Journal of Molecular Graphics and Modelling 54 (2014) 62–79

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Topical Perspectives

Challenges in computational studies of enzyme structure, function and dynamics

Alexandra T.P. Carvalho[a], Alexandre Barrozo[a], Dvir Doron[b], Alexandra Vardi Kilshtain[b], Dan Thomas Major[b,*], Shina Caroline Lynn Kamerlin[a,**]

[a] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden
[b] Department of Chemistry and the Lise Meitner-Minerva Center of Computational Quantum Chemistry Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

Article history:
Accepted 16 September 2014
Available online 28 September 2014

Keywords:
Computational enzymology
QM/MM
Free energy simulations
Reaction coordinates
Conformational sampling

ABSTRACT

In this review we give an overview of the field of Computational enzymology. We start by describing the birth of the field, with emphasis on the work of the 2013 chemistry Nobel Laureates. We then present key features of the state-of-the-art in the field, showing what theory, accompanied by experiments, has taught us so far about enzymes. We also briefly describe computational methods, such as quantum mechanics-molecular mechanics approaches, reaction coordinate treatment, and free energy simulation approaches. We finalize by discussing open questions and challenges.

1. Introduction

Enzymes are the guardians of life, reducing the timescales of the chemical reactions so crucial to biology from millions of years to fractions of seconds [1]. Therefore, to understand how enzymes work is to understand how life works, at the most fundamental, molecular level. In addition to their biological roles, there is also great interest in understanding how to use enzymes as artificial catalysts for a range of processes, from chemical synthesis to the generation of novel biofuels [2]. The history of contemporary enzymology is a long and distinguished one, which, despite its seeming experimental dominance, is nevertheless inseparable from theoretical contributions. The earliest work in enzymology dates back to the mid-to-late 1800s, with studies of fermentation (in fact, the word enzyme literally translates to "in yeast" [3]), followed by Emil Fischer's "lock-and-key" theory of enzyme–ligand binding in 1894 [4]. However, in the absence of any structural data, it was hard to infer what kind of molecules enzymes even *are*, let alone how they actually work. Here, the development of transition state theory (TST) by Eyring and Polanyi [5] proved crucial, as it paved the way for Linus Pauling's brave (for the time) assertion that the tremendous catalytic power of enzymes comes from the highly specific tight binding of the transition states (TS) during the chemical reaction [6]. This observation is particularly impressive in light of the extremely limited information available about enzymes at the time. A major turning point in enzymology came in 1965, with the publication of the 2 Å resolution crystal structure of hen egg white lysozyme by Phillips and coworkers [7]. This small 14.3 kDa enzyme (mature form, Fig. 1) was the third protein and the first enzyme to have its structure solved [7], and as such, finally allowed an enzyme structure to be examined at atomic resolution, giving birth to structure-function studies of enzymes and allowing us to think of enzyme mechanism in structural terms. Increasing computational power and methodological developments at the time [8–13] meant that it was becoming possible to perform computer simulations on molecular structures. In subsequent years, *in silico* protein folding by minimization with normal mode jumps [14], and the earliest molecular dynamics (MD) simulation on a biological system [15], demonstrated that proteins are not in fact static structures, but rather dynamic entities. With time such protein dynamics have been accepted to be critical for their function. These major experimental and theoretical advances created a huge paradigm shift in the field, and marked the beginning of what is now contemporary enzymology.

Since these early days, computational enzymology has become an invaluable tool for studying enzyme activity. However, despite the many advances in this field, and the corresponding experimental progress [16–19], we still do not have a complete understanding of how enzymes really work. In the present perspective, we begin with a historical overview, showing the birth of computational

* Corresponding author.
** Corresponding author. Tel.: +46 18 471 4423.
E-mail addresses: majort@biu.ac.il (D.T. Major), kamerlin@icm.uu.se (S.C.L. Kamerlin).

http://dx.doi.org/10.1016/j.jmgm.2014.09.003

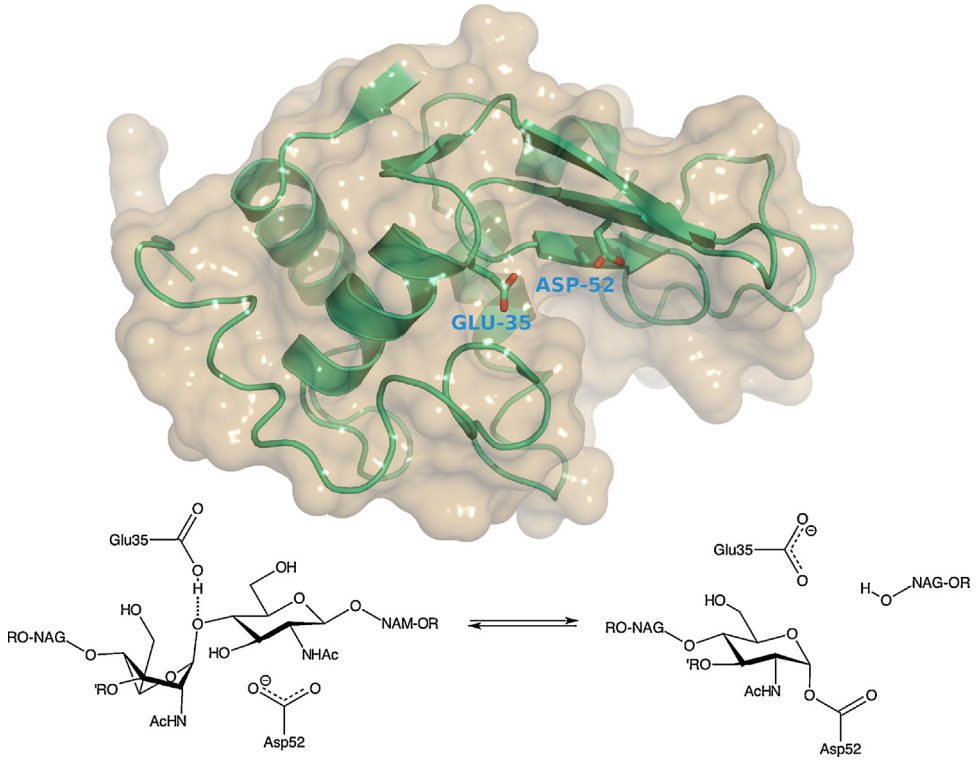

Fig. 1. Overall structure [285] and proposed mechanism [235] of hen egg-white lysozyme. The figure shows two key residues in the cleavage of the ester bond holding two sugar rings. In the first step, the Michaelis complex adopts a skew-boat or distorted envelope conformation. The reaction proceeds through an undistorted covalent intermediate.

enzymology, with an emphasis on the work that provided the basis for awarding the 2013 Chemistry Nobel prize to Karplus, Levitt and Warshel [20–22]. Following this, we provide a concise summary of the current state-of-the-art in the field, showing what experiment and, in particular, theory, has taught us so far about enzymes. We also discuss some of the most popular theoretical techniques to study these biological machines. We finalize by highlighting and discussing open questions and challenges that we believe still need addressing.

2. The birth of a field

Choosing an exact event for the birth of a field is a difficult task. As discussed in the Introduction, a good "time-zero" for Computational Biochemistry could go back as far as the development of the first computer code for molecular mechanics (MM) methods by Allinger et al. [10,11], which was based on previous work on intermolecular potentials. Subsequently, Némethy and Scheraga developed simplified versions of these potentials for use in statistical mechanics simulations [9]. Concurrently, MD simulations (which date back to the early 1950s [23,24]), were gradually increasing in popularity at this point with the advent of the first supercomputers. However, it was still limited to small system sizes pertinent to problems in chemical and theoretical physics, such as the first simulations of liquids, a molten salt and a small organic molecule [25–29]. The first step enabling the leap from chemical physics to computational biology came in the late 1960s, when Lifson and Warshel developed the first consistent force field (CFF) [13]. This was combined with the fact that the Weizmann Institute had recently acquired what was then a powerful machine, aptly named the Golem (after the automaton from Jewish folklore) [30], which allowed them to perform comparatively computationally demanding calculations. While the first of these still, as would be expected, focused on small molecules [12], in 1969, Levitt and Lifson used CFF to perform the first energy minimization of simple protein structures (myoglobin and lysozyme) [31]. At the time, Karplus, who was an established theoretical chemist in areas ranging from NMR to reaction dynamics, joined the Lifson group at the Weizmann Institute for a six-month leave and the first contact between the Nobel Prize trio was made [32]. In 1970, Warshel and Bromberg then published a QM(VB)+MM study of the oxidation of 4a,4b-dihydrophenanthrene oxidation [33] which is the first published QM+MM simulation of a chemical process and laid the foundation for subsequent QM/MM calculations of biological systems. Note that the key difference between a QM+MM and a QM/MM simulation is that the latter includes treatment of the coupling between the QM and MM regions, whereas the former does not, as also highlighted in [34].

At this stage, further progress was to some extent limited not just by computational power, but also by the lack of available structural data. The Protein Data Bank was only established in 1971 [35], at which point it only contained 7 protein structures. Another limitation was that traditional MM approaches, such as the ones used in the study of myoglobin and lysozyme, do not describe electrons explicitly, and cannot be used to describe bond making-and-breaking processes, making their usefulness for understanding chemical reactivity limited. However, numerous QM models were available at the time that *could* treat electronic structure properties. Thus, the ability to do separate QM and MM calculations existed in these early days, although a QM treatment was only possible for small molecules [33]. Still, there was a tremendous gap that needed to be bridged before it became possible to move to enzymes.

In 1971, Karplus, together with Barry Honig, were working on the retinal chromophore, and employed a hybrid Hamiltonian relying on a Hückel one-electron term for the π-electrons, and a pairwise non-bonded energy function for the sigma bond framework [36]. Subsequently, Warshel joined the Karplus group as a