RECENT DEVELOPMENTS IN THEORETICAL STUDIES OF PROTEINS

Advanced Series in Physical Chemistry – Vol. 7

RECENT DEVELOPMENTS IN THEORETICAL STUDIES OF PROTEINS

Editor

Ron Elber

Department of Physical Chemistry
Hebrew University
Israel

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

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

British Library Cataloguing-in-Publication Data

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

RECENT DEVELOPMENTS IN THEORETICAL STUDIES OF PROTEINS

ISBN 981-02-2196-7

Printed in Singapore.

RECENT DEVELOPMENTS IN THEORETICAL STUDIES OF PROTEINS

ADVANCED SERIES IN PHYSICAL CHEMISTRY

INTRODUCTION

Many of us who are involved in teaching a special-topic graduate course may have the experience that it is difficult to find suitable references, especially reference materials put together in a suitable text format. Presently, several excellent book series exist and they have served the scientific community well in reviewing new developments in physical chemistry and chemical physics. However, these existing series publish mostly monographs consisting of review chapters of unrelated subjects. The modern development of theoretical and experimental research has become highly specialized. Even in a small subfield, experimental or theoretical, few reviewers are capable of giving an in-depth review with good balance in various new developments. A thorough and more useful review should consist of chapters written by specialists covering all aspects of the field. This book series is established with these needs in mind. That is, the goal of this series is to publish selected graduate texts and stand-alone review monographs with specific themes, focusing on modern topics and new developments in experimental and theoretical physical chemistry. In review chapters, the authors are encouraged to provide a section on future developments and needs. We hope that the texts and review monographs of this series will be more useful to new researchers about to enter the field. In order to serve a wider graduate student body, the publisher is committed to making available the monographs of the series in a paperbound version as well as the normal hardcover copy.

Cheuk-Yiu Ng

PREFACE

Modeling, simulations and theoretical studies of biological molecules have expanded tremendously in the last few years. Some of the growth was attributed to the rapid advances in computer technology. Faster and more economical computers made it possible to simulate larger systems for longer periods of time at lower costs. The impact on simulations of biological macromolecules such as proteins was thus profound.

However, this was only part of the story. Equally significant has been the explosion in the number of computational tools available today to investigators in the field. This is especially striking considering that the field was overwhelmed, just a few years ago, by essentially one methodology — the Molecular Dynamics (MD) approach. The strength of MD is its simplicity and generality. The prime weakness of the MD approach is the restriction on the time scale and the limited conformation space that one can sample.

A few recent uses of the MD approach are described by K. Kuczera. The studies of dynamics and thermodynamics of the globins are of special interest. This is due to the wealth of experimental data available, which makes detailed atomic studies worthwhile. Furthermore, since many processes in the globins occur rapidly (e.g. nitric oxide diffusion and recombination), the limited time scale of Molecular Dynamic simulations is less of a problem.

Due to the specificity of proteins and their complexity, it was difficult to come up with an analytical theory of proteins. It is desirable of course to have methods of sufficient generality to address more than one or a few special cases.

Difficulty is not impossibility. Motivated by a problem (protein folding) of utmost importance in biochemistry and biophysics, researchers from different fields entered the area of theoretical biophysics. Armed with a set of tools new to biophysics of proteins, researchers understood a number of beautiful and general principles of protein design. The "polymer physicists"' approach to proteins is discussed by Garel, Orland, and Thirumalai.

The title of the chapter "Analytical theories of protein folding" clearly indicates that this is one contribution that is independent of the revolution in computational power.

Another development motivated by the protein folding problem is the work by Luthey-Schulten, Goldstein and Wolynes. This investigation aims at the design of an effective potential that will fold a protein to its correct structure at maximum speed. The principle of the design of such a potential came from the "physicists"' approach to proteins and spin glass models. In fact, Peter Wolynes should be credited for many early studies and ideas on the relation between spin glasses and proteins.

Even if an exact potential energy surface is provided, the problem is not yet solved. It is still a non-trivial problem; how to find the global (free) energy minimum among all of the possible alternative conformations. Fortunately, a number of original algorithms appeared in the field of "global optimization", suggesting that the problem previously believed to be too difficult to solve may be more tractable after all. Straub describes the philosophy and the applications of the new class of optimization techniques.

Low energy minima provide information on probable structures. However, they cannot tell us about kinetics. To bridge the gap between static information and dynamics it is necessary to compute properties related to transitions between alternative minima. Possible approaches to the problem are described in Chapter 2.

Finally we are concerned with electrostatic and ion channels. Ion channels belong to a specific type of proteins that transport ions through membranes. While electric field is important almost everywhere in molecular biology, its importance cannot be overemphasized in channel, in which a charged entity (the ion) is passed through a low dielectric medium (the membrane) using a special biological machinery (the channel). The discussion on electrostatic properties in channels and beyond, in dynamics and in statics of biomolecules, is provided in the last chapter by Eisenberg.

Ron Elber

CONTENTS

CHAPTER 1

DYNAMICS AND THERMODYNAMICS OF GLOBINS

KRZYSZTOF KUCZERA
Departments of Chemistry and Biochemistry,
University of Kansas,
2010 Malott Hall, Lawrence, KS 66045
USA

Contents

1. Introduction

Molecular dynamics simulations of macromolecules were launched over 15 years ago[1] and have since then become an increasingly used and powerful tool in understanding basic physico-chemical properties, structure, dynamics and functions of these complex and interesting systems. In this chapter I review our current progress in this understanding, illustrated with results obtained for proteins of the globin family — mainly myoglobin and hemoglobin.

The power of computer simulations of molecular dynamics lies in the enormous detail of information they provide by probing motions of individual atoms on a subpicosecond time scale. Analysis of this information leads to important insights into the microscopic underpinnings of observed molecular properties. Although approximate, simulation results are currently of sufficient accuracy to be complementary to experimental studies. In this complementary relationship, simulation methods are tested by comparison with available experimental data, used to interpret the data in terms of atomic-level events, and finally to predict novel phenomena and suggest new experiments. A growing number of studies employ the joint experimental-simulation approach to macromolecular systems.[2–6]

A molecular dynamics simulation consists of solving Newton's equations of motion for a system of particles on a computer, and analyzing the resulting trajectory — the collection of snapshots of the particles' positions and velocities at consecutive points in time. To solve Newton's equations we need two important pieces of information about our system: the initial conditions (initial positions and velocities of particles) and the forces acting on the particles in any given configuration. The forces are usually derived from semi-empirical analytical potential energy functions. The trajectory is analyzed to yield quantities of interest — including time evolution and averages for structural and energetic properties. Finally, the results are interpreted in terms of observable quantities and compared with available experimental data. After a short introduction to the properties of myoglobin and hemoglobin, the remaining parts of the Introduction deal with different aspects of simulations in detail. For excellent reviews of molecular dynamics simulations see Ref. 7, for biomolecular applications — Ref. 8, and for free energy simulations — Refs. 9–11.

1.1. *Myoglobin and Hemoglobin Structure and Function*

Myoglobin (Mb) and hemoglobin (Hb) are an "oxygen carrying team" in vertebrate organisms.[12] Hemoglobin binds oxygen in the lungs and transports it through the bloodstream to other parts of the organism, while myoglobin stores oxygen in muscle cells. Both Mb and Hb bind the oxygen to an iron atom in prosthetic heme groups. The heme is covalently bound to the protein and is tightly surrounded by the polypeptide chain.

Myoglobin is a monomer, consisting of a single polypeptide chain and one heme group. The protein has a compact globular structure, with eight alpha-helical fragments labeled A to H. In the case of sperm whale Mb, for which most of the simulations were performed, the protein has 153 amino acid residues, and is roughly spherical in shape, with a radius of ca. 15 Å. The structure of sperm whale myoglobin is shown in Fig. 1.

Hemoglobin is a tetramer, consisting of two α and two β subunits. The subunits differ slightly in size and sequence, but have spatial structures very similar to each other, and to Mb. Normal human hemoglobin (HbA) consists of 576 amino acid residues and four heme groups. It is roughly spherical in shape, with a radius of ca. 30 Å. The structure of human hemoglobin is shown in Fig. 2.

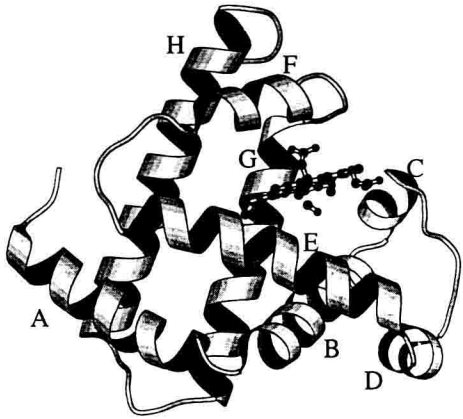


Fig. 1. Three-dimensional structure of myoglobin. Protein backbone shown as helical ribbons, labeled A–H, connected by loops; heme and CO ligand shown in ball-and-stick representation. Structure of sperm whale MbCo[18] taken from PDB file 1 mbc.[114,115] Drawing created using program MOLSCRIPT.[116]

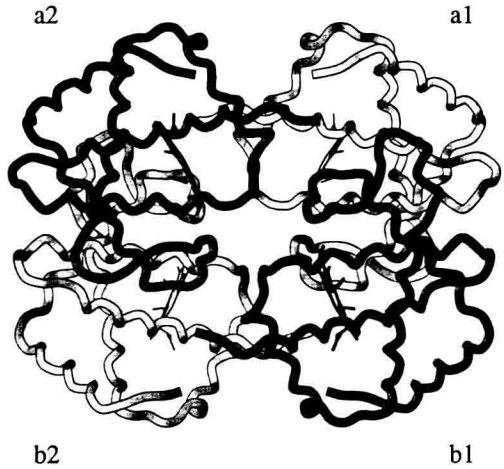


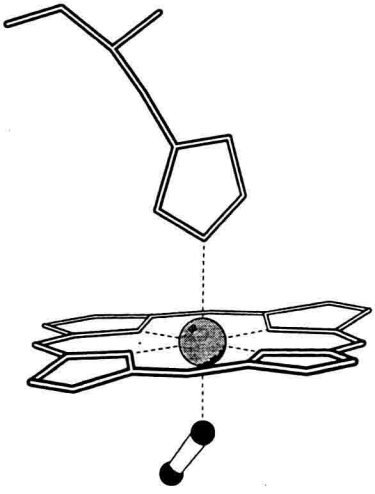
Fig. 2. Three-dimensional structure of hemoglobin. Protein backbone traced out for the four subunits, α (upper right), β (lower right), α_2 (upper left), and β_2 (lower left). Structure of human deoxy-Hb[15] taken from PDB file 3hhb.[114,115] The $\alpha_2\beta_2$ coordinates were generated by rotating $\alpha_1\beta_1$ by 180° around the y (vertical) axis. Drawing created using program MOLSCRIPT.[116]

Although they are not enzymes, myoglobin and hemoglobin participate in chemical reactions — ligand binding and dissociation at the heme. The possible ligands include oxygen (O_2), carbon monoxide (CO), and nitric oxide (NO). What makes myoglobin and hemoglobin especially interesting systems for molecular dynamics simulations is that motion plays an important role in their biological function. Two types of motion are relevant, involving equilibrium fluctuations and a nonequilibrium structural transition.

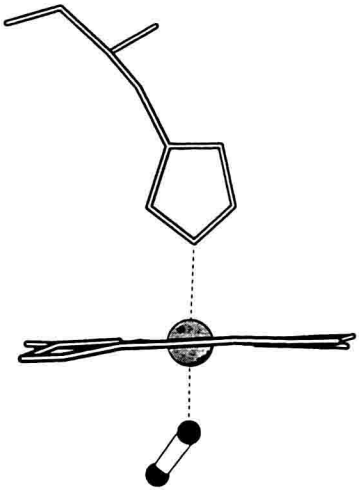
From the refined X-ray structures of both liganded and unliganded Mb[13,14] and Hb,[15,16] it can be seen that there is no evident pathway for ligand travel between the heme binding site and the external environment. Thus, structural fluctuations must accompany ligand diffusion through the protein. Crystal structures of liganded and unliganded heme systems — protoheme, Mb or Hb,[17] indicate that a structural change occurs upon ligand dissociation. In the liganded or oxy state, the heme iron atom has six ligands — four are pyrrole nitrogens of the heme, one is a nitrogen atom ($N_{\epsilon 2}$) of a histidine sidechain (the proximal histidine) which covalently links the heme to the protein, and the sixth is a small, mobile molecule — O_2, CO, NO etc. The liganded heme is planar, with the Fe atom in the mean heme plane[17] (see Fig. 3a and b). After dissociation of the sixth ligand, the heme assumes the unliganded or deoxy structure; the Fe atom moves out of the mean heme plane by about 0.5 Å, while the heme itself becomes domed[17] (see Fig. 3c and d).

In myoglobin this shift results in a small-scale tertiary structural change involving mostly a shift of the F helix (containing the proximal histidine residue) relative to the heme. The backbone rms deviation between liganded and unliganded structures is only about 0.3 Å.[18] In hemoglobin the tertiary structural change of the subunits is of similar scale as in myoglobin; however, there is a relatively large quaternary structural change, consisting of a rotation of the two halves of the molecule by about 15° and a relative translation by about 1 Å.[12] This structural transition is intimately involved in the regulation of Hb affinity for oxygen, and also plays an important role in Hb sickling, which will be described in Sec. 4.

Clearly, knowledge of the dynamics of structural fluctuations, ligand diffusion and the structural transition at the atomic level is important if we wish to understand the mechanism of the biological function of heme proteins.

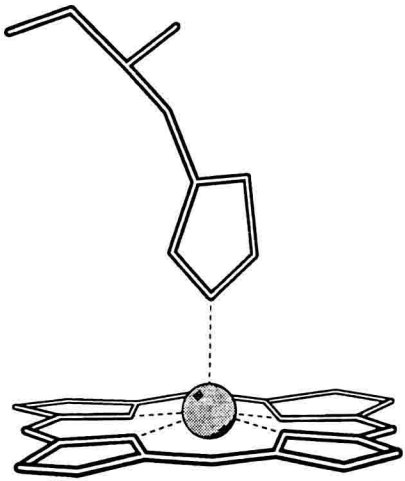


(a)

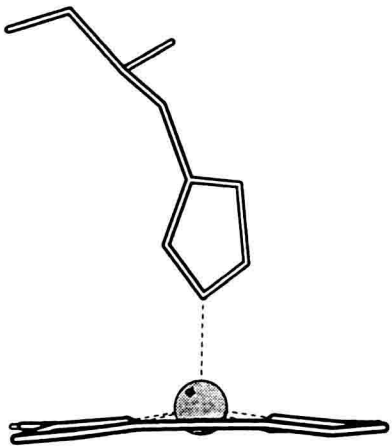


(b)

Fig. 3. Comparison of liganded and unliganded heme structures. (a) and (b) — two views of the planar six-liganded heme, including the proximal histidine (His 93) and CO, from sperm whale MbCO[18] (PDB file 1 mbc[114,115]).



(c)



(d)

Fig. 3. (c) and (d) — two views of the domed five-liganded heme, including the proximal histidine (His 93), from sperm whale deoxy-Mb[14] (PDB file 1 mbd[114,115]). For clarity, only heavy atoms of heme porphyrin skeleton and His residue are shown. The central Fe atom is displayed as a sphere of 0.5 Å radius; bonds between Fe and ligands shown as dashed lines. Drawings created using program MOLSCRIPT.[116]

1.2. *Potential Energy Functions*

In classical mechanics the answers to all questions concerning molecular structure, interactions and dynamics lie in the potential energy function. The potential energy determines the relative stabilities of different molecular structures. The dynamics of the system is determined by forces acting on the atoms, which are the first derivatives of the potential with respect to atomic coordinates. The second derivatives of the potential energy with respect to atomic coordinates are the harmonic force constants, which enable the calculation of molecular vibrational spectra.

At present the only viable approach for macromolecular systems is to use empirical energy functions, which can usually be expressed in a simple form, such as:

$$\begin{aligned} U = & \frac{1}{2} \sum_{\text{bonds}} k_b(b - b_0)^2 + \frac{1}{2} \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\ & + \frac{1}{2} \sum_{\text{impropers}} k_\omega(\omega - \omega_0)^2 \\ & + \frac{1}{2} \sum_{\text{dihedrals}} k_\phi[1 + \cos(n\phi + \delta)] \\ & + \sum_{\text{atom pairs } ij} \left[\frac{q_i q_j}{R_{ij}} - \frac{A_{ij}}{R_{ij}^6} + \frac{B_{ij}}{R_{ij}^{12}} \right], \end{aligned} \tag{1}$$

which is employed in CHARMM[19] and other macromolecular simulation programs. In this equation b, θ, ω, ϕ and R_{ij} are the actual bond lengths, valence bond angles, improper and proper dihedral angles and atom-atom distances for a given configuration of the system, respectively; $k_b, b_0, k_\theta, \theta_0, k_\omega, \omega_0, k_\phi, n, \delta, q_i, A_{ij}$ and B_{ij} are parameters. One way of looking at potential energy functions of this type is to treat them simply as mathematical models leading to certain predictions. However, both the method of derivation and the form of the function gives the parameters physical meaning.

The first three terms in Eq. (1) describe energies of deformation from equilibrium of chemical bond length (b), bond angles (θ) and improper dihedrals (ω) and angle deformations from equilibrium values. These distortions are usually small under normal conditions, allowing the energy to be described in the form of a sum of harmonic (Hooke's law) terms. The

presence of these terms assures that throughout the simulation the molecule preserves its basic chemical structure.

The fourth term in Eq. (1) describes the energy of torsion around chemical bonds — it is periodic and is modeled as a cosine term (or sum of several cosines with different periodicities). The presence of the $\cos(n\phi)$ term shows that the molecule will sample n distinct potential energy minima as the complete rotation around the chemical bond is performed.

The last line of Eq. (1) describes the nonbonded interactions between pairs of atoms, and has two contributions: electrostatic interactions between point charges and a Lennard–Jones 6–12 potential which models the interplay of weak attraction between neutral atoms at moderate distances and strong repulsion on close contact.

The force constants k_b, k_θ, k_ϕ and k_ω can be determined from vibrational spectroscopy, equilibrium values for bond lengths and angles b_0, θ_0 and ω_0 — from crystallography and microwave spectroscopy of model compounds. Dihedral angle torsional potentials can be obtained from infrared and microwave spectroscopic data. Nonbonded interaction parameters q, A and B can be determined by a wide variety of methods, both experimental (e.g. crystallography, thermodynamic measurements, gas phase scattering) and theoretical (quantum mechanical calculations). Generally, parameter values are determined in studies of small molecule models such as amino acids or their fragments. The assumption that the same parameters can be used to describe macromolecules is verified by comparison of calculated and observed properties.

The representation of the potential energy of a macromolecular system by Eq. (1) involves a number of approximations, some of which are discussed in Sec. 5. For a more complete discussion of potentials, see Ref. 8.

For a peptide or protein system, the potential energy would thus involve the terms describing deformations of all the bonds, angles and dihedrals from their minimum values, as well as nonbonded interactions between atoms which are not chemically bonded, but are brought close together due to the three-dimensional folding of the molecule (tertiary structure). To simulate a solvated protein, we add additional terms in the potential, describing bond and angle deformations of the water molecules as well as water-water and water-protein nonbonded interactions.

It is truly amazing that such a simple functional form of the potential can describe the wide range of interactions occurring in macromolecules —