 \text{ ms}$, $\tau_c = 1.07 \text{ ms}$, $\tau_d = 3.125 \text{ ms}$, $\tau_e = 0.625 \text{ ms}$, $\tau_f = 1.75 \text{ ms}$. The gradients lengths and strengths are; $g1 = (300 \mu\text{s}, 20.420 \text{ G/cm})$; $g2 = (1.0 \text{ ms}, 30.630 \text{ G/cm})$; $g3 = (500 \mu\text{s}, 4.084 \text{ G/cm})$; $g4 = (500 \mu\text{s}, 51.050 \text{ G/cm})$.

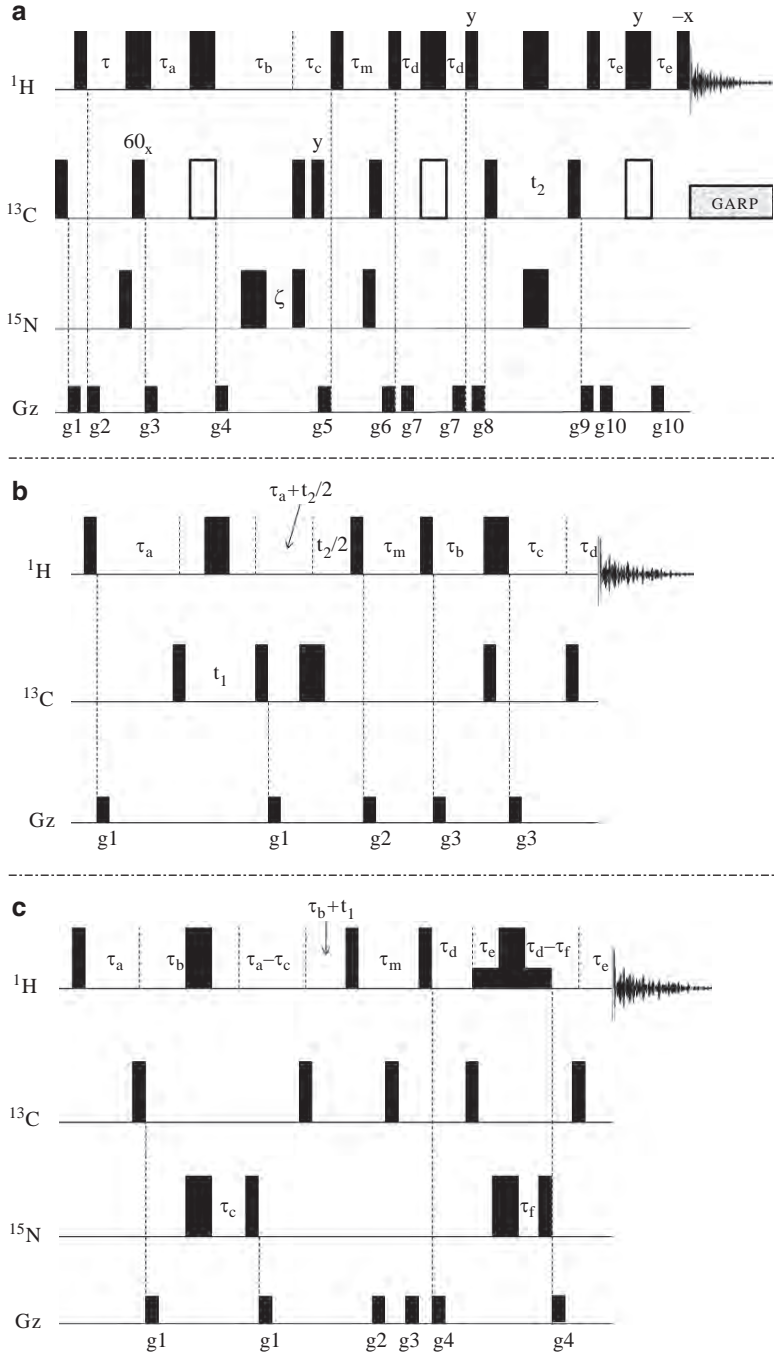


Fig. 2. Examples of pulse sequences that filter signals from isotropically labeled proteins. For all of the pulse sequences, the narrow rectangles represent 90° pulses, and the wide rectangles

Theory

The two ^{13}C edited/filtered experiments, the ^{13}C -edited/filtered HMQC-NOESY and the ^{13}C -edited/filtered NOESY-HSQC as well as the ^{13}C , ^{15}N -filtered NOESY are illustrated in Fig. 2. In the text below a brief description of each of the experiments is provided. The product operator analysis on the purging technique used in the ^{13}C -edited/filtered HMQC-NOESY is in the appendix.

^{13}C -EDITED/FILTERED NOESY-HSQC (F_1 -FILTERED, F_3 -EDITED)

In this experiment, the general pulse scheme (Fig. 2a) proceeds as follows. Initially a filter removes any signal originating from either ^1H - ^{13}C or ^1H - ^{15}N . The filter period uses a semi-constant time technique to frequency label a ^1H attached to ^{12}C . Following the initial filter and semi-CT time, a mixing period allows for the build-up of NOEs between unlabeled ligand and labeled protein.

After the NOE mixing period, ^1H magnetization is transferred to ^{13}C of the labeled protein by the standard INEPT (employing composite 180° ^{13}C pulses). The ^{13}C spins are frequency labeled and the magnetization is transferred back to the protons via a reverse INEPT for detection during acquisition (t_3). Since detection occurs on the labeled protein, the ^{13}C must be decoupled, therefore the length of the acquisition time/resolution in t_3 is limited.

The main concept behind the filter applied in the ^{13}C -edited/filtered NOESY-HSQC is that the sequence uses composite rotations generated by the ^1H - ^{13}C J -coupling \hat{H} to compensate for a range of $^1J_{\text{HC}}$ instead of relying on the precise selection of delays to filter signals from the labeled protein.

The efficiency (ϵ) of the filter is given by the following expression [1]:

$$\epsilon = \cos(\pi^1 J_{\text{HC}} \tau) \cos(2\pi^1 J_{\text{HC}} \tau) + 0.5 \sin(\pi^1 J_{\text{HC}} \tau) \sin(2\pi^1 J_{\text{HC}} \tau) = \cos^3(\pi^1 J_{\text{HC}} \tau)$$

where the filter is tuned to $^1J_0 = 1/(2\tau)$. The efficiency of the filter (ϵ) is equal to 1 when $^1J_{\text{HC}} = 0$ (uncoupled proton), and equal to 0 when $^1J_{\text{HC}} = ^1J_0$. The filter has a high filtering efficiency for a range of 50 Hz; so that for the selection of $^1J_0 = 140$ Hz, the filter exhibits good signal suppression from 115–165 Hz. For more details regarding the efficiency of the isotope filter or other aspects of the pulse sequence see Stuart et al. [19].

^{13}C -EDITED/FILTERED HMQC-NOESY (F_1 -EDITED, F_3 -FILTERED)

The ^{13}C -edited/filtered HMQC-NOESY represents a different method to detect intermolecular NOE contacts between a labeled protein and unlabeled ligand (refer to Fig. 2b for the pulse scheme). The magnetization is first transferred from the ^1H to the attached ^{13}C via an HMQC transfer, and is subsequently

labeled with the ^{13}C chemical shift during t_1 . The delay τ_a is set to $1/(2^1J_{\text{HC}})$, where $^1J_{\text{HC}} = 140\text{ Hz}$ for aliphatic ^1H - ^{13}C . Following t_1 , the magnetization is transferred back to ^1H and the chemical shifts are recorded during t_2 .

After the ^1H chemical shift labeling, the magnetization is rotated along the z-axis for NOE transfer wherein the ^1H - ^{13}C magnetization is transferred to the ^1H - ^{12}C . After the mixing time the magnetization is returned to the transverse plane, now labeled with ^{13}C attached ^1H (t_2) and ^{13}C (t_1) chemical shifts, as well as the ^1H - $^{13}\text{C} \leftrightarrow ^1\text{H}$ - ^{12}C NOE. The final stage prior to detection involved a purging period tuned to two different coupling constants (120 Hz, and 150 Hz). This double-tuned filter removes isotope labeled signals leaving behind only the uncoupled protons for detection (t_3).

The disadvantage of the double-tuned filter is that the coupling constants do not cover the entire range of $^1J_{\text{HC}}$ possibly present in a protein. In the pulse sequence put forth by Palmer and colleagues [19] discussed above, the selection of a single coupling constant provides filtration for a range of $^1J_{\text{HC}}$ values. The advantage of the ^{13}C edited/filtered HMQC-NOESY is that since no ^{13}C decoupling is present during t_3 , there is no limitation on the acquisition time (t_3) and therefore the resolution for the unlabeled ligand.

$^{13}\text{C}, ^{15}\text{N}$ FILTERED NOESY AND $^{13}\text{C}, ^{15}\text{N}$ FILTERED TOCSY

The pulse sequence employed for the detection of intramolecular NOE contacts within the unlabeled peptide is displayed in Fig. 2c. The magnetization transfer up until the mixing time is the exact same in the $^{13}\text{C}, ^{15}\text{N}$ filtered TOCSY and therefore theoretical discussion will focus on the filtering in the $^{13}\text{C}, ^{15}\text{N}$ filtered NOESY. The first ^1H 90_x pulse flips I_z into the transverse plane. If the delay τ_a is tuned correctly, after the evolution period ($\tau_a = 1/2^1J_{\text{HC}}$) the magnetization is in complete anti-phase with respect to the 1J -coupled ^{13}C . In the scenario where the delay is not tuned accurately, the magnetization will be a mixture of anti-phase and in-phase operators. The $^1J_{\text{HN}}$ coupling will also generate proton nitrogen anti-phase/in-phase mixture that has evolved under the ^{15}N - ^1H scalar coupling for the duration τ_a .

The subsequent ^{13}C 90_x pulse produces the unobservable MQC for the properly tuned delay. The gradient will de-phase all transverse magnetization. The next delay, $\tau_b = \tau_a - 1/2^1J_{\text{HC}2}$ is tuned to a slightly different $^1J_{\text{HC}}$, so that if τ_a was inaccurately tuned, by the second ^{13}C 90_x pulse the in-phase magnetization not purged previously, will have evolved into anti-phase magnetization. Following the ^{13}C 90_x pulse, the magnetization is converted into MQC; which is not detected during the acquisition (t_3). After the delay τ_c ($\tau_c = 1/2^1J_{\text{HN}} - \tau_a - \tau_b$), the ^{15}N one-bond coupled protons will have evolved into anti-phase magnetization. So the applied 90_x pulse on the ^{15}N will yield MQC, which is subsequently de-phased by a gradient.

In order to refocus the proton chemical shift, time periods $\tau_a - \tau_c$ and τ_b are followed. The delay $\tau_a - \tau_c$ is also necessary to convert the magnetization not completely removed in the first purging period (τ_a) to anti-phase prior to the second carbon 90_x pulse. Finally, just preceding the mixing period, ^1H (on ^{12}C nuclei) chemical shift is recorded during t_1 . After the mixing time the signal is in the transverse plane, and there is another filtering period (in the $^{13}\text{C}, ^{15}\text{N}$ filtered NOESY, but not in the $^{13}\text{C}, ^{15}\text{N}$ filtered TOCSY). In this period, larger coupling constants are chosen for the delays in order to filter signals originating from aromatic couplings, $^1J_{\text{HCl}}$ and $^1J_{\text{HC2}} \sim 160$ and 200 Hz . In this fashion, one can obtain an artifact-free “finger-print” spectrum with purging of aliphatic signals in the F_1 dimension, and purging of amide and aromatic protein signals in the F_2 dimension. In the $^{13}\text{C}, ^{15}\text{N}$ filtered TOCSY pulse sequence (pulse scheme not shown), the $^1\text{H}-^{13}\text{C}$ magnetization is initially filtered as in the $^{13}\text{C}, ^{15}\text{N}$ filtered NOESY, and then the $^1\text{H}-^{12}\text{C}$ signal undergoes a DIPSI mixing [1] in order to observe $^1\text{H}-^1\text{H}$ correlations from intraresidue ^1H spin systems.

Results and Discussion

The structure of the calcium saturated cCTnC bound with the N-terminal region of cTnI (cTnI₄₀₋₇₁) is displayed in Fig. 3a [20]. The tight association and known structure of this complex makes this system a perfect model to study the efficacy of the various filtering techniques. An alignment of the X-ray crystal structure of cCTnC bound to cTnI (cTnI₄₀₋₇₁) with the solution structure of the chicken cCTnC [18] (Fig. 3b) demonstrates the high degree of similarity between the two structures (RMSD = 1.5 \AA for the backbone C α). The minimal structural change between the two proteins highlights the significance of these different techniques in solving protein structure. If the structural change of a protein is small when binding a ligand, then the elucidation of the protein–ligand contacts in the binding interface represents a crucial aspect in determining an accurate complex structure. Another important principle in protein–ligand complexes is exemplified in Fig. 4. The interface between the cCTnC–cTnI₄₀₋₇₁ contains almost exclusively hydrophobic methyl–methyl interactions. The prevalence of aliphatic ^1H in the protein–peptide interface allows for the selection of small sweep widths, and the use of a double-tuned filter (as is in the ^{13}C -edited/filtered HMQC–NOESY) is sufficient to cover the expected range of aliphatic $^1J_{\text{HC}}$ -couplings ($120\text{--}150\text{ Hz}$).

Before discussing the edited/filtered experiments we will briefly review the $^{13}\text{C}, ^{15}\text{N}$ filtered experiments. Figure 5 shows the filtered (a) and unfiltered NOESY (b) spectra. The signal removed by employing the simple approach

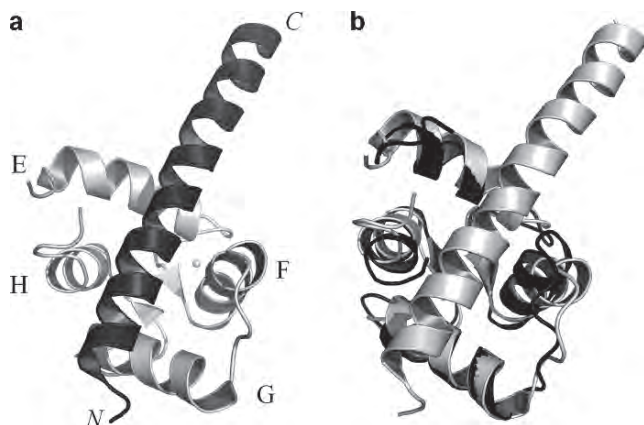


Fig. 3. A ribbon diagram of the C domain of the human cardiac Troponin C in complex with cardiac Troponin I (40–71) (pdb: 1J1E). **(a)** In light grey is the calcium saturated cCTnC, and in black is cTnI_{40–71}. The C domain diagram is taken from the X-ray crystal structure of the core domain of the troponin complex (cTnC-cTnI-cTnT) [20]. The helices are labeled on the cCTnC, and the N and C termini are labeled on the cTnI_{40–71}. **(b)** The cCTnC-cTnI_{40–71} complex from **(a)** is colored completely in light grey. The structure is aligned with the chicken C domain of cTnC (in black) described by Sia et al. [18] (pdb: 3CTN) in the absence of cTnI_{40–71}. The only sequence difference between the two domains is in amino acid 93 (S93 in cCTnC-cTnI_{40–71} and T93 in cCTnC).

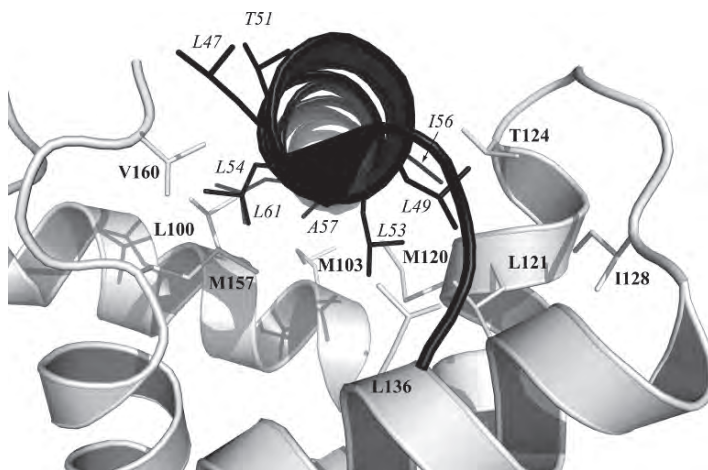


Fig. 4. A close up view of the ribbon diagram in Fig. 3a, rotated by 90° along the horizontal axis to emphasize the cCTnC-cTnI_{35–72} interface. The color scheme employed in Fig. 3a is adopted here. An interesting aspect of the cCTnC-cTnI_{40–71} interface is the predominance of methyl–methyl hydrophobic interactions. This feature of the protein–peptide interaction was considered when selecting the region to focus on for the ¹³C-edited/filtered experiments (see text).

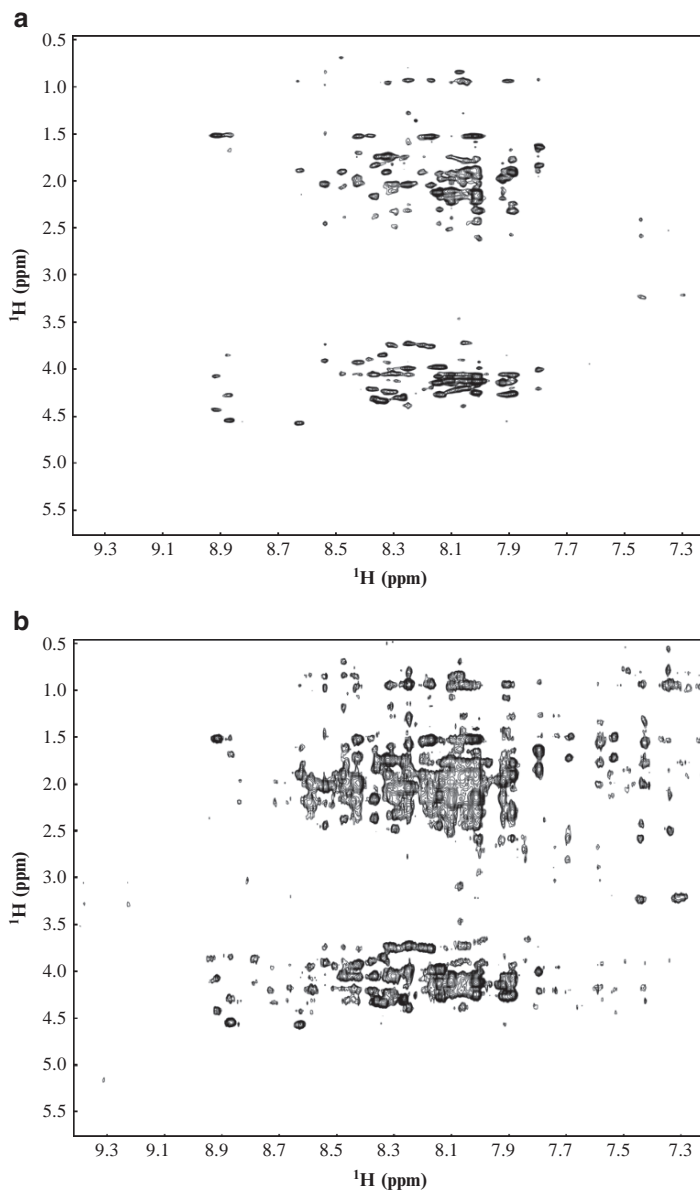


Fig. 5. The efficiency of the ^{13}C , ^{15}N filtered NOESY of the cTnC-cTnI₃₅₋₇₂. The pulse scheme is presented in Fig. 2c. The different $^1\text{J}_{\text{HC}}$ chosen for filtering out were 125 Hz, 140 Hz (both selected for filtering of aliphatic protons), 160 Hz and 200 Hz (for removal of ^{13}C aromatic protons). (a) A two-dimensional spectrum of the ^{13}C , ^{15}N filtered NOESY, with the filtering turned on, and in (b) the filtering was turned off. The number of points acquired in t_3 was 4096 and the number of increments was 256. The total number of transients was 64. (c) The filtering efficiency is visualized by the overlay of one-dimensional spectra acquired with 4096 points in detected dimension and 128 transients. In light grey is the unfiltered NOESY, ^{13}C and ^{15}N pulses have been turned to 0 dB, and in black is the ^{13}C , ^{15}N filtered NOESY. Notice the loss of signal within the amide region (inset in (c)), and the $\text{H}_{\text{N}}\text{--}\text{H}_{\text{sidechain}}$ region of the ^{13}C , ^{15}N filtered NOESY spectrum (a).

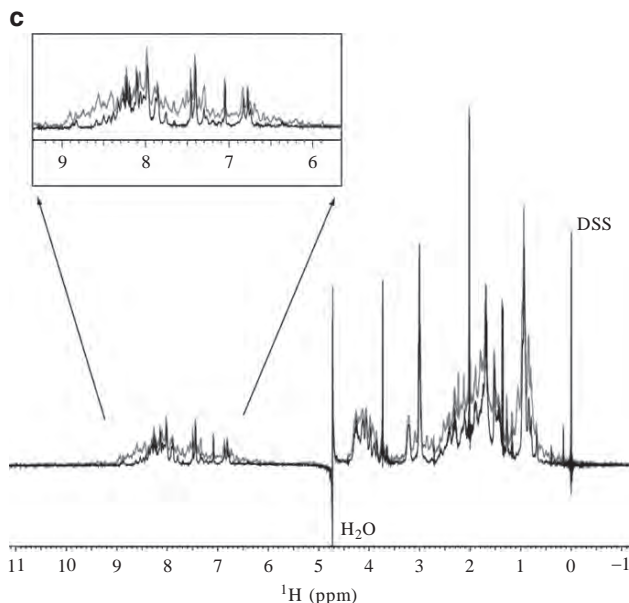


Fig. 5. (continued)

described in the theory section is evident. The selection of four different delays for four different coupling constants (125 Hz, 140 Hz, 160 Hz, and 200 Hz), and pulsed field gradients allows for exquisite filtering in the 'finger print' region. In the ^{13}C , ^{15}N filtered TOCSY the aromatic purging period was not included, but still displayed good filtering (data not shown). In Fig. 6 the H_N – H_α region of the ^{13}C , ^{15}N filtered NOESY and ^{13}C , ^{15}N filtered TOCSY are overlaid and the sequential assignment strategy for the bound unlabeled peptide is illustrated. For a smaller ligand like a drug the assignment of the resonances would be simpler and the use of these filtered experiments may not be required if a one-dimensional spectrum can be unambiguously assigned.

The two edited/filtered experiments chosen for comparison of NOE transfer between unlabeled ligands and labeled proteins are the classical approach [4, 6, 9, 15, 16], which utilizes the evolution of the heteronuclear scalar coupling \hat{H} to filter out signals from the isotropically labeled proteins, and the more recent approach discussed in Stuart et al. [19], which uses composite pulses to rotate the operators without relying on a specific ^1H – ^{13}C coupling constant. See the theory section for a detailed description of the filtering strategies employed by the different methods. In Fig. 7, an expanded view of the two-dimensional spectrum from the ^{13}C -edited/filtered NOESY-HSQC is shown. The diagonal represents the signal that was not filtered during acquisition. Although not shown, the ^{13}C -edited/filtered HMQC-NOESY

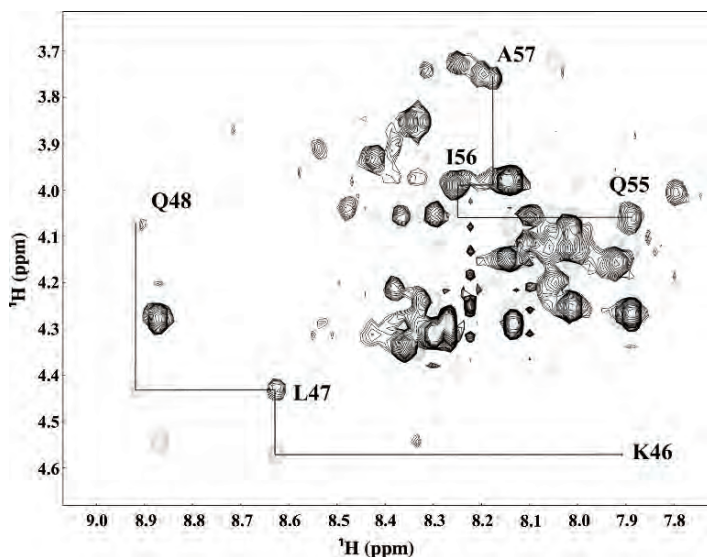


Fig. 6. The $^{13}\text{C},^{15}\text{N}$ filtered NOESY (light grey) and $^{13}\text{C},^{15}\text{N}$ filtered TOCSY (black) overlapped to illustrate the assignment of the unlabeled peptide when bound to the labeled protein. Both experiments had 32 transients, 512 increments and 4096 points acquired. The mixing times in the above spectra were, 80 ms for the $^{13}\text{C},^{15}\text{N}$ filtered TOCSY, and 100 ms for the $^{13}\text{C},^{15}\text{N}$ filtered NOESY. The region viewed here designates the amide protons coupled to the intraresidue $\text{C}\alpha$, and the interresidue $\text{C}\alpha$. Several small chains are labeled at the TOCSY peaks (intraresidue $\text{H}_\text{N} - \text{H}_\alpha$).

reveals an almost identical spectrum verifying the labeled peaks in Fig. 7 are NOEs. The labeled protein methyls were unambiguously assigned by the use of a 3D- ^{13}C -edited/filtered HMQC-NOESY experiment and the conventional triple resonance experiments, such as the 3D-CBCACONNH [14], 3D-HNCACB [14], 3D-CCONH [5], and 3D-HCCONH [13]. Care must be taken, however when attempting to assign peaks along the diagonal. In Fig. 7b the view from Fig. 4 is rotated by 180° about the vertical axis, so that now we are looking down the C terminus of the cTnI₄₀₋₇₁ peptide. The figure confirms the assigned contact in Fig. 7a between the L100 terminal methyls and the β -methyl of A57 of the peptide. Not shown in Fig. 7a is the assigned contact also present between the H_ϵ of M120 and the H_α of A57 of cTnI₄₀₋₇₁ which is illustrated in Fig. 7b.

In order to address the potential advantages of the two filtering techniques, identical experiments were set up (see Fig. 8). The same number of increments and transients for each experiment were acquired. The only difference between the two experimental parameters lied within the number of points in the acquired dimension (t_3). This is due to the fact that in the

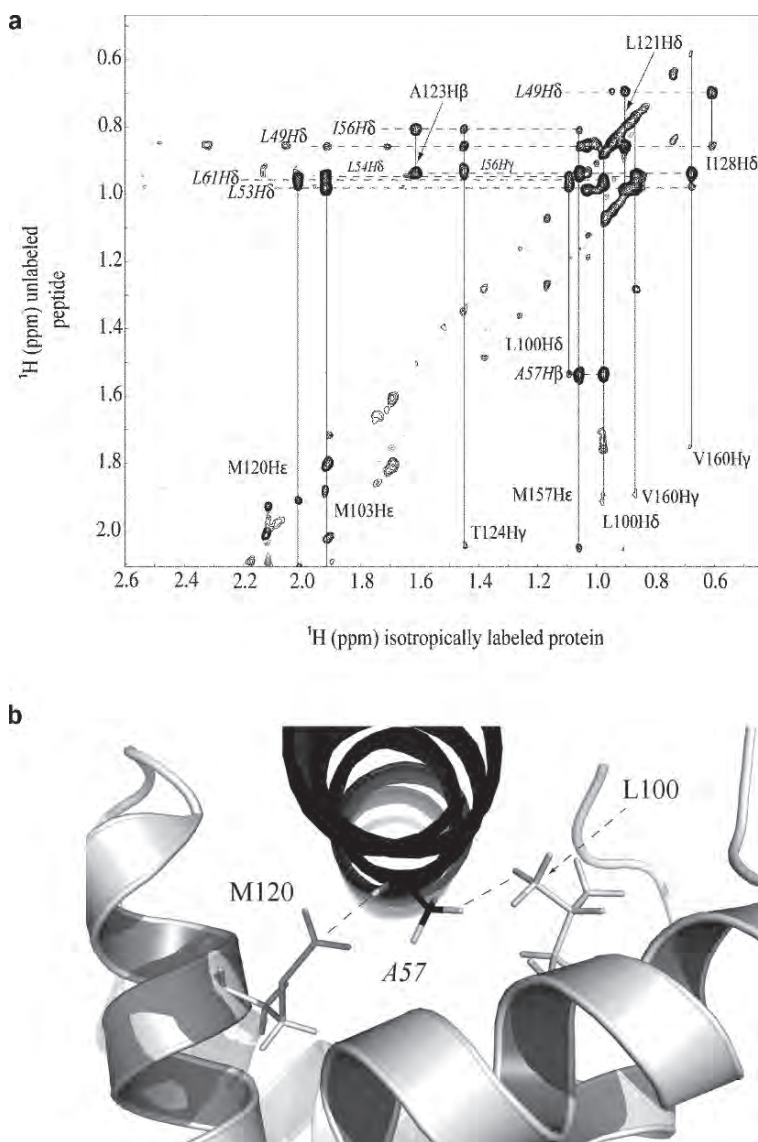


Fig. 7. (a) The assigned spectrum acquired from the ^{13}C -edited/filtered NOESY-HSQC experiment. The assigned peaks represent intermolecular NOEs between the labeled cCTnC and the unlabeled cTnI₃₅₋₇₂. Vertical lines represent peaks from the labeled protein, while the horizontal dotted lines are from the peptide. Peak labels shown in italics originate from the peptide. The number of transients taken was 64, 256 increments during F_1 and 392 points during the acquisition. The NOE mixing time was set to 150 ms. The sweep widths were 5.0 ppm in the F_1 and 5.0 ppm in F_2 . **(b)** The coloring scheme is the same as in Fig. 3a. The peptide is in black, while the labeled protein is in grey. The NOE contacts between the H α and H β of A57 from the peptide and the H ϵ of M120 and the terminal H δ s of L100 from the peptide are assigned in (a) and then identified in (b) using the structure from Takeda et al. (pdb: 1J1E) [20].

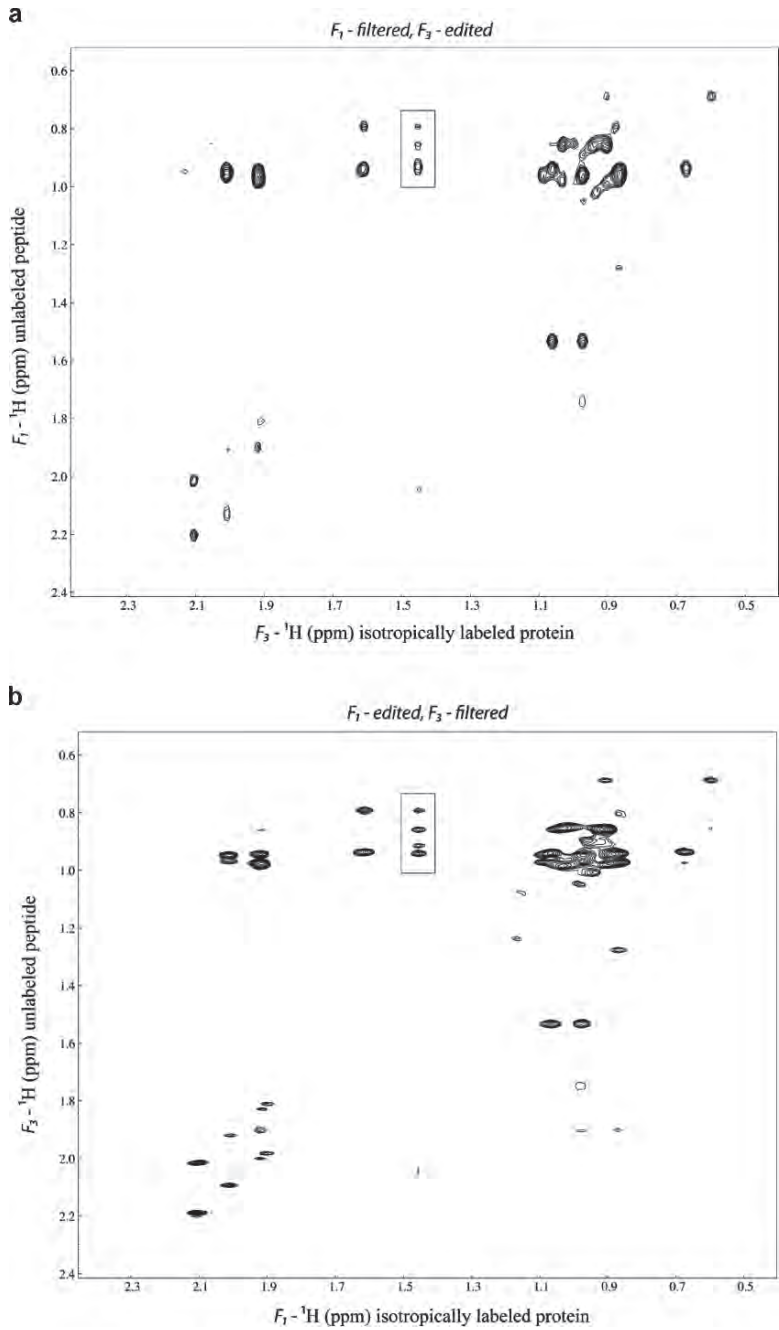


Fig. 8. A comparison of the two ¹³C-edited/filtered experiments. Both experiments were acquired at 30°C on a Varian Inova 800 MHz with identical sweep widths (5.0 ppm × 5.0 ppm). The carrier frequency was set to 2.42 ppm, and in the three-dimensional experiment (not shown) the carbon carrier was 20.0 ppm with a sweep width of 20 ppm. The mixing time for both

^{13}C -edited/filtered NOESY-HSQC experiment decoupling along the carbon channel is required during acquisition. In the ^{13}C -edited/filtered HMQC-NOESY no decoupling during detection permits the number of points acquired to depend solely on the resolution desired. The increase in resolution in the F_3 dimension (unlabeled ligand) is evident in the ^{13}C -edited/filtered HMQC-NOESY spectrum (Fig. 8b). There is an improvement on the labeled protein resolution of the ^{13}C -edited/filtered NOESY-HSQC when compared with the ^{13}C -edited/filtered HMQC-NOESY experiment. The improvement is because, even though there is decoupling during t_3 in the ^{13}C -edited/filtered NOESY-HSQC experiment, the time restriction prevents the experimentalist to acquire as many increments in the ^{13}C -edited/filtered HMQC-NOESY during F_1 .

The filtering efficiency is not immediately clear when comparing the two spectra in Fig. 8. Since there is a high proportion of methyl–methyl contacts in the interface between the cTnC and cTnI_{35–72} (see Fig. 4), the sweep widths and carrier frequencies selected focus on the aliphatic region. Thus the increased efficiency of the filter as discussed in Stuart et al. [19], is of less value in the overall performance of the experiments. Instead, other considerations such as resolution and acquisition time become more relevant when selecting the appropriate pulse sequence.

Conclusion

In the present review, an overview of several of the different methods applied to filter signals from a uniformly labeled protein, while maintaining the signal from an unlabeled bound ligand is provided. The large range in the one-bond ^1J -couplings between ^1H and ^{13}C and ^{13}C chemical shift makes the efficient purging of ^{13}C , ^{15}N isotropically labeled protein signals challenging. More recent approaches have been described [8, 21, 22] which attempt to improve on the filtering efficiency of the approaches described here. In many situations

← experiments was set to 200 ms. (a) ^{13}C -edited/filtered NOESY-HSQC experiment with 392 points in the F_3 dimension (labeled protein), 128 increments in the F_1 (unlabeled peptide) dimension, the 32 transients. (b) ^{13}C -edited/filtered HMQC-NOESY experiment also acquired 32 transients with 128 increments in the F_1 dimension. The large difference between the two experiments is 1,024 points were acquired in the F_3 dimension, rather than the 392 points in the ^{13}C -edited/filtered NOESY-HSQC. The ^{13}C -edited/filtered NOESY-HSQC provides slightly higher resolution in the labeled protein dimension, but greatly diminished resolution in the unlabeled ligand dimension when compared to the ^{13}C -edited/filtered HMQC-NOESY (see the boxed in area in the above figure).

involving protein–ligand interactions, however, the binding interface is predominated with methyl–methyl contacts. Given this aspect, often the selection of small sweep widths and specific coupling constants that focus on the relevant interactions render a wide filtering range less crucial. The system used as a model to describe the filtering techniques was the cCTnC-cTnI_{35–72} complex; however, the utility of purging isotropically labeled protein signals can also provide rapid identification of the binding location of unlabeled drugs to their target proteins. The knowledge of the binding location between a drug and its target protein may greatly enhance the design of new more potent drugs.

Acknowledgements We thank David Corson for the preparation of isotropically labeled cCTnC, Monica Li for helpful discussions regarding the troponin complex, Jeff DeVries for his assistance in the operation and upkeep of the spectrometers, and Ryan Hoffman and Olivier Julien for help with the programs NMRView and NMRPipe. Finally we thank Ryan McKay at the Canadian National Nuclear Magnetic Resonance Facility (NANUC) for support and helpful discussion. This study was supported by the Canadian Institutes for Health Research, the National Heart, Lung and Blood Institute of the NIH and the Heart and Stroke Foundation of Canada.

Appendix

The purging period employed in the ¹³C-edited/filtered HMQC-NOESY experiment (Fig. 2b) is a common filtering method and will be discussed in detail using the product operator formalism. The product operator representing the ¹H magnetization after the mixing period (τ_m) is I_z and following the 90° ¹H pulse it becomes transverse magnetization (I_y). During this final filtering section of the pulse sequence any signal from ¹H one-bond ¹J-coupled to ¹³C will be purged. In principle, ¹H ¹J-coupled to ¹⁵N could also be removed by making minor revisions to the pulse sequence; however, for simplicity these will be omitted from the discussion.

The pulse sequence was designed so that the two delays, τ_b and τ_c are tuned to different ¹J_{HC} couplings. The two values selected were 120 Hz and 150 Hz to cover the range of aliphatic ¹H–¹³C ¹J-couplings. The assumption made was that most protein–protein interactions involve methyl–methyl contacts, thus we are not concerned with protons coupled to aromatic residues (see *results and discussion*).

We will examine three separate scenarios during the purging period. The situation for an incorrectly tuned delay is not discussed, but would result in incomplete signal suppression.

- (a) The delay τ_b is tuned correctly for ¹J_{HC}
- (b) The delay τ_b is incorrectly tuned for ¹J_{HC}, but τ_c is tuned correctly for ¹J_{HC}

(c) ^1H is not ^1J -coupled to any ^{13}C

Since the ^1H 180_x pulse is flanked by equal delays (τ_b , and $\tau_c + \tau_d$), where τ_d is a delay after the final ^{13}C 90_x pulse tuned to completely refocus the proton chemical shift ($\tau_d = \tau_b - \tau_c$). The gradients (g4) surrounding the ^1H 180_x pulse function to clean up artifacts, as well as dephase multiple quantum coherences.

$$(a) \tau_b = 1/(2J_{\text{HC}}): -I_y \xrightarrow{2\pi J_{\text{HC}} I_z S_z \tau_b} -[\cos(\pi J_{\text{HC}} \tau_b) I_y - \sin(\pi J_{\text{HC}} \tau_b) 2I_x S_z]$$

Given $\tau_b = 1/(2J_{\text{HC}})$, the operator prior to the ^{13}C 90_x pulse is the anti-phase term:

$$2I_x S_z$$

Since the S_z commutes with the ^{13}C chemical shift Hamiltonian (\hat{H}), it is not considered. Given the ^1H 180_x pulse refocuses the ^1H chemical shift, it will also be neglected. After the ^{13}C 90_x pulse the term generated is the multiple quantum operator.

$$-2I_x S_z \longrightarrow 2I_x S_y$$

The MQC term is unobservable and also subsequently de-phased by the g4 pulse, and the following time period (τ_c) is of no significance, given all the signal has been eliminated.

(b) $\tau_b \neq 1/(2J_{\text{HC}})$; $\tau_c = 1/(2J_{\text{HC}})$:

$$-I_y \xrightarrow{2\pi J_{\text{HC}} I_z S_z \tau_b} -[\cos(\pi J_{\text{HC}} \tau_b) I_y - \sin(\pi J_{\text{HC}} \tau_b) 2I_x S_z]$$

The next ^{13}C 90_x and ^1H 180_x pulses will produce both in-phase ^1H magnetization and MQC.

$$-\cos(\pi J_{\text{HC}} \tau_b) I_y + \sin(\pi J_{\text{HC}} \tau_b) 2I_x S_z \xrightarrow{90_x, 180_y} +\cos(\pi J_{\text{HC}} \tau_b) I_y - \sin(\pi J_{\text{HC}} \tau_b) 2I_x S_y$$

The gradient pulse (g4) will de-phase the multiple quantum operators, but re-phase the in-phase ^1H operators. If we follow the in-phase term,

$$+\cos(\pi J_{\text{HC}} \tau_b) I_y \xrightarrow{2\pi J_{\text{HC}} I_z S_z \tau_c} +\cos(\pi J_{\text{HC}} \tau_b) [\cos(\pi J_{\text{HC}} \tau_c) I_y - \sin(\pi J_{\text{HC}} \tau_c) 2I_x S_z]$$

Given $\tau_c = 1/(2J_{\text{HC}})$ the signal just prior to the ^{13}C 90_x pulse is:

$$-\cos(\pi J_{\text{HC}} \tau_b) 2I_x S_z$$

As in the previous example, the ^{13}C 90_x pulse will generate MQC:

$$\cos(\pi J_{\text{HC}} \tau_b) 2I_x S_y$$

During the remaining delay ($\tau_d = \tau_b - \tau_c$) and the detection period the multiple quantum operators are not observable.

(c) The signal is not coupled to ^{13}C :

In the final example, where the signal did not originate from a proton on a ^{13}C the case is rather simple. The ^1H 180°_x pulse will refocus the chemical shift during the purging period, as well as invert the sign of the ^1H operator. The final in-phase ^1H magnetization is detected during acquisition (t_3).

$$-I_y \xrightarrow{180^\circ_x} +I_y$$

References

1. Cavanagh, J., and Rance, M. (1992) Suppression of Cross-Relaxation Effects in Tocsy Spectra Via a Modified Dipsi-2 Mixing Sequence, *Journal of Magnetic Resonance* 96, 670–678.
2. Chandra, M., Dong, W. J., Pan, B. S., Cheung, H. C., and Solaro, R. J. (1997) Effects of Protein Kinase A Phosphorylation on Signaling Between Cardiac Troponin I and the N-terminal Domain of Cardiac Troponin C, *Biochemistry* 36, 13305–13311.
3. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRpipe – a Multidimensional Spectral Processing System Based on Unix Pipes, *Journal of Biomolecular NMR* 6, 277–293.
4. Gemmecker, G., Olejniczak, E. T., and Fesik, S. W. (1992) An Improved Method for Selectively Observing Protons Attached to C-12 in the Presence of H-1-C-13 Spin Pairs, *Journal of Magnetic Resonance* 96, 199–204.
5. Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in $^{13}\text{C}/^{15}\text{N}$ -Enriched Proteins by Isotropic Mixing of ^{13}C Magnetization, *Journal of Magnetic Resonance Series B* 101, 114–119.
6. Ikura, M., and Bax, A. (1992) Isotope-Filtered 2d NMR of a Protein Peptide Complex – Study of a Skeletal-Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin, *Journal of the American Chemical Society* 114, 2433–2440.
7. Johnson, B. A., and Blevins, R. A. (1994) NMR View – a Computer-Program for the Visualization and Analysis of NMR Data, *Journal of Biomolecular NMR* 4, 603–614.
8. Kupče, E., and Freeman, R. (1997) Compensation for Spin-Spin Coupling Effects During Adiabatic Pulses, *Journal of Magnetic Resonance* 127, 36–48.
9. Lee, W., Revington, M. J., Arrowsmith, C., and Kay, L. E. (1994) A Pulsed Field Gradient Isotope-Filtered 3D ^{13}C HMQC-NOESY Experiment for Extracting Intermolecular NOE Contacts in Molecular Complexes, *FEBS letters* 350, 87–90.
10. Li, M. X., Gagné, S. M., Tsuda, S., Kay, C. M., Smillie, L. B., and Sykes, B. D. (1995) Calcium-Binding to the Regulatory N-Domain of Skeletal-Muscle Troponin-C Occurs in a Stepwise Manner, *Biochemistry* 34, 8330–8340.
11. Li, M. X., Wang, X., Lindhout, D. A., Buscemi, N., Van Eyk, J. E., and Sykes, B. D. (2003) Phosphorylation and Mutation of Human Cardiac Troponin I Differentially Destabilize the Interaction of the Functional Regions Of Troponin I with Troponin C, *Biochemistry* 42, 14460–14468.
12. Li, M. X., Wang, X., and Sykes, B. D. (2004) Structural Based Insights into the Role of Troponin in Cardiac Muscle Pathophysiology, *Journal of Muscle Research and Cell Motility* 25, 559–579.
13. Lyons, B. A., and Montelione, G. T. (1993) An Hcnh Triple-Resonance Experiment Using ^{13}C Isotropic Mixing for Correlating Backbone Amide and Side-Chain Aliphatic

- Resonances in Isotopically Enriched Proteins, *Journal of Magnetic Resonance Series B* 101, 206–209.
14. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-Enhanced Triple-Resonance 3-Dimensional NMR Experiments with Improved Sensitivity, *Journal of Magnetic Resonance Series B* 103, 203–216.
 15. Ogura, K., Terasawa, H., and Inagaki, F. (1996) An Improved Double-Tuned and Isotope-Filtered Pulse Scheme Based on a Pulsed Field Gradient and a Wide-Band Inversion Shaped Pulse, *Journal of Biomolecular NMR* 8, 492–498.
 16. Otting, G., and Wüthrich, K. (1990) Heteronuclear Filters in Two-Dimensional [1H,1H]-NMR Spectroscopy: Combined Use with Isotope Labelling for Studies of Macromolecular Conformation and Intermolecular Interactions, *Quarterly Reviews of Biophysics* 23, 39–96.
 17. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) Computer-Optimized Decoupling Scheme for Wideband Applications and Low-Level Operation, *Journal of Magnetic Resonance* 64, 547–552.
 18. Sia, S. K., Li, M. X., Spyropoulos, L., Gagné, S. M., Liu, W., Putkey, J. A., and Sykes, B. D. (1997) Structure of Cardiac Muscle Troponin C Unexpectedly Reveals a Closed Regulatory Domain, *The Journal of Biological Chemistry* 272, 18216–18221.
 19. Stuart, A. C., Borzilleri, K. A., Withka, J. M., and Palmer, A. G. (1999) Compensating for Variations in H-1-C-13 Scalar Coupling Constants in Isotope-Filtered NMR Experiments, *Journal of the American Chemical Society* 121, 5346–5347.
 20. Takeda, S., Yamashita, A., Maeda, K., and Maeda, Y. (2003) Structure of the Core Domain Of Human Cardiac Troponin in the Ca(2+)-Saturated Form, *Nature* 424, 35–41.
 21. Valentine, E. R., Ferrage, F., Massi, F., Cowburn, D., and Palmer, A. G., 3rd. (2007) Joint Composite-Rotation Adiabatic-Sweep Isotope Filtration, *Journal of Biomolecular NMR* 38, 11–22.
 22. Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage Lambda N-Peptide/Box B RNA Complex, *Journal of the American Chemical Society* 119, 6711–6721.

RIBOSOME: AN ANCIENT CELLULAR NANO-MACHINE FOR GENETIC CODE TRANSLATION

ADA YONATH

*The Weizmann Institute Of Science, Department Of Structural
Biology, Helen and Milton A. Kimmelman Building, Rehovot,
Israel 76100*

E-mail: ada.yonath@weizmann.ac.il

Abstract The ribosome is a ribozyme whose active site, the peptidyl transferase center (PTC), is situated within a highly conserved universal symmetrical region that connects all ribosomal functional centers involved in amino-acid polymerization. The linkage between this elaborate architecture and A-site tRNA position revealed that the A- > P-site passage of the tRNA terminus in the peptidyl-transferase center is performed by a rotatory motion, synchronized with the overall tRNA/mRNA sideways movement. Guided by the PTC the rotatory motion leads to stereochemistry suitable for peptide bond formation as well as for substrate mediated catalysis, consistent with quantum mechanical calculations illuminating the transition state mechanism for peptide bond formation and indicating that the peptide bond is being formed during the rotatory motion.

Analysis of substrate binding modes to inactive and active ribosomes illuminated the significance of PTC mobility and supported the hypothesis that the ancient ribosome produced single peptide bonds and non-coded chains, utilizing nucleotide conjugated amino acids. Genetic control of the reaction evolved after polypeptides capable of enzymatic function were created, and an ancient stable RNA fold was converted into tRNA molecules. As the symmetry relates only the backbone fold and nucleotides orientations, but not nucleotide sequence, it emphasizes the superiority of functional requirement over sequence conservation, and indicates that the PTC has evolved by gene fusion, presumably by taking advantage of similar RNA fold structures.

The increase in antibiotic resistance among pathogenic bacterial strains poses a significant health threat. Therefore, improvement of existing antibiotics and the design of advance drugs are urgently needed. Ribosomes provide binding sites for many antibiotic families, utilizing their inherent functional flexibility, which triggers induced fit mechanism by remote interactions, and facilitates antibiotics synergism as well as reshaping less suitable binding pockets,

leading to clinical usefulness even for antibiotics that bind to conserved functional regions. Exploitation of the diverse properties of antibiotics binding and benefiting from the detailed structural information that keeps emerging, should result in significant antibiotics improvement.

Introduction

Ribosomes, the universal cellular riboprotein assemblies, are the nano-machines which translate the genetic code into proteins. The translation process requires a complex apparatus composed of many components, among them the ribosome is the key player, as it provides the framework for the proper positioning of all other participants and is actively involved in the translation process. Ribosomes operate in each living cell continuously since the constant programmed cell death, which implies constant proteins degradation, requires simultaneous production of proteins. For example, hundreds of thousands of ribosomes are present in typical mammalian cells. Fast replicating cells, e.g. liver cells, may contain a few millions ribosomes. Even bacterial cells may contain to 100,000 ribosomes during their log period.

Within the framework of living cells, ribosomes are giant assemblies, composed of many different proteins (r-proteins) and long ribosomal RNA (rRNA) chains. The ratio of rRNA to r-proteins (~2:1) is maintained throughout evolution, with the exception of mammalian mitochondrial ribosome (mitoribosome) in which almost half of the bacterial rRNA is replaced by r-proteins, and consequently in mitoribosome the ratio of RNA to proteins is ~1/1. All ribosomes are constituted by two unequal subunits (Table 1). In prokaryotes, the small subunit, denoted as 30S, contains an RNA chain (16S) of about 1500 nucleotides and 20–21 different proteins, whereas the large subunit (called 50S in prokaryotes) has two RNA chains (23S and 5S RNA) of about 3,000 nucleotides in total, and different 31–35 proteins. In all organisms the two subunits exist independently and associate to form functionally active ribosomes.

The process of mRNA-encoded protein synthesis requires a complex apparatus composed of the ribosome, transfer RNA molecules (tRNA) and accessory protein factors. The mRNA chains are produced by the transcription of the segments of the DNA that should be translated. The mRNA chains carry the genetic information to the ribosomes, and tRNA molecules bring the cognate amino acids to the ribosome. For increasing efficiency, a large number of ribosomes act simultaneously as polymerases synthesizing proteins by one-at-a-time addition of amino acids to a growing peptide chain, while translocating along the mRNA template, producing proteins on

a continuous basis in an incredible speed (namely >20 new peptide bonds per second). While the elongation of the nascent chain proceeds, the two subunits perform cooperatively. The small subunit provides the path along which the mRNA progresses, the decoding center and the mechanism controlling translation fidelity, and the large subunit contains the site for the main ribosomal

TABLE 1. Ribosome composition.

| |
|---|
| <i>Prokaryotic ribosome: Sedimentation coefficient: 70S</i> |
| ◦ Small subunit: 30S |
| ▪ One rRNA molecule (16S with 1500 nucleotides) |
| ▪ ~21 different proteins, called S1–S21 |
| ◦ Large subunit: 50S |
| ▪ Two rRNA molecules (5S and 23S, with 120 and 2,900 nucleotides, respectively) |
| ▪ ~31 different proteins, called L1–L31, among which only L12 is present in more than a single copy |
| <i>Eukaryotic ribosomes: Sedimentation coefficient: 80S</i> |
| ◦ Small subunit: 40S |
| ▪ One rRNA molecule (18S with 1,900 nucleotides) |
| ▪ ~33 different proteins, called S1–S33 |
| ◦ Large subunit: 60S |
| ▪ Three rRNA molecules (5S, 5.8S and 28S, with 120, 156 and 4,700 nucleotides, respectively) |
| ▪ ~50 different proteins, called L1–L50 |

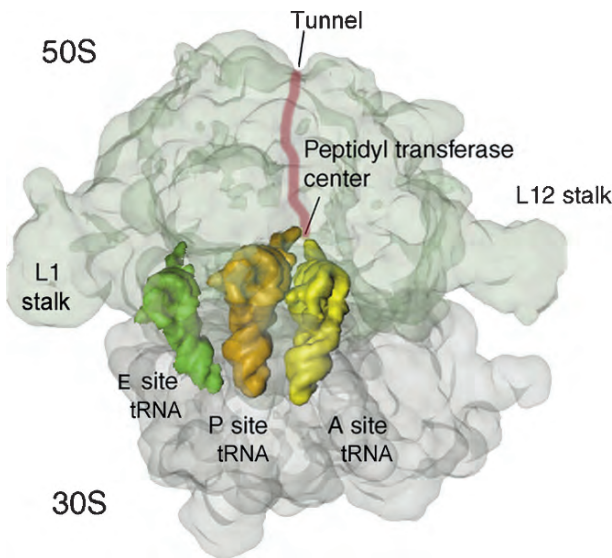


Fig. 1. Structure and functional sites of bacterial ribosomes.

catalytic function, polymerization of the amino acids and provides the protein exit tunnel (Fig. 1).

The three-dimensional structures and secondary diagrams of all tRNA molecules from all living cells across evolution are alike, although each of them is specific to its amino acid. They are all built mainly of double helical L-shape molecules, made by a stem–elbow–stem organization, and contain an anticodon loop that matches its complementary three-nucleotide codes on the mRNA on one of their edges. About 70 Å away, at their 3' end, the tRNA molecules contain a single strand with the universal sequence CCA, to which the cognate amino acid is bound by an ester bond. The tRNA molecules are the non-ribosomal entities combining the two subunits, as each of their three binding sites, A-(aminoacyl), P-(peptidyl), and (exit), resides on both subunits. At the A- and P-sites the tRNA anticodon loops interact with the mRNA on the small subunit, and the acceptor stem with the aminoacylated or peptidylated 3' end are located on the large subunit. However, so far the involvement of the existing tRNA in codon-anticodon interactions remains unclear. Peptide bond is being formed during A- to P-site tRNA translocation, comprised of sideways shift and a ribosomal navigated rotatory motion. Nascent proteins progress along a dynamic tunnel and emerge into a shelter formed by ribosome-bound trigger-factor, acting as a chaperone preventing aggregation and misfolding.

The recent availability of over two dozens of crystal structures of bacterial ribosome and their complexes (see below), enabled a quantum jump in the understanding of the machinery of protein biosynthesis. These structures showed that in each of the two subunits the ribosomal proteins are entangled within the complex rRNA conformation, thus maintaining a striking dynamic architecture that is ingeniously designed for their functions: precise decoding; substrate mediated peptide-bond formation and efficient polymerase activity. This chapter focuses on the structural and dynamic properties of the ribosome that facilitate its function as an efficient nano-machine, addresses issues relating to the ribosome origin and attempt to answer the question: “Can structures lead to improved antibiotics?”

Historical Comments

Owing to the vast unpredictable expansion of studies on the various aspects of ribosome function and their significant clinical impact, time has come to shed light on selected points in the history of ribosome structural research. The involvement of RNA rich particles in genetic expression was suggested over 5 decades ago, when the so-called ‘Palade particles’ were located within RNA rich regions, in close association with the membrane of the endoplasmic

reticulum (Palade, 1955; Watson, 1963). The localization of the cellular translation site and the extensive biochemical, biophysical and genetic research that followed, led to illuminating findings about the overall nature of ribosome function. Nevertheless, because the structure of the ribosome was not available, detailed functional information could not be obtained. Moreover, owing to the huge size and the complexity of the ribosome, it was widely assumed that it would never be crystallized, and even if it was its structure would be extremely hard to be elucidated. Consequently, a series of hypothesis was accumulated, based on common wisdom combined with results of biochemical experiments. For example, originally, it was assumed that the decoding of the genetic information as well as the enzymatic activity of the ribosome is performed mainly (or solely) by the ribosomal proteins and that the r-RNA provides the scaffold holding the numerous ribosomal proteins in the stereochemistry allowing for performing their tasks (Garrett and Wittmann, 1973a, b).

Challenging the assumption that the r-proteins catalyze peptide bond formation by a series of biochemical experiments (Noller et al., 1992) was first met with some skepticism. However, as the major role played by RNA in various life processes (not only protein biosynthesis) became evident at approximately the same time, the notion that the main actor in ribosome function is rRNA was accepted within a relatively short period. In contrast, the latter example, namely that the nascent proteins progress on the ribosome surface until its maturation, was well implanted in researcher's mind. Hence, even after biochemical experiments indicated clearly protection of nascent chains by the ribosome (Malkin and Rich, 1967; Sabatini and Blobel, 1970) and three dimensional image reconstructions of electron microscopical data obtained from ordered array of ribosomes visualized the tunnel (Milligan and Unwin, 1986) or their large subunits (Yonath et al., 1987) doubt was publicly expressed (Moore, 1988; Ryabova et al., 1988). The notion that the ribosome indeed contains a long tunnel along which the nascent proteins progress on their way out became more acceptable only after verification by cryo electron microscopy (Frank et al., 1995; Stark et al., 1995), which was performed almost a decade after the tunnel first visualization (Milligan and Unwin 1986; Yonath et al., 1987) and 3 decades after the initial suggestion. Remarkably, even once a tunnel with dimensions matching accurately those predicted in the 1960s, namely about 100 Å in length (Malkin and Rich, 1967; Sabatini and Blobel, 1970) and observed experimentally and detected in a high resolution crystal structure, it was assigned a Teflon-like passive character, with no dynamic and/or chemical properties that allow its interactions with the progressing nascent chains (Ban et al., 2000; Nissen et al., 2000). This idea was conflicting previous observation (e.g. Crowley et al., 1993; Walter and Johnson, 1994; Nagano et al., 1991), and hence was soon challenged by various experimental and computational methods, including biochemistry,

electron microscopy and crystallography, as can be seen in the following partial ensemble of papers (e.g. Gabashvili et al., 2001; Nakatogawa and Ito, 2002; Gong and Yanofsky, 2002; Berisio et al., 2003a; Woolhead et al., 2004; Gilbert et al., 2004; Johnson, 2004; Ziv et al., 2005; Amit et al., 2005; Baram and Yonath, 2005; Mankin, 2006; Deane et al., 2007; Petrone et al., 2008; Mitra et al., 2006; Voss et al., 2006; Schaffitzel and Ban, 2007).

Twenty years passed from the first indications for potential organization of ribosomes by the growth of microcrystals diffracting to relatively high resolution, namely 3.5 Å (Yonath et al., 1980) to the vast expansion in the availability of ribosome structures. The shift from poorly diffracting microcrystals to high-resolution structures was achieved gradually, based on the presumption that the higher the sample conformational homogeneity the better the crystals, and that the preferred conformation is that of functionally active ribosomes. Assuming that the ribosomes of bacteria that grow under robust conditions are less sensitive to external conditions, we focused on such sources, and, indeed, the first three dimensional microcrystals were obtained from the large ribosomal subunits from *Bacillus stearothermophilus* (Yonath et al., 1980), a source considered to be almost an extremophile at the beginning of the 1980s. This hypothesis was verified thoroughwort, as by extensive systematic explorations for suitable bacterial sources it was found that the key for obtaining crystals suitable for crystallographic studies is to use ribosomes from relatively robust bacteria (Gluehmann et al., 2001) or ribosomes mutated in a fashion that minimizes their mobility, so that they can be trapped at a specific, albeit not necessarily functional, conformation (Schuwirth et al., 2005). An alternative strategy is to crystallize complexes of ribosomes with substrates, inhibitors and/or factors that can trap them at preferred orientations. Among such complexes are the initial crystals of the whole ribosome from *Thermus thermophilus* (T70S) with mRNA and tRNA molecules (Hansen et al., 1990) and of the large subunit from *Deinococcus radiodurans* (D50S) with the antibiotic clindamycin (Schlunzen et al., 2001).

Efforts aimed at the improvement of existing crystals included a through examination of the influence of the relative concentrations of mono- and di-valent ions (von Bohlen et al., 1991), on the growth and properties of the crystals from the large ribosomal subunits form *Haloarcula marismortui* led to dramatic improvements in the quality of these crystals. Furthermore, a constant refinements of bacterial growth pathways and its linkage to crystal growth (Auerbach-Nevo et al., 2005) alongside a thorough investigation on crystallization conditions in relation to those required for ribosome functional activity, indicated a noteworthy correlation between the conditions under which these ribosomes function and the quality of the resulting crystals. Remarkably, reasonable conformational rearrangements were observed for

flexible functional regions in ribosome crystals grown under conditions mimicking their physiological environment (Harms et al., 2001), whereas in ribosome crystals obtained under far from their physiological environment such regions may be highly disordered (Ban et al., 2000).

Alongside the improvement of the ribosome crystals, ribosome crystallography required the development of innovative methodologies. Examples are the bio-crystallography at cryogenic temperatures, which was introduced because of the extreme radiation sensitivity of the ribosomal crystals (Hope et al., 1989), and an unconventional use of multi-heavy atom cluster, which was found to play a dual role in the determination of the structure of the small ribosomal subunit from *Thermus thermophilus* (T30S). Thus, post crystallization treatment with these clusters, originally used for providing anomalous phasing power (Thygesen et al., 1996) increased dramatically the resolution of the X-ray diffraction from the initial 7–9 to 3 Å (Schlunzen et al., 2000) presumably by minimizing the internal flexibility involved naturally in mRNA binding to the ribosome and its progression through the ribosome (See below and in Bashan and Yonath, 2008a; Zimmerman and Yonath, 2009).

The currently available high resolution crystallographic structures of ribosomal particles and their complexes with substrate analogs, factors, inhibitors and antibiotics that target the ribosome (Ban et al., 2000; Nissen et al., 2000; Schlunzen et al., 2000, 2001, 2003, 2004, 2005; Wimberly et al., 2000; Carter et al., 2000; Brodersen et al., 2000; Pioletti et al., 2001; Carter et al., 2001; Yusupov et al., 2001; Harms et al., 2001, 2004; Clemons et al., 2001; Ogle et al., 2001, 2002; Brodersen et al., 2002; Hansen et al., 2002a, b, 2003; Bashan et al., 2003; Berisio et al., 2003a, b; Murphy et al., 2004; Murphy and Ramakrishnan, 2004; Schuwirth et al., 2005; Ferbitz et al., 2004; Petry et al., 2005; Yusupova et al., 2006; Korostelev et al., 2006; Dunham et al., 2007; Selmer et al., 2006; Wilson et al., 2005a, b; Baram et al., 2005; Baram and Yonath, 2005; Schmeing et al., 2005a, b; Davidovich et al., 2007, 2008; Pyetan et al., 2007; Weixlbaumer et al., 2007, 2008; Blaha et al., 2008; Ippolito et al., 2008; Tu et al., 2005; Schroeder et al., 2007; Evans et al., 2008; Simonovic and Steitz, 2008; Auerbach et al., 2009).

The crystal structures were soon accompanied by impressive advances in cryo electron microscopy (cryo EM) of ribosomal particles, which revealed additional elements of ribosomes functional dynamics (e.g. Agrawal et al., 1999; Frank et al., 2000; Stark et al., 2000; Gao et al., 2003; 2005; Valle et al., 2003a, b; Passmore et al., 2007) and shed more light on many issues concerning ribosomal mode of function. These include the initiation of the translation, which is among the most complex and the most divergent, tRNA selection, decoding, translational fidelity, peptide bond formation, polymerase activity, the energetic of the elongation cycle, translation termination etc. These, in turn, led to numerous studies, documented

in large number of original publication as well in many summarizing reviews, some of which are cited below (Thompson et al., 2001; Barta et al., 2001; Xiong et al., 2001; Bayfield et al., 2001; Polacek et al., 2001, 2003; Ramakrishnan and Moore, 2001; Ramakrishnan and Moore, 2002; Harms et al., 2002; Katunin et al., 2002; Ramakrishnan, 2002; Moore and Steitz, 2002, 2003, 2005; Yonath, 2002, 2003a, b, 2005a, b; 2006; Steitz and Moore, 2003; Rodnina and Wintermeyer, 2003; Beringer et al., 2003, 2005, 2007; Yonath and Bashan 2004; Thompson and Dahlberg, 2004; Auerbach et al., 2004; Zarivach et al., 2004; Agmon et al., 2005; Bashan and Yonath, 2005, 2008a, b; Ogle et al., 2003; Marintchev and Wagner, 2004; Nakatogawa and Ito, 2004; Fujiwara et al., 2004; Sievers et al., 2004; Cochella and Green, 2004; Blanchard et al., 2004; Gromadski and Rodnina, 2004; Weinger et al., 2004; Weingner and Strobel, 2006; Maier et al., 2005; Ogle and Ramakrishnan, 2005; Ogle and Ramakrishnan, 2005; Saguy et al., 2005; Nilsson and Nissen, 2005; Polacek and Mankin 2005; Sharma et al., 2005; Diaconu et al., 2005; Bieleng et al., 2006; Kaiser et al., 2006; Sato et al., 2006; Gindulyte et al., 2006; Trobro and Aqvist, 2006; Wohlgemuth et al., 2006; Brunelle et al., 2006, 2008; Rodnina et al., 2007; Anderson et al., 2007; Shaw and Green, 2007; Fabbretti, et al., 2007; Uemura et al., 2007; Konevega et al., 2007; Wekselman et al., 2008; Petry et al., 2008; Steitz, 2008; Lang et al., 2008; Johansson et al., 2008; Beringer and Rodnina, 2007; Youngman et al., 2006, 2007, 2008; Lang et al., 2008; Zimmerman and Yonath 2009).

Architectural and Dynamic Aspects of the Initiation Step

The initiation of protein biosynthesis is an incredibly intricate and a highly divergent process (Marintchev and Wagner, 2004) This dynamic step is triggered by initiation factors and matures in constructing the initiation complex composed of the small ribosomal subunit, mRNA and an initiator tRNA. The performance of the initiation complex hinges on accurate positioning of messenger RNA on the small ribosomal subunit, which is a prerequisite for correct translation. Specific ribosome architectural elements allows for accurate tRNA selection by the ribosome in addition to codon-anticodon pairing that leads to a closure of the domains of the small subunit around cognate tRNA, believed to trigger subsequent steps in tRNA recognition and selection (Ogle et al., 2003). Similarly, mRNA positioning is navigated and controlled by the elaborate ribosomal architectural-design, a purine-rich sequence made of a few (up to ten) nucleotides, called “Shine–Dalgarno sequence” in prokaryote. This sticky sequence anchors the mRNA 5'-end onto the small subunit by forming numerous interactions including base pairing (Yusupova et al., 2006) mRNA binding involves a latch-like closing-opening mechanism (Schluenzen et al., 2000; Schuwirth et al., 2005), performed within the small subunit by

coordinated motions, which seem to be minimized by the entrance of the mRNA to its groove on the small subunit.

As mentioned above and in (Schluenzen et al., 2000; Bashan and Yonath, 2008a) the motions involved in mRNA attachment and progression could also be minimized chemically by post crystallization treatment using minute amounts of the heteropolytungstate $[(\text{NH}_4)_6(\text{P}_2\text{W}_{18}\text{O}_{62})14\text{H}_2\text{O}]$, which led to a increased dramatically the resolution of the X-ray diffraction (from 7–9 to 3 Å). Analysis of the tungstenated structure T30S structure indicating firm and quantitative attachment of 14 clusters to well defined locations on each subunit, within crystals containing more than two different conformations of the small subunit. This binding was later found to stabilize the selected functional conformation, called “open latch” (Schluenzen et al., 2000) that enables the attachment of the mRNA chain to the small subunit. Thus, four cluster that bind to the r-protein (S2), which is exposed on the ribosome surface, across from the flexible region, by making numerous interactions with the positively charged side chains of its extremely flexible C- and N-positively charged termini, hampered the motions of the entire region in a manner that consequently influenced the rigidity of the subunit and controlled the internal flexibility of the region surrounding the mRNA chain. Local ribosome mobility plays a major role also the selection of the following incoming aminoacylated tRNAs (aa-tRNAs) and in proofreading in translation (Blanchard et al., 2004; Murphy et al., 2004; Murphy and Ramakrishnan, 2004; Cochella and Green, 2004), exploiting the inherent flexibility of the decoding center for monitoring base pairing at the first two positions of each codon, but tolerating non-canonical base pairs at the third position.

Once the initiator tRNA binds to the start codon (AUG) at the P-site, downstream of the sticky sequence that anchors and direct the mRNA binding, the construction of the initiation complex is completed. The two subunits associate and intersubunit bridges are being formed by their conformational rearrangements (Bashan and Yonath, 2008b; Harms et al., 2001; Selmer et al., 2006; Korostelev et al., 2006; Yusupov et al., 2001). Among the newly formed intersubunit bridges, B1a, which is located near the entrance region of A-site tRNA, hence called the A-finger, and bridge B2a, the connecting element between the environs of the ribosomal active site, the peptidyl transferase center (PTC), in the large subunit with the decoding center in the small subunit, play key roles in substrate accommodation and positioning. Bridge B2a (helix H69 of the large subunit) which is located at the heart of the ribosome, is a highly flexible multi-task feature. In the assembled ribosome, regardless of its source, T70S or E70S (namely the *E. coli* ribosome) and/or and of the functional state, it stretches out towards the small subunit and interacts with both the A-site and the P-site tRNA molecules (Schuwirth et al, 2005; Selmer et al., 2006; Korostelev et al., 2006; Yusupov et al., 2001) whereas in

the unbound large subunit of, D50S, it is positioned at the subunit interface with a distinctly different conformation (Harms et al., 2001). The significant flexibility of the bridges that allows their conformational rearrangements seems also to facilitate deactivation of ribosomes under unfavorable condition. Thus, these bridges can become readily disordered under conditions far from physiological, as observed in crystal structures of the large ribosomal subunit from *Haloarcula marismortui*, H50S (Ban et al., 2000).

It is likely that the ribosome benefits from the flexibility of bridge B2a (helix H69) are beyond its participation in intersubunit bridging. Helix H69 and its extension, Helix H70, originate near the peptidyl transferase center (PTC) in the large subunit, and reach the vicinity of the decoding site in the small subunit. Connecting the two ribosomal active sites, this bridge seems to be designed for transmitting signals between them. Its proximity to both the A- and the P-site tRNAs in the large subunit suggests that it may also participate in translocation by assisting the shift of the A-site tRNA acceptor stem into the P-site. Helix H69 is also one of the constituents of the PTC upper side making crucial contributions to the remote interactions that govern the precise positioning and accurate orientation of the tRNA substrates (Bashan et al., 2003).

Motions of bridge B1a, which appears to act also as a guide for eliminating the entrance of A-site tRNA to uncontrolled directions, are limited by a particular r-protein, called L25, TL5 and CTC in *E. coli*, *T. thermophilus* and *D. radiodurans*, respectively. This protein seems to have evolved according to environmental conditions. In the mesophile *E. coli*, the single domain ribosomal protein (r-protein) L25 is located on the solvent side of this bridge, in a position enabling preventing slippage of this bridge towards the solution. In the thermophile *T. thermophilus* one of the two domains of r-protein TL5 replaced L25 and its second domain interacts with B1a from its other solvent side, so that the two domains can hinder almost all uncontrolled motions of this bridge towards the ribosome periphery, thus providing additional stability at high temperature. In *D. radiodurans*, an extremely robust organism that can survive under extreme stress conditions, including starvation, irradiation, high as well as low temperatures, this protein (called CTC, after a general shock protein) replaces the *E. coli* r-proteins L25 and its homologue TL5 in T50S. Within the known members of the CTC protein-family, that from *D. radiodurans* is the longest. It contains 253 residues, about 150 residues longer than L25 and 60 residues longer than TL5 and has three domains. The CTC N-terminus is similar to the entire L25, the middle domain is similar to the C-terminal domain of TL5, and the third domain can reach the A-site and restricts the space available for the tRNA molecules. As it is built of three long alpha helices connected by a pointed slim end that can function as a flexible hinge, CTC can control or even eliminate A-site RNA binding under

unfavorable conditions, hence may be parts of the mechanisms that *D. radiodurans* developed for survival under extremely stressful conditions (Yonath, 2002).

In prokaryotes, three non-ribosomal protein factors are involved in initiation. Among them initiation factor 3 (IF3) binds to the small subunit in proximity to the mRNA progressing channel so that its C-terminal domain interacts extensively with the flexible termini of r-protein S18 that perturb into the solution. Interestingly, one of the heteropolytungstate molecules that bind to the small subunit stabilizes the conformation of the flexible termini of the r-protein S18 in a fashion mimicking this r-protein involvement in the binding of the C-terminal domain of initiation factor IF3 and competes with its binding (Pioletti et al., 2001; Yonath and Bashan, 2004). In this position the C-terminal domain of IF3 can interfere with subunits association and promote the fidelity of P-site codon–anticodon interactions, by exploiting its inherent flexibility alongside the formation of transient perturbations in the 30S subunit. Once the initial P-site tRNA binds the two subunits associate to form the functional ribosome concurrently with IF3 dissociation (Fabbretti et al., 2007).

The Ribosome is an RNA Nano-machine Acting as a Polymerase

Simultaneously with the advancement of the mRNA/tRNA along the path in the small subunit, peptide bonds are being formed in the large subunit. This process involves also the translocation of the tRNA 3' end from the A- to the P-site, the detachment of the P-site tRNA from the growing polypeptide chain, the passage of the deacylated tRNA molecule to the E-site and its subsequent release. In all currently available crystal structures of assembled ribosomes (Yusupov et al., 2001; Schuwirth et al., 2005; Selmer et al., 2006; Korostelev et al., 2006) as well as in the unbound subunits (Ban et al., 2000; Schlutzen et al., 2000; Wimberly et al., 2000; Harms et al., 2001) the intersubunit interface surfaces are rich in rRNA, and based on the high sequence conservation of these ribosomal regions, it seems that the intersubunit interfaces of all ribosomes, throughout evolution, are rich in rRNA. Since both the decoding center and the site of peptide bond formation (called the peptidyl-transferase-center or PTC) reside in proximity to the intersubunit interface, they are located in rRNA predominant environments. Hence, unlike typical polymerases, which are protein enzymes, the major player in ribosomes activities is RNA. However, although in addition to the structural observations, protease resistance and structural data implicated rRNA as the catalytic entity for peptide bond formation, and although it is likely that single peptide bonds can be produced solely by RNA, the polymerase activity of the ribosome, namely subsequent occurrence of peptidyl transfer by rRNA has not been fully demonstrated (Anderson et al., 2007).

Peptide bond formation, the detachment of the P-site tRNA from the growing polypeptide chain, the passage of the deacylated tRNA molecule to the E-site, where it is engaged by codon–anticodon interactions, and its subsequent release, occur simultaneously with the advancement of the mRNA. Amino-acylated tRNA molecules are the natural substrates of ribosomes. However, ‘minimal substrates’ or ‘fragment reaction substrates’ were, and still are the substrate analogs commonly used biochemically, although the time required for peptide bond formation is much longer than the time required by the real substrates, namely tRNA molecules. Importantly, the mystery of the increased duration of peptide bond formation by these single-bond substrate analogs was recently partially clarified, as it was shown that the excessive time is due to conformational rearrangements of the substrates, as well as of specific PTC components (Selmer et al., 2006; Yonath, 2003a). Consistently, it was found that the peptidyl transfer reaction is modulated by conformational changes at the active site, and this process consumes time (Schmeing et al., 2005a and b; Beringer and Rodnina, 2007; Beringer et al., 2005; Brunelle et al., 2006).

The ‘fragment reaction substrates’ analogs are basically derivatives of puromycin, and although they are capable of producing single peptide bonds, they were overestimated to be suitable to mimic the natural ribosome function. They were, and still are, being used in a series of crystallographic studies on complexes of the large ribosomal subunit from *Haloarcula marismortui* (H50S) with such minimal substrates. These complexes were obtained under far from optimal functional conditions and contain key PTC components in orientations that differ significantly from those observed in functional complexes of T70S ribosome. Noteworthy, the initial suggestion, that three specific rRNA nucleotides catalyze peptide bond formation by the general acid/base mechanism that was based on the crystal structure of complexes of the H50S with such minimal substrates (Nissen et al., 2000) was challenged almost instantaneously by a battery of biochemical and mutational studies (e.g. Polacek et al., 2001; Barta et al., 2001; Thompson et al., 2001; Polacek et al., 2001).

The current consensus view is consistent with ribosomal positional catalysis that is assisted by its P-site tRNA substrate. The importance of the accurate positioning of the substrates within the ribosome frame, accompanied by the key role that the tRNA interactions with 23S rRNA play in peptide bond formation on the ribosome, are currently widely accepted (e.g. Beringer et al., 2007; Bashan and Yonath 2008b) even by those who suggested originally that the ribosome catalyze peptide bond formation by acid/base mechanism (Simonovic and Steitz 2008). Nevertheless, the detailed mechanism of peptide bond formation is still a mystery, waiting for availability of structures of ribosomes containing both the aminoacylated and the peptidyl tRNA molecules

in the PTC, trapped at the midst of this reaction. Furthermore, so far very few studies deal with the process allowing the ribosome to act not only as the frame for peptide bond formation, but also as the entity providing the machinery for amino-acid polymerization. In contrast, most biochemical studies focused on either peptide bond formation or P-site tRNA release and the termination of the process.

It appears that the choice of substrate analogs is partially responsible to the misinterpretation. Thus, analysis of the structure of a complex of the large ribosomal subunit form D50S with a substrate analog mimicking the whole tRNA acceptor stem and its aminoacylated 3' end, called ASM, which is the entire fraction of A-site tRNA that binds to the large subunit. This D50S-ASM complex was obtained under conditions similar to those found to be suitable for optimal ribosome function, advanced the comprehension of peptide bond formation by showing that rather than participating chemically in peptide bond formation, the ribosomal architecture positions the substrates in stereochemistry suitable for peptide bond formation, thus providing the machinery for peptide bond formation and tRNA translocation. (Bashan et al., 2003; Agmon et al., 2005). Furthermore, the ribosomal architecture that dictates "positional catalysis" to peptide bond formation revealed by this complex, also promotes substrate mediated chemical acceleration of peptide bond formation, in accord with the finding that rapid and smooth peptide bond formation requires full-length tRNA in both A and P sites, observed by various methods, including the usage of chemical (Brunelle et al., 2006; Weinger et al., 2004; Weinger and Strobel, 2006) mutagenesis (Youngman et al., 2006; Sato et al., 2006), computational (Trobroy and Aqvist, 2006; Sharma et al., 2005; Gindulyte et al., 2006) and kinetic procedures (Beringer et al., 2005; Wohlgemuth et al., 2006; Beringer and Rodnina, 2007). Notably, the D50S-ASM complex (Bashan et al., 2003) is so far the only complex with A-site tRNA mimic extending beyond the tip of the tRNA 3' end, although crystals supposed to contain assembled ribosomes with two and/or three tRNA molecules have been subjected to crystallographic analysis (Selmer et al., 2006; Korostelev et al., 2006).

Thus, it is clear that the ribosome provided the suitable stereochemistry and the geometrical means for substrate mediated catalysis. Nevertheless, despite intensive efforts to elucidate the details of peptide bond formation by various non-crystallographic methods, the elucidation of the detailed mechanisms that connect peptide bond formation to the polymerase function of the ribosome should benefit significantly from the availability of structures of complexes of ribosomes containing its two substrates at the A- and the P-sites (namely both the aminoacylated and the peptidyl tRNAs in the PTC) and a nascent protein trapped in the exit tunnel.

Is the Ribosomal Core an Optimized Ancient Entity?

Different from enzymes catalyzing a single chemical reaction and similar to polymerases, the ribosome provides the means not only for the mere formation of peptide bonds, but also, actually mainly, for the processivity of the reaction, namely for amino acid polymerization. Both tasks are governed by the ribosomal striking architecture, as it contains a highly conserved region of 180 nucleotides, relating by pseudo two-fold symmetry the rRNA fold and the orientations of its nucleotides, but not their sequences. This sizable intra-ribosomal pseudo symmetrical region is located within the otherwise asymmetric ribosome, and has been identified in all known structures of the whole ribosome or the large ribosomal subunits, regardless of their source, their functional state, or their kingdom of life. Particularly, the same sub-structure was identified in the cores of ribosomes from mesophilic, thermophilic, radiophilic and halophilic bacteria from eubacteria and archaea; in assembled empty or in complexes of them with substrates; in unbound and complexed large subunit, including complexes with ribosomal antibiotics and non ribosomal factors involved in protein biosynthesis (Agmon et al., 2005, 2006). Thus, despite size differences between the various kingdoms of life, the functional regions of the ribosomes are rather well conserved, with the highest level of sequence conservation at their central core, and the largest structural differences are located at the periphery. Remarkably, this core contains the symmetrical region in which 98% of the nucleotides are found in >95% of sequences (from 930 different species from the three domains of life), whereas only 36% of all *E. coli* nucleotides, excluding the symmetrical region, can be categorized as such. Importantly, 75% of the 27 nucleotides lying within 10 Å from the symmetry axis are highly conserved. Among them, seven are completely conserved (Agmon et al., 2006). This exceptional high conservation of the ribosome core, namely the symmetrical region, throughout evolution indicates that its structure is not sensitive environmental conditions, contrary to the large differences detected at the periphery of the ribosome, far away from the central core (Mears et al., 2002).

This symmetrical region includes the PTC and its environs, and connects all ribosomal functional regions involved in amino-acid polymerization, namely the tRNA entrance/exit dynamic stalks, the PTC, and the bridge connecting the PTC cavity with the vicinity of the decoding center in the small subunit and the nascent protein exit tunnel. Hence it appears that the internal symmetrical region is a universal feature facilitating peptide bond formation. Furthermore, as it is located at the heart of the ribosome, it can serve as the central feature for signaling between all the functional regions involved in protein biosynthesis, that are located remotely from each other (up to 200 Å away), but must “talk” to each other during elongation (Agmon et al., 2005; 2006). A suitable

example is the direct connection between peptide bond formation in the large subunit and the formation of the Shine–Dalgarno interactions on the small one (Uemura et al., 2007).

The PTC is located at the midst of this symmetrical region in the bottom of a V-shaped cavity and is built as an arched void. The complex D50S-ASM demonstrates the significance of the interactions between the tRNA acceptor stem and the cavity's walls. These key interactions seem to govern the accurate substrate positioning in the PTC, which has dimensions suitable for accommodating the 3' ends of the A- and P-site tRNA molecules (Bashan et al., 2003) with significant tolerance (Yonath, 2003a) and a shape apt for accommodating the acceptor stems of both A- and P-tRNAs in configuration allowing for peptide bond formation. The linkage between the elaborate architecture of the symmetrical region and the position of the A-site tRNA, as observed crystallographically (Bashan et al., 2003) indicates that the translocation of the tRNA 3' end is performed by a combination of two independent, albeit synchronized motions: a sideways shift, performed as a part of the overall mRNA/tRNA translocation, and a rotatory motion of the A-tRNA 3' end along a path confined by the PTC walls.

This rotatory motion is navigated and guided by the ribosomal architecture, mainly the PTC rear wall that confines the rotatory path and the flexible nucleotides 2602 and 2585 (*E. coli* nomenclature) that anchor and seem to propel it. Hence, it appears that the ribosomal architecture and its mobility provides all structural elements enabling ribosome function as an amino acid polymerase, including the formation of two symmetrical universal base pairs between the tRNAs and the PTC (Bashan et al., 2003; Agmon et al., 2005), a prerequisite for substrate mediated acceleration (Weinger and Strobel, 2006) and for the direction of the nascent protein into the exit tunnel. Importantly, all nucleotides involved in this rotatory motion have been classified as essential by a comprehensive genetic selection analysis (Sato et al., 2006). Furthermore, the rotatory motion positions the proximal 2'-hydroxyl of P-site tRNA A76 in the same position and orientation found in crystals of the entire ribosome with mRNA and tRNAs, as determined independently in two laboratories (Selmer et al., 2006; Korostelev et al., 2006), and allows for chemical catalysis of peptide bond formation by A76 of the P-site tRNA (Weinger and Strobel, 2006).

Simulation studies indicated that during this motion the rotating moiety interacts with ribosomal components confining the rotatory path, along the 'PTC rear wall' (Agmon et al., 2005, 2006). Consistently, quantum mechanical calculations, based on D50S structural data, indicated that transition state (TS) of this reaction, namely peptide bond formation, is being formed during the rotatory motion and is stabilized by hydrogen bonds with rRNA nucleotides (Gindulyte et al., 2006) and is located between the A- and the P-sites at a position similar

to that found experimentally in the crystal structure of a complex made of the large subunit from a ribosome from a different source, H50S, with a chemically designed TS analog (Schmeing et al., 2005). The correlation between the rotatory motion and amino acid polymerization rationalize the apparent contradiction associated with location of the growing protein chain. Thus, the traditional biochemical methods for the detection of ribosome activity were based on the reaction between substrate analogs designed for producing a single peptide bond and do not involve A- to P-site translocation, whereas nascent protein elongation by substrates suitable to perform the A- to P-site passage occurs close to the P-site in a position close to that of properly designed TS analogs (Schmeing et al., 2005), near the P-site. Consistently, the difference between the formation of single peptide bond by minimal substrates and possessive amino acid polymerization highlights the PTC remarkable ability to rearrange itself upon substrate binding (Selmer et al., 2006; Yonath, 2003a) and with the finding that the peptidyl transfer reaction is modulated by conformational changes at the active site (Schmeing et al., 2005; Beringer et al., 2005; Beringer and Rodnina, 2007; Brunelle et al., 2006).

The rotatory motion machinery explains the remarkable difference between the formation of single peptide bond by minimal substrates and by full size tRNA, which can undergo subsequent amino acid polymerization. It also connects the fragment reaction with the PTC remarkable ability to rearrange itself upon substrate binding (Selmer et al., 2006) by indicating that the peptidyl transfer reaction is modulated by conformational changes at the active site (Yonath, 2003a, Schmeing et al., 2005). Additionally, this mechanism indicates that dipeptides can be detected in the A-site simply because minimal substrates do not possess the structural element that can undergo the rotational motion, as they are too short. In addition, by their nature, the minimal substrate reaction are capable of a single peptide bond formation, hence, have cannot elongate. Thus, these finding contradict the notion that actual nascent chains are elongated at the A-site, and supports the finding that peptide bond formation that can lead to nascent chain elongation occurs during the rotatory motion (Gindulyte et al., 2006).

Remarkably, the high level of conservation of components of the symmetrical region that was detected even in mitochondrial ribosomes, in which half the ribosomal RNA is replaced by proteins and the ability of the symmetrical region to provide all structural elements required for performing polypeptide elongation, suggest that the modern ribosome evolved from a simpler entity that can be described as a pro-ribosome, by gene fusion or gene duplication. In particular, the preservation of the three-dimensional structure of the two halves of the ribosomal frame regardless of the sequence demonstrates the rigorous requirements of accurate substrate positioning in stereochemistry supporting peptide bond formation. This, as well as the universality of the

symmetrical region led to the assumption that the ancient ribosome contained a pocket confined by two RNA chains, which formed a dimer.

A feasible sequence of events, which is based on the hypothesis that the two halves of a structure resembling the symmetrical region were the core of the ancient ribosome, may shed light on the evolving complexity of the ribosome. The main assumption for this hypothesis is that the ancient machinery that could form peptide bonds was made exclusively from RNA chains that dimerize. It is conceivable that among other substituents, the primordial soup contained single nucleotides, short RNA segments, and RNA chains of significant size, of around 50–60 nucleotides, which survived since they acquired a stable conformation. These surviving ancient RNA chains were the ancestors of the RNA chains that construct the symmetrical region, and underwent dimerization as they possess the tendency to dimerize, similar to small RNA molecules that form dimers (Agmon et al., 2006).

The products of the dimerization yielded three-dimensional structures with a symmetrical pocket that could accommodate two amino acids facing each other, hence could trap the amino acids once they became available. The spontaneous reaction of peptide bond formation followed, and the structures that include the appropriate pockets for accommodating this reaction survived. As RNA chains can act as gene-like molecules coding for their own reproduction, it is conceivable that the surviving ancient pockets became the templates for the ancient ribosomes. In later stage these primitive RNA genes underwent initial optimization to produce a more defined, relatively stable pocket, and when a clear distinction was made between the amino acid and the growing peptidyl sites, each of the two halves was further optimized for its task so that their sequences evolved differently.

In parallel, the substrates of the ancient ribosomes, which were initially single amino acids, dipeptides or short polypeptides, evolved to allow accurate binding. Compounds mimicking the termini of the modern tRNA molecules (part or the entire amino acylated CCA 3' end) provide more control on the reaction, its reactant and its products. Consequently, these were converted into compounds with a contour that can complement the inner surface of the reaction pocket. The binding affinity could be increased by enlarging the nucleic acid component, exploiting double or triple nucleotides, similar to the universal CCA end of the modern tRNA molecules. Later, for increasing specificity, these short RNA segments were extended to larger structures by their fusion with RNA stable features, to form the ancient tRNA. Later, RNA chains capable of storing, selecting and transferring instructions for producing useful proteins became available. Subsequently, the decoding process was combined with peptide bond formation. A single molecule evolved, capable of not only carrying the amino acids while bound to them, but also translating the genomic instructions, by adding a feature similar to the modern anticodon arm to the ancient tRNA structure.

To conclude: the ancient ribosome appears to be a simple ribozyme that produces peptide bonds, utilizing nucleotide conjugated or otherwise activated amino acids. The formation of single peptide bonds seems to be accidental or sporadic, thus not controlled by genetic instructions. Since the products of this reaction may be substrates for it, elongation of the dipeptides could occur. Only when these polypeptides acquired capacity to perform enzymatic tasks, the information about their desired structure was stored in genes. Consequently, molecules capable of decoding this information simultaneously with transporting the cognate substrates evolved and the pro-ribosome acquired properties enabling the smooth translation of genetic information into proteins.

Chaperoning Events Within the Tunnel and at Its Opening

As mentioned above, the nascent proteins progress towards their emerging site through their exit tunnel, a universal feature of the large ribosomal subunit that lies adjacent to the PTC (Harms et al., 2001; Nissen et al., 2000) and has the dimensions of (~ 120 Å in length and varying diameter, 10–25 Å). Lined primarily by rRNA with a few r-proteins reaching its walls from its exterior, this tunnel may provide a crevice adjacent to the tunnel wall that may provide space for cotranslational transient folding (Amit et al., 2005), which is predominantly suitable for hydrophobic regions of the translated proteins (e.g. Woolhead et al., 2004). Additionally, this tunnel possesses the dynamics required for gating and discriminating events and plays an active role in sequence-specific arrest of nascent chains and in responding to cellular signals. Examples include: (i) The secretion monitor (SecM) system, which contains a 17 amino acids sequence that under specific cellular conditions can interact with the tunnel gating components, thereby interfering with their own translation elongation (Nakatogawa and Ito, 2004; Woolhead et al., 2006); (ii) the 24 amino acid sequence of TnaC that appears to be involved in gene expression regulation by influencing the RF2-mediated cleavage, or by introducing PTC rearrangements (Cruz-Vera et al., 2006); (iii) hydrophobic trans membrane sequences that seem to controls the operational mode of the translocon at the ER membrane while being synthesized (Woolhead et al., 2004). Nevertheless, despite recent intensive research, performed by several groups (e.g. Petrone et al., 2008; Voss et al., 2006) the mechanisms facilitating tunnel's recognition are still to be revealed. Generation of homogenous and stable complexes of ribosome with nascent chains suitable for crystallographic studies (Schaffitzel and Ban, 2007) should provide some of the answers. So far the tunnel wall elements that appear to sense nascent-peptide specific sequences include the r-proteins L22 (Berisio et al., 2003a) that forms the tunnel's constriction together with protein L4 and L23 (Baram et al.,

2005; Baram and Yonath, 2005), within a cascade of rRNA features (Mitra et al., 2006).

At the end of their way, before gaining sufficient length to acquire the final fold, nascent proteins emerge from their protective exit tunnel into the crowded cellular environment. To assist correct folding and prevent aggregation under the cellular unfavorable conditions, all cells contain molecular chaperones, exerting various effects. In eubacteria, the first chaperone uncounted by the emerging nascent chain, called trigger factor (TF), binds to the translating ribosome and prevents aggregation and misfolding at 1:1 stoichiometry, by interacting with ribosomal proteins L23 and L29 (Baram et al., 2005; Schlunzen et al., 2005). Protein L23 belongs to the small group of ribosomal proteins that display a significant divergence from conservation and only in eubacteria it possesses a sizable elongated loop, which extends from ribosome exterior all the way into the tunnel walls. At this position the L23 extended loop can undergo allosteric conformational changes that, in turn, can modulate the shape of the tunnel, in a fashion implying a trafficking of the nascent protein (Baram and Yonath, 2005; Baram et al., 2005).

Studies aimed at revealing the mode of function of TF were carried out using a chimeric (Ferbitz et al., 2004; Maier et al., 2005; Rutkowska et al., 2008), as well as a physiological (Baram et al., 2005; Schlunzen et al., 2005) complexes of TF binding domain and large ribosomal subunits, alongside a complex of the entire TF with full length short proteins (Martinez-Hackert and Hendrickson, 2007). The crystal structures of physiologically meaningful complexes of D50S with TF binding domain (called TFa) (Baram et al., 2005; Schlunzen et al., 2005) revealed that TFa provides a shelter for the emerging nascent protein and undergoes conformational rearrangements. As a result of these rearrangements a sizable hydrophobic region becomes exposed thus acquiring a configuration suitable to adhere to hydrophobic patches on the nascent chain. Consistent with dynamic studies (Kaiser et al., 2006), it appears that TFa prevents the aggregation of the emerging nascent chain by transiently masking exposed hydrophobic surfaces on the elongating polypeptide chains until they can be buried in the interior of the mature protein (Yonath, 2006).

The Dual Contributions of Ribosomal Proteins

The significance of the contribution of the ribosomal proteins to ribosome function and to the integrity of the ribosomal particles has undergone dramatic alterations. As described above, originally the proteins were thought to carry out the catalytic tasks of the ribosome, but when it became clear that the ribosome is a ribozyme, the proteins were suggested to be mainly the

entities stabilizing the rRNA structure (Brodersen et al., 2002; Klein et al., 2004). Previous and recent biochemical experiments, combined with careful analysis of the ribosome crystal structures, indicated that r-proteins, which are situated in proximity to ribosomal functional regions, may support these functions by controlling their fidelity and/or facilitating specific activities, such as influencing ribosomal accurate function. Proteins S5, S6, S12 that affect mRNA binding accuracy reside in the small subunit (Cochella and Green, 2004), whereas proteins L2 and L27 seem to be involved in the ribosomal polymerase function (Cooperman et al., 1995; Maguire et al., 2005).

Proteins S12 and L2 are among the few proteins that are located partially on the rRNA rich intersubunit interface. Remarkably, computational methods showed that the most ancient ribosomal proteins are these two proteins (Sobolevsky and Trifonov, 2005), consistent with the hypothesis that the first proteins were incorporated into the ribosome in order to increase the fidelity and efficiency of the ancient RNA-only ribosome. These two proteins belong to the group of proteins that were found to be involved in the catalytic reaction of the ribosome. This group includes also L16 and L27, which seem to assist efficient peptide bond formation. Protein L16 contribute to correct positioning of the A-site tRNA acceptor stem (Harms et al., 2001; Bashan et al., 2003) and hence plays an important role in acquiring the right orientation of its aminoacylated CCA end within the PTC. Notably, similar to the symmetrical region, L16 displays conserved tertiary structure but diverged primary sequence. In addition to L16, the finding that the N-terminal tail of L27 can be cross-linked to A76 of P-site tRNA, the observation that deletion of L27 gene results in severe defects in cell growth (Maguire et al., 2005), and the likelihood that this L27 resides close to the PTC (Selmer et al., 2006) indicate its possible involvement in peptide bond formation. Protein L2 it is not required for peptide bond formation, but seems to play a key role in the processivity of the amino acid polymerization (Cooperman et al., 1995). As L2 is the only protein interacting with both the A- and the P-regions (Agmon et al., 2005), it appears that its main function is related to its interactions with the symmetrical region.

Several r-proteins facilitate ribosomal dynamic functions. Proteins L22 and L4 that line the tunnel walls and appear to be involved in tunnel gating (Berisio et al., 2003a), and protein L23 that may control nascent protein trafficking (Baram et al., 2005). Large scale rearrangements in the tRNA configuration, as well as lateral movements of the two large subunit stalks, called L7/L12 (the entrance door) and L1 (the exiting gate) have been detected, based on comparisons of crystals structures of ribosomes trapped at various functional states and results of various methods, including cryo electron microscopy and single-molecule fluorescence spectroscopy. Proteins L1 and L12, the main protein component of the L1 and the L7/L12 stalks, are dynamically

involved in tRNA translocation. L12, which is the only r-protein that appears in more than a single copy in the ribosome, binds directly to several elongation factors (Kavran and Steitz, 2007; Helgstrand et al., 2007; Zavialov et al., 2005; Wilden et al., 2006; Diaconu et al., 2005).

The task of L11, another component of this stalk, seems to be associated with rejection of unfavorable compounds. Thus it was found to interact with the antibiotic thiostrepton (Bowen et al., 2005), but it is not vital for cell life and its depletion causes thiostrepton resistance. Additional examples for proteins that bind to non-ribosomal factors or inhibitors are proteins S18 that interacts with initiation factor IF3 (Pioletti et al., 2001). Notably, the translational factors functioning at the subunit-subunit interactions are conserved in phylogenetically distant species despite highly divergent environments to which these species have adapted (Thompson and Dahlberg, 2004). Of interest is protein L25 (called also CTC) that, as described above, seems to control the creation of one of the intersubunit bridges (called the A-finger or H38) as well as to regulate A-site tRNA incorporation (Harms et al., 2001; Yonath, 2002).

Can Structures Lead to Improved Antibiotics?

The intensive research on ribosomes has some practical aspects; one of them has clinical relevance since many antibiotics target the ribosome. The increasing incidence of antibiotic resistance and toxicity creates serious problems in modern medicine, combating resistance to antibiotics has been a major concern in recent years. The vast amount of structural data on ribosomal antibiotics accumulated recently may be useful for this aim.

Antibiotics are compounds used in clinical medicine for treating bacterial infections inhibit selectively bacterial ribosomes and not the eukaryotic ones. More than 40% of the useful antibiotics interfere with the biosynthetic machinery and most of them target the ribosomes at distinct locations within functionally relevant sites. These act by diverse mechanism, many of them were revealed by analysis of crystallographic results (for review see Auerbach et al., 2004; Yonath and Bashan 2004; Yonath, 2005a; Tenson and Mankin, 2006; Poehlsgaard and Douthwaite, 2005; Bottger, 2006; 2007; Mankin, 2001, 2006, 2008). These analyses showed that the ribosomal antibiotics hamper protein synthesis in bacterial pathogens by diverse mechanism, including causing miscoding, minimizing essential mobility, interfering with substrate binding at the decoding center and at the PTC, or blocking the protein exit tunnel. Common to ribosomal antibiotic is that their binding pockets coincide with functionally critical centers of the ribosome, and as most of the ribosomal functions are performed by the rRNA, the antibiotics binding

sites are composed primarily of rRNA. Numerous structural, biochemical and genetic studies provided indispensable information that illustrated the basic mechanisms of ribosomal antibiotics activity and synergism; provided the structural basis for antibiotic resistance and enlightened the principles of antibiotics selectivity, namely the discrimination between pathogens and humans, the key for therapeutical usefulness.

By its nature, X-ray crystallography should be the choice method for investigating ribosome–antibiotics interactions. However, since X-ray crystallography requires diffracting crystals, and since so far no ribosomes from pathogenic bacteria could be crystallized, currently the crystallographic studies are confined to the currently available crystals of suitable pathogen models. Currently available are high-resolution structures of complexes of antibiotics with ribosomal particles from the eubacterium *Deinococcus radiodurans*, a species suitable to serve as a pathogen model; as well as complexes obtained from antibiotics bound to ribosomes from the Dead Sea archaeon *Haloarcula marismortui* that resembles eukaryotes in respect to antibiotics binding site. Importantly, enormously high concentrations were needed for obtaining these complexes, consistent with their high similarity to eukaryotic ribosomes. As seen below, comparisons between the two types of complexes proved indispensable for increasing our understanding on antibiotics action.

The structures of its large ribosomal subunit D50S complexed with various antibiotics determined so far (Schluenzen et al., 2001, 2004; Berisio et al., 2003a, b; Auerbach et al., 2004, 2009; Yonath and Bashan 2004; Harms et al., 2004; Pyetan et al., 2007; Yonath, 2005a and b; Davidovich et al., 2007, 2008; Vazquez-Laslop et al., 2008) revealed that among the modes of action of ribosomal antibiotics are causing miscoding, minimizing essential mobility, interfering with substrate binding at the decoding center and at the PTC, or blocking the protein exit tunnel. Common to ribosomal antibiotic is that their binding sites are composed primarily of rRNA and coincide with functionally critical centers of the ribosome. Furthermore, comparisons between these structures demonstrated that members of antibiotic families possessing common chemical elements with minute differences might bind to ribosomal pockets in significantly different modes, and that the nature of seemingly identical mechanisms of drug resistance may be dominated by the antibiotics' chemical properties.

A major issue concerning the clinical usefulness of ribosomal antibiotics is their selectivity, namely their capabilities in the discrimination between the ribosomes of the eubacterial pathogens and those of eukaryotes. As described above, although prokaryotic and eukaryotic ribosomes differ in size (~2.4 and 4 Mega Dalton, respectively), their functional regions, which are the targets for the antibiotics, are highly conserved. Therefore the imperative distinction between eubacterial pathogens and mammals, the key for antibiotics

usefulness, is achieved generally, albeit not exclusively, by subtle structural difference within the antibiotics binding pockets of the prokaryotic and eukaryotic ribosomes. A striking example for discrimination between pathogens and humans is the immense influence of the minute difference between adenine and guanine in position 2058. This small difference was found to govern the binding of macrolides, a prominent antibiotics family (Lai and Weisblum, 1971) that obstructs the progression of the nascent proteins within the tunnel. However, although 2058 identity determines the antibiotic affinity, this analysis showed that the mere binding of the antibiotics is not sufficient for obtaining efficient therapeutical effectiveness. Comparisons between crystal structures of antibiotics bound to the eubacterial large ribosomal subunit, D50S to structures of complexes of the large ribosomal subunit from H50S, an archaeon sharing properties with eukaryotes, which required either extremely high antibiotics concentrations (Hansen et al., 2002a, 2003) or G2058A mutations, which facilitates macrolides/ketolides binding (Tu et al., 2005), indicated the significance of additional structural elements (Yonath and Bashan, 2004) of the binding pocket, which dictate inhibitory activity. Similar observations were made by mutagenesis in the yeast *Saccharomyces cerevisiae* at a position equivalent to *E. coli* A2058, which allows erythromycin binding but does not confer erythromycin susceptibility (Bommakanti et al., 2008).

The fine details of binding, resistance and selectivity of the members of the macrolide family and its off springs, namely the azalides and ketolides, presents additional issues, as the sequence specificity that determines the susceptibility and the fitness cost of the ketolides (Schlunzen et al., 2003; Pfister et al., 2005). Another intriguing issue, which led to continuous expansion of research and consequently to new insights, relates to the nature of the contributions of two ribosomal proteins, namely L4 and L22. The two proteins line the exit tunnel at its constriction, and do not interact directly with most of the members of the macrolides family, yet several types of mutations at their tip acquire resistance to them (Gregory and Dahlberg, 1999; Davydova et al., 2002; Lawrence et al., 2008; Berisio et al., 2006; Zaman et al., 2007; Moore and Sauer, 2008), presumably by perturbing the rRNA structure at the tunnel walls (Gregory and Dahlberg, 1999; Lawrence et al., 2008).

Attempts aimed at alleviating the resistance problem include the developments of synergetic antibiotics. An example is the very potent family of the streptogramins, a two-component antibiotics drug family, each of which is a rather weak drug. The impressive synergetic effect of this family can be understood by examining the mechanism exploited by the rather recent antibiotic drug, synercid. This mechanism is based on the binding of one of the components to the PTC that causes a dramatic alteration in the orientation of the very flexible nucleotide, U2585, which plays a principal role in the

A- to P-site rotatory motion, and the fixation of the altered orientation by the second compound that binds at the tunnel entrance (Harms et al., 2004; Yonath and Bashan, 2004).

Even subtle differences, such as the identity of nucleotide 2058 (A in eubacteria, G elsewhere) hardly exist in the PTC. Therefore, obtaining selectivity in antibiotics binding to the PTC, the core catalytic center of the ribosome, is more complex. Nevertheless, some of the PTC antibiotics bind to the PTC of eubacterial ribosomes with high affinity and great specificity, without significant effect on the eukaryotic hosts. The crystal structures of ribosomal complexes with antibiotics indicated that the PTC provides binding sites to several clinically useful antibiotics, called below the super family of PTC antibiotics, shed light on general as well as specific properties of the interactions of the members of this family with their binding pockets in the PTC. This family includes chloramphenicol, clindamycin, pleuromutilins, streptogramins_A and oxazolidinones and lankacidins. Although basically all PTC antibiotics act by blocking part or the entire PTC, they utilize different binding modes and consequently they possess various inhibitory mechanisms. Thus, chloramphenicol was found to hamper the binding of the A-site tRNA, like (Schlunzen et al., 2001), the pleuromutilins, linezolid and streptogramins_A occupy both the A- and the P-site tRNAs (Harms et al., 2004; Hansen et al., 2003; Ippolito et al., 2008), and clindamycin interfere with the peptide bond formation (Schlunzen et al., 2001).

Contribution of several PTC flexible nucleotides to productive binding was also observed by investigating the mode of action of the pleuromutilin family, which revealed a unique inhibitory mechanism alongside a novel selectivity and resistance strategies. In particular, the elaborate pleuromutilins binding mode demonstrates how selectivity and resistance are acquired despite almost full conservation (Davidovich et al., 2007, 2008; Schlunzen et al., 2004). As all nucleotides in the immediate vicinity of the binding site are highly conserved, pleuromutilins selectivity is determined by nucleotides that are not located in the immediate vicinity of the antibiotic binding site, hence are less conserved. Thus, pleuromutilins binding triggers an induced-fit mechanism by exploiting the flexibility of the rRNA nucleotides residing in and around the PTC, as well as a network of interactions with contacts with less conserved remote nucleotides, hence allowing for drug selectivity (Davidovich et al., 2007, 2008). In particular, this family exploits the remote interactions that effect the positioning of the extremely flexible nucleotide U2506, as well as of U2585 that participate in navigating and anchoring the rotatory motion. These interactions evacuate the binding region and at the same time tighten the binding pocket on the bound antibiotic molecule, without interacting with it. As mutations within the PTC should be lethal,

resistance to pleuromutilins requires mutations or modifications of nucleotides residing either in PTC components with identity that is less crucial for ribosome function, or in the PTC environs rather than within the core of the binding pocket, therefore should occur in a relatively slow pace. Remarkably, these crystallographic studies led the way to attempts to produce advanced compounds (Lolk et al., 2008). Indeed, cross resistance was detected between all PTC antibiotics, regardless of their mode of binding, and the nucleotides mediating it are residing only on one side of the PTC, similar or in close proximity to those acquiring resistance to the pleuromutilins.

Current attempts to overcome antibiotics resistance and increase their selectivity are being made (e.g. Yassin et al., 2005; Wilson et al., 2005a; Bottger, 2007). Those include several strategies, including the insertions of moieties that should compensate for the lost interactions of the resistant strains, benefiting from synergism of known or novel compounds possessing inhibitory properties of various levels of potency, and reviving “forgotten” antibiotics families, such as the lankacidins. Furthermore, as for each antibiotic family most eubacteria belonging to it utilize similar structural principles for selectivity and resistance, comprehending the factors allowing for selectivity should provide powerful tools to understand many of the mechanisms exploited for acquiring resistance. Therefore, the lessons learned from ribosome crystallography concerning combating resistance to antibiotics targeting the ribosome, are rather optimistic, as these studies opened new paths for antibiotics improvement. Thus, the elucidation of common principles of antibiotics action, combined with the variability in binding modes; with the revelation of the common as well as unique structural bases to antibiotics resistance mechanism; with the discovery that remote interactions can govern induced fit mechanisms enabling species discrimination even in highly conserved regions, and with the identification of deleterious mutations in the ribosomal RNA that can become potential targets for antibiotics, justify expectations for structural based improved properties of existing compounds as well as for the development of novel drugs.

Acknowledgments Thanks are due to all members of the ribosome group at the Weizmann Institute for their experimental efforts and illuminating discussion. Support was provided by the US National Inst. of Health (GM34360), and the Kimmelman Center for Macromolecular Assemblies. AY holds the Martin and Helen Kimmel Professorial Chair.

The currently available crystal structures are of native and complexed ribosomal subunits from two eubacteria and one archaeon with substrate-analogs or inhibitors (including antibiotics). These are the small ribosomal subunit from *Thermus thermophilus* (called T30S), the large subunit from *Deinococcus radiodurans* (called D50S), and the large ribosomal subunit of the archaeon *Haloarcula marismortui* (called H50S). Also available are two structures of assembled ribosome complexed with their substrates from *Thermus thermophilus* (called T70S) and of empty ribosomes from *E. coli* (called E70S).

References

- Agmon, I., Bashan, A., Zarivach, R., and Yonath, A. (2005). Symmetry at the active site of the ribosome: structural and functional implications. *Biol Chem* 386, 833–844.
- Agmon, I., Bashan, A., and Yonath, A. (2006). On ribosome conservation and evolution. *Isr J Ecol Evol* 52, 359–379.
- Agrawal, R. K., Heagle, A. B., Penczek, P., Grassucci, R. A., and Frank, J. (1999). EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nat Struct Biol* 6, 643–647.
- Amit, M., Berisio, R., Baram, D., Harms, J., Bashan, A., and Yonath, A. (2005). A crevice adjoining the ribosome tunnel: hints for cotranslational folding. *FEBS Lett* 579, 3207–3213.
- Anderson, R. M., Kwon, M., and Strobel, S. A. (2007). Toward ribosomal RNA catalytic activity in the absence of protein. *J Mol Evol* 64, 472–483.
- Auerbach, T., Bashan, A., and Yonath, A. (2004). Ribosomal antibiotics: structural basis for resistance, synergism and selectivity. *Trends Biotechnol* 22, 570–576.
- Auerbach, T., Mermershtain, I., Bashan, A., Davidovich, C., Rozenberg, H., Sherman, D. H., and Yonath, A. (2009). Structural basis for the antibacterial activity of the 12-membered-ring mono-sugar macrolide methymycin. *Biotechnologia*, 84, 24–35.
- Auerbach-Nevo, T., Zarivach, R., Peretz, M., and Yonath, A. (2005). Reproducible growth of well diffracting ribosomal crystals. *Acta Crystallogr D Biol Crystallogr* 61, 713–719.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920.
- Baram, D., and Yonath, A. (2005). From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. *FEBS Lett* 579, 948–954.
- Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A., and Yonath, A. (2005). Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc Natl Acad Sci U S A* 102, 12017–12022.
- Barta, A., Dorner, S., and Polacek, N. (2001). Mechanism of ribosomal peptide bond formation. *Science* 291, 203.
- Bashan, A., Agmon, I., Zarivach, R., Schlutzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H. A. S., et al. (2003). Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol Cell* 11, 91–102.
- Bashan, A., and Yonath, A. (2005). Ribosome crystallography: catalysis and evolution of peptide-bond formation, nascent chain elongation and its co-translational folding. *Biochem Soc Trans* 33, 488–492.
- Bashan, A., and Yonath, A. (2008a). The linkage between ribosomal crystallography, metal ions, heteropolytungstates and functional flexibility. *J Mol Struct* 890 289–294.
- Bashan, A., and Yonath, A. (2008b). Correlating ribosome function with high-resolution structures. *Trends Microbiol* 16, 326–335.
- Bayfield, M. A., Dahlberg, A. E., Schulmeister, U., Dorner, S., and Barta, A. (2001). A conformational change in the ribosomal peptidyl transferase center upon active/inactive transition. *Proc Natl Acad Sci U S A* 98, 10096–10101.
- Beringer, M., Adio, S., Wintermeyer, W., and Rodnina, M. (2003). The G2447A mutation does not affect ionization of a ribosomal group taking part in peptide bond formation. *RNA* 9, 919–922.
- Beringer, M., Bruell, C., Xiong, L., Pfister, P., Bieling, P., Katunin, V. I., Mankin, A. S., Bottger, E. C., and Rodnina, M. V. (2005). Essential mechanisms in the catalysis of peptide bond formation on the ribosome. *J Biol Chem* 280, 36065–36072.
- Beringer, M., and Rodnina, M. V. (2007). Importance of tRNA interactions with 23S rRNA for peptide bond formation on the ribosome: studies with substrate analogs. *Biol Chem* 388, 687–691.
- Beringer, M., and Rodnina, M. V. (2007). The ribosomal peptidyl transferase. *Mol Cell* 26, 311–321.

- Berisio, R., Schlutzen, F., Harms, J., Bashan, A., Auerbach, T., Baram, D., and Yonath, A. (2003a). Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat Struct Biol* 10, 366–370.
- Berisio, R., Harms, J., Schlutzen, F., Zarivach, R., Hansen, H. A., Fucini, P., and Yonath, A. (2003b). Structural insight into the antibiotic action of telithromycin against resistant mutants. *J Bacteriol* 185, 4276–4279.
- Berisio, R., Corti, N., Pfister, P., Yonath, A., and Bottger, E. C. (2006). 23S rRNA 2058A->G alteration mediates ketolide resistance in combination with deletion in L22. *Antimicrob Agents Chemother* 50, 3816–3823.
- Bieling, P., Beringer, M., Adio, S., and Rodnina, M. V. (2006). Peptide bond formation does not involve acid-base catalysis by ribosomal residues. *Nat Struct Mol Biol* 13, 424–428.
- Blaha, G., Gurel, G., Schroeder, S. J., Moore, P. B., and Steitz, T. A. (2008). Mutations outside the anisomycin-binding site can make ribosomes drug-resistant. *J Mol Biol* 379, 505–519.
- Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., and Puglisi, J. D. (2004). tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* 11, 1008–1014.
- Bommakanti, A. S., Lindahl, L., and Zengel, J. M. (2008). Mutation from guanine to adenine in 25S rRNA at the position equivalent to *E. coli* A2058 does not confer erythromycin sensitivity in *Saccharomyces cerevisiae*. *RNA* 14, 460–464.
- Bottger, E. C. (2006). The ribosome as a drug target. *Trends Biotechnol* 24, 145–147.
- Bottger, E. C. (2007). Antimicrobial agents targeting the ribosome: the issue of selectivity and toxicity – lessons to be learned. *Cell Mol Life Sci* 64, 791–795.
- Bowen, W. S., Van Dyke, N., Murgola, E. J., Lodmell, J. S., and Hill, W. E. (2005). Interaction of thiostrepton and elongation factor-G with the ribosomal protein L11-binding domain. *J Biol Chem* 280, 2934–2943.
- Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* 103, 1143–1154.
- Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002). Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16S RNA. *J Mol Biol* 316, 725–768.
- Brunelle, J. L., Youngman, E. M., Sharma, D., and Green, R. (2006). The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity. *RNA* 12, 33–39.
- Brunelle, J. L., Shaw, J. J., Youngman, E. M., and Green, R. (2008). Peptide release on the ribosome depends critically on the 2' OH of the peptidyl-tRNA substrate. *RNA* 14, 1526–1531.
- Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407, 340–348.
- Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. (2001). Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291, 498–501.
- Clemons, W. M., Jr., Brodersen, D. E., McCutcheon, J. P., May, J. L., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2001). Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination. *J Mol Biol* 310, 827–843.
- Cochella, L., and Green, R. (2004). Wobble during decoding: more than third-position promiscuity. *Nat Struct Mol Biol* 11, 1160–1162.
- Cooperman, B. S., Wooten, T., Romero, D. P., and Traut, R. R. (1995). Histidine 229 in protein L2 is apparently essential for 50S peptidyl transferase activity. *Biochem Cell Biol* 73, 1087–1094.
- Crowley, K. S., Reinhart, G. D., and Johnson, A. E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. *Cell* 73, 1101–1115.

- Cruz-Vera, L. R., Gong, M., and Yanofsky, C. (2006). Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. *Proc Natl Acad Sci U S A* *103*, 3598–3603.
- Davidovich, C., Bashan, A., Auerbach-Nevo, T., Yaggie, R. D., Gontarek, R. R., and Yonath, A. (2007). Induced-fit tightens pleuromutlins binding to ribosomes and remote interactions enable their selectivity. *Proc Natl Acad Sci U S A* *104*, 4291–4296.
- Davidovich, C., Bashan, A., and Yonath, A. (2008) Structural basis for cross resistance to ribosomal PTC antibiotics, *Proc Natl Acad Sci U S A*, *105*, 20665–20670.
- Davydova, N., Streltsov, V., Wilce, M., Liljas, A., and Garber, M. (2002). L22 ribosomal protein and effect of its mutation on ribosome resistance to erythromycin. *J Mol Biol* *322*, 635–644.
- Deane, C. M., Dong, M., Huard, F. P., Lance, B. K., and Wood, G. R. (2007). Cotranslational protein folding – fact or fiction? *Bioinformatics* *23*, i142–148.
- Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G., Stark, H., Rodnina, M. V., and Wahl, M. C. (2005). Structural basis for the function of the ribosomal L7/12 Stalk in factor binding and GTPase activation. *Cell* *121*, 991–1004.
- Dunham, C. M., Selmer, M., Phelps, S. S., Kelley, A. C., Suzuki, T., Joseph, S., and Ramakrishnan, V. (2007). Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit. *RNA* *13*, 817–823.
- Evans, R. N., Blaha, G., Bailey, S., and Steitz, T. A. (2008). The structure of LepA, the ribosomal back translocase. *Proc Natl Acad Sci U S A* *105*, 4673–4678.
- Fabbretti, A., Pon, C. L., Hennelly, S. P., Hill, W. E., Lodmell, J. S., and Gualerzi, C. O. (2007). The real-time path of translation factor IF3 onto and off the ribosome. *Mol Cell* *25*, 285–296.
- Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* *431*, 590–596.
- Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K., and Agrawal, R. K. (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* *376*, 441–444.
- Frank, J., Penczek, P., Agrawal, R. K., Grassucci, R. A., and Heagle, A. B. (2000). Three-dimensional cryoelectron microscopy of ribosomes. *Methods Enzymol* *317*, 276–291.
- Fujiwara, T., Ito, K., Yamami, T., and Nakamura, Y. (2004). Ribosome recycling factor disassembles the post-termination ribosomal complex independent of the ribosomal translocase activity of elongation factor G. *Mol Microbiol* *53*, 517–528.
- Gabashvili, I. S., Gregory, S. T., Valle, M., Grassucci, R., Worbs, M., Wahl, M. C., Dahlberg, A. E., and Frank, J. (2001). The polypeptide tunnel system in the ribosome and its gating in erythromycin resistance mutants of L4 and L22. *Mol Cell* *8*, 181–188.
- Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A., et al. (2003). Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement. *Cell* *113*, 789–801.
- Gao, N., Zavialov, A. V., Li, W., Sengupta, J., Valle, M., Gursky, R. P., Ehrenberg, M., and Frank, J. (2005). Mechanism for the disassembly of the posttermination complex inferred from cryo-EM studies. *Mol Cell* *18*, 663–674.
- Garrett, R. A., and Wittmann, H. G. (1973a). Structure and function of the ribosome. *Endeavour* *32*, 8–14.
- Garrett, R. A., and Wittmann, H. G. (1973b). Structure of bacterial ribosomes. *Adv Protein Chem* *27*, 277–347.
- Gilbert, R. J., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., Robinson, C. V., Dobson, C. M., and Stuart, D. I. (2004). Three-dimensional structures of translating ribosomes by Cryo-EM. *Mol Cell* *14*, 57–66.
- Gindulyte, A., Bashan, A., Agmon, I., Massa, L., Yonath, A., and Karle, J. (2006). The transition state for formation of the peptide bond in the ribosome. *Proc Natl Acad Sci U S A* *103*, 13327–13332.

- Gluehmann, M., Zarivach, R., Bashan, A., Harms, J., Schlutzen, F., Bartels, H., Agmon, I., Rosenblum, G., Pioletti, M., Auerbach, T., et al. (2001). Ribosomal crystallography: from poorly diffracting microcrystals to high-resolution structures. *Methods* 25, 292–302.
- Gong, F., and Yanofsky, C. (2002). Instruction of translating ribosome by nascent Peptide. *Science* 297, 1864–1867.
- Gregory, S. T., and Dahlberg, A. E. (1999). Erythromycin resistance mutations in ribosomal proteins L22 and L4 perturb the higher order structure of 23 S ribosomal RNA. *J Mol Biol* 289, 827–834.
- Gromadski, K. B., and Rodnina, M. V. (2004). Streptomycin interferes with conformational coupling between codon recognition and GTPase activation on the ribosome. *Nat Struct Mol Biol* 11, 316–322.
- Hansen, H. A., Volkmann, N., Piefke, J., Glotz, C., Weinstein, S., Makowski, I., Meyer, S., Wittmann, H. G., and Yonath, A. (1990). Crystals of complexes mimicking protein biosynthesis are suitable for crystallographic studies. *Biochim Biophys Acta* 1050, 1–7.
- Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., and Steitz, T. A. (2002a). The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol Cell* 10, 117–128.
- Hansen, J. L., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2002b). Structural insights into peptide bond formation. *Proc Natl Acad Sci U S A* 99, 11670–11675.
- Hansen, J. L., Moore, P. B., and Steitz, T. A. (2003). Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit. *J Mol Biol* 330, 1061–1075.
- Harms, J., Schlutzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* 107, 679–688.
- Harms, J., Schlutzen, F., Zarivach, R., Bashan, A., Bartels, H., Agmon, I., and Yonath, A. (2002). Protein structure: experimental and theoretical aspects. *FEBS Lett* 525, 176.
- Harms, J., Schlutzen, F., Fucini, P., Bartels, H., and Yonath, A. (2004). Alterations at the peptidyl transferase centre of the ribosome induced by the synergistic action of the streptogramins dalbapristin and quinupristin. *BMC Biol* 2, 4;1–10.
- Helgstrand, M., Mandava, C. S., Mulder, F. A., Liljas, A., Sanyal, S., and Akke, M. (2007). The ribosomal stalk binds to translation factors IF2, EF-Tu, EF-G and RF3 via a conserved region of the L12 C-terminal domain. *J Mol Biol* 365, 468–479.
- Hope, H., Frolow, F., von Bohlen, K., Makowski, I., Kratky, C., Halfon, Y., Danz, H., Webster, P., Bartels, K. S., Wittmann, H. G., and et al. (1989). Cryocrystallography of ribosomal particles. *Acta Crystallogr B* 45, 190–199.
- Ippolito, J. A., Kanyo, Z. F., Wang, D., Franceschi, F. J., Moore, P. B., Steitz, T. A., and Duffy, E. M. (2008). Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit. *J Med Chem* 51, 3353–3356.
- Johansson, M., Bouakaz, E., Lovmar, M., and Ehrenberg, M. (2008). The kinetics of ribosomal peptidyl transfer revisited. *Mol Cell* 30, 589–598.
- Johnson, A. E. (2004). Functional ramifications of FRET-detected nascent chain folding far inside the membrane-bound ribosome. *Biochem Soc Trans* 32, 668–672.
- Kaiser, C. M., Chang, H. C., Agashe, V. R., Lakshmipathy, S. K., Etchells, S. A., Hayer-Hartl, M., Hartl, F. U., and Barral, J. M. (2006). Real-time observation of trigger factor function on translating ribosomes. *Nature* 444, 455–460.
- Katunin, V. I., Savelsbergh, A., Rodnina, M. V., and Wintermeyer, W. (2002). Coupling of GTP hydrolysis by elongation factor G to translocation and factor recycling on the ribosome. *Biochemistry* 41, 12806–12812.
- Kavran, J. M., and Steitz, T. A. (2007). Structure of the base of the L7/L12 stalk of the *Haloarcula marismortui* large ribosomal subunit: Analysis of L11 Movements. *J Mol Biol* 371, 1047–1059.
- Klein, D. J., Moore, P. B., and Steitz, T. A. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 340, 141–177.

- Konevega, A. L., Fischer, N., Semenov, Y. P., Stark, H., Wintermeyer, W., and Rodnina, M. V. (2007). Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. *Nat Struct Mol Biol* 14, 318–324.
- Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H. F. (2006). Crystal Structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126, 1065–1077.
- Lai, C. J., and Weisblum, B. (1971). Altered methylation of ribosomal RNA in an erythromycin-resistant strain of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 68, 856–860.
- Lang, K., Erlacher, M., Wilson, D. N., Micura, R., and Polacek, N. (2008). The role of 23S ribosomal RNA residue A2451 in peptide bond synthesis revealed by atomic mutagenesis. *Chem Biol* 15, 485–492.
- Lawrence, M., Lindahl, L., and Zengel, J. M. (2008). Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel. *J Bacteriol* 190, 5862–5869.
- Lolk, L., Pohlsgaard, J., Jepsen, A. S., Hansen, L. H., Nielsen, H., Steffansen, S. I., Sparving, L., Nielsen, A. B., Vester, B., and Nielsen, P. (2008). A click chemistry approach to pleuromutilin conjugates with nucleosides or acyclic nucleoside derivatives and their binding to the bacterial ribosome. *J Med Chem* 51, 4957–4967.
- Maguire, B. A., Beniaminov, A. D., Ramu, H., Mankin, A. S., and Zimmermann, R. A. (2005). A protein component at the heart of an RNA machine: the importance of protein 127 for the function of the bacterial ribosome. *Mol Cell* 20, 427–435.
- Maier, T., Ferbitz, L., Deuerling, E., and Ban, N. (2005). A cradle for new proteins: trigger factor at the ribosome. *Curr Opin Struct Biol* 15, 204–212.
- Malkin, L. I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol* 26, 329–346.
- Mankin, A. S. (2001). Ribosomal antibiotics. *Mol Biologia* 35, 509–520.
- Mankin, A. S. (2006). Nascent peptide in the “birth canal” of the ribosome. *Trends Biochem Sci* 31, 11–13.
- Mankin, A. S. (2008). Macrolide myths. *Curr Opin Microbiol* 11, 414–421.
- Marintchev, A., and Wagner, G. (2004). Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys* 37, 197–284.
- Martinez-Hackert, E., and Hendrickson, W. A. (2007). Structures of and interactions between domains of trigger factor from *Thermotoga maritima*. *Acta Crystallogr D Biol Crystallogr* 63, 536–547.
- Mears, J.A., Cannone, J.J., Stagg, S.M., Gutell, R.R., Agrawal, R.K. and Harvey, S.C (2002). Modeling a minimal ribosome based on comparative sequence analysis. *J Mol Biol.* 321, 215–234.
- Milligan, R. A., and Unwin, P. N. (1986). Location of exit channel for nascent protein in 80S ribosome. *Nature* 319, 693–695.
- Mitra, K., Schaffitzel, C., Fabiola, F., Chapman, M. S., Ban, N., and Frank, J. (2006). Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol Cell* 22, 533–543.
- Moore, P. B. (1988). The ribosome returns. *Nature* 331, 223–227.
- Moore, P. B., and Steitz, T. A. (2002). The involvement of RNA in ribosome function. *Nature* 418, 229–235.
- Moore, P. B., and Steitz, T. A. (2003). After the ribosome structures: How does peptidyl transferase work? *RNA* 9, 155–159.
- Moore, P. B., and Steitz, T. A. (2005). The ribosome revealed. *Trends Biochem Sci* 30, 281–283.
- Moore, S. D., and Sauer, R. T. (2008). Revisiting the mechanism of macrolide-antibiotic resistance mediated by ribosomal protein L22. *Proc Natl Acad Sci U S A* 105, 18261–18266.
- Murphy, F. V., and Ramakrishnan, V. (2004). Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* 11, 1251–1252.
- Murphy, F. V., Ramakrishnan, V., Malkiewicz, A., and Agris, P. F. (2004). The role of modifications in codon discrimination by tRNA(Lys)(UUU). *Nat Struct Mol Biol* 11, 1186–1191.

- Nagano, K., Takagi, H., and Harel, M. (1991). The side-by-side model of two tRNA molecules allowing the alpha-helical conformation of the nascent polypeptide during the ribosomal transpeptidation. *Biochimie* 73, 947–960.
- Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell* 108, 629–636.
- Nakatogawa, H., and Ito, K. (2004). Intra ribosomal regulation of expression and fate of proteins. *ChemBioChem* 5, 48–51.
- Nilsson, J., and Nissen, P. (2005). Elongation factors on the ribosome. *Curr Opin Struct Biol* 15, 349–354.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930.
- Noller, H. F., Hoffarth, V., and Zimmnick, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256, 1416–1419.
- Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292, 897–902.
- Ogle, J. M., Carter, A. P., and Ramakrishnan, V. (2003). Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci* 28, 259–266.
- Ogle, J. M., Murphy, F. V., Tarry, M. J., and Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* 111, 721–732.
- Ogle, J. M., and Ramakrishnan, V. (2005). Structural insights into translational fidelity. *Annu Rev Biochem* 74, 129–177.
- Palade, G. E. (1955). A small particulate component of the cytoplasm. *J Biophys Biochem Cytol* 1, 59–68.
- Passmore, L. A., Schmeing, T. M., Maag, D., Applefield, D. J., Acker, M. G., Algire, M. A., Lorsch, J. R., and Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* 26, 41–50.
- Petrone, P. M., Snow, C. D., Lucent, D., and Pande, V. S. (2008). Side-chain recognition and gating in the ribosome exit tunnel. *Proc Natl Acad Sci U S A* 105, 16549–16554.
- Petry, S., Brodersen, D. E., Murphy, F. V. t., Dunham, C. M., Selmer, M., Tarry, M. J., Kelley, A. C., and Ramakrishnan, V. (2005). Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* 123, 1255–1266.
- Petry, S., Weixlbaumer, A., and Ramakrishnan, V. (2008). The termination of translation. *Curr Opin Struct Biol* 18, 70–77.
- Pfister, P., Corti, N., Hobbie, S., Bruell, C., Zarivach, R., Yonath, A., and Bottger, E. C. (2005). 23S rRNA base pair 2057–2611 determines ketolide susceptibility and fitness cost of the macrolide resistance mutation 2058A → G. *Proc Natl Acad Sci U S A* 102, 5180–5185.
- Pioletti, M., Schlutzenzen, F., Harms, J., Zarivach, R., Gluehmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., et al. (2001). Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *Embo J* 20, 1829–1839.
- Poehlsgaard, J., and Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. *Nat Rev Microbiol* 3, 870–881.
- Polacek, N., Gaynor, M., Yassin, A., and Mankin, A. S. (2001). Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* 411, 498–501.
- Polacek, N., Gomez, M. J., Ito, K., Xiong, L., Nakamura, Y., and Mankin, A. (2003). The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol Cell* 11, 103–112.
- Polacek, N., and Mankin, A. S. (2005). The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit Rev Biochem Mol Biol* 40, 285–311.
- Pyetan, E., Baram, D., Auerbach-Nevo, T., and Yonath, A. (2007). Chemical parameters influencing fine-tuning in the binding of macrolide antibiotics to the ribosomal tunnel. *Pure Appl Chem* 79, 955–968.
- Ramakrishnan (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557–572.

- Ramakrishnan, V., and Moore, P. B. (2001). Atomic structures at last: the ribosome in 2000. *Curr Opin Struct Biol* 11, 144–154.
- Rodnina, M. V., Beringer, M., and Wintermeyer, W. (2007). How ribosomes make peptide bonds. *Trends Biochem Sci* 32, 20–26.
- Rodnina, M. V., and Wintermeyer, W. (2003). Peptide bond formation on the ribosome: structure and mechanism. *Curr Opin Struct Biol* 13, 334–340.
- Rutkowska, A., Mayer, M. P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E., and Bukau, B. (2008). Dynamics of trigger factor interaction with translating ribosomes. *J Biol Chem* 283, 4124–4132.
- Ryabova, L. A., Selivanova, O. M., Baranov, V. I., Vasiliev, V. D., and Spirin, A. S. (1988). Does the channel for nascent peptide exist inside the ribosome? Immune electron microscopy study. *FEBS Lett* 226, 255–260.
- Sabatini, D. D., and Blobel, G. (1970). Controlled proteolysis of nascent polypeptides in rat liver cell fractions. II. Location of the polypeptides in rough microsomes. *J Cell Biol* 45, 146–157.
- Saguy, M., Gillet, R., Metzinger, L., and Felden, B. (2005). tmRNA and associated ligands: a puzzling relationship. *Biochimie* 87, 897–903.
- Sato, N. S., Hirabayashi, N., Agmon, I., Yonath, A., and Suzuki, T. (2006). Comprehensive genetic selection revealed essential bases in the peptidyl-transferase center. *Proc Natl Acad Sci U S A* 103, 15386–15391.
- Schaffitzel, C., and Ban, N. (2007). Generation of ribosome nascent chain complexes for structural and functional studies. *J Struct Biol* 158, 463–471.
- Schlunzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* 102, 615–623.
- Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., and Franceschi, F. (2001). Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* 413, 814–821.
- Schlunzen, F., Harms, J. M., Franceschi, F., Hansen, H. A., Bartels, H., Zarivach, R., and Yonath, A. (2003). Structural basis for the antibiotic activity of ketolides and azalides. *Structure* 11, 329–338.
- Schlunzen, F., Pyetan, E., Fucini, P., Yonath, A., and Harms, J. (2004). Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin. *Mol Microbiol* 54, 1287–1294.
- Schlunzen, F., Wilson, D. N., Tian, P., Harms, J. M., McInnes, S. J., Hansen, H. A., Albrecht, R., Buerger, J., Wilbanks, S. M., and Fucini, P. (2005). The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure (Camb)* 13, 1685–1694.
- Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A., and Steitz, T. A. (2005a). Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol Cell* 20, 437–448.
- Schmeing, T. M., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005b). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* 438, 520–524.
- Schroeder, S. J., Blaha, G., Tirado-Rives, J., Steitz, T. A., and Moore, P. B. (2007). The structures of antibiotics bound to the E site region of the 50 S ribosomal subunit of *Haloarcula marismortui*: 13-deoxytetracycline and girodazole. *J Mol Biol* 367, 1471–1479.
- Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science* 310, 827–834.
- Selmer, M., Dunham, C. M., Murphy IV, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.
- Sharma, P. K., Xiang, Y., Kato, M., and Warshel, A. (2005). What are the roles of substrate-assisted catalysis and proximity effects in peptide bond formation by the ribosome? *Biochemistry* 44, 11307–11314.

- Shaw, J. J., and Green, R. (2007). Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol Cell* 28, 458–467.
- Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004). The ribosome as an entropy trap. *Proc Natl Acad Sci U S A* 101, 7897–7901.
- Simonovic, M., and Steitz, T. A. (2008). Peptidyl-CCA deacylation on the ribosome promoted by induced fit and the O3'-hydroxyl group of A76 of the unacylated A-site tRNA. *RNA* 14, 2372–2378.
- Sobolevsky, Y., and Trifonov, E. N. (2005). Conserved sequences of prokaryotic proteomes and their compositional age. *J Mol Evol* 61, 591–596.
- Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R., and van Heel, M. (1995). The 70S *Escherichia coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. *Structure* 3, 815–821.
- Stark, H., Rodnina, M. V., Wieden, H. J., van Heel, M., and Wintermeyer, W. (2000). Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* 100, 301–309.
- Steitz, T. A. (2008). A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* 9, 242–253.
- Steitz, T. A., and Moore, P. B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* 28, 411–418.
- Tenson, T., and Mankin, A. (2006). Antibiotics and the ribosome. *Mol Microbiol* 59, 1664–1677.
- Thompson, J., and Dahlberg, A. E. (2004). Testing the conservation of the translational machinery over evolution in diverse environments: assaying *Thermus thermophilus* ribosomes and initiation factors in a coupled transcription-translation system from *Escherichia coli*. *Nucleic Acids Res* 32, 5954–5961.
- Thompson, J., Kim, D. F., O'Connor, M., Lieberman, K. R., Bayfield, M. A., Gregory, S. T., Green, R., Noller, H. F., and Dahlberg, A. E. (2001). Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc Natl Acad Sci U S A* 98, 9002–9007.
- Thygesen, J., Weinstein, S., Franceschi, F., and Yonath, A. (1996). The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. *Structure* 4, 513–518.
- Trobro, S., and Aqvist, J. (2006). Analysis of predictions for the catalytic mechanism of ribosomal peptidyl transfer. *Biochemistry* 45, 7049–7056.
- Tu, D., Blaha, G., Moore, P. B., and Steitz, T. A. (2005). Structures of MLSBK Antibiotics Bound to Mutated Large Ribosomal Subunits Provide a Structural Explanation for Resistance. *Cell* 121, 257–270.
- Uemura, S., Dorywalska, M., Lee, T. H., Kim, H. D., Puglisi, J. D., and Chu, S. (2007). Peptide bond formation destabilizes Shine-Dalgarno interaction on the ribosome. *Nature* 446, 454–457.
- Valle, M., Gillet, R., Kaur, S., Henne, A., Ramakrishnan, V., and Frank, J. (2003a). Visualizing tmRNA entry into a stalled ribosome. *Science* 300, 127–130.
- Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003b). Locking and unlocking of ribosomal motions. *Cell* 114, 123–134.
- Vazquez-Laslop, N., Thum, C., and Mankin, A. S. (2008). Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell* 30, 190–202.
- von Bohlen, K., Makowski, I., Hansen, H. A., Bartels, H., Berkovitch-Yellin, Z., Zaytzev-Bashan, A., Meyer, S., Paulke, C., Franceschi, F., and Yonath, A. (1991). Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from *Haloarcula marismortui* diffracting to 3 Å resolution. *J Mol Biol* 222, 11–15.
- Voss, N. R., Gerstein, M., Steitz, T. A., and Moore, P. B. (2006). The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* 360, 893–906.
- Walter, P., and Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* 10, 87–119.
- Watson, J. D. (1963). Involvement of RNA in the synthesis of proteins. *Science* 140, 17–26.

- Weinger, J. S., Parnell, K. M., Dorner, S., Green, R., and Strobel, S. A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nat Struct Mol Biol* 11, 1101–1106.
- Weinger, J. S., and Strobel, S. A. (2006). Participation of the tRNA A76 Hydroxyl Groups throughout Translation. *Biochemistry* 45, 5939–5948.
- Weixlbaumer, A., Murphy, F. V. t., Dziergowska, A., Malkiewicz, A., Vendeix, F. A., Agris, P. F., and Ramakrishnan, V. (2007). Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat Struct Mol Biol* 14, 498–502.
- Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R. M., Petry, S., Kelley, A. C., and Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* 322, 953–956.
- Wekselman, I., Davidovich, C., Agmon, I., Zimmerman, E., Rosenberg, H., Bashan, A., Berisio, R., and Yonath, A. (2008). Ribosome's mode of function: myths and facts. *J Pept Sci* 15, 122–130.
- Wilden, B., Savelsbergh, A., Rodnina, M. V., and Wintermeyer, W. (2006). Role and timing of GTP binding and hydrolysis during EF-G-dependent tRNA translocation on the ribosome. *Proc Natl Acad Sci U S A* 103, 13670–13675.
- Wilson, D. N., Harms, J. M., Nierhaus, K. H., Schlunzen, F., and Fucini, P. (2005a). Species-specific antibiotic-ribosome interactions: implications for drug development. *Biol Chem* 386, 1239–1252.
- Wilson, D. N., Schlunzen, F., Harms, J. M., Yoshida, T., Ohkubo, T., Albrecht, R., Buerger, J., Kobayashi, Y., and Fucini, P. (2005b). X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit. *Embo J* 24, 251–260.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.
- Wohlgemuth, I., Beringer, M., and Rodnina, M. V. (2006). Rapid peptide bond formation on isolated 50S ribosomal subunits. *EMBO Rep* 7, 699–703.
- Woolhead, C. A., McCormick, P. J., and Johnson, A. E. (2004). Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* 116, 725–736.
- Woolhead, C. A., Johnson, A. E., and Bernstein, H. D. (2006). Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell* 22, 587–598.
- Xiong, L., Polacek, N., Sander, P., Bottger, E. C., and Mankin, A. (2001). pKa of adenine 2451 in the ribosomal peptidyl transferase center remains elusive. *RNA* 7, 1365–1369.
- Yassin, A., Fredrick, K., and Mankin, A. S. (2005). Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc Natl Acad Sci U S A* 102, 16620–16625.
- Yonath, A., Leonard, K. R., and Wittmann, H. G. (1987). A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science* 236, 813–816.
- Yonath, A., Muessig, J., Tesche, B., Lorenz, S., Erdmann, V. A., and Wittmann, H. G. (1980). Crystallization of the large ribosomal subunit from *B. stearothermophilus*. *Biochem Int* 1, 315–428.
- Yonath, A. (2002). High-resolution structures of large ribosomal subunits from mesophilic eubacteria and halophilic archaea at various functional states. *Curr Protein Peptide Sci* 3, 67–78.
- Yonath, A. (2003a). Ribosomal tolerance and peptide bond formation. *Biol Chem* 384, 1411–1419.
- Yonath, A. (2003b). Structural insight into functional aspects of ribosomal RNA targeting. *ChemBioChem* 4, 1008–1017.
- Yonath, A., and Bashan, A. (2004). Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics. *Annu Rev Microbiol* 58, 233–251.

- Yonath, A. (2005a). Antibiotics targeting ribosomes: resistance, selectivity, synergism, and cellular regulation. *Annu Rev Biochem* 74, 649–679.
- Yonath, A. (2005b). Ribosomal crystallography: peptide bond formation, chaperone assistance and antibiotics activity. *Mol Cells* 20, 1–16.
- Yonath, A. (2006). Molecular biology: triggering positive competition. *Nature* 444, 435–436.
- Youngman, E. M., Cochella, L., Brunelle, J. L., He, S., and Green, R. (2006). Two distinct conformations of the conserved RNA-rich decoding center of the small ribosomal subunit are recognized by tRNAs and release factors. *Cold Spring Harb Symp Quant Biol* 71, 545–549.
- Youngman, E. M., He, S. L., Nikstad, L. J., and Green, R. (2007). Stop codon recognition by release factors induces structural rearrangement of the ribosomal decoding center that is productive for peptide release. *Mol Cell* 28, 533–543.
- Youngman, E. M., McDonald, M. E., and Green, R. (2008). Peptide release on the ribosome: mechanism and implications for translational control. *Annu Rev Microbiol* 62, 353–373.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883–896.
- Yusupova, G., Jenner, L., Rees, B., Moras, D., and Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. *Nature* 444, 391–394.
- Zaman, S., Fitzpatrick, M., Lindahl, L., and Zengel, J. (2007). Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in *Escherichia coli*. *Mol Microbiol* 66, 1039–1050.
- Zarivach, R., Bashan, A., Berisio, R., Harms, J., Auerbach, T., Schlutzenzen, F., Bartels, H., Baram, D., Pyetan, E., Sittner, A., et al. (2004). Functional aspects of ribosomal architecture: symmetry, chirality and regulation. *J Phys Org Chem* 17, 901–912.
- Zavialov, A. V., Hauryliuk, V. V., and Ehrenberg, M. (2005). Guanine-nucleotide exchange on ribosome-bound elongation factor G initiates the translocation of tRNAs. *J Biol* 4, 9.
- Zimmerman, E., and Yonath, A. (2009). Biological Implications of the Ribosome's Stunning Stereochemistry. *ChemBioChem* 10, 63–72.
- Ziv, G., Haran, G., and Thirumalai, D. (2005). Ribosome exit tunnel can entropically stabilize {alpha}-helices. *Proc Natl Acad Sci U S A* 102, 18956–18961.

COURSE ABSTRACTS

COMPARATIVE BINDING ENERGY (COMBINE) ANALYSIS OF THE SUBSTRATE SPECIFICITY OF HALOALKANE DEHALOGENASE FROM *BRADYRHIZOBIUM JAPONICUM* USDA110

JAN BREZOVSKÝ¹, MARTA MONINCOVÁ¹,
ZBYNĚK PROKOP¹, YUKARI SATO^{2,3},
TOSHIYA SENDA³, YUJI NAGATA²,
REBECCA C. WADE⁴ AND JIŘÍ DAMBORSKÝ¹

¹*Loschmidt Laboratories, Faculty of Science, Masaryk University, Kamenice 5/A4, 625 00 Brno, Czech Republic*

²*Environmental Life Sciences, Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Sendai 980-8577, Japan*

³*National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Kotoku, Tokyo 135-0064, Japan*

⁴*EML Research, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany*

Abstract Haloalkane dehalogenase DbjA from *Bradyrhizobium japonicum* USDA110 is the novel enzyme [1] capable of catalyzing dehalogenating reactions with high enantioselectivity. One of the structural bases for distinguishing between *R* and *S* form of a substrate may arise from differential binding. Comparative Binding Energy (COMBINE) analysis [2] was employed for identification of amino acid residues involved in substrate binding. Twenty enzyme–substrate complexes were modeled using automated docking method implemented in program AUTODOCK 3.05 and refined with molecular mechanics energy minimization by program AMBER 8.1. Intermolecular interaction energy between substrate and DbjA enzyme were decomposed into van der Waals and electrostatic terms for individual amino acid residues. Partial least-square projection to latent structures (PLS) analysis [3] was used to establish relationships between energy contributions and apparent dissociation constants K_m measured for this set of substrates in our laboratory.

The final model was composed of three statistically significant principal components and explained 93.5% (88.1% cross-validated) of quantitative variance in K_m . Binding of substrates to the active site of DbjA is primarily governed by van der Waals interactions with amino acid residues forming the active site, but electrostatic interactions with amino acid residues located

near the protein surface were also identified as important for explanation of binding of substrates with multiple strongly electronegative atoms.

Overall 25 amino acid residues provide most important interactions for the substrate specificity of DbjA: (i) first-shell residues Asn38, Asp103, Trp104, Ile129, Phe138, Phe157, Phe160, Phe176, Gly184, Pro214, Leu217, Leu254, His280 and Tyr281, (ii) second-shell residues Gly37, Pro39, Ala101, Gln102, Glu127, Leu181, Val210 and Phe213, (iii) tunnel-forming residues Val180 and Gly183, and (iv) distant charged residue Arg161. Interactions with only 12 out of 305 of protein residues (4%) were able to quantitatively describe differences in binding of majority of investigated substrates with 90.3% (81.7% cross-validated) quantitative variance in K_m . The residues belonging to the groups ii–iv represent suitable targets for site-directed mutagenesis or directed evolution aimed to modulate substrate specificity and enantioselectivity of DbjA.

References

1. Sato, Y., Monincova, M., Chaloupkova, R., Prokop, Z., Ohtsubo, Y., Minamisawa, K., Tsuda, M., Damborsky, J., Nagata, Y., 2005, *Appl. Environ. Microbiol.* 71: 4372–4379.
2. Ortiz, A. R., Pisabarro, M. T., Gago, F., Wade, R. C., 1995, *J. Med. Chem.* 38: 2681–2691.
3. Geladi, P., Kowalski, B. R., 1986, *Anal. Chim. Acta* 185: 1–17.

DISSECTING GENETICS OF HOST-PATHOGEN INTERACTIONS

PETER HRABER

Los Alamos National Laboratory

*Theoretical Biology & Biophysics, MS K710, Los Alamos NM
87545*

Abstract Methods in computational and molecular biology generally study the genetic processes of one model organism at a time. In reality, life is anything but a pure culture. Pathosystems extend the single-species paradigm across replicating organisms and present new problems, because infected cells house both host and pathogen. Computational solutions to challenges in the analysis of pathosystem data will be reviewed.

THE CY3/CY5 FRET PAIR FROM NUISANCE TO NEW SENSE

ASIF IQBAL^{1,*} DAVID NORMAN¹, BURAK OKUMUS²,
SINAN ARSLAN², CATHERINE THOMPSON³,
TAEKJIP HA⁴, AND DAVID LILLEY¹

¹*University of Dundee, College of Life Sciences, Dundee DD1 5EH*

²*University of Illinois at Urbana-Champaign*

³*School of Biological and Chemical Sciences, Birkbeck
University of London*

⁴*University of Illinois at Urbana-Champaign*

**Corresponding author: asif-1.iqbal@novartis.com*

Abstract Fluorescence resonance energy transfer (FRET) provides valuable long-range distance information about biological molecules in solution. Cyanine 3 and Cyanine 5 are an important donor–acceptor pair of fluorophores which are predominantly used in single molecule FRET experiments. Two of the major uncertainties in the use of distance information arising from FRET measurements are a lack of knowledge concerning the location of the fluorophores relative to the nucleic acid that they are attached to and also the lack of knowledge regarding the orientation of the fluorophores relative to each other. Understanding and defining the positions of these fluorophores is crucial for exact distance measurements and accurate analysis of FRET data. Knowing their exact locations will further allow to measure distances from a known fixed point with greater certainty which will in turn enhance the accuracy of fluorescence measurements.

In this study a NMR analysis of a self-complementary DNA duplex 5' labeled with Cy5 shows that the fluorophore is stacked onto the end of the helix in a manner similar to that of an additional base pair. This structure determined for Cy5-DNA bears close resemblance with an earlier study that defined the position of the Cy3-DNA which also stacked in a similar manner. Since the cyanine dyes are situated on the helical axis in a relatively fixed position on the end of DNA duplex this could have implications for the orientation factor which could lead to a FRET efficiency which is not only distance dependent but also orientation dependent. To see whether this characteristic modulation exists steady-state FRET measurements were done on a series of duplexes of varying length. This revealed that the FRET efficiency is orientation dependent however the observed modulation for both polymers is less than that calculated for a fully rigid attachment of the fluorophores suggesting that the FRET is averaged by some dynamic process.

MOLECULAR INTERACTION MAP (MIM) OF P73

EZGI KARACA AND MELDA YAVUZTURK

Bogazici University, Polymer Research Center (PRC), Chemical Engineering Department, 34342, Bebek, Istanbul

Abstract The protein of interest in this project is p73, a homolog of p53. The aim is to construct the interaction map of p73, and compare its features with the interaction map of p53. In the construction of molecular interaction map of p73, notations proposed by Kohn et.al. are used. p73 has many isoforms, constructed by truncation of N-terminus, C-terminus or both. Due to existence of these isoforms, the domain specificity of interactions became more significant. Therefore, an additional symbol is used in combination with the domain representation of Kohn et.al.

The molecular interaction map (MIM) constructed for p73 is consisted of 55 proteins and 122 interactions at this moment. A comparison of the functions of common proteins in p53 and p73 MIMs is made. This is the first step of a series of analyses that are to be made comparing p53 and p73. In the future work the similar and distinctive motifs of two networks will be searched. In addition to comparison with p53, proteins that are thought to be significant in p73 network are also to be further analyzed, both sequence wise and structure wise.

A METHOD TO RECOGNIZE MISFOLDED MODELS OF α -HELICAL TRANSMEMBRANE PROTEIN DOMAINS

NOVOSELETSKY V. N., CHUGUNOV A. O AND EFREMOV R. G.

*Laboratory of Biomolecular Modeling
Department of Structural Biology, Shemyakin & Ovchinnikov
Institute of Bioorganic Chemistry,
Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10,
GSP Moscow 117997, Russia*

Abstract Integral membrane proteins (MP) are objects of exceptional importance, but experimental methods often fail to determine their spatial structures. To solve the problem, one can use computational modeling techniques, however the low quality of the MP models constructed in silico is a major shortcoming of these methods. And once a model is proposed, a common problem of assessing its quality arises. Such an assessment is necessary for subsequent model refinement and application to biomedical studies.

Earlier we have described the “membrane score” method [1] for assessment of packing quality of transmembrane domains of α -helical MPs. In brief, the score of the whole structure is $S = \sum S_j^i(F_p^{-1}, F_{np}^{-1})$, where S_j^i is the score of a residue j of type i , and F_p^{-1}, F_{np}^{-1} are the fractions of its side chain surface area that are covered by polar (nitrogen or oxygen) and nonpolar atoms of neighboring helices in the transmembrane (TM) domain, respectively. In our previous work we used the concept of “classes of environment” [2] and a discrete function $S_i(F_p^{-1}, F_{np}^{-1})$. Here we present the further improvement of the method.

To enhance the accuracy of the “membrane score” we introduce a set of continuous functions $S_i(F_p^{-1}, F_{np}^{-1})$ for each type of residues. This eliminates possible sharp changes in S corresponding to slight variations in F_p^{-1}, F_{np}^{-1} , thus making the method more robust.

Furthermore we tested the ability of our method to distinguish between correct and misfolded structures. With this aim in view we built a number of theoretical models for some MP of various TM-length for which experimental structures are known. This was done using templates of different sequence relativeness. Assessment of the models with the new version of “membrane score” revealed strong negative correlation between the score value and rmsd. Based on this observation we introduce a “3Å-line” as a criterion for validity of the model structure. This line can be thought of as a separation between regions in “score-rmsd” coordinate plane where theoretical models demonstrate rmsd from reference structure less and more than 3 Å, respectively.

References

1. Chugunov AO, Novoseletsky VN, Nolde DE, Arseniev AS, Efremov RG. *J Chem Inf Model*. 2007 May 29;47(3):1150–1162.
2. Bowie JU, Luthy R, Eisenberg D. *Science*. 1991 Jul 12;253(5016):164–170.

FRAGMENT-BASED ATP-ORIENTED SCORING CRITERION TO RE-RANK DOCKING SOLUTIONS

TIMOTHY V. PYRKOV, YURI A. KOSINSKY AND
ROMAN G. EFREMOV

*Laboratory of Biomolecular Modeling
Department of Structural Biology, Shemyakin & Ovchinnikov
Institute of Bioorganic Chemistry, Russian Academy of Sciences,
ul. Mikhukho-Maklaya 16/10, GSP Moscow 117997, Russia*

Abstract Numerous biochemical reactions in living cells involve interactions of ATP and its analogues with different proteins. Considerable insight into the problem of molecular recognition of this substrate may be addressed by molecular docking simulations. Yet standard scoring functions are often insufficient to distinguish between correct and misleading solutions generated by docking algorithms. To improve the selection of correct ATP poses post-docking re-ranking criteria may be applied.

In this work a new ranking method for ATP-protein complexes is described. The method is based on detailed analysis of the intermolecular interactions in 50 high-resolution 3D-structure ATP-protein complexes. We found that the most important factor governing the recognition of ATP is the complementarity of hydrophobic and hydrophilic properties between the adenine moiety and the phosphate tail of the ligand, respectively (Pyrkov et al., *PROTEINS* 66, 388–398, 2007).

A distinctive new feature of the proposed method is that the ligand molecule is divided into fragments that differ in their physical properties. The placement of each of them is then judged separately by different criteria. Such approach avoids undesirable averaging of the scoring function terms by highlighting those relevant for particular fragments.

The scoring performance of the new criteria was tested with the docking solutions for ATP-protein complexes generated with GOLD (Jones et al., *J Mol Biol* 267, 727–748, 1997) and a significant improvement in the selection of correct docking poses was observed as compared to the standard scoring function.

References

- Pyrkov, T.V., Kosinsky, Y.A., Arseniev, A.S., Priestle, J.P., Jacoby, E., Efremov, R.G. (2007). *Proteins*. Feb 1;66(2):388–398.
Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. (1997). *J Mol Biol*. Apr 4;267 (3):727–748.

PROTEIN RECOGNITION OF VANADYL MACROCYCLES

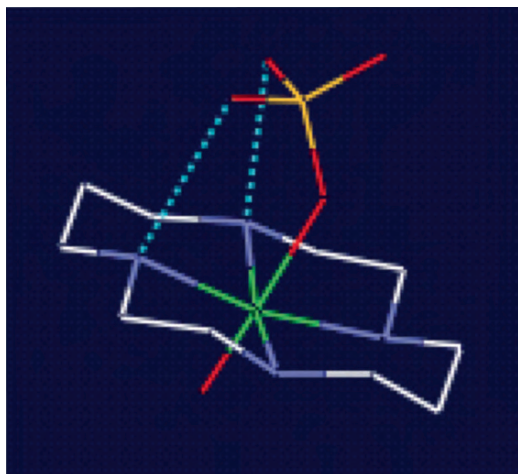
ALLISON ROSS¹, DINESH C. SOARES¹, NEIL ROBERTSON¹, TINA M. HUNTER, SIMON PARSONS¹ AND PETER J. SADLER^{1,2}

¹*School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK*

²*Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL UK*

Abstract Some macrocycles, including cyclam derivatives, show promise as stem cell mobilizers and antivirals, including activity against HIV and AIDS. The specific configurations adopted by metal cyclam complexes may be important for receptor recognition and biological activity.

We have synthesised novel vanadium(IV) cyclam and bicyclam complexes with unusual axial sulfate groups. The monocyclam complex crystallises in the thermodynamically stable *trans*-III configuration and contains a characteristic V = O unit with a *trans* sulfate group, stabilised via H-bonding to ring NH protons. Mass spectrometry confirmed the presence of two vanadyl sulfate groups for the bicyclam complex. The complexes have been studied by EPR, UV-Vis and resonance Raman spectroscopy. Magnetic susceptibility measurements confirmed that the monocyclam complex contains V(IV). Protein recognition is being investigated since the CXCR4 coreceptor is involved in HIV entry into cells and stem cell anchoring.



Keywords: Vanadium; Macrocyclic; Antiviral; Protein recognition

Acknowledgements We thank the BBSRC (studentship for A Ross) for support, and Drs. Abraha Habtemarian, Fiona MacKay, Ana Pizarro and Arindam Mukherjee for assistance and advice.

References

1. Hunter T M, McNae I W, Liang X, Bella J, Parsons S, Walkinshaw M D, Sadler P J (2005) *Proc Natl Acad Sci USA* 102: 2288–2292.
2. Liang X, Parkinson J A, Weishaupl M, Gould R O, Paisey S J, Park H, Hunter T M, Blindauer C A, Parsons S, Sadler P J (2002) *J Am Chem Soc* 124: 9105–9112.

HOW TO MEASURE STRUCTURAL KINETICS IN BIO-MACROMOLECULE?

IRIT SAGI

*The Weizmann Institute Of Science, Department Of Structural
Biology, Helen and Milton A. Kimmelman Building, Rehovot,
Israel 76100*

Abstract This lecture introduces the importance of real-time quantification of biophysical properties and structural kinetic behavior applied to the study of enzymes and metalloenzymes to dissect their molecular mechanisms. Protein motion and dynamics manifest in backbone and side-chain mobility seem to play a crucial role in protein stability, function, and reactivity. Dynamic processes occur in proteins on many different time scales. Surprisingly, very limited information is available on how to correlate protein motions occurring on short time scales (pico to nanoseconds) and longer time scales (micro to milliseconds) with protein function and enzyme catalysis.

Being proteins, enzymes are flexible moieties whose structures exhibit dynamic fluctuations and structural kinetics on a wide range of time scales. The inherent mobility of a protein fold results in protein conformational transitions and has been shown to be manifested in the various steps constituting the catalytic cycle. The nature of this linkage between protein structure movement and function is undoubtedly complex and might involve the formation of a coupled network of interactions [1]. However, the molecular basis to link protein dynamics and catalytic chemistry to key kinetic,

electronic, and structural events have remained elusive because of the difficulties associated with probing time-dependent structure-function aspects of enzymatic reactions.

The use of conventional and real-time structural-spectroscopic tools to study the reactive metal site during metalloenzyme catalysis is reviewed. Stopped-Flow X-ray absorption spectroscopy and transient kinetic analyses are presented as novel experimental approaches to study enzyme conformational transitions during catalysis [2]. Approaches such as time-resolved structural analysis may be useful in the design of specific inhibitors as drug candidates.

References

1. Sagi, I. and Milla, M. (2008) Application of structural-dynamic approaches provide novel insights into the enzymatic mechanism of the tumor necrosis factor- α converting enzyme (TACE). *Analytical Biochemistry*, 372, 1–10.
2. Kleinfeld, O., Frenkel, A., Martin, J.M.L., Sagi, I. (2003) Active Site Electronic Structure and Dynamics during Metalloenzyme Catalysis. *Nature Structural Biology* 10, 98–103.

STRUCTURAL-KINETICS OF CANCER PROTEASES: APPLICATION TO DRUG DESIGN

IRIT SAGI

*The Weizmann Institute Of Science, Department Of Structural
Biology, Helen and Milton A. Kimmelman Building, Rehovot,
Israel 76100*

Abstract Zinc dependent metalloproteinases comprise a large family of structurally homologous enzymes with a wide variety of biological roles. Originally described as proteinases involved in extracellular matrix (ECM) catabolism, these enzymes were later found to serve major roles as initiators of signaling pathways in many aspects of biology, ranging from cell proliferation, differentiation and communication, to pathological states associated with tumor metastasis, inflammation, tissue degeneration and cell death. From these enzymes, TACE stands out as a central shedding activity mediating the regulated release of a host of cytokines, receptors and other cell surface molecules [1–2]. Selective drugs targeted at blocking TACE for treatment of rheumatoid arthritis and other disease indications are highly sought [3–4]. Yet, the structural and chemical knowledge underlying its enzymatic

activity is very limited. This is in part due to the fact that the catalytic zinc atom of metalloproteinases is usually spectroscopically silent and hence difficult to study using conventional spectroscopic and analytical tools [5]. Most structural and biochemical studies, as well as medicinal chemistry efforts carried out so far were limited to non-dynamic structure/function characterization. Thus, to date, our mechanistic knowledge comes from theoretical calculations derived from static crystal structures from family members that are highly similar in their amino acid sequence and three-dimensional structure.

We have recently introduced a structural-dynamic experimental strategy to the field of metalloproteinases. Specifically, this strategy employs stopped-flow freeze-quench X-ray absorption spectroscopy (XAS) in conjunction with transient kinetic studies to probe changes in the structure and reactivity of the catalytic zinc-protein complex residing in the catalytic cleft of metalloproteinases in real-time. XAS reports accurately on metal-protein bond distances, coordination numbers and total effective charges of designated metal centers residing in metalloproteins. XAS can be conducted on diluted protein solutions and in real-time [5]. Applying this methodology to TACE, we have quantified the structure, electronics, and lifetime of the evolving zinc-protein reaction intermediates during the peptide hydrolysis reaction using a peptide substrate [6–7]. This allowed us to visualize changes in a catalytic center that researchers have, until now, been unable to investigate. Importantly, we correlated for the first time the distinct kinetic phases involved in the proteolytic reaction and key structural and electronic intermediates evolving at TACE active site.

Biophysical quantification of large, slow-moving protein conformational changes is of major importance for understanding molecular mechanisms and biological function. In addition, this information may be applied to the design of novel inhibitors aimed at selective conformation transitions rather than active site pocket chemistry.

References

1. Moss, M.L., Jin, C., Milla, M.E., Bickett, D. M., Buckhart, W., Carter, H.L., Chen, W.-J., Clay, W.C., Didsbury, J., Hassler, D., Kost, T.A., Lambert, M.H., Leesnitzer, M.A., McCauley, P., McGeehan, G., Moyer, M., Pahel, G.L., Rocque, W., Seaton, T., Su, J.-L., Warner, J. and Willard, D. Becherer, D. (1997) Cloning of a disintegrin metalloproteinase that processes precursor tumor-necrosis factor- α . *Nature* 385, 733–736.
2. Black, R.A., Rauch, C.T., Kozlosky, C.J., Peschon, J.J., Slack, J.L., Wolfson, M.F., Castner, B.J., Stocking, K.L., Reddy, P., Srinivasan, S., Nelson, N., Boiani, N., Schooley, K.A., Gerhart, M., Davis, R., Fitzner, J.N., Johnson, R.S., Paxton, R.J., March, C.J. and Cerretti, D.P. (1997) A metalloproteinase disintegrin that releases tumour necrosis factor- α from cells. *Nature* 385, 729–733.
3. Old, L.J. (1985) Tumor necrosis factor. *Science* 230, 630–632.

4. Beutler, B. and Cerami, A. (1988) Tumor necrosis, cachexia, shock and inflammation: a common mediator. *Annu. Rev. Biochem.* 57, 505–518.
5. Kleifeld, O., Frenkel, A., Martin, J.M.L., Sagi, I. (2003) Active Site Electronic Structure and Dynamics during Metalloenzyme Catalysis. *Nature Structural Biology* 10, 98–103.
6. Solomon, A., Akabayov, B., Milla, M., Sagi, I. (2007). Key feature of the catalytic cycle of TNF- α converting enzyme involves communication between distal protein sites and the enzyme catalytic core. *Proc Natl Acad Sci U S A*, 104, 4931–4936.
7. Sagi, I. and Milla, M. (2008) Application of structural-dynamic approaches provide novel insights into the enzymatic mechanism of the tumor necrosis factor- α . converting enzyme (TACE). *Analytical Biochemistry*, 372, 1–10.

ELECTRIC FIELDS IN BIOMOLECULES: ARE THEY EXPERIMENTALLY MEASURABLE?

ALEKSANDR B. SAHAKYAN, ALEKSAN G.
SHAHKHATUNI, ASTGHIK A. SHAHKHATUNI AND
HENRY A. PANOSYAN

*Molecule Structure Research Center, National Academy of
Sciences, Azatutian Avenue 26, Yerevan 375014, Armenia*

Abstract Electric fields play pivotal role in biomolecular processes. They drive the folding process, are important for molecular recognition, particularly for macromolecule–ligand and protein–protein recognition. Despite the importance of the issue, there are very few experimental methods for electric field detection in biomolecules. Unfortunately, even that few methods lack of universality and have many restrictions. Thus, the main source of electric field information in proteins is continuing to be calculations but they are too rigid against changes.

In this study we have tried to look for such kind of detector among NMR parameters. Taking into account the results of DFT computations and some observations from the NMR of aligned molecules, we have been concentrated on one bond J couplings in amino acid residues. Comparison between the deviations of abovementioned parameters measured in several proteins and calculation show an interesting correlation.

It is very challenging to find an experimental method for electric field detection among NMR parameters, which can be more convenient to extract from standard NMR experiments during protein geometry determination. From the other side, NMR provides us an opportunity to study protein dynamics and folding, thus such kind of detector can lead to simultaneous determination of electrostatic potential evolution during those processes.

UDP-D-MANNAC3NACA BIOSYNTHESIS IN PSEUDOMONAS SPP. AND BORDETELLA SPP.: IDENTICAL PATHWAY DESPITE DISTINCT GENOMIC CONTEXT

ERIN L WESTMAN AND JOSEPH S LAM

Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada

Abstract *Pseudomonas aeruginosa* serogroup O2 and *Bordetella pertussis* produce lipopolysaccharide (LPS) that contains 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid (UDP-D-ManNAc3NAcA) as a constituent of the B-band O antigen or band-A trisaccharide, respectively. Genetic and biochemical studies using *P. aeruginosa* provided the basis for a proposed biosynthetic pathway consisting of five enzymes, encoded in the LPS O antigen biosynthetic cluster.

The band-A LPS biosynthetic cluster in *B. pertussis* contains four genes that encode four of the five enzymes thought to be required for UDP-D-ManNAc3NAcA synthesis. Surprisingly, a gene encoding a putative dehydrogenase, which is required to catalyze the first step in the pathway in *P. aeruginosa*, was missing from this cluster. Thus, it was unclear whether synthesis of UDP-D-ManNAc3NAcA in *B. pertussis* could follow the same scheme established for *P. aeruginosa*. Cross-complementation analysis demonstrated that the four initial sequence-based homolog assignments were correct, as the relevant *B. pertussis* genes from the band-A biosynthetic cluster were supplied *in trans* to the relevant *P. aeruginosa* knockout mutants and were shown to restore B-band LPS production.

Further analysis of the *B. pertussis* genome sequence documented two putative dehydrogenases outside of the band-A biosynthesis cluster that might be able to provide the missing activity. Both putative dehydrogenases were individually capable of cross-complementing the *P. aeruginosa* dehydrogenase knockout. Taken collectively, these results suggest that *B. pertussis* produces UDP-D-ManNAc3NAcA through the same pathway proposed in *P. aeruginosa*, despite differences in the genomic context of the genes involved.

COURSE POSTERS**¹³C-DETECTION IN RNA BASES: REVEALING BASE-PAIRING FROM SMALL CHEMICAL SHIFT PERTURBATION**

CRISTOPHE FARES, IRENE AMATA
AND TERESA CARLOMAGNO

*Max Planck Institute for Biophysical Chemistry, Department
of NMR-Based Structural Biology, Am Fassberg 11, D-37077,
Goettingen, Germany*

**PRACTICAL METHODS OF APPLICATION OF QUANTITATIVE
NMR SPECTROSCOPY**

KH.M. BOBAKULOV, M.G. LEVKOVICH,
N.D. ABDULLAEV AND S. YUNUSOV

*Institute of Chemistry of Plant Substances, Academy of Sciences
Ruz, 77, Kh. Abdullaev Street, 100170, Tashkent, Uzbekistan*

**CHANGES IN DYNAMICS AND INTERACTION OF CHICKEN
LIVER BILE ACID BINDING PROTEIN INDUCED BY THE
PRESENCE OF A DISULPHIDE BRIDGE: MUTAGENESIS
AND NMR STUDY**

CLELIA COGLIATI¹, SIMONA TOMASELLI¹,
LAURA RAGONA¹, MICHAEL ASSFALG², MARA
GUARIENTO², MASSIMO PEDÒ², SERENA ZANZONI²,
MARIAPINA D'ONOFRIO², LUCIA ZETTA¹ AND
HENRIETTE MOLINARI²

¹*Laboratorio NMR, ISMAC, CNR, via Bassini 15, 20133
Milano, Italy*

²*Dipartimento Scientifico e Tecnologico, Università di Verona,
Strada Le Grazie 15, 37134 Verona, Italy*

NMR STUDIES OF A SPLICEOSOME RELATED LARIAT FORMING RIBOZYME

MELANIE FALB, JÖRG FOHRER, CLAUDIA SCHWIEGK AND TERESA CARLOMAGNO

Max Planck Institute for Biophysical Chemistry, Department of NMR-Based Structural Biology, Am Fassberg 11, D-37077, Goettingen, Germany

TB SHIKIMATE KINASE DYNAMICS BY NMR

NICOLE E. FREEDMAN¹, BENJAMIN E. RAMIREZ²,
ANDREW D. MESECAR¹, SCOTT G. FRANZBLAU¹
AND MICHAEL CAFFREY²

University of Illinois at Chicago

¹ *Department of Medicinal Chemistry and Pharmacognosy*

² *Department of Biochemistry and Molecular Genetics*

Center for Pharmaceutical Biotechnology, Molecular Biology Research Building 900 S Ashland Avenue – M/C 870, Chicago, Illinois USA 60607-7173

Abstract Tuberculosis (TB) is considered a worldwide health threat. Approximately 2 billion people, or a third of the world's population, are infected with the TB bacilli. The high susceptibility of immunodeficient populations and the emergence of multi-drug resistant (MDR) and extensively drug resistant (XDR) TB contribute strongly to the reemergence of TB as a World health threat. These facts underscore the need for new drug development and new drug targets to combat TB.

The shikimate pathway provides attractive and new drug targets for anti-TB drug development for many reasons. The shikimate pathway synthesizes chorismate which is the precursor for many essential metabolites including the aromatic amino acids, folate, para-aminobenzoic acid, and quinones. The shikimate pathway is found in bacteria, plants, fungi, and apicomplexan parasites, but is not found in mammals. Shikimate kinase catalyzes the fifth step in the shikimate pathway. The gene encoding shikimate kinase, AroK,

has been shown by homologous recombination knockout studies to be essential in *Mycobacterium tuberculosis*.

We have cloned the AroK gene into a pET 17b vector, to produce the Shikimate kinase enzyme with no tags. We have purified and isotope labeled the enzyme. Backbone assignments were made and side chain assignments are underway. Drug discovery by NMR is underway on a small library of 172 compounds based on an oxidole scaffold.

STRUCTURAL STUDIES ON TWO HEME BINDING PROTEINS FROM THE HBP/SOUL FAMILY; AN INTEGRATED STUDY BY CRYSTALLOGRAPHY AND NMR

FILIPPE FREIRE¹, ANA L. CARVALHO¹, JORGE S. DIAS¹,
GLORIA C. FERREIRA², BRIAN F. VOLKMAN³,
BRIAN J. GOODFELLOW⁴, MARIA J. ROMAO¹
AND ANJOS L. MACEDO¹

¹*REQUIMTE, Det. de Química, Fac. Ciências e Tecnologia,
Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*

²*Department of Biochemistry and Molecular Biology, College
of Medicine, USF, Tampa, FL 33612-4799, USA*

³*Department of Biochemistry, Medical College of Wisconsin,
Milwaukee, WI 53226, USA*

⁴*CICECO, Department de Química, Universidade de Aveiro,
3810-193 Aveiro, Portugal*

E-mail: filipefreire@dq.fct.unl.pt

COMPARISON OF X-RAY AND NMR RESOLVED STRUCTURES OF THE SAME PROTEINS

SERGIY O. GARBUZYNSKIY

*Russian Academy of Sciences, Laboratory of Protein Physics,
Institute of Protein Research, Institutskaya Street, 4, Pushchino,
142290, Moscow Region, Russia*

IN SEARCH OF NEW BILE ACID-BINDING PROTEINS

MARA GUARIENTO¹, MICHAEL ASSFALG¹,
DOMENICO RAIMONDO², SERENA ZANZONI¹,
PATRIZIA PESENTE³, LAURA RAGONA⁴, LUCIA
ZETTA⁴, ANNA TRAMONTANO² AND HENRIETTE
MOLINARI¹

*¹Scientific and Technological Department, University of
Verona, Strada Le Grazie 15, 37134 Verona, Italy*

*²Department of Biochemical Sciences, University of Rome, La
Sapienza' 00185 Rome, Italy*

*³Laboratorio Tre Valli, Corte Pellegrina, San Martino Buon
Albergo, 37036 Verona, Italy*

*⁴Laboratorio NMR, ISMAC, CNR, via Bassini 15, 20133
Milano, Italy*

HOST RECOGNITION BY COMPLEMENT FACTOR H

HERBERT AP, DEAKIN JA, SCHMIDT CQ, BLAUM BS,
EGAN C, FERREIRA VP, PANGBURN MK, LYON M,
UHRIN D AND BARLOW PN.

*Department of Chemistry, University of Edinburgh, West Mains
Road, Edinburgh, UK, EH9 3JJ*

THE CY3/CY5 FRET PAIR FROM NUISANCE TO NEW SENSE

ASIF IQBAL, DAVID NORMAN AND DAVID LILLEY

*University of Dundee, College of Life Sciences, Dundee DD1
5EH*

LINKING THE ALLOSTERIC TRANSITIONS OF GROEL WITH PROTEIN FOLDING

YAKOV KIPNIS

Department of Structural Biology, Weizmann Institute of Science, P.O. Box 26, Rehovot 76100, Israel

CONFORMATIONAL CHANGE OF N-HEXYL-7-(3-OCTYLUREIDO)-1H-INDOLE-2-CARBOXAMIDE UPON ANION INTERACTION

DAMJAN MAKUC AND JANEZ PLAVEC

Slovenian NMR Centre, National Institute of Chemistry, P.O. Box 660, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

INVESTIGATION OF PATHOLOGIC TISSUES AT MARFAN SYNDROME

J. MONASELIDZE³, N. ESIPOVA¹, V. TUMANYAN¹, E. KATOVSKAIA², M. GORGOSHIDZE³, SH. BARBAKADZE³ AND I. MESROPYAN³

¹Enhelgardt Institute of Molecular Biology, RAS, 32 Vavilov Street, Moscow, 119991, Russia

²State Establishment Medical Social Committee of Experts, Moscow, Russia

³E. Andronikashvili Institute of Physics, 6 Tamarashvili Street, Tbilisi, 0177, Georgia

NMR STRUCTURAL INVESTIGATION OF VEGF AND VAMMIN DERIVED PEPTIDES

YASMINA MIRASSOU ARGÜELLO¹, JOSÉ M. PÉREZ-CAÑADILLAS¹, M. JESÚS PÉREZ DE VEGA², ROSARIO GONZÁLEZ-MUÑIZ², M. ANGELES JIMÉNEZ¹

*¹Instituto de Química-Física Rocasolano, CSIC, Madrid
E-mail: yasmına@iqfr.csic.es*

²Instituto de Química Médica, CSIC, Madrid

STRUCTURAL STUDIES ON PROTEINS OF THE CALYCIN SUPERFAMILY: HUMAN GLYCODELIN AND CHICKEN LIVER BILE ACID BINDING PROTEIN

PEDÒ M., D'ONOFRIO M., ANJALI A.KARANDE,
MONACO H.L AND MOLINARI H.

*University of Verona, Nuclear Magnetic Resonance Laboratory,
Strada le Grazie 15 Cà Vignal 1*

REGULATOR ROLE OF PLASTOQUINONES IN STRESS REACTION OF PLANTS UNDER HEAT TREATMENT

NATALLIA L. PSYBYTKO¹, JERZY KRUK²,
LIUDMILA F. KABASHNIKOVA¹ AND KAZIMIERZ
STRZALKA²

*¹Institute of Biophysics and Cell Engineering, National
Academy of Sciences of Belarus, Minsk, Belarus*

*²Faculty of Biochemistry, Biophysics and Biotechnology,
Jagiellonian University, Krakow, Poland*

PROTEIN RECOGNITION OF VANADYL MACROCYCLES

ALLISON ROSS¹, DINESH SOARES¹,
NEIL ROBERTSON¹, SIMON PARSONS¹
AND PETER J SADLER^{1,2}

¹School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ

²Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL

**THE LA PROTEIN: THE NEW FEATURES INVOLVED
IN RNA BINDING**

DOMENICO SANFELICE

*Dipartimento di Chimica, Università di Napoli "Federico II",
Stanza 2MC-25, Via Cintia 80126, Napoli, Italia*

**THE QUANTUM CHEMICAL STUDY OF SOME
BISBENZYLISOQUINOLINE DERIVATIVES COMPLEXES
WITH ALKALI EARTH CATIONS**

IOANA STANCULESCU¹, CRISTINA MANDRAVEL¹,
LEONARDO BERNASCONI² AND EVERT JAN
BAERENDS²

*¹Department of Physical Chemistry, University of Bucharest, 4-12 Bd. Regina Elisabeta, District 3, 030018 Bucharest, Romania
E-mail: ioana@gw-chimie.math.unibuc.ro, <http://gw-chimie.math.unibuc.ro>*

*²Free University of Amsterdam, Faculty of Sciences,
Department of Chemistry and Pharmaceutical Sciences,
Theoretical Chemistry Group, De Boelelaan 1083, 1081 HV
Amsterdam, The Netherlands*

**FUNCTIONAL ANALYSIS OF CONSERVED ACTIVE SITE
ARGININES IN THE HHA1 DNA METHYLTRANSFERASE**

ZDISLAV STASEVSKIJ, LINA LEINARTAITĖ, DALIA
DAUJOTYTĖ AND SAULIUS KLIMASAUŠKAS

*Laboratory of Biological DNA Modification, Institute of
Biotechnology, LT-02241 Vilnius, Lithuania*

**GEOMETRY OF WATER MOLECULES HYDROGEN BONDS
AND COLLAGEN-DNA COMPLEX**

DAVID SVINTRADZE

*Department of Physics, Tbilisi State University Chavchavadze
av. 3, Tbilisi 0128, Georgia*

**INVESTIGATION DNA RESPONSE ELEMENT BY ANDROGEN
RECEPTOR DNA BINDING DOMAIN USING NMR**

Y.H. WANG¹, A. MONGE², J. CÉRALINE², R.A.
ATKINSON¹, AND B. KIEFFER¹

¹ *Institut de Génétique et de Biologie Moléculaire et Cellulaire,
UMR 7104, 1 rue Laurent Fries, BP 10142, Illkirch Cedex
67404, France*

² *Laboratoire de Cancérologie Expérimentale et de Radiobiologie,
EA 3430, Université Louis Pasteur de Strasbourg, IRCAD,
BP426 Strasbourg Cedex, 67091, France*

**BIOSYNTHESIS OF UDP-D-MANNAC3NACA
IN PSEUDOMONAS AERUGINOSA**

ERIN L WESTMAN AND JOSEPH S LAM

*Department of Molecular and Cellular Biology, University of
Guelph, Guelph, Ontario, Canada*

**STRUCTURAL STUDIES OF ALCOHOL AND ODORANT
BINDING PROTEINS**

BRIAN ZIEMBA AND DR. DAVID N. M. JONES

*University of Colorado Health Sciences Center Dept of
Pharmacology RC1-South, L18-6402G 12801 East 17th
Avenue, P.O. BOX 6511, MS 8303 1775 N Ursula Street,
Aurora, CO 80045*

AUTHOR INDEX

A

Aitken, C.E. 83

C

Chugunov, A.O. 21

D

Daujotyte, D. 51

E

Eaton, W.A. 1

Efremov, R.G. 21

G

Gronenborn, A.M. 43

H

Henry, E.R. 1

K

Klimasauskas, Saulius 51

Krylov, N.A. 21

L

Liu, C.W. 65

Liutkeviciute, Z. 51

M

Marshall, R.A. 83

N

Nolde, D.E. 21

P

Puglisi, J.D. 83

Pyrkov, T.V. 21

R

Robertson, I.M. 101

S

Spyracopoulos, L. 101

Sykes, B.D. 101

Y

Yonath, A. 121