
# Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines

Nobel Lecture, 8 December 2013

by Arieh Warshel

University of Southern California, Los Angeles, CA, USA.

## ABSTRACT

A detailed understanding of the action of biological molecules is a prerequisite for rational advances in health sciences and related fields. Here, the challenge is to move from available structural information to a clear understanding of the underlying function of the system. In light of the complexity of macromolecular complexes, it is essential to use computer simulations to describe how the molecular forces are related to a given function. However, using a full and reliable quantum mechanical representation of large molecular systems has been practically impossible. The solution to this (and related) problems has emerged from the realization that large systems can be spatially divided into a region where the quantum mechanical description is essential (*e.g.* a region where bonds are being broken), with the remainder of the system being represented on a simpler level by empirical force fields. This idea has been particularly effective in the development of the combined quantum mechanics / molecular mechanics (QM/MM) models. Here, the coupling between the electrostatic effects of the quantum and classical subsystems has been a key to the advances in describing the functions of enzymes and other biological molecules. The same idea of representing complex systems in different resolutions in both time and length scales has been found to be very useful in modeling the action of complex

systems. In such cases, starting with coarse grained (CG) representations that were originally found to be very useful in simulating protein folding, and augmenting them with a focus on electrostatic energies, has led to models that are particularly effective in probing the action of molecular machines. The same multiscale idea is likely to play a major role in modeling even more complex systems, including in describing cells and collections of cells.

## INTRODUCTION

The ability to model complex molecular systems is crucial for advances in understanding biological systems and in rational progress in molecular medicine, as well as in the rational design of new materials and devices. However, progress in this direction was hindered by the fact that rigorous modeling of complex systems requires enormous computational power. That is, a reliable quantum mechanical description [1] of more than a few atoms was practically impossible for a very long time. Even now, it is still too computationally expensive to use high-level quantum calculations to obtain convergent sampling on the many configurations needed to reliably describe the free energies of even medium sized systems. The solution to this challenge (and related problems) has emerged from the realization that a description of the properties of complex systems does not require the representation of all parts of the system at the same level of detail. For example, the interactions of a water molecule with a charge center that is 10 Å away can be treated classically instead of quantum mechanically. Similarly, a bond that does not participate in a chemical reaction can be represented as a classical spring. Thus, it is possible to decompose the system to parts where the quantum mechanical description is essential (*e.g.* parts where bonds are being broken), and other parts that can be represented on a simpler level with empirical force fields. This idea, which may seem obvious in retrospect, led to the development of the combined quantum mechanics/molecular mechanics (QM/MM) model [2]. Here, the coupling between the electrostatic effects of the quantum and classical subsystems has eventually become a key to advances in describing the function of enzymes and other biological molecules.

The emergence of the QM/MM approach allowed one to ask for the first time, in a well-defined and logical way, what the origin of the catalytic power of enzymes actually *is*. That is, although landmark works (see discussion in [3],[4],[5]) suggested various ways by which enzymes can accelerate reactions, none of these could directly relate the structure of the enzyme to its catalytic effect, nor could any approach reliably predict the rate constants of enzymatic reactions. Here, the QM/MM approach (and, in particular its empirical valence

bond (EVB) version) has provided what is probably the best solution to this long-standing fundamental puzzle. The idea of dealing with complex systems by treating different parts of the system on different scales is very general, and has found applications in many areas, and, in particular, in studies of complex biological systems. An early example of this has been our simplified coarse-grained (CG) model for protein folding [6]. Subsequent focus on electrostatic models has led to CG models that are particularly effective in probing the action of molecular machines.

Overall, the philosophy that has emerged from our studies is that the description of complex molecular systems requires computers to bridge between structural and functional information, and that computational scientists should carefully consider the resources available when choosing optimal models for describing the simulated systems. Here, using multiscale strategies is almost always a powerful way to explore different systems with different time and length scales. In describing the emergence of multiscale modeling, I will start by some recollections of the early developments in the field, and then move to specific examples, starting with enzyme action all the way through to the action of molecular machines.

## EARLY JOURNEYS IN MULTISCALE COMPUTER MODELING

Growing up in a Kibbutz in Israel, I did not have much scientific experience, but I liked to experiment with hot air balloons and building handguns, as well as in other subjects that have no relationship to chemistry. Nevertheless, after being accepted to the Technion (Israel Institute of Technology), I rather randomly chose to study chemistry. Eventually in 1964–1965, during my third year, I became interested in understanding how enzymes can accelerate chemical reactions, sometimes by up to twenty orders of magnitude. I started an experimental project that resulted in perhaps the first NMR measurement of a very fast step in the catalytic reaction of chymotrypsin, but this experiment did not provide any reasonable clues about the origin of the catalytic effect. In fact, although I did suspect that electrostatic effects are making the reaction go faster, my experiments showed that changing the ionic strength does not influence catalysis in a major way [7]. This result (incorrectly) indicated that electrostatic effects do not contribute significantly to catalysis.

After the Technion, I joined Shneior Lifson, who was the scientific director of the Weizmann Institute, and was starting to move from statistical mechanics of helix coil transitions, to modeling molecules with digital computers. In the fall of 1966, I started my PhD trying to develop what became known as

the consistent force field (CFF) [8, 9]. My general suggested direction was to represent molecules as balls and springs (which became known as molecular mechanics [MM] or a "force field" approach) and to reproduce energies, structures, and perhaps vibrations. This was supposed to be done by a consistent refinement of the MM parameters that will force the calculated and observed properties to be as close as possibly to each other. However, we had no clue how to actually do so. As a start, I attempted to treat cyclic amides, on the way to parameterizing amino acids' potential functions, by extending the internal coordinate approach of Mordechai Bixon [10], who was the previous student of Lifson. Unfortunately, this approach, which involved analytical derivatives of complex interdependent transformation matrices, became basically impossible to formulate and implement. The same internal coordinate treatment had been the key to practical conformational analysis programs of that time (e.g. [11, 12]), which incidentally could not obtain convergent minimization, because this required the first and second derivatives. In desperation, I tried to abandon the common description of molecules in terms of bond lengths and angles, and to move to a Cartesian coordinate description, where suddenly all the problems with analytical derivatives seemed to disappear. For example, obtaining the analytical first and second derivatives needed for minimizing the energy of a cyclic molecule in a converging way, which was close to impossible in internal coordinates (because each internal coordinate depends on all other coordinates), became trivial in Cartesian coordinates. Similarly, obtaining vibrational modes, which previously demanded spending half a year on reading Bright Wilson's molecular vibrations book [13], and then almost (at least for cyclic molecules) hopeless programming, required only the use of one simple equation in terms of the Cartesian second derivatives.

Fortunately, the Weizmann Institute had a specialized computer called the Golem (named after the "robot" from Jewish legend that helps the famous Prague rabbi), which had a remarkable double precision. Thus, I was able to obtain very accurate first and second numerical derivatives, and to prove that I was on the right track in obtaining exact minima and molecular vibrations in a general molecule. At that point I started to write a program with Cartesian analytical derivatives and a least squares force field refinement (using the numerical derivatives in pinpointing errors), stopping for a while during the Six-Day War, and then moving back to the program. At the end of the war, I returned to the Weizmann Institute, and around this time Michael Levitt appeared. Guided by Schneior's insight on obtaining consistent force field parameters, and his insistence that these parameters can describe reality regardless of whether they are derived from experiment or theory, we developed the general CFF Cartesian

force field programs [8, 14]), that allowed one to use MM to find exact local minima and vibrations of any medium sized molecule. The program also allowed for a fully consistent refinement of the MM parameters, by fitting the calculated and observed properties of molecules and molecular crystals. At any rate, our CFF program eventually became the basis of all modern MM molecular simulation programs [14]. The CFF parameter refinement turned out to be quite a demanding job (as it required automatic fitting to many independent properties), including inventing automatic frequency assignments and developing a general way of refining parameters that would reproduce known unit cell dimension of molecular crystals [9]. During 1968, in what turned out to be eventually significant, I also started experimenting with combining my newly developed CFF method (with the spring-like description of bonds with localized electrons) and a valence bond (VB) quantum model [15]. This QM (VB) + MM model helped to describe the extremely large isotope effect in a chemical reaction between oxygen and a medium sized organic molecule, and indicated to me that such a combination can be useful.

While still keeping enzymes at the back of my mind, I started a postdoctoral position at Harvard with Martin Karplus at the beginning of 1970, hoping to make the QM + MM CFF more general. Karplus and his postdoc Barry Honig were at that time making important advances in the study of retinal (the chromophore of the visual pigment) [16], which involves a 12 $\pi$-electron system. This seemed to be a good rationale to start developing the CFF for $\pi$-electron systems. Indeed, I succeeded in connecting the molecular orbital (MO) description of atoms with $\pi$-electrons with an MM description of $\sigma$-bonds with localized electrons [17], and in consistently refining the corresponding parameters for a unified CFF description. This QM (MO) + MM model included only the bonding between the QM and MM region, and thus ignored all key (*e.g.* electrostatic) coupling between the MM and QM regions. Nevertheless, the model provided a very powerful and general way to treat large conjugated molecules. During this project, I also figured out how to get the exact analytical forces from the QM treatment, by fixing the molecular orbitals and differentiating only the integrals. As usual, I made this fundamental advance by guessing it, then (as before) I confirmed my idea by using numerical derivatives and then finding the exact mathematical proof [18]. Here again it was shown that the combination of intuition and numerical validation is a powerful tool.

At any rate, the QM + MM treatment of delocalized electron systems still did not help me to move towards studying enzymes. Thus, upon returning to the Weizmann Institute in 1972, I started to develop a very effective hybrid orbital quantum program (QCFF/ALL), that represented all atoms in a relatively

small part of a molecular quantum mechanically, while representing the rest classically. I felt that this should allow me to finally make a progress towards my old dream of studying enzymes. At that time, Mike returned from his PhD at the Medical Research Council (MRC) to the Weizmann Institute, and I started to explore the possibility of combining my quantum mechanical model with his MM calculations on lysozyme (see below). While still struggling with the development of my QCFF/ALL approach, I found myself one day in the computer center discussing the protein-folding problem with Mike. This discussion turned to a strange idea of studying mechanical models of molecules on a gravitation-less spacecraft, and we suddenly came out with the idea of simplified protein models where spheres would represent amino acid side chains, and started to work on this project. This folding project started to move in a remarkable way, and it appeared that the drastic simplifications we had suggested allowed us to fold the small protein BPTI without using an enormous number of minimization steps. This simplified CG model [6] appeared to resolve the so-called "Levinthal paradox" [6], where the observation that proteins actually fold appeared to contradict the fact that they have an astronomical number of possible conformations so that they could never find a path for folding in a reasonable timescale. In fact, our simulations showed that the number of relevant coordinates is relatively small, and that the protein folding process can be effectively simulated. The progress on the folding problem helped me to obtain an EMBO fellowship so that I could collaborate with Mike when he moved back to the MRC. My time at the MRC turned out to lead to the culmination of several key advances pushing the frontiers in the understanding of biological function.

## ENZYMES, ELECTROSTATICS AND QM/MM

The three-dimensional structures of the enzyme lysozyme, which were solved by Phillips and coworkers in 1967 [19], provided the first glimpse of the structure of the enzyme-substrate complex. These breakthroughs offered enormous hope that enzyme catalysis would now be finally understood. For example, Phillips suggested that enzymes work by applying steric strain that pushes the substrate to a structure that is closer to the structure of the so-called "transition state," where the crucial bond between the carbon and oxygen atom in the sugar substrate is broken. This idea was due in part to the observation of what looks a distorted sugar ring and to the assumption that the protein can induce a significant strain. The strain was argued to reduce the barrier for bond breaking, and thus the activation barrier for the reaction. However, as Mike demonstrated, the strain idea was problematic since enzymes are flexible [20], and it seemed clear

to me that any further progress would require actually modeling the chemical reaction in the enzyme. Therefore, upon my arrival to the MRC in the autumn of 1974, I started to focus on modeling enzymatic reactions, still attempting to somehow combine my QCFF/ALL program with Mike's energy minimization of lysozyme [21]. The first attempt to combine the programs resulted in ridiculously high activation energies (so the reaction would never happen), and I realized that something must be completely wrong with my modeling direction. Eventually, it became clear that the work of breaking the bond between the carbon and oxygen atoms in the sugar substrate was being described incorrectly. The problem was that the bond is broken to a positive carbon and a negative oxygen ($C^+$ $O^-$), and that these charges must be stabilized by the electrostatic environment of the protein + solvent system (see Fig. 1).

The introduction of the effect of external charges was not so simple, since practically all earlier work that tried to add the effect of the environment started from the complicated configuration-interaction (CI) picture, which gave the overall molecular dipole moment and then used unreliable continuum cavity models (where the cavity radius is basically a free parameter) to describe the

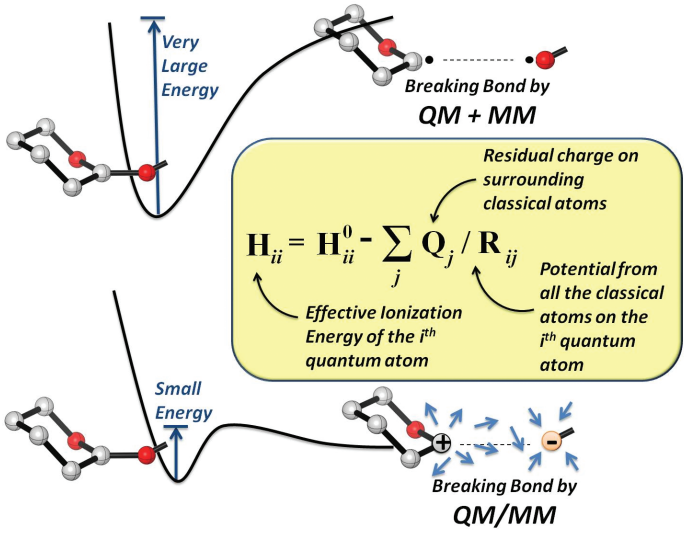


**FIGURE 1.** Showing the energetics of breaking a C-O bond in an uncoupled QM + MM (upper diagram) and when the electrostatic and steric effects of the environment are included in a coupled QM/MM (lower diagram) The dipoles designate the effect of the surrounding residual charges. As seen from the figure it is very hard to break the bond without including the coupling between the QM and MM regions.

environment. Instead, I realized that one can start from the general expression of the quantum mechanical self-consistent Hamiltonian (see *e.g.* [22]) :

$$F_{\mu\mu}^{ii} \equiv U_{\mu\mu} + 1/2 P_{\mu\mu}\gamma_{\mu\mu} - \sum_{\nu-\mu} P_{\nu\nu}\gamma_{\mu\nu} - \sum_{i'\neq i} Q_i\gamma_{ii'} - \sum_j Q_j\gamma_{ij} \tag{1}$$

where U is the core Hamiltonian, P is the quantum mechanical bond order, Q is the net atomic charge, $\gamma$ is the electronic repulsive integral, and $\mu$ and $\nu$ are atomic orbitals on atom i. Now, assigning atoms I to the part of the system that should be treated quantum mechanically indicated that the other atoms (denoted by j) can be treated classically, assuming that their charge is constant. That is, replacing $\gamma$ by $e^2$/r gives:

$$F_{\mu\mu}^{ii} \equiv \left(F_{\mu\mu}^{ii}\right)_0 - \sum_{\sigma} \frac{e^2 q_B}{r_{AB}} = \left(F_{\mu\mu}^{ii}\right)_0 - U_A \tag{2}$$

where $\mu \in$ A and $(\,)_0$ designates the contribution from the quantum atoms (typically the "solute"), and $U_A$ designates the total electrostatic potential from the classical atoms (typically the "solvent" molecules) at the site of atom A. This equation can be generalized to cases where the charge distribution of the classical atoms it not fixed and can be polarized by the field of the quantum atoms [23]. Thus, the leading term in the solute-solvent coupling Hamiltonian is obtained by adding the potential from the solvent atoms to the solute Hamiltonian. The total potential energy is then given by:

$$V_{\text{total}} = E_S\left(\mathbf{F}^{\mathbf{S}}\right) + E'_{Ss} + E_{ss} \tag{3}$$

In this equation, $E_S(\mathbf{F}^{\mathbf{S}})$ is the energy that is quantum mechanically obtained with the **F** matrix that includes the given electrostatic potential from the solvent (the vector of all the $U_A$'s). $E'_{Ss}$ is the non-electrostatic solute-solvent interaction term, and $E_{ss}$ is the solvent-solvent classical force field. At this level of approximation, the non-electrostatic term is evaluated by the standard classical van der Waals potential function. In studies of very large solute molecules, we sometimes divide the solute region in quantum and classical parts. The "connection" between the quantum and classical regions is treated by a classical force field (which is included in $E'_{Ss}$), where the quantum atoms at the boundaries are connected to dummy hydrogen-like atoms in order to balance the electrons in the quantum system. The main problem we faced in 1975 was how to evaluate the magnitude and positions of the charges in the environment (*e.g.* water

molecules). Eventually, after spending several months in the library and talking to eminent experts on electromagnetic theory, I realized that none of the textbooks or the experts could tell us much about how to computationally model electrostatic effects in proteins or solution. Thus, I turned to what I learned from my experience with developing force fields: forget about what is in the books that were written before the emergence of computers, and just go to the basic molecular level, using simplified models if needed.

I concluded that we would be unable to progress consistently as long as we thought in terms of the standard electrostatic theory, where all the details of the protein or the surrounding solvent are included with an elusive dielectric constant. Obviously, the computer power of the time was insufficient for modeling a protein surrounded by atomistic models of water molecules, while obtaining meaningful energetics. Thus, after considering several options with Mike, we decided to represent the water molecules as a grid of polarizable Langevin type dipoles (the LD model), and self-consistently evaluated the interaction of these dipoles with the charges in the protein-substrate system and with each other. Of course, the key to the success of this approach was the calibration of the LD model to observed solvation free energies. A similar self-consistent treatment was then introduced for the induced dipoles on the protein atoms [2]. This LD water model led to the first microscopic description of protein electrostatics, evading all the conceptual traps of the past and future continuum dielectric descriptions. Apparently, this model looked problematic to those who were trained with the idea that the special, highly symmetric structure of water molecules must be very relevant to their enormous solvation effects. However, the LD grid model eventually turned out to be an excellent approximation for studying solvation effects, long before any other microscopic model, and also before the development of macroscopic models that tried to consider the protein shape.

The introduction of a realistic electrostatic model for the enzyme and its surrounding water molecules, together with the incorporation of this effect in a quantum Hamiltonian, finally for the first time yielded the energy of heterolytic bond breaking processes in enzymes and in solution. This QM/MM approach reflected the realization that we cannot treat large systems quantum mechanically, and we cannot describe the chemistry without a quantum treatment. Thus, we used Eqs. 2 and 3 and described only the reacting region quantum mechanically, while treating the rest of the protein and the solvent classically (Fig. 2). This approach, along with related models that we subsequently introduced, has become known as "multiscale modeling." The QM/MM model suggested that enzymes work by using electrostatic fields to reduce the activation barriers for

bond breaking (see below). At any rate, the use of our QM/MM approach in modeling the catalytic reaction of lysozyme paved the way for the current direction in modeling enzyme action [24], and has become a major direction in theoretical chemistry and biophysics.

Our QM/MM studies also eventually led to what I believe is a true understanding of the origin of enzyme catalysis, which turned out to be associated with the electrostatic preorganization effect [25]. More specifically, my subsequent (mainly EVB) studies led to the non-trivial finding that enzyme catalysis is not due to the interaction between the enzyme and substrate (which is what was believed by most people), but rather to a large free-energy penalty for the reorganization of the solvent in the reference reaction without the enzyme (the work of rotating the water molecules towards the transition state charges). As described in Fig. 3, the reorganization energy increases the activation barrier in solution, whereas in the enzyme, the polar groups that stabilize the transition state do not have to rotate, since they are already folded with correctly polarized dipoles. In subsequent years, I was also able to prove that the change in the electrostatic reorganization energy accounts for almost the entire catalytic power of enzymes [26]. Although this elusive origin of catalysis was not envisioned during our studies at the MRC, I had benefited from the general feeling that it should be somehow associated with electrostatic effects (see Max Perutz' insightful review [27]). Interestingly, while Max intuitively recognized the importance of this effect, he originally felt that it was like the assumed stabilization

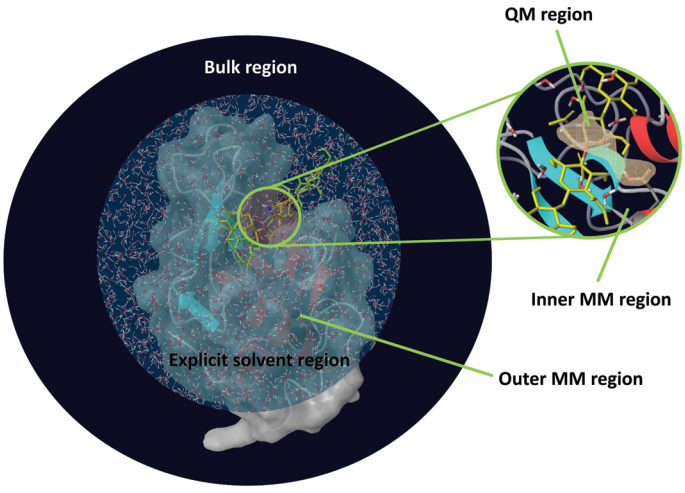


**FIGURE 2.** A QM/MM model of the lysozyme active site. The enzyme is divided into a small reactive QM region and the rest of the system, which is described by a classical MM model.

of ion pairs in a low dielectric environment, while I found that enzyme active sites are in fact very polar and pointed out to him that ion pairs would not be stable in oil surrounded by water. This explanation eventually led to a paper that he communicated for me to PNAS [25]. The electrostatic models conceived in 1975 became the basis for consistent microscopic treatment of biological models, the understanding of the true nature of protein dielectric constants [28, 29], and the simulation of key functional properties, including $pK_a$ values, redox potentials, binding free energies, ion and proton conductance [29], and protein stability [30].

In subsequent years, my coworkers and I drastically simplified the QM/MM approach, using a valence bond description of the different steps of the reaction, in what I called the "Empirical Valence Bond" (EVB) method [31]. This approach, which exploits the clear physics of the diabatic reactant and product states, has allowed us to take a considerable leap towards approaching my early vision, and to finally quantitatively model enzyme catalysis and explore enzyme design. This also helped me to explore (and frequently to eliminate) popular suggestions of factors that presumably lead to enzyme catalysis, such as entropic effects, ground state destabilization by desolvation, dynamical effects, orbital steering and more (see discussions in [32] and [26]). The key to the ability to figure out the secret of enzyme catalysis has been the ability to model the actual chemical reaction in the enzyme active site, and to dissect the different contributions to the rate constant, which is close to impossible when one is just using experimental approaches. Overall, the QM/MM studies provided a solution to the

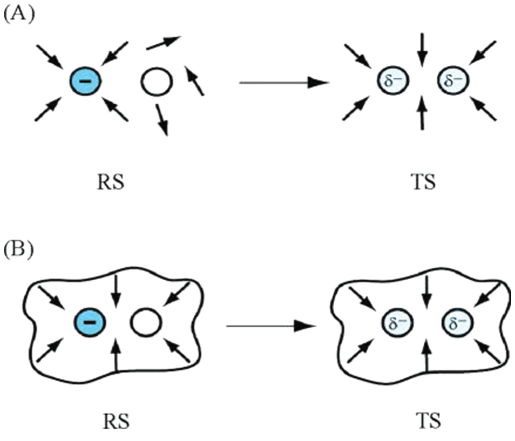


**FIGURE 3.** Schematic demonstration of the reorganization of the environment dipoles in an $S_N2$ reaction, where the charges change from being on one atom in the reactant state (RS) to being delocalized in the transition state (TS) in (A) water and (B) an enzyme active site [26].

long-standing puzzle of the origin of the catalytic power of enzymes and paved the way for quantitative studies of enzymatic reactions [26]. This strategy also allowed one to start to explore the issue of enzyme design in a rational way [33].

QM/MM approaches with an *ab initio* QM Hamiltonian (QM(ai)/MM calculations) have advanced in recent years to a level where they can be used with proper sampling to obtain reliable free energy surfaces in the condensed phase [34]. Nevertheless, it seems to me that at the time of writing this paper, it is still preferable to calibrate the EVB on QM(ai)/MM calculations in solution and then move to studies in proteins with the EVB approach [35]. However, it is clear that in the future, one will be able to obtain convergent QM(ai)/MM surfaces also for reactions occurring in enzyme active sites. Finally, when talking about multiscale modeling in the context of QM/MM and related approaches, it is important to emphasize that the general idea can be described as an *embedding approach*, where one is looking for the best way to incorporate the effect of the surrounding of the system that is the focus of the given study. Here, one of the most promising strategies is the use of the frozen DFT (FDFT) and constrained DFT (CDFT) approaches (*e.g.* [36]). These approaches treat the entire system on the quantum mechanical DFT level, with a formalism that is in principle rigorous [37]. However, the density around the main region is not subject to self-consistent optimization, and the corresponding electron densities are determined by approximate considerations (including a freeze-and-thaw strategy). The CDFT approach can be described as a QM/QM approach, but, again, the main idea is to have a less rigorous and less demanding description of part of the system in order to save computational cost.

## THE PRIMARY EVENT IN VISION AND THE DAWN OF MOLECULAR DYNAMICS SIMULATIONS IN BIOLOGY

In 1971, I realized that the optimal way to study photochemical reactions of any medium size or large molecules was to forget about the traditional description of crossing between energy levels, and to adopt the surface-hopping semi-classical trajectory approach that was introduced for treating gas-phase reactions of very small molecules [38]. This advance, which turned out to be a conceptual breakthrough, was only published in 1975 [39], using the photoisomerization of butene as an example. Fortunately, my conviction that this was the key to quantitative studies of photobiology gave me the courage to look at the most important problem of biological photochemistry, namely the primary photoisomerization of retinal in the visual process. More specifically, I became interested in retinal during my postdoc time (see *e.g.* [40]), but this interest was

focused on spectroscopic and geometric properties, and not on the most exciting problem of what is happening during the first step of the vision process (a problem that seems to be completely inaccessible to theoretical studies with the standard strategies). At that time, it was known that when light strikes the eye, it is absorbed in the Schiff base of retinal, which is embedded in a protein called rhodopsin. After absorption of light, the retinal molecule isomerizes from its initial 11-cis structure to an all-trans structure, forcing structural changes of the protein, where the new form of the protein (metarhodopsin) activates the transfer of the visual signal to the brain. Later, it was found that the metarhodopsin activates a G-protein called transducing, and that rhodopsin is in fact a G protein-coupled receptors (GPCR) [41]. It was also known that the primary absorption of light leads to a photoisomerization of the retinal molecule in less than six picoseconds (which was the shortest time that could be measured in the early 1970s). Furthermore, the absence of structural information seemed to introduce an even bigger challenge. Although I considered binding retinal to chymotrypsin, I decided to model the protein's effect by a steric cavity plus an assumed internal counter ion, and used the semiclassical surface hopping approach with a Schiff base of retinal, constrained to be in the starting 11-cis conformation [42]. My molecular dynamics (MD) simulations, depicted in Fig. 4, predicted that the primary process takes about 100 femtoseconds, with an enormous probability of jumping from the excited state to the ground state due to very large non-adiabatic coupling (a phenomenon that was later identified as the effect of conical intersections). Remarkably, the results of these simulations that represented the first use of MD simulations in biology have since been confirmed both experimentally [43] and theoretically [44].

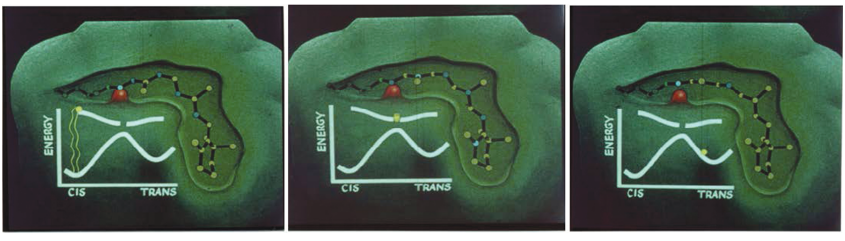


**FIGURE 4.** Snapshots from the simulated MD trajectory of the primary event in the vision process. The trajectory starts with 11-cis retinal in the ground state, and, upon absorption of light, the system moves to the excited state where the 11–12 torsional angle rotates without a barrier to 90°, and the trajectory crosses to the ground state in the trans direction. The motion involves only a small change in the overall structure, since the other torsional angles move in the opposite direction to the 11–12 torsional angle. The snapshots are taken from a movie that used the original trajectory presented in [42].

The molecular motion that emerged from these computer simulations resolved the problem of fast movement in a restricted protein cavity without strongly clashing with it. That is, it was found that the isomerization occurs with a concerted rotation of several bonds, which I called the "bicycle pedal" motion. To see if the bicycle pedal model made sense, I borrowed model building parts from Max Perutz's structural biology lab, and built a model that appeared to reproduce the concerted motion without any large structural changes. Interestingly, about 30 years after my original model, this motion has been confirmed by *ab initio* studies [44].

Over the following years, my long-time collaborator Bill Parson and our coworkers [45] used the structure of a bacterial reaction center (RC) and the same semi-classical approach to model the primary electron transfer event in photosynthesis, establishing that the observed 3 ps process involves a sequential hopping from the primary chlorophyll dimer (P*) to one monomer (B), and then to a second monomer (H). This was again done before the confirmation of our findings by decisive experimental studies (e.g.[46]), and at a time where most workers assumed that the primary event cannot be stepwise and assumed that it is a single step super-exchange process. Here, the ability to determine the correct electrostatic energy of each intermediate has been a major advantage over related attempts that did not involve experience in the conversion of protein structures to model electrostatic energies and redox potentials. Instructively, in this case, the advantage of working with and developing tools for studies of biological functions has been demonstrated in an effective way. That is, although we waited four years to get the RC coordinates, all the computer programs were ready and tested for a long time, and it took us only two weeks at the end of 1987 to convert the structure of the RC to a detailed (and correct) functional mechanism.

## FREE ENERGY CALCULATIONS AND THERMODYNAMIC CYCLES

One of the most remarkable advances that resulted from the emergence of computer modeling of biological molecules has been the ability to evaluate the relevant free energies, and, in particular, the energetics of charged groups in proteins. Arguably, this started with the very rough attempt in the original 1976 paper (see Fig. 8 of Ref. [2]), and continued with more quantitative free energy considerations and the introduction of well-defined microscopic based thermodynamic cycles, using the PDLD model that paved the way to evaluation of $pK_a$s [28], redox energies [47], ion transfer energies [48] and drug binding free energies [26, 29]. In 1977–78, after reading Valleau and Toerrie's masterful review [49], I started free energy perturbation (FEP) calculations of the charging of

an ion in my surface-constrained soft sphere dipole (SCSSD) water model [50]. This was mainly in order to show the referees of the SCSSD paper that the entropic contribution to the solvation free energies of ions is small. The calculations gave a reasonable trend, but I was so busy trying to fight the referees on other trivial issues that I did not publish the preliminary entropy study. Eventually, the increase in computer power allowed us to move into free energy calculations of charges in all atom solution models in 1982 [51], as well as starting free energy calculations of proteins in 1983–4 [52]. The microscopic free energy perturbation calculations and the corresponding free energy cycles have become a major part of the field, in part due to the excitement from rather trivial changes of a few solute atoms in solution [53] and then in proteins [54]. In light of my conviction in the importance of electrostatic energies, I did not consider these so-called "alchemical changes" to be a real challenge and continued to focus on evaluating the *large* absolute solvation free energy, providing the first FEP studies of the free energy of ionizing acids in proteins and of redox process (for a review see [29]), as well as the free energies of countless enzymatic reactions [26]. My coworkers and I also tried to educate the community about the enormous risks of looking at the so-called potential of mean force (PMF) in studies of biological charge transport and related problems. Here, we pointed out that looking at the PMF of, say, ion penetration in ion channels can be extremely misleading, since it does not tell you much about the error in getting the absolute solvation free energy and can lead to enormous problems. On the other hand, insisting on obtaining the absolute free energy is the best way to know if the model captures the correct physics [26]. This issue is strongly related to the tendency to confuse formal rigor with actual reliability. Here, the realization that the proper boundary conditions are key to the reliability of the results and the speed of the convergence, took a rather long time to reach the community.

## BRIDGING TIME AND LENGTH SCALES: COARSE-GRAINED (CG) SIMPLIFIED MODELS OF THE FUNCTION OF COMPLEX MOLECULAR MACHINES

While the MD studies of ultrafast photobiological processes of the type discussed above have been very effective [55], the simulations of functional properties that involve longer time steps and larger systems have presented a much more serious challenge. In fact even today, despite the exponential growth of MD simulations of proteins and related systems, and the enormous progress in computer power (e.g. [56, 57]), the ability to capture functional properties has been limited. Here, one faces enormous sampling problems that, (as discussed in [58])), are not necessarily reduced by using sophisticated formulations such as that of [59]. Of course, running one very long trajectory to represent a

functional property suffers from the problem of having a single observation, which might not correspond to the overall action. Furthermore, having a single long trajectory can still be considered as an experiment that needs careful interpretation and an analysis by a general reduced model. Thus, our point of view has been that in simulating complex systems, we clearly have the need to bridge the time and length scales by simple models.

Here, we have reverted back to the CG idea of the protein folding days [6], and asked how it can be used to study protein functions. It was clear that this task requires an improved treatment of electrostatic energies (which appear to be the key for structure-function correlations), and thus we undertook a major project, generating an improved electrostatic model and calibrating it on absolute protein stabilities [30]. The resulting CG model appeared to provide what is arguably the best current tool for moving from the structure to the function of molecular machines (see below). Another challenge that we had to address has been the requirement that the long time-scale behavior of the simplified model would reproduce the corresponding trend in the full model. The solution came with our renormalization method [60], where we apply strong external forces in MD simulations of the full model (thus inducing large conformational changes in short time) and also apply the same forces in the reduced model, which is simulated by Langevin dynamics. We then change the effective friction in the Langevin dynamics simulations until both the full and the reduced model produced the same time-dependent response to the applied force. The resulting friction is then used as the optimal friction for long timescale simulations with the reduced model, in the absence of the external force. This renormalization approach appeared to reliably reproduce the long timescale microscopic simulation [61], and allowed us to explore the long timescale behavior of complex molecular machines [60, 62].

Significantly, in developing the above CG and multiscale models, one faces the question of how to relate the simplified free energy surface to the corresponding results that would be obtained with the full explicit model. Here, we recruited the paradynamics (PD) philosophy, first evaluating the CG free energy, and then performing a perturbation between the CG and full surfaces at different key regions on the landscape [60]. These developments allowed us to use our CG model in simulations of molecular machines and other complex biological systems, and I will consider some of the most instructive recent examples below.

$F_1F_0$-ATPsynthase is a ubiquitous cellular engine composed of two rotational motors, the cytoplasmic $F_1$ coupled to the membrane embedded $F_0$ units. The $F_0$ rotor uses the energy of the proton transport across the cellular membrane to rotate the membrane embedded c-ring, while the $F_1$ couples the rotation of

the c-ring with its central stalk (g subunit) to generate ATP from ADP and $P_i$. In spite of numerous simulation and phenomenological studies (*e.g.* [63, 64]), the origin of the coupling between chemical and mechanical events in the $F_1F_0$-ATPsynthase has not been elucidated or simulated in a consistent and unbiased way. More specifically, several single molecule experiments [65] have discovered the amazing presence of the 80°/40° stepwise rotation of the system and noticed that the chemical step occurs after the 80° step (the delay before the chemical step has been called the "catalytic dwell"). Unfortunately, it has been especially difficult to understand the origin and significance of the stepwise coupling from a structural perspective. This difficulty has been in part due to the large system size, and the very long timescale of the process, which extends beyond the millisecond regime. Remarkably, the CG electrostatic free energy surface coupled to the ATP hydrolysis and product release free energies could successfully reproduce the observed behaviour of the system. This included generating electrostatic landscape that has a high energy region after the 80° $\gamma$-stalk rotation (see figure 5) and there upon addition of the chemical landscape, reproduce (see ref. 62) a functional landscape, where the 80° barrier is coupled to the chemical coordinate of the ATP hydrolysis and generate the catalytic dwell. This reproduced the experimentally observed catalytic dwell at 80°/40°. The details of our CG modeling and the corresponding analysis are given in [62].

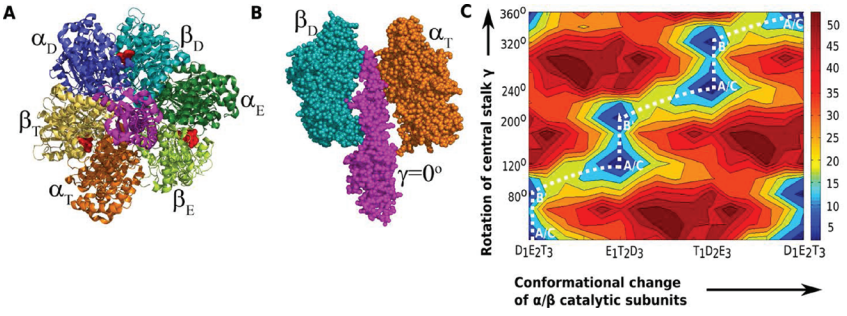


**FIGURE 5.** Exploring the coupling between the rotation of the $\gamma$-stalk to ATP hydrolysis in $F_1$-ATPase. The relevant system (namely $F_1$-ATPase) is shown from the membrane side (A), and along the vertical direction parallel to the central $\gamma$-stalk (B). The $\alpha$ catalytic subunits are shown in deep blue, deep green and orange, while the $\beta$ units are shown in cyan, light green and yellow. The $\gamma$-stalk is shown in magenta. The nucleotide occupancies of the $\beta$ subunits are depicted as T (ATP bound), D (ADP bound) or E (empty) states. (C) The CG electrostatic free energy surface of the rotation of the $\gamma$-stalk coupled to the $\alpha/\beta$ conformational changes. This landscape reflects the stepwise 80°/40° features discussed in the main text. The combination of the diagram of (C) with the energetic of the chemical steps (which is given in [62]) provides a structure-based description of the action of $F_1$-ATPase. This figure is taken from [62].

An additional encouraging CG study [66] has for the first time reproduced the directionality of the coupling between the protomotive force and the rotation of the c ring in $F_0$-ATPase. Phenomenological models have been used in attempts to understand the action of the C-ring rotation coupled to the proton transfer from the low to high pH reservoirs across the membrane [67, 68]. However, a quantitative structure-function relationship that elucidates the physical nature of the directional rotation has been completely missing. Our CG model has generated the electrostatic free energy surface of the c-ring rotation coupled to the proton transport from the P side (pH = 5) to the N side (pH = 8) of the membrane. The generated landscape has shown that the molecular origin of the directional c-ring rotation is mostly due to the asymmetry of the proton transport path on the N and P sides of the $F_0$ unit, rather than being driven by the energetics of the centrally placed salt bridge between the c-ring and the stator subunit a [66].

Another interesting biological system that was explored with our CG model is the translocon complex that controls the translocation of polypeptides across the membrane. We used the CG model to address several key questions about this system, starting with the mechanism of membrane insertion of charged residues [69]. We then made significant advances in exploring the energetics of the translocon-assisted protein insertion, where we challenged ourselves to obtain the complete free energy profile for the protein translocation through the translocon and the partition to the water and membrane phases. By applying several constraints on the system, we were able to obtain a free energy profile [70] that was used to investigate the effect of different mutations and the ribosome binding. Comparison with experimental data led to the conclusion that the insertion process is most likely a non-equilibrium process, and that the insertion barrier into the translocon controls the peptide topology. The obtained free energy profile allowed us to approach extremely challenging and fundamental questions regarding the nature of the coupling between two large biological systems: the translocon and ribosomes. That is, we investigated the origin of the experimentally observed [71] biphasic pulling force from the translocon that releases the stalling of some elongated nascent peptide chain from the ribosome. By combining the estimates of the chemical barriers of peptide bond formation for the regular and stalled peptide sequences with the profile for the translocon-assisted protein membrane integration and performing Langevin dynamics simulations of the ribosome/translocon model, we were able to reproduce the experimental effect ([72] and Fig. 6). Our simulation of the action of voltage activation ion channels [73] provides another instructive case study. The above examples highlight the importance of obtaining the relevant free energy profiles for a thorough understanding of the mechanisms underlying different biological processes.

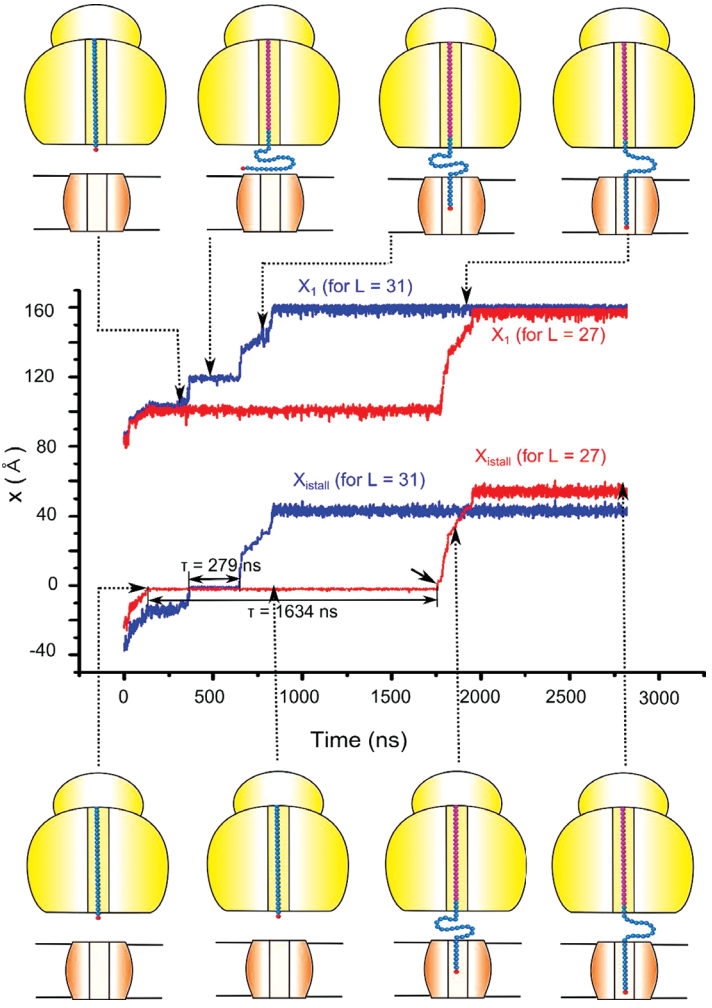


**FIGURE 6.** Simulations of the coupling between the ribosome and the translocon (TR). The simulation addresses the effect of the TR on stalled peptides, where for some lengths of the linker, L, the coupling to the TR helps to release the stalled peptide. The time dependence of $x_{istall}$ and $x_1$ for a peptide chain with 40 and 36 units is shown here, which corresponds to L = 31 (blue) and 27 (red), respectively. The x coordinate designates the insertion coordinate and is defined in [72]. The barriers used for the LD simulations were obtained by scaling down the energy terms by 0.43. This allowed for the simulation of the insertion process in a relatively short timescale, and then estimating the relevant time for the actual barriers by using the corresponding Boltzmann probability. The snapshots on the top and bottom of the plot show the configuration of the nascent peptide chain for L = 31 and L = 27, respectively. The ribosome and TR are shown schematically, the starting configuration of the nascent chain is in cyan, the leading particle ($x_1$) is in red, and all other particles added to the growing chain are shown in magenta. The interpolated times (which should be obtained without scaling) for L = 31 and L = 27 are 6 min and 36 min, respectively. This figure is taken from [72], which also gives a complete description of the problem and the simulations performed.

## FUTURE DIRECTIONS

The enormous increase in computer power makes it virtually certain that computer simulations will increasingly become the key tool in modeling complex systems. Although it is hard to predict the exact future direction of the field, it may be useful to consider some promising directions. One clear trajectory is the field of fighting drug resistance. That is, at the turn of the 20th century, we had a short life-span due in part to the effect of deadly diseases. The discovery of penicillin and other drugs helped to protect against major diseases. However, in recent years, the phenomenon of drug resistance has started to reverse the picture. It appears that there is no magic bullet: key drugs become ineffective due in part to excessive irresponsible use of antibiotics. In other cases, we have diseases like HIV that are hard to combat due to the inherent fast mutations of the pathogen, or diseases like malaria, where we also have drug resistance. Thus, it has become essential to pursue new drug design strategies. Here, the challenge is to predict the moves of the pathogen in response to different drugs. Of course, one can try to explore the actual experimental response of the pathogen to different drugs, but this is obviously not a predictive approach. Thus, it would be tremendously helpful to use computational strategies for studies of drug resistance, but such a strategy must drastically reduce the options for effective mutations.

One such strategy is to exploit the fact that a virus fighting against a given drug must reduce the affinity to this drug, while still maintaining a reasonable catalytic efficiency towards the native substrates. Thus, it would be useful to find a way to out-maneuver the virus by designing inhibitors, whose binding to the target enzyme cannot be reduced by mutations without significantly reducing its $k_{cat}/K_M$ value. In other words, the drug resistant mutants must increase $K_i$ for the drug, while maintaining a reasonable $k_{cat}/K_M$ value for their native substrate. Thus, an effective strategy can exploit the ability to calculate the vitality value, $\gamma (\gamma = K_i\, k_{cat}/K_M)$, and to determine the chance that the virus will mutate in a given way. Combining the vitality value and other constraints (such as maintaining reasonable protein stability) will provide the survival value, which is the chance that the given mutant will survive in the presence of the specific drug. Our ability to evaluate the vitality value has already been demonstrated in preliminary studies [74, 75], and thus we are confident that it will be possible to develop a robust ability to predict the survival of the virus mutants, and thus to design drugs that would reduce the resistance problem. Other constraints such as mutation tendency and other factors can be introduced by bioinformatics approaches. It is quite likely that an aggressive use of computer simulations will provide a way to beat pathogens in their own game.

Another exciting direction can involve the design of drugs that interfere with protein-protein interactions. Here, the idea is to learn about key interactions between partners in signal transduction networks (*e.g.* Ras/RAF [76]), and then designing molecules based on the regions with the strongest interaction (see the strategy in Fig. 49 of [77]). Yet another direction that will gradually mature is the field of truly rational enzyme design. Here, it seems obvious to me that the design approaches must involve actual modeling of the catalytic effect of different design options. It is unlikely that unverified ideas of how enzymes may work (*e.g.* the idea that enzyme catalysis is due to dynamical effects), or ideas that are based on gas-phase modeling, would lead to artificial enzymes with large catalytic effects. On the other hand, approaches that can reproduce the catalytic effects of known enzymes must eventually be very powerful in screening different design options. Multiscale modeling of the action of molecular complexes is likely to be used in describing signal transduction, and allowing one to have a clearer and clearer understanding of cellular action. Finally, it should also be mentioned that multiscale modeling provides a very powerful tool in modeling non-biological systems. Promising directions here include the design of catalysts for a wide range of applications, the design of advanced materials, and the optimization of nanotechnological devices. Overall, the use of computer modeling is likely to increase enormously in any branch of molecule science, as well as in modeling very large systems that can be considered as macroscopic systems.

## ACKNOWLEDGMENT

I would like to express my gratitude to my wife Tamar, who followed me along my scientific life, and helped me both to endure difficulties and to enjoy successes.

## REFERENCES

1. Pople, J.A., "Quantum chemical models (Nobel lecture)," *Angewandte Chemie-International Edition*, 1999. **38**(13–14): p. 1894–1902.
2. Warshel, A. and M. Levitt, "Theoretical Studies of Enzymic Reactions—Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme," *Journal of Molecular Biology*, 1976. **103**(2): p. 227–249.
3. Fersht, A., *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*. 1999, New York: W.H. Freeman. xxi, 631 p.
4. Warshel, A., *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. 1997, Wiley-Interscience Imprint John Wiley & Sons: Hoboken. 256 p. ill.
5. Jencks, W.P., *Catalysis in chemistry and enzymology*. McGraw-Hill series in advanced chemistry. 1969, New York,: McGraw-Hill. xvi, 644 p.

6. Levitt, M. and A. Warshel, "Computer Simulation of Protein Folding," *Nature*, 1975. **253**(5494): p. 694–698.
7. Warshel. A. and Y. Shalitin, "Abstract of the 34th meeting: On the interaction of Chymotrypsin with ionized substrates," *Israel Journal of Chemistry*, 1966. **4**: p. 85.
8. Lifson, S. and A. Warshel, "Consistent Force Field for Calculations of Conformations Vibrational Spectra and Enthalpies of Cycloalkane and N-Alkane Molecules," *Journal of Chemical Physics*, 1968. **49**(11): p. 5116–&.
9. Warshel, A. and S. Lifson, "Consistent Force Field Calculations .2. Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes," *Journal of Chemical Physics*, 1970. **53**(2): p. 582–&.
10. Bixon, M. and S. Lifson, "Potential Functions and Conformations in Cycloalkanes," *Tetrahedron*, 1967. **23**(2): p. 769–&.
11. Hendrickson, J.B., "Molecular Geometry .4. Medium Rings," *Journal of the American Chemical Society*, 1964. **86**(22): p. 4854–&.
12. Allinger, N.L. and J.T. Sprague, "Conformational-Analysis .84. Study of Structures and Energies of Some Alkenes and Cycloalkenes by Force Field Method," *Journal of the American Chemical Society*, 1972. **94**(16): p. 5734–&.
13. Wilson, E.B., Decius, J.C. and Cross, P.C., *Molecular vibrations; the theory of infrared and Raman vibrational spectra*. 1955, New York,: McGraw-Hill. 388 p.
14. Levitt, M., "The birth of computational structural biology," *Nature Structural Biology*, 2001. **8**(5): p. 392–393.
15. Warshel, A. and A. Bromberg, "Oxidation of 4a,4b-Dihydrophenanthrenes .3. A Theoretical Study of Large Kinetic Isotope Effect of Deuterium in Initiation Step of Thermal Reaction with Oxygen," *Journal of Chemical Physics*, 1970. **52**(3): p. 1262–+.
16. Honig, B. and M. Karplus, "Implications of Torsional Potential of Retinal Isomers for Visual Excitation," *Nature*, 1971. **229**(5286): p. 558–&.
17. Warshel, A. and M. Karplus, "Calculation of Ground and Excited-State Potential Surfaces of Conjugated Molecules .1. Formulation and Parametrization," *Journal of the American Chemical Society*, 1972. **94**(16): p. 5612–&.
18. Warshel, A., "Quantum-Mechanical Consistent Force-Field (Qcff/Pi) Method—Calculations of Energies, Conformations and Vibronic Interactions of Ground and Excited-States of Conjugated Molecules," *Israel Journal of Chemistry*, 1973. **11**(5): p. 709–717.
19. Phillips, D.C., "Hen Egg-White Lysozyme Molecule," *Proceedings of the National Academy of Sciences of the United States of America*, 1967. **57**(3): p. 484–&.
20. Levitt, M., "On the Nature of the Binding of Hexa-N-Acetyl Glucosamine Substrate to Lysozyme," *Peptides, Polypeptides and Proteins*, 1974: p. 99–113.
21. Levitt, M., "Energy Refinement of Hen Egg-White Lysozyme," *Journal of Molecular Biology*, 1974. **82**(3): p. 393–420.
22. Pople, J.A. and D.L. Beveridge, *Approximate molecular orbital theory*. McGraw-Hill series in advanced chemistry. 1970, New York,: McGraw-Hill. viii, 214 p.
23. Luzhkov, V. and A. Warshel, "Microscopic Models for Quantum Mechanical Calculations of Chemical Processes in Solutions: LD/AMPAC and SCAAS/AMPAC Calculations of Solvation Energies," *J. Comp. Chem.*, 1992. **13**: p. 199–213.

24. Kamerlin, S.C.L., M. Haranczyk, and A. Warshel, "Progress in Ab Initio QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pK(a), Redox Reactions and Solvation Free Energies," *Journal of Physical Chemistry B*, 2009. **113**(5): p. 1253–1272.
25. Warshel, A., "Energetics of Enzyme Catalysis," *Proc. Natl. Acad. Sci. USA.*, 1978. **75**: p. 5250–5254.
26. Warshel, A., et al., "Electrostatic basis for enzyme catalysis," *Chem. Rev.*, 2006. **106**(8): p. 3210–35.
27. Perutz, M.F., "Electrostatic Effects in Proteins," *Science*, 1978. **201**(4362): p. 1187–1191.
28. Warshel, A. and S.T. Russell, "Calculations of Electrostatic Interactions in Biological Systems and in Solutions," *Q. Rev. Biophys.*, 1984. **17**: p. 283–421.
29. Warshel, A., et al., "Modeling electrostatic effects in proteins," *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 2006. **1764**(11): p. 1647–1676.
30. Vicatos, S., M. Roca, and A. Warshel, "Effective approach for calculations of absolute stability of proteins using focused dielectric constants," *Proteins-Structure Function and Bioinformatics*, 2009. **77**(3): p. 670–684.
31. Warshel, A. and R.M. Weiss, "An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes," *J. Am. Chem. Soc.*, 1980. **102**(20): p. 6218–6226.
32. Warshel, A., "Computer simulations of enzyme catalysis: Methods, progress, and insights," *Ann. Rev. Biophys. Biomol. Struct.*, 2003. **32**: p. 425–443.
33. Frushicheva, M.P., et al., "Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase," *Proceedings of the National Academy of Sciences of the United States of America*, 2010. **107**(39): p. 16869–16874.
34. Plotnikov, N.V. and A. Warshel, "Exploring, Refining, and Validating the Paradynamics QM/MM Sampling," *Journal of Physical Chemistry B*, 2012. **116**(34): p. 10342–10356.
35. Prasad, B.R., et al., "Quantitative exploration of the molecular origin of the activation of GTPase," *Proceedings of the National Academy of Sciences of the United States of America*, 2013. **110**(51): p. 20509–20514.
36. Wesolowski, T.A., "Quantum chemistry 'without orbitals'—An old idea and recent developments," *Chimia*, 2004. **58**(5): p. 311–315.
37. Wesolowski, T.A. and A. Warshel, "Frozen Density-Functional Approach for Ab-Initio Calculations of Solvated Molecules," *J. Phys. Chem.*, 1993. **97**: p. 8050–8053.
38. Tully, J.C. and R.M. Preston, "Semiclassical Trajectory Surface Hopping Approach to Nonadiabatic Molecular Collisions: The reaction of $H^+$ with D2," *J. Phys. Chem.*, 1971. **56**: p. 562–572.
39. Warshel, A. and M. Karplus, "Semiclassical Trajectory Approach to Photoisomerization," *Chemical Physics Letters*, 1975. **32**(1): p. 11–17.
40. Warshel, A. and M. Karplus, "Calculation of Pi Pi] Excited-State Conformations and Vibronic Structure of Retinal and Related Molecules," *Journal of the American Chemical Society*, 1974. **96**(18): p. 5677–5689.
41. Kobilka, B.K. and R.J. Lefkowitz, "Nobel Prizes 2012: Cells and Sensitivity-GPCR Heroes Honored," *Angewandte Chemie-International Edition*, 2012. **51**(45): p. 11199.

42. Warshel, A., "Bicycle-Pedal Model for 1st Step in Vision Process," *Nature*, 1976. **260**(5553): p. 679–683.
43. Mathies, R.A., et al., "Direct Observation of the Femtosecond Excited-State Cis-Trans Isomerization in Bacteriorhodopsin," *Science*, 1988. **240**(4853): p. 777–779.
44. Frutos, L.M., et al., "Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry," *Proceedings of the National Academy of Sciences of the United States of America*, 2007. **104**(19): p. 7764–7769.
45. Creighton, S., et al., "Simulating the Dynamics of the Primary Charge Separation Process in Bacterial Photosynthesis," *Biochemistry*, 1988. **27**: p. 774–781.
46. Holzwarth, A.R. and M.G. Muller, "Energetics and kinetics of radical pairs in reaction centers from Rhodobacter sphaeroides: A femtosecond transient absorption study," *Biochemistry*, 1996. **35**(36): p. 11820–11831.
47. Churg, A.K. and A. Warshel, "Control of Redox Potential of Cytochrome c and Microscopic Dielectric Effects in Proteins," *Biochemistry*, 1986. **25**: p. 1675.
48. Aqvist, J. and A. Warshel, "Energetics of Ion Permeation through Membrane Channels—Solvation of Na+ by Gramicidin-A," *Biophysical Journal*, 1989. **56**(1): p. 171–182.
49. Valleau, J.P. and G.M. Torrie, *A Guide to Monte Carlo for Statistical Mechanics: 2. Byways*, in *Statistical Mechanics*, B. Berne, Editor. 1977, Springer US. p. 169–194.
50. Warshel, A., "Calculations of Chemical Processes in Solutions," *Journal of Physical Chemistry*, 1979. **83**(12): p. 1640–1652.
51. Warshel, A., "Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions," *J. Phys. Chem.*, 1982. **86**: p. 2218–2224.
52. Warshel, A., *Simulating the Energetics and Dynamics of Enzymatic Reactions*, in *Specificity in Biological Interactions*, C. Chagas and B. Pullman, Editors. 1984, Springer Netherlands. p. 59–81.
53. Jorgensen, W.L. and C. Ravimohan, "Monte Carlo Simulation of Differences in Free Energies of Hydration," *Journal of Chemical Physics*, 1985. **83**(6): p. 3050–3054.
54. Rao, S.N., et al., "Free Energy Perturbation Calculations on Binding and Catalysis after Mutating Asn 155 in Subtilisin," *Nature*, 1987. **328**: p. 551–554.
55. Warshel, A., "Molecular dynamics simulations of biological reactions," *Accounts of Chemical Research*, 2002. **35**(6): p. 385–395.
56. Shaw, D.E., "Millisecond-long molecular dynamics simulations of proteins on a special-purpose machine," *Abstracts of Papers of the American Chemical Society*, 2012. **243**.
57. Voelz, V.A., et al., "Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1–39)," *Journal of the American Chemical Society*, 2010. **132**(5): p. 1526–+.
58. Dryga, A. and A. Warshel, "Renormalizing SMD: The Renormalization Approach and Its Use in Long Time Simulations and Accelerated PAU Calculations of Macromolecules," *Journal of Physical Chemistry B*, 2010. **114**(39): p. 12720–12728.
59. Jarzynski, C., "Nonequilibrium equality for free energy differences," *Physical Review Letters*, 1997. **78**(14): p. 2690–2693.
60. Kamerlin, S.C., et al., "Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems," *Ann. Rev. Phys. Chem.*, 2011. **62**: p. 41–64.

61. Dryga, A. and A. Warshel, "Renormalizing SMD: The Renormalization Approach and Its Use in Long Time Simulations and Accelerated PMF Calculations of Macromolecules," *J. Phys. Chem. B*, 2010. **114**(39): p. 12720–12728.
62. Mukherjee, S. and A. Warshel, "Electrostatic origin of the mechanochemical rotary mechanism and the catalytic dwell of F1-ATPase," *Proc. Natl. Acad. Sci. USA*, 2011. **108**(51): p. 20550–5.
63. Wang, H.Y. and G. Oster, "Energy transduction in the F-1 motor of ATP synthase," *Nature*, 1998. **396**(6708): p. 279–282.
64. Pu, J.Z. and M. Karplus, "How subunit coupling produces the gamma-subunit rotary motion in F-1-ATPase," *Proc. Natl. Acad. Sci. USA*, 2008. **105**(4): p. 1192–1197.
65. Noji, H., et al., "Direct observation of the rotation of F-1-ATPase," *Nature*, 1997. **386**(6622): p. 299–302.
66. Mukherjee, S. and A. Warshel, "Realistic simulations of the coupling between the protomotive force and the mechanical rotation of the F0-ATPase," *Proc. Natl. Acad. Sci. USA*, 2012. **109**(37): p. 14876–81.
67. Dimroth, P., C. von Ballmoos, and T. Meier, "Catalytic and mechanical cycles in F-ATP synthases—Fourth in the cycles review series," *Embo Reports*, 2006. **7**(3): p. 276–282.
68. Junge, W., H. Sielaff, and S. Engelbrecht, "Torque generation and elastic power transmission in the rotary F0F1-ATPase," *Nature*, 2009. **459**(7245): p. 364–370.
69. Rychkova, A., S. Vicatos, and A. Warshel, "On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes," *Proc. Natl. Acad. Sci. USA*, 2010. **107**(41): p. 17598–17603.
70. Rychkova, A. and A. Warshel, "Exploring the nature of the translocon-assisted protein insertion," *Proc. Natl. Acad. Sci. USA*, 2013. **110**(2): p. 495–500.
71. Ismail, N., et al., "A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration," *Nature Structural & Molecular Biology*, 2012. **19**(10): p. 1018–U68.
72. Rychkova, A., et al., "Simulating the pulling of stalled elongated peptide from the ribosome by the translocon," *Proc. Natl. Acad. Sci. USA*, 2013. **110**(25): p. 10195–200.
73. Dryga, A., et al., "Realistic simulation of the activation of voltage-gated ion channels," *Proc. Natl. Acad. Sci. USA*, 2012. **109**(9): p. 3335–3340.
74. Ishikita, H. and A. Warshel, "Predicting drug-resistant mutations of HIV protease," *Angew. Chem. Int. Ed.*, 2008. **47**(4): p. 697–700.
75. Singh, N., M.P. Frushicheva, and A. Warshel, "Validating the vitality strategy for fighting drug resistance," *Proteins-Structure Function and Bioinformatics*, 2012. **80**(4): p. 1110–1122.
76. Muegge, I., T. Schweins, and A. Warshel, "Electrostatic contributions to protein-protein binding affinities: Application to Rap/Raf interaction," *Proteins-Structure Function and Genetics*, 1998. **30**(4): p. 407–423.
77. Kamerlin, S.C.L., et al., "Why nature really chose phosphate," *Quarterly Reviews of Biophysics*, 2013. **46**(1): p. 1–132.

Portrait photo of Arieh Warshel by photographer Alexander Mahmoud.