Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems

Shina C.L. Kamerlin,[1,2] Spyridon Vicatos,[1] Anatoly Dryga,[1] and Arieh Warshel[1]

[1]Department of Chemistry, University of Southern California, Los Angeles, California 90089; email: vicatos@usc.edu, warshel@usc.edu

[2]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-10691 Stockholm, Sweden; email: l.kamerlin@gmx.com

Annu. Rev. Phys. Chem. 2011. 62:41–64

First published online as a Review in Advance on October 28, 2010

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

This article's doi: 10.1146/annurev-physchem-032210-103335

0066-426X/11/0505-0041$20.00

Keywords

simplified models, reference potential approaches, free-energy landscapes, QM/MM approaches, long-timescale simulations, renormalization

Abstract

Recent years have witnessed an explosion in computational power, leading to attempts to model ever more complex systems. Nevertheless, there remain cases for which the use of brute-force computer simulations is clearly not the solution. In such cases, great benefit can be obtained from the use of physically sound simplifications. The introduction of such coarse graining can be traced back to the early usage of a simplified model in studies of proteins. Since then, the field has progressed tremendously. In this review, we cover both key developments in the field and potential future directions. Additionally, particular emphasis is given to two general approaches, namely the renormalization and reference potential approaches, which allow one to move back and forth between the coarse-grained (CG) and full models, as these approaches provide the foundation for CG modeling of complex systems.

1. INTRODUCTION

Coarse graining: a means to study complex systems by smoothing away the fine details of the full explicit system (e.g., using pseudo-atoms)

CG: coarse-grained

Computer modeling of the function of macromolecules and related systems presents a problem of enormous complexity. Although available computer power has increased rapidly, there are, nevertheless, still many cases for which the use of brute-force simulations is clearly not the best approach. Furthermore, there exist many systems whose nature was correctly elucidated even before the emergence of the current level of computing power. This issue can be most dramatically illustrated by considering the possibility of using ab initio representations of the entire protein in the study of the action of molecular motors. Obviously, this has enormous difficulties. However, there also exist far less dramatic examples of cases in which one cannot (and perhaps should not) progress without the use of a simplified model, and, in fact, recent years have witnessed a growing appreciation of the fact that the simulation of complex systems, and, in particular, the modeling of biological function, can greatly benefit from physically sound simplifications. The idea of such coarse graining first appeared in protein folding studies and has since become a common, well-accepted, and powerful strategy. This review considers the developments in the field, covering both advances and new directions. We try to emphasize that often focusing on minute details is not the best way to model a complex system. We also emphasize the importance of capturing the relevant physical features, as well as the strategies that allow one to move back and forth between the coarse-grained (CG) and explicit models.

The idea of using a simplified model in computational studies of proteins dates back to Levitt & Warshel's (LW) simplified model for protein folding (1), as well as the much simpler Gō model (2), which emerged at the end of the same year as the LW model. The field of multiscale modeling of proteins and related systems has grown tremendously since that time, and this review considers developments in the field. We also emphasize the critical difference between a simplified complete model and an explicit but incomplete model, as well as approaches that allow one to move between models of different degrees of sophistication.

In general, one may start from Einstein's advice to "make everything as simple as possible, but not simpler." Here, one should of course define what the question is, and what level of understanding is desired. It is also important to keep in mind the available computer power and its ability to give a stable (converging) result at that given moment in the history of the field. An excellent example is provided by the early difficulties with acceptance of the role of simplified models. People hoped for a complete Hamiltonian representation of subsystems, where in fact a soft sphere dipolar model for the solvation of molecules in water was suitable (3). This is representative of a time during which studies of two water molecules and an ion were tractable (e.g., 4), with a complete Hamiltonian (but not, of course, with a Hamiltonian relevant for the complete solute-solvent system). A system with a simplified model for each water molecule, but with a physical representation of the surface between the explicit system and the bulk, turns out to be an excellent model for microscopic studies of solvation effects (see, e.g., 5, 6). A preoccupation with minute details might prevent the realization of this point, thus slowing progress in the field.

Similarly, sometimes a dangerous assumption is that the availability and capability of running long trajectories (such as a millisecond trajectory for an ion channel) represent a major breakthrough that will yield great insight into a system. This can be quite problematic, for example, because understanding selectivity requires running multiple simulations and exploring the effect of different parameters on the overall current (e.g., 7). The same holds true for a single long trajectory that explores protein folding (8). Of course, with increasing computational power, it will be possible to run multiple long simulations using all-atom models. However, by that time, many of the key problems will probably have already been resolved by the use of simpler models.

This review discusses the broad usage of CG models. Here, some recent relevant reviews of this topic (which include 9–12) can also be useful. In addition to pointing out the wider scope of the field, we also emphasize two general approaches, namely the renormalization approach, which allows one to move from a full to a simplified model, and the reference potential approach, which allows one to go from the simplified to the full model. We believe that keeping these approaches in mind provides an excellent way to understand the foundations of CG modeling.

Reference potential: a simplified (e.g., coarse-grained) potential that can be used as a reference for a full explicit potential, thus providing significant savings in computational cost

2. PROTEIN FOLDING AS AN EXAMPLE OF THE NEED FOR COARSE-GRAINED MODELS

Probably the earliest example of the CG idea in biology is the development of the simplified protein folding model (1). That is, protein folding presented an enormous challenge, in light of what came to be known as the Levinthal paradox (13) where it seemed that it was close to impossible to rationalize how a protein with so many degrees of freedom is capable of folding within any reasonable timescale. In 1974, we tried to attack this fundamental problem, and, realizing that even the minor energy minimization of a protein took an extremely long time, we moved to a seemingly drastic simplification (while still retaining the main physics of the problem). This resulted in the replacement of the protein side chains by spheres with an effective potential that implicitly represented the average potential of the solvated side chains. The main chain was represented by virtual bonds between the C_α's. This model was surprisingly effective and in fact provided the first reasonable physically based solution of the Levinthal paradox, by finding several native structures while starting from the unfolded state (see **Figure 1** and the discussion in 1). The success of this model led to significant criticism (see, e.g., 14). Today, the best response to such criticism has simply been the widespread adaptation of this model (see below). At any rate, a further useful simplification that emerged at the same time was a model that kept the helices of the simplified model in a fixed helical configuration (15).

A related model was introduced by Gō and coworkers (2) shortly after our model. In its early versions, this model considered a chain of nonintersecting units of a given length on the two-dimensional (2D) square lattice and explored interesting formal issues such as the partition function of the simplified model. However, the early version of this model sacrificed too much physics to be considered a realistic model of a protein and thus could not be used (before the introduction of major changes) to explore the protein folding puzzle. While this is clearly an interesting approach that can provide some useful basic information, constraining the system to a 2D square lattice simplifies and limits the correspondence between this 2D model and the actual complex structure of a protein, making it somewhat primitive to produce a real protein topology (17–20). This brings us to the point that different simplifications are needed for different problems, and the relationship between the CG model and the full model should always be considered. In fact, as discussed below, it is crucial to be able to move between the CG and full models.

The study of protein folding by simplified models has become a major research field. A significant amount of work (e.g., 21–25) was done using the simple Gō model (which has became known as a lattice model). In contrast to this, other works that tried to be more realistic used the LW model, which has been termed an off-lattice model. The selection of which model to use depends on the question being asked, and both models have provided enormous insight into the folding process, on the corresponding landscape and timescales (26), as well as encouraging more experimental studies, and the initial attempts at all-atom simulations (e.g., 27, 28).

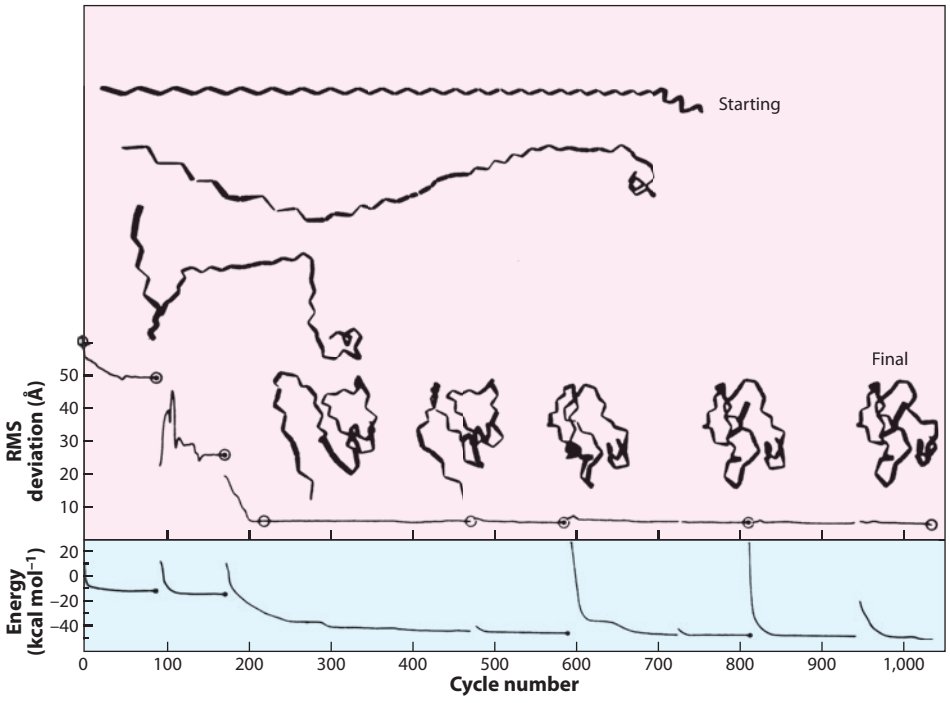


Figure 1

The folding trajectory produced by the coarse-grained model of Reference 1. This simulation was initiated from an extended starting conformation, with α of all terminal helices set to 180°, with the exception of residues 48 to 58, where $\alpha = 45°$. No other knowledge whatsoever was used about the native protein during the simulation. At the end of each minimization cycle, the conformation was thermalized; i.e., thermal fluctuations were reintroduced, and the conformation was considered to be vibrating around the minimum (such that each mode has an average kinetic energy of $kT/2$, where k is the Boltzmann constant and T is the absolute temperature). The thermal vibration is then suddenly stopped, and a new starting conformation for the next pass of energy minimization is generated from the structure at this point. Such normal-mode thermalization avoids nonproductive changes in the protein conformation because it knows which combination of angle changes should cause the greatest change in conformation for a given energy increase. The normal-mode treatment presents what is probably the first realistic dynamical treatment of large-amplitude protein motions. This figure was originally presented in Reference 1. Adapted with permission from Macmillan Publishers Ltd: *Nature*, copyright 1975.

Despite the emergence of a powerful simplified model for protein folding, the lingering question remains as to what the relationship between the simplified model and the corresponding explicit model actually is. This appears to be crucial in light of the increasing interest in protein landscapes, which is an important element in attempts to explore the energetics and kinetics of the folding process (for a review, see 21, and references therein). Furthermore, the ability to explore and sample large regions in the accessible conformational space can help investigators improve the description of functional properties, as well as explore the possible relationships between landscape and function (e.g., 21, 25, 29). Unfortunately, the detailed sampling of protein landscapes requires enormous computational resources. Thus it is important to develop multiscale approaches that allow one to effectively generate the free-energy surface for folding and related processes. A general way to resolve this problem was introduced some time ago (30), in which we

proposed the use of a simplified model as a reference potential for explicit calculations of folding free energies. This approach is discussed further in Section 3.

QM/MM: quantum mechanical/molecular mechanical

PMF: potential of mean force

3. USING THE COARSE-GRAINED MODEL AS A REFERENCE POTENTIAL FOR EXPLICIT CALCULATIONS

The use of CG models leaves one with the question of how capable the given model is at reproducing the corresponding full model. Fortunately, it is possible to systematically resolve this issue by generating the results (i.e., the free energy and some dynamical features) of the explicit model using the simplified model as a reference potential. More specifically, following the idea introduced in Reference 30, we can use the CG model as a reference for the full model. In this approach, which is described in **Figure 2**, we evaluate the free energy of moving between two states in the explicit model by evaluating the corresponding free energy with the CG model and then just evaluating the free energy of moving from the CG surface to the surface of the full model at the end points.

The above idea has been demonstrated to work in a protein folding study (30), and other workers (e.g., 25, 31, 32) have also recently explored related strategies. As clarified further below, we have exploited the same idea in a wide range of problems, including the acceleration of quantum mechanical/molecular mechanical (QM/MM) calculations (33–37; see also Section 4), as well as path integral calculations of nuclear quantum mechanical effects (38, 39). In subsequent sections, we also consider several key implementations of the reference potential idea.

One possible application of our reference potential approach is the evaluation of the effect of mutations on protein stability. Although this can be done by evaluating the folding potential of mean force (PMF) for both systems, it is much simpler to use a thermodynamic cycle of the type presented in **Figure 2**, as was done in Reference 40, in which we studied mutations of ubiquitin.

The reference potential idea considered above provides a promising strategy in the field of enzyme design, in which it can be used to evaluate the binding free energy of rate-determining transition states. This can be done by focusing on the electrostatic free-energy contribution, while

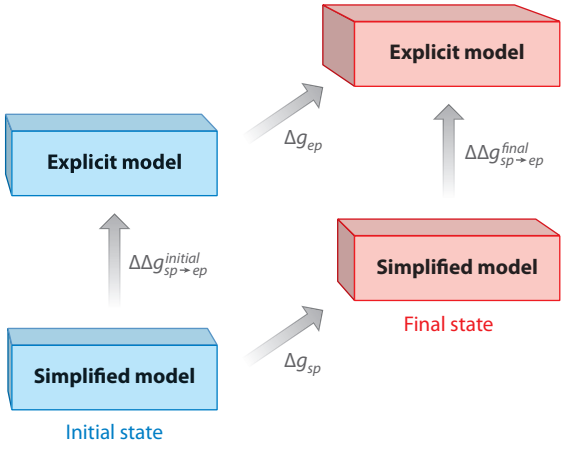


Figure 2

The thermodynamic cycle used to calculate the change (Δg_{ep}) in free energy for a generic process in an explicit system. Having calculated the free-energy change of the corresponding simplified model, Δg_{sp}, umbrella sampling can be used to calculate the free-energy change $\Delta\Delta g_{sp\to ep}$ for the initial and final states to obtain Δg_{ep}.

using the cycle described in Reference 40. Reference 41 discusses the potential of this approach in enzyme design.

4. QM/MM AND RELATED MULTISCALE MODELS

One of the best demonstrations of the multilevel strategy for modeling biological systems involves the development of the QM/MM approach (16) with its crucial electrostatic embedding idea (for reviews see, e.g., 33, 42). Over the years, different versions of the QM/MM model have emerged, although all approaches share the idea of treating the reactive part of the system using a quantum mechanical approach and embedding this part in a system that is treated on a simpler level. This model has rapidly become one of the most popular approaches for studying chemical reactivity in general and enzyme function in particular (e.g., 33, and references therein; 43–46). The main strategies currently being adopted for performing QM/MM calculations are summarized in **Figure 3**, and here we provide a brief overview of each.

Although QM/MM approaches have clearly become an essential tool for modeling enzymatic reactions (at least until it is possible to represent all of the enzyme quantum mechanically), it is clearly important to use this tool correctly. Here an important issue is to perform extensive configurational sampling during the course of the simulation, as the use of QM/MM simulations without proper sampling is not so effective. The difficulties arising from performing only limited

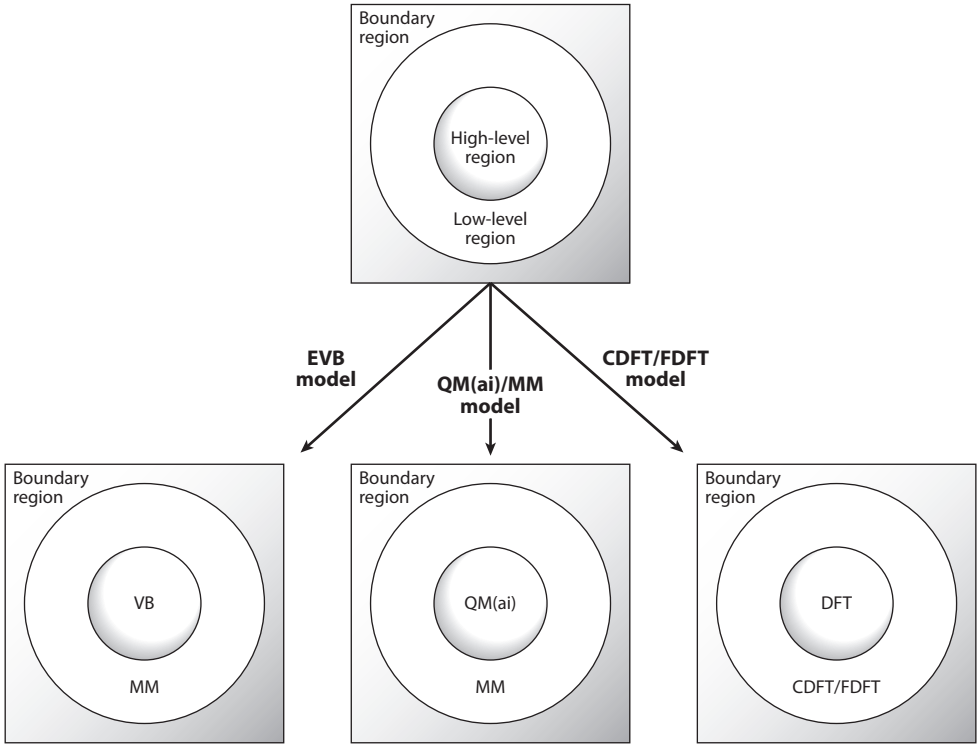


Figure 3

An overview of different quantum mechanical/molecular mechanical (QM/MM) approaches. Illustrated here are the empirical valence bond (EVB) approach, an ab initio QM(ai)/MM approach, and the constrained/frozen density functional theory (CDFT/FDFT) approaches.

energy minimization have been highlighted elsewhere (33), but we would like to point out that one of the most significant shortcomings of such an approach is that, as the enzyme active-site landscape is quite complex, estimating the QM/MM reaction path along a fixed reaction coordinate can reflect artificial minima. This issue is particularly challenging when dealing with ab initio QM systems in QM(ai)/MM simulations. Addressing the need for properly sampling ab initio surfaces has led to several important advances (e.g., 33, 35, 47–53), many of which (e.g., 33, 35, 49, 50, 52, 53) exploit our idea (34, 54) of utilizing a classical potential as a reference for QM/MM calculations.

The key point in the use of a reference potential for QM/MM calculations is the evaluation of the free energy of moving from the reference potential to the ab initio potential. This free energy can be evaluated either by means of a single-step free-energy perturbation (FEP) approach or by means of the linear response approximation (LRA). Although both approaches are viable in principle, the LRA approach is particularly powerful in that it allows one to obtain a reasonable result even in cases in which the ab initio and reference potentials are significantly different. Here the empirical valence bond (EVB) approach [which has been discussed in detail elsewhere (e.g., 55, 56) and reflects the knowledge of a tremendous amount of chemical information] provides a particularly powerful reference potential (for discussion see, e.g., 56). Such an approach has been effectively and successfully applied to the study of activation barriers both in solution and in proteins (e.g., 34, 36, 54, 57), most notably in the case of Reference 36, which used the EVB as a reference potential for QM(ai)/MM calculations to successfully resolve the highly controversial issue (58) of the energetics of the reference reaction for haloalkane dehalogenase in solution. Specifically, this study clarified that the earlier EVB estimate (58) of the catalytic effect and the effect of the enzyme is quantitatively correct, and it is presently the only true QM(ai)/MM study that considers the free-energy surface of the haloalkane dehalogenase reaction in the protein and in solution.

Recently, we also advanced the idea of performing QM(ai)/MM-FEP calculations of solvation free energies using a classical reference potential, by means of a powerful approach in which the solute environment is represented by an average solvent potential, which is then added to the solute Hamiltonian (33, 35). This approach has been demonstrated to lead to computational time savings of up to 1,000 times in QM(ai)/MM-FEP calculations of solvation free energies of simple systems in which the solute structure is kept fixed during the simulation.

The EVB approach is much more that just an effective reference potential; it is in fact probably the most powerful current QM/MM approach when one is interested in long-timescale simulations and extensive sampling (see, e.g., 55, 56). Here, however, we only mention the nature of the energy coordinate, x, which is taken as the energy gap between the diabatic states. This selection (55, 59) is particularly powerful when one tries to represent the entire many-dimensional solvent space by a single coordinate (see 60), as it guarantees accelerated convergence for processes in condensed phases because it captures the main physics of the solvent response. The energy gap as a reaction coordinate has been successfully used to study a wide range of complex problems, such as in the case of ATP synthase, in which the energy gap was capable of describing the coupling between the mechanical and chemical steps (61), or in the cases of chorismate mutase (29) and adenylate kinase (62). We also note that the power of the EVB energy-gap mapping is increasingly being appreciated by other workers (63–65).

The idea of embedding the QM model into a simpler model has been advanced on a level that can be called QM/QM/MM, which is best demonstrated by the so-called frozen density functional theory (FDFT) and constrained density functional theory (CDFT) approaches (37, 47, 54, 57, 66, 67). In the FDFT approach, a very large part of the entire system is represented by a QM(DFT) approach. However, the region immediately surrounding the internal region is represented by fixed DFT densities (37). In contrast, the CDFT approach allows the surroundings to relax by a freeze-and-thaw approach (and thus constrains the surrounding densities rather than freezes

FEP: free-energy perturbation

Linear response approximation (LRA): this approximation assumes that a system following it has the same solvent force constants in the initial and final states

EVB: empirical valence bond

Constrained/ frozen density functional theory (CDFT/FDFT): these approaches split the system into two regions, both treated using ab initio DFT; however, the electron densities of atoms in the outer region are either frozen (FDFT) or constrained (CDFT)

them). The CDFT and FDFT approaches have been recently adopted by key research groups (e.g., 68–70) and are reviewed in detail in, e.g., Reference 71. The FDFT approach also provides an ideal way to embed the central region and its surroundings without the problems associated with the link-atom treatment.

The CDFT approach has been demonstrated to be extremely effective for the evaluation of the diabatic free-energy functional as well as for the exploration of the mixing between diabatic states (66) (i.e., the H_{ij} term) and for the evaluation of QM(ai)/MM free-energy surfaces that take into account the free energy associated with both the substrate and solvent motions, which in turn allows one to obtain a free-energy barrier that properly reflects the solute entropy (72).

As what could perhaps be considered a testimony to the increasing popularity of this approach, there are several recent examples in the literature of some of the most important CDFT ideas reappearing in other forms. The work of Wu & Van Voorhis (73), which might be seen as a breakthrough (74–77), is effectively an adaptation of key CDFT ideas (37, 54, 57, 67) for a somewhat different approach (73) that emphasizes fixing the diabatic densities by Lagrange multipliers (73), rather than by the physically based approach used, e.g., in References 37, 47, 54, 57, 66, and 67. Here, it is also important to clarify a critical point, which is easily misunderstood: That is, diabatic states are never (and should not be) unique. They are simply a useful mathematical representation for solving the physics of the real adiabatic system. In general, it is important to force the diabatic states to reproduce the physics of the reactant and products, and, at least in the case of the crucial charge-transfer reactions, this approach is much more effective than the seemingly rigorous use of Lagrange multipliers (73). The CDFT approach follows the EVB philosophy and considers the wave functions as being Löwdin orthogonalized, and, of course, the corresponding H_{ij} is different from the one used by Wu & Van Voorhis (73), with both providing correct physics if one uses the corresponding diabatic states (see 78).

Another recent development in the field, which has achieved great popularity, is the metadynamics approach of Laio & Parrinello (79). This approach, which is in many ways similar to earlier ideas (e.g., 30, 34, 80, 81) and in some respects to the idea of using a CG reference potential, has been reviewed in detail in Reference 82, so here we only cover it briefly, pointing out its similarities, as well as its shortcomings, relative to the reference potential approach.

At its core, the strategy of the metadynamics model involves attempting to build the best potential, whose addition to the actual potential will result in a flat surface, i.e., the potential that is the closet to $-E(r)$. This is done by iteratively building successive Gaussian potentials that fill the deepest wells along a small number of user-defined chemically relevant collective variables. In this way, the system is allowed to escape over the lowest transition state as soon as the growing biasing potential and the underlying free energy well exactly counterbalance each other, which, as the title of the original Laio & Parrinello work suggests (79), effectively allows the simulation to escape free-energy minima.

Although the metadynamics approach has been elegantly formulated, the philosophy behind it is almost identical to our earlier approach of using a reference potential (see 33 for background). That is, constructing a potential that makes the landscape flat (as the metadynamics approach does) is similar to using a simplified folding model as a reference potential (30) and is additionally quite similar to the general use of a reference potential for accelerated sampling, which has been part of our QM/MM-FEP studies for a long time. Furthermore, although it is currently quite popular to use approaches in which the best reaction coordinate is not assumed a priori (83–85), using chemical knowledge could be superior to a blind search (even though many workers prefer black-box approaches that limit user input). More specifically, using an approach that we call paradynamics, we take a physically based reference potential and make it as close as possible to $E(r)$. This is done by first evaluating the real potential on a rough grid of $n \times m$ points [a search

Metadynamics/paradynamics: alternative approaches for examining an entire complex system using high-level ab initio calculations; the core of both approaches is to identify the best potential that represents the full explicit potential

Simplified folding model: simplified approach to study protein folding by modeling the side chains as, e.g., beads instead of modeling the full explicit system

that can be done by very short molecular dynamics (MD) simulations with a constraint on each grid point or, even better, by evaluating $E(r_{nm})$] and then subsequently fitting the EVB potential to the grid. The fitted EVB is further refined by minimizing $\langle E_{EVB} - E_{real} \rangle$, which is evaluated by running trajectories at the reactant and transition-state regions of both the EVB and the QM/MM surfaces. This refinement is done automatically using the derivatives of the energy gap with regard to the EVB parameters. Once the average energy gap has been minimized, it is rather easy to get the LRA estimate of the free energy for moving from the EVB to the QM/MM surfaces, as these two surfaces are quite similar. We recently (86) demonstrated that this approach is a highly powerful strategy that requires less effort than current implementations of the metadynamics approach.

The success of the paradynamics approach is a reflection of the fact that that the EVB approach not only is a powerful semiempirical QM/MM approach in its own right, but also makes an ideal reference potential for higher-level simulations. This is because it stores a tremendous amount of chemical information, while simultaneously facilitating extensive sampling. Therefore, although metadynamics is in principle a useful tool, it is also an expensive approach, both in terms of computational cost (due to the need for repeated calls to the QM) and in terms of manpower (due to the need for identifying the correct collective variables to obtain meaningful results, which can pose a nontrivial challenge).

Free-energy landscape: the detailed dependence of the free energy of a system on its coordinates; the corrugation of such landscapes is currently a topic of great interest in studies of catalysis and folding

5. REDUCING THE DIMENSIONALITY OF THE FREE-ENERGY LANDSCAPE

One of the key issues when using reduced models is the ability to capture the main physics of the given system. A case in point is the description of the free-energy landscapes for reacting enzymes. An excellent example (29) has been provided by a study of the catalytic landscape of chorismate mutase. This study was performed by using a CG model to generate the free-energy landscape of the enzyme, followed by an explicit EVB evaluation of the activation barriers for the chemical step in different regions of the landscape, defined by the conformational and chemical coordinates. Additionally, the approach introduced in Reference 29 has also been used in a detailed study of adenylate kinase (62; see Section 6).

Now of course an important issue when modeling the catalytic landscape is the choice of a proper reaction coordinate. At present, the most effective possibility is arguably provided by the EVB formalism, which is considered in Section 4.

An excellent example of the use of a reduced coordinate is the description of the folding landscape in terms of an order parameter, ρ (87), which serves as a measure of the degree of similarity of each state with the native conformation. The energy has been demonstrated, on average, to be a decreasing function of ρ (87), in line with the assumption of minimal frustration.

6. RENORMALIZING LONG-TIMESCALE PROCESSES

One of the most interesting problems in CG modeling is the challenge of simulating long-timescale events. The most obvious approach is the use of a friction term to reflect the effect of the implicit thermal bath, using the Einstein (88) or Wang & Uhlenbeck (89) formulations. In fact, the use of frictional models to describe biological and chemical problems has a long history (e.g., 90–93). However, when focusing on obtaining similar physics in the full and CG systems, and on studying long-timescale biological processes, we cannot safely use the frictions obtained by standard frictional models (see below).

Our main approach for studying long-timescale processes involves taking an explicit all-atom potential model and the corresponding simplified model (with its simplified free-energy landscape)

Renormalization approach: a means of moving from a simplified model to a full explicit model, by ensuring that there is a correspondence between the simplified and full models

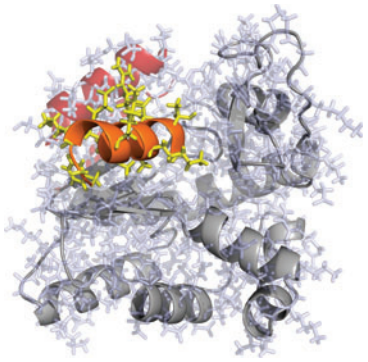
and running MD simulations and Langevin dynamics (LD) on each model, respectively. This is done while imposing a series of constraints on both models, which force these systems to move along a given reaction coordinate on different timescales. Here, larger constraints force faster motion (this approach is illustrated in **Figure 4** and described more rigorously in 94). The optimal friction is obtained by using the same constraints for the simplified and explicit models and adjusting the friction in the simplified model until the timescale for the motion (for each constraint) becomes equivalent in the two models. However, the main point of the renormalization process is that we can force the explicit model to undergo large structural changes within reasonable computational time by using a large constraint, but it is essential to also interpolate the results to cases without constraints (which would require enormous computational time). The use of our renormalization approach has been recently validated on the well-defined test case of a simplified ion channel and expanded to provide an extremely powerful approach, not only for long-timescale simulations, but also to obtain free-energy surfaces (94).

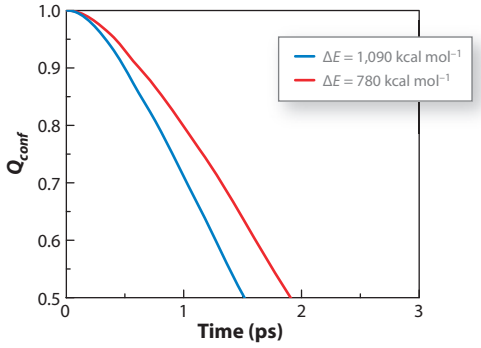
Our validation of the renormalization approach (94) is important, in light of the possibility that it may be perceived that elegant nonequilibrium-type approaches (such as those of 95 and 96) can provide an effective way to obtain free-energy profiles (see the discussion in 94). Here our point has been that the ultimate validation of a simulation approach does not come from the formal elegance of the approach, but rather from its ability to obtain converging results for the systems of interest. In this respect, considering the difficulties with attempts to estimate effective friction using a standard treatment, we insist on the idea that there must be some correspondence between the full and CG models and that the best way to obtain this correspondence is by simultaneously adjusting the friction and PMF of the CG model until the best agreement between the two models is reached. We believe that this seemingly pedestrian approach is actually far more promising than other approaches with more sophisticated formulations.

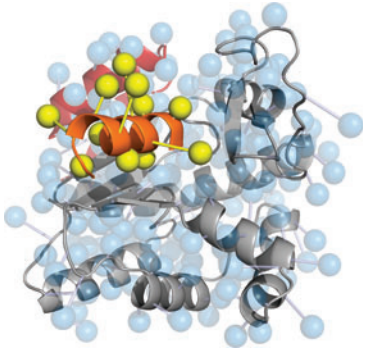
Here it is also useful to consider our recent study (62), which explored the idea that slow-timescale conformational motions play a major role in enzyme catalysis (see the discussion in 62). Although there are clear logical flaws with this argument, which have been discussed in detail in, e.g., References 62 and 97, it is nevertheless crucial to explore this proposal by simulating the relevant processes on the millisecond timescale. Such a study was performed by moving from the full explicit model to a simplified CG model, and then to an even simpler 2D model, which represents the landscape and dynamics in the space defined by the conformational and chemical coordinates of the enzyme under study (62). However, this was achieved by means of a simple single barrier potential in the 2D model, rather than by first evaluating the actual PMF in the full or CG models (because the conclusions were not restricted to any specific barrier shape). The renormalization results for this system from a more recent study with weaker forces are depicted in **Figure 4** (94). As seen from this figure, we can find friction constants that satisfy the

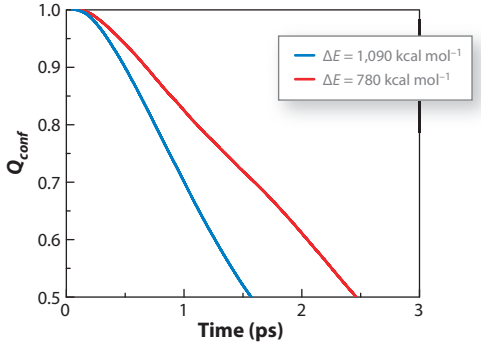
Figure 4

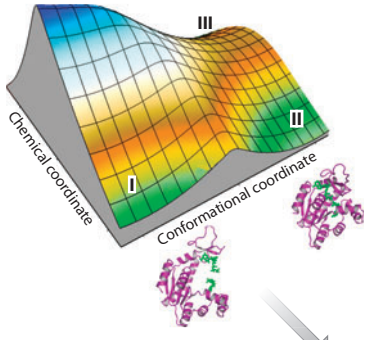
An outline of our renormalization approach for studying long-timescale processes. (*Left column*) Representations of (*top*) the all-atom explicit model, (*middle*) a coarse-grained model in which the protein side chains are represented as spheres, and (*bottom*) our 2D simplified model, in which the system is described by two effective dimensionless coordinates, Q_1 and Q_2, which describe the conformational transition and chemical step, respectively. (*Right column*) Plots showing the corresponding time required for crossing the conformational barrier for each model, using constraints and friction coefficients of different magnitudes. The time required to cross the conformational barrier is similar in all three models (for the same constraint). (*Bottom center*) An actual renormalized surface, obtained using the 2D model, on which it is then possible to run Langevin dynamics without a constraint (illustrated by the transparent plot above the surface). For further information on this approach, we refer readers to Reference 62 and the main text.

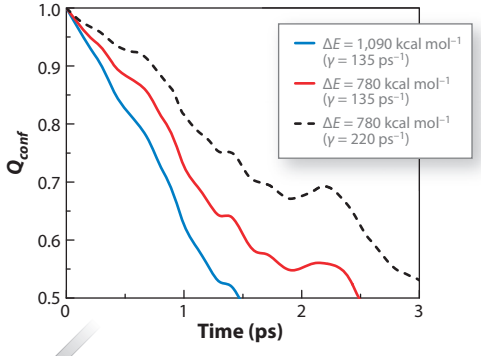


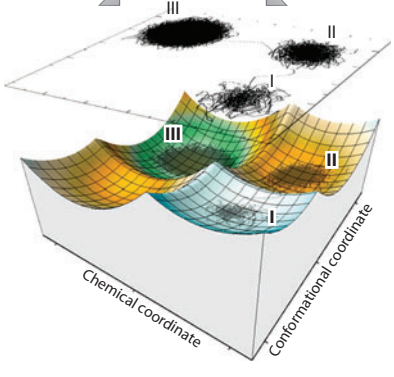












renormalization requirement. However, it seems that also in the challenging case of conformational changes, it is important and possible to further refine the renormalization procedure (see 94).

At any rate, generating a reasonable 2D surface with a reasonable range of frictions allowed us to explore the time dependency of long-timescale processes. This was found to be particularly important in providing the first direct proof that that the dynamical proposal is invalid (62). In this respect it is important to point out that, regardless of the fact that the use of the renormalization treatment for examining large conformational changes still needs further refinement and validation, nevertheless, our finding of the absence of dynamical coupling between the chemical and conformational coordinates in enzymes is still completely valid. That is, our conclusions were obtained by examining all the ranges of reasonable frictions, which included changing the corrugation of the 2D model. It was also confirmed using the much more explicit CG model.

The power of the renormalization approach has also been demonstrated when exploring the selectivity of the KcsA ion channel (7), as discussed in Section 7, and the same approach has been effectively applied to studies of proton transport (PTR) (see Section 9), as well as to the study of vectorial translocation discussed in Section 8. It is also useful to mention that the long-timescale behavior of the conformational coordinates has been effectively explored by the use of CG models (e.g., 98).

At this point, it is important to clarify some recent confusion (99, 100) with regard to the potential of approaches such as transition path sampling, and the assumption that relatively short runs can be used to study millisecond processes (101). In fact, running short reactive trajectories, which may be useful for exploring the reaction coordinate, cannot tell us much about the probability of climbing high activation barriers. The same misunderstandings have appeared in a recent proposal (99) that downhill trajectories can be used to explore the long-timescale coupling between conformational motions and chemical catalysis. Of course, there is no way to explore these issues without simulating long-timescale barrier climbing processes. In conclusion, perhaps the ultimate renormalization approach would involve the emergence of a method that can use the CG simulations to obtain trajectories of the explicit models directly, and there are several options for such a strategy.

7. SIMULATING LONG-TIMESCALE PROTON AND ION TRANSPORT

There exists significant current interest in MD simulations that account for changes in ionization states during the simulated process (e.g., 102, 103). However, the current models do not consider the time dependency of the PTR process. To advance in this challenging field, we combined our approach of time-dependent Monte Carlo (MC) simulations of PTR processes (104) and the simplified protein model in studies of pH-dependent MD (40).

Our model uses a simplified version of the EVB approach, which takes the energetics of any possible proton transfer (PT) step into account. Here, the MC moves are based on the electrostatic energies of the CG model and are then scaled by the characteristic PT time to correspond to the rate constant predicted by transition-state theory. The barrier for the PT moves is then given by a modified Marcus expression (105). This allows us to convert an MC procedure to a time-dependent simulation by exploiting the isomorphism between the probability obtained from the MC procedure and the probability factor of transition-state theory (see 40 for details).

Renormalized simulations have been used to study PTR in carbonic anhydrase (106) and gramicidin (105). Both simulation studies established that PTR in proteins is controlled by the electrostatic free-energy barrier, rather than by the Grotthus mechanism (see 107 for discussion). However, when simulation studies become too time-consuming, one can move to the MC approach. This approach has been used in Reference 104 to study PTR in cytochrome *c* oxidase.

The above renormalization approaches have also been exploited and demonstrated to be highly powerful when exploring the selectivity of the KcsA potassium channels (7). This was done by evaluating the free energy for the penetration of a single ion by means of the PDLD/S-LRA model and then evaluating the ion-ion interaction on the fly by using a dielectric function, $\varepsilon_{eff}(r)$, which represents the effective dielectric for charge-charge interactions by means of a distance-dependent function that typically changes from approximately 20 at short distances to the bulk value (see 108). The friction constant was evaluated by the same renormalization strategy mentioned above, and our treatment of the charge-charge interaction appears to provide an extremely effective way to study long-timescale processes in ion channels. We note that practically the same charge-charge treatment was also used recently by Coalson and coworkers (109) (which can be verified by examining the nature of the dielectric constant).

PDLD/S: semi-microscopic protein dipoles–Langevin dipoles approach

QCFF/PI: quantum mechanical consistent force field method for pi electron systems

8. SIMULATING VECTORIAL PROCESSES

The use of CG models has been effective in situations in which the details of the simulated system are not completely clear. A case in point is a recent study (110) of the nature of the vectorial translocation of a single-stranded DNA by translocases. This study focused on the electrostatic interaction between the protein and the DNA main-chain-ionized phosphate group. The use of the PDLD/S-LRA electrostatic potential for the simplified system generated a unique free-energy surface with a clear valley that leads in one direction, thus supporting a vectorial process. Running LD simulations on the corresponding system, and recovering unidirectional translocation (**Figure 5**), where the energy of ATP hydrolysis is coupled to the translocation process, verified the simple insight provided by inspecting the surface. It should be noted that our CG simulations are, in fact, the first fully consistent simulations of a biological vectorial process in which the results are not assumed a priori or introduced by phenomenological parameters (see the discussion in 110).

9. SIMULATING LIGHT-INDUCED PHOTOBIOLOGICAL ELECTRON AND PROTON TRANSPORT

The study of light-induced electron transport and PTR can be divided into two limits. On the short timescale (i.e., from a few femtoseconds up to 100 ps), one can use explicit all-atom simulations to explore the nature of the corresponding primary events [see, e.g., studies of the primary photochemical event in bacteriorhodopsin (111)] and photosynthesis (112). On the long timescale (i.e., from 100 ps to seconds), one can use CG models to explore the behavior of the system. Here, it would be useful to perform LD simulations of the PTR, following the initial LD treatment of the change in the pK_a of the key ionizable group. An ideal system for such a study is bacteriorhodopsin, in which the nature of the activation barriers for the primary PT was recently explored (113) by combining the QCFF/PI and the EVB methods in a unified QM/MM framework. It was established that the initial charge-separation process, which leads to the primary PT, has sufficient excess free energy to drive the subsequent PT process and, in doing so, provided the first glimpse into the energetics of protein conformational change. Obviously, the use of the LD simulations described in Section 6, and even MC simulations, can be effective in exploring the subsequent steps and the overall pumping process.

The simulation of electron transport processes poses another important challenge. Here we introduced many of the key approaches (see 112) of getting the relevant activation barriers from direct simulations. As far as the present review is concerned, it is instructive to note that one can explore long-range electron transport steps by the same MC used for the PT, but with a characteristic time that reflects the corresponding pre-exponential factor (see 104).

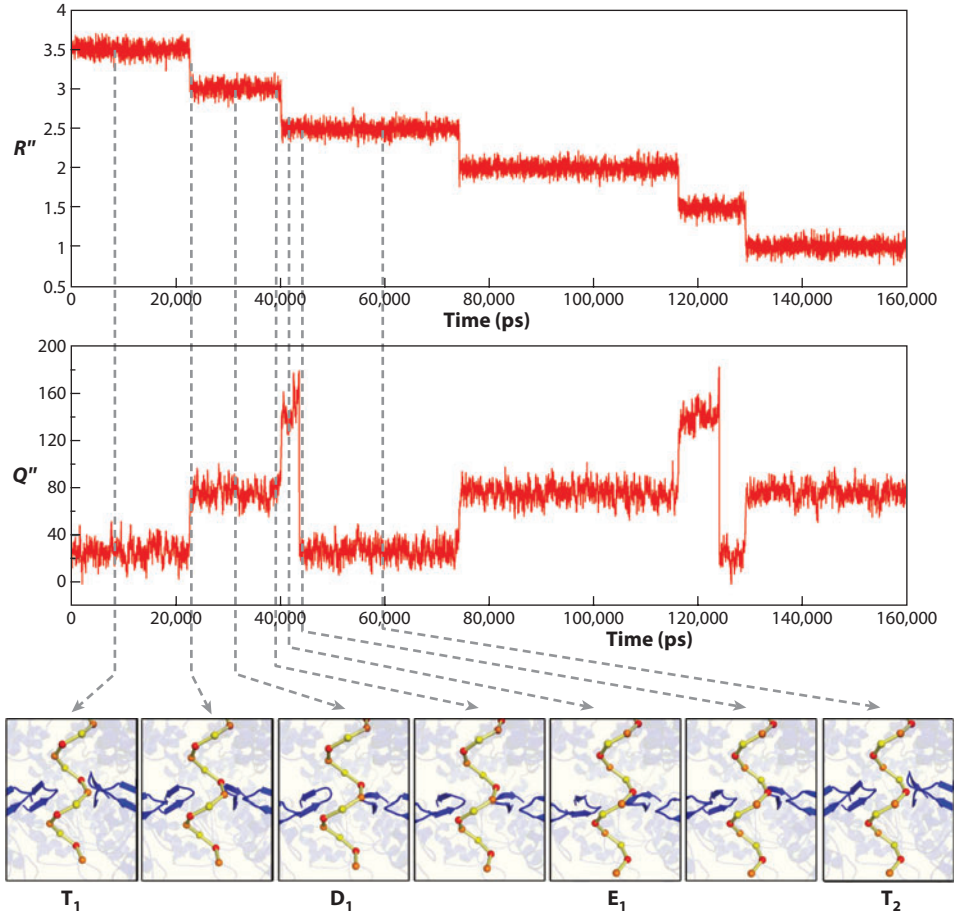


Figure 5

Simulating vectorial translocation in hexameric helicases. Shown here is the simulated time dependency of the R'' and Q'' coordinates corresponding to the DNA and the protein, respectively (for details see 110), as well as snapshots along the translocation path, for a case with a low barrier of 4 kcal mol^{-1} for the relevant transition (note that simulations with higher barriers gave similar results, albeit with longer translocation times). Figure adapted from Reference 110.

10. ELECTROSTATIC MODELING AS A MULTILEVEL STRATEGY

As pointed out in our recent review, electrostatic effects provide what is arguably the most important correlation between structure and function. A crucial question is what level of simplification would still be able to capture the correct physics of electrostatic effects in macromolecules.

In principle, there are three general ways the solvent and/or protein can be modeled. The first is to represent all solvent and/or protein atoms explicitly, using a fully microscopic model (114, 115). However, such an explicit solvation model is computationally expensive, particularly for larger systems [although with increasing computer power, it can be used for many applications (108)]. An effective CG simplification involves representing the solvent molecules as polarizable Langevin dipoles on a grid (16, 108) or as soft-sphere dipoles (3). The limit of the simplification is obtained by using continuum models (for discussion see 108, 116). Unfortunately, such models are based on phenomenological considerations, which make them completely dependent on procedure in

studies of proteins (108). Thus one of the best compromises, with the most consistent connection between macro- and microscopic models, is the PDLD/S-LRA model (see 117, 118 for discussion). This model provides a clear definition of the protein dielectric, in terms of the few factors that are not treated explicitly (namely, the limited protein relaxation and the limited representation of water penetration during the charging process).

Quantum classical path: a multiscale approach that evaluates the nuclear quantum mechanical correction to free-energy surfaces by means of a perturbation from a classical trajectory to a path integral centroid potential, allowing for efficient calculation of the path integral results

11. ADDITIONAL APPROACHES AND SYSTEMS

As the realization of the power of the CG approach in molecular modeling is rapidly growing, there are an increasing number of studies and directions that are being taken (for a recent review, see, e.g., 9). In light of space limitations, we only mention some select approaches here.

Recently, we have witnessed a major productive effort in modeling membranes by CG models by Marrink and coworkers (119), who developed the MARTINI force field, which uses extensive calibration of the building blocks of the CG force field against thermodynamic data. At present, the model shows reasonable behavior for lipid bilayers, in terms of the stress profile across the bilayer and its tendency to form pores, as well as accurate agreement with all-atom simulations for the free energies of lipid desorption and, to some extent, flip-flopping across the bilayer.

Another recent approach is the dynamic linear response theory by Essiz & Coalson (120). This approach has common elements with our LRA approach for free-energy calculations and in exploring the energetics of large conformational changes in proteins (61), as well as with our use of the LRA in simulating fast relaxation processes (60, 112), although the LRT formulation is elegant and independent of our formulation. The linear response theory approach is aimed at studies of the response of a macromolecular system, such as a protein, to a change in the potential of the system, such as a change that is induced when a ligand bound to a well-defined binding pocket within a protein dissociates from the binding pocket. Such studies are formally similar to our use of the linear response theory approach in studies of the response to light-induced charge-transfer processes (112, 121).

Another recent development is Calderon and coworkers' (122) idea of summarizing the state of complex systems by the time series of low-dimensional system observables, such as the use of nonequilibrium trajectories to extract both the equilibrium quantities and kinetic parameters, which are sometimes used to describe the dynamics occurring over longer timescales than those explored in the simulations, by means of a surrogate process approximation method. This interesting approach uses time-series techniques (123, 124) to estimate a low-dimensional stochastic differential equation. These equations approximate the dynamics of an observed signal, which can come from either a computer simulation or an experiment. In their recent study (122), the authors applied such stochastic differential equations to approximate the various statistical properties associated with steered MD simulations of ion transport across a channel protein, which, in many respects, resembles our earlier renormalization approach considered in Section 6.

Savelyev & Papoian (125, 126) used an approach whose essence is to match correlators obtained from atomistic and CG simulations, for observables that explicitly enter the CG Hamiltonian, which leads to the equivalency of the corresponding partition functions. This resulted in a one-step renormalization process to reach a consensus between the two models, which allows for the reproduction of many-body effects at low computational cost, while increasing the likelihood of finding unique solutions for the CG force-field parameters values.

Finally, it might be useful to mention here the quantum classical path approach (38, 127–130) as a powerful multiscale approach, which exploits our idea of transferring between two potentials (which is basically the scenario illustrated in **Figure 2**). This approach evaluates the nuclear quantum mechanical corrections to free-energy surfaces by using a perturbation from a classical

trajectory to a path integral centroid potential. The effectiveness of this approach has led others to try to adopt it (128–130), and even to try to attribute it to other workers (131), as well as to the development of closely related methods such as those of Reference 132.

12. CONCLUDING REMARKS AND PERSPECTIVE

Computers are constantly increasing in power, and with this increase comes the possibility to address ever more challenging and complex problems, such as detailed sampling of protein landscapes to explore the energetics and kinetics of folding processes, as well as to explore possible relationships between landscape and function. This in turn should help improve the description of functional properties. However, such complex problems require enormous computational power; moreover, as discussed in Section 1, brute-force approaches are not necessarily the best way to address such issues. Therefore, a powerful compromise can be obtained by using multiscale approaches that can allow one to effectively explore complex systems, to model long-timescale processes, and to capture the main physics of the given process. When doing so, it is essential to focus on the main issue: What should be simplified? This problem can be highlighted by recent progress in the promising MARTINI CG models for membranes (119). That is, taking the idea put forth in Reference 1 and expanding it into membranes would have been the most logical direction three decades ago, as in many cases the effects of the membranes are second order. For example, modeling the membrane by a grid of induced dipoles (133, 134) probably captures the most important effect of membranes on the function of membrane proteins (short of establishing the given folding). However, until recently, most effort in modeling membrane effects has been invested into detailed membrane models (135–139), thus utilizing major computer time, which was not necessarily effective (especially in studies of relevant functional membrane proteins). From our point of view, it would have been more useful to start with simplified models and then to move to greater detail as time progresses. However, even now, the advantages of using CG models in studies of membrane proteins are enormous, including in studies that are currently underway in our lab, which simulate complex processes like the gating of ion channels, and even in studies of protein insertion through the translocon.

This review emphasizes the use of a simplified folding model as a reference potential for all-atom simulations. This idea is quite crucial as it provides a clear link between the simplified and complex models and thus allows one to judge the validity of the given simplification. It also allows one to systematically refine the CG model. Of course, the closer the simplified and explicit potentials are, the faster the convergence of the free energy is for the transfer between these potentials. The utility of the powerful reference potential idea has been demonstrated here with many different examples.

BRUTE-FORCE SIMULATIONS ARE NOT ENOUGH

The challenge of obtaining relevant results on complex systems with a given amount of computer power is the central issue in biological modeling and applies to almost any simulation problem. As we do not have an infinite amount of computer time, it is essential to keep in mind that seemingly rigorous options (which may superficially seem to be the preferable ones) are often not the best way to move forward. Thus the use of CG multilevel approaches is recommended in almost any simulation of biophysical systems, ranging from enzyme design to energy and signal transduction.

The power of our treatment has been repeatedly illustrated in calculations of a wide range of problems, including studies of folding, evaluation of protein-ligand interactions (including transition-state binding free energies), studies of translocation processes, long-timescale simulations, and, of course, multilevel approaches such as the QM/MM method. More systematic options of using the reference potential to explore the dynamics of the full model have not been fully formalized, but developments in this crucial direction are clearly expected.

Overall, it is clear that the multiscale modeling of proteins has advanced significantly from its early days in 1975, finding ever more exciting applications. As there will always be concern about the validity of simplified treatments, it is important to have a bridge between the simplified and explicit models, and more rigorous treatments should focus on minimizing the difference between the average of the difference between the simplified and explicit potentials [i.e., $\langle(U_{sp} - U_{\exp})\rangle$] (40). In fact, minimizing this functional with respect to the parameters of the simplified model is probably the most promising way to refine the CG model.

SUMMARY POINTS

1. Early studies of protein folding in the mid-1970s jump-started the highly powerful use of simplified models in simulating complex biological problems.
2. This field was further advanced by the use of hybrid QM/MM approaches, which save computational cost by splitting the system into different regions being treated at different levels of theory. This review touches on some of the various advances over the past 35 years.
3. A crucial point when using a simplified model is that it allows one to obtain a faithful representation of the real (full) system. One way of approaching this issue is to use a CG reference potential for the full explicit model. This approach allows one to move from the simplified to the full model.
4. Another way of approaching this issue is by means of the renormalization process outlined in Section 6, which allows one to move from the full to the simplified model and is a powerful tool when examining long-timescale processes.
5. Multiscale approaches have allowed for great progress in the study of PTR and ion transport (e.g., across membranes). Particularly, such studies have been instrumental in establishing that the electrostatic free-energy barrier controls PTR in proteins, rather than the Grotthus mechanism (as was popularly believed). Such studies were also instrumental in probing the origin of ion channel selectivity.
6. Multiscale approaches have been effective in situations in which the details of the simulated system are unclear or unknown, such as in the case of vectorial translocation by translocases. In this way, it is possible to model biological vectorial translation processes consistently with no prior assumption as to the results, and with no phenomenological parameters.
7. Whereas explicit all-atom simulations can be effective in the exploration of the primary events in light-induced electron transport and PTR, studying long-timescale processes (i.e., the range from 100 ps to seconds) is more complicated. Section 9 covers recent advances in the use of CG models in this area.

8. Finally, electrostatic effects have been argued to provide the most important correlation between structure and function, and thus a crucial question is what level of simplification would still be able to capture the correct physics of such effects in macromolecules. The multiple ways in which this issue can be addressed are discussed in Section 10.

FUTURE ISSUES

1. Although semi-empirical QM/MM approaches have proven to be a powerful tool for simulating complex biological problems, it is nevertheless desirable to move to ab initio representations. A major challenge is the handling of the high computational cost associated with the need for extensive sampling to obtain meaningful results from QM(ai)/MM calculations. Thus, clearly further advances (several of which are underway) are required in this area to be able to examine complex systems using high-level approaches.
2. The use of a CG potential as a reference potential for an explicit model, introduced in Section 3, has particularly shown promise in the field of enzyme design, in which it can be used for the evaluation of the binding free energy of the rate-determining transition state. Further advances in this biologically critical area are expected in the next few years.
3. Future advances in the simulation of long-timescale events include approaches that can use CG simulations directly to obtain trajectories of the explicit models. This will be useful in exploring proposals about long-timescale dynamical coupling and long-timescale dynamical memory effects.
4. Although the field is rapidly expanding with many new innovations (some of which are presented in Section 11), it is important to also adapt existing approaches to move to ever more complex problems (such as the gating of ion channels or protein insertion through the translocon), and such studies are currently underway.
5. Finally, the most crucial future direction is to develop approaches that can minimize the difference between the simplified and explicit models. In this way, one can ensure one is modeling highly complex chemical and biological problems using a simplified model that is physically meaningful.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by NCI grant 1 U19 CA105010-01, NSF grant MCB-03442276, and NIH grants GM24492 and GM40283. The authors would like to thank Hans Ågren (KTH) for organizing a workshop on multiscale modeling that led in part to this review.

LITERATURE CITED

1. **Levitt M, Warshel A. 1975. Computer simulation of protein folding. *Nature* 253:694–98**
2. **Taketomi H, Ueda Y, Gō N. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Res.* 7:445–59**
3. Warshel A. 1979. Calculations of chemical processes in solutions. *J. Phys. Chem.* 83:1640–50
4. Pullman A, Pullman B. 1975. New paths in the molecular orbital approach to solvation in biological molecules. *Q. Rev. Biol.* 7:506–66
5. Liu Y, Ichiye T. 1996. Soft sticky dipole potential for liquid water: a new model. *J. Phys. Chem.* 100:2723–30
6. Papazyan A, Warshel A. 1997. Continuum and dipole-lattice models of solvation. *J. Phys. Chem.* 101:11254–64
7. Burykin A, Kato M, Warshel A. 2003. Exploring the origin of the ion selectivity of the KcsA potassium channel. *Proteins* 52:412–26
8. Freddolino PL, Liu F, Gruebele M, Schulten K. 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* 94:L75–77
9. Ayton GS, Noid WG, Voth GA. 2007. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* 17:192–98
10. Saiz L, Klein ML. 2002. Computer simulation studies of model biological membranes. *Acc. Chem. Res.* 2002:482–89
11. Aksimentiev A, Brunner R, Cohen J, Comer J, Cruz-Chu E, et al. 2008. Computer modeling in biotechnology: a partner in development. *Methods Mol. Biol.* 474:181–234
12. Stein M, Gabdoulline RR, Wade RC. 2007. Bridging from molecular simulation to biochemical networks. *Curr. Opin. Struct. Biol.* 17:166–72
13. Levinthal C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* 65:44–45
14. Hagler AT, Honig B. 1978. On the formation of protein tertiary structures on a computer. *Proc. Natl. Acad. Sci. USA* 75:554–58
15. Warshel A, Levitt A. 1976. Folding and stability of helical proteins: carp myogen. *J. Mol. Biol.* 106:421–37
16. **Warshel A, Levitt M. 1976. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103:227–49**
17. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. 2006. Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106:3210–35
18. Nymeyer H, García AE, Onuchic JN. 1998. Colloquium paper: folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* 95:5921–28
19. Thirumalai D, Klimov DK. 1999. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* 10:197–207
20. Clementi C. 2008. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* 18:10–15
21. **Wolynes PG. 2005. Recent successes of the energy landscape theory of protein folding and function. *Q. Rev. Biophys.* 38:405–10**
22. Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–55
23. Hinds DA, Levitt M. 1992. A lattice model for protein-structure prediction at low resolution. *Proc. Natl. Acad. Sci. USA* 89:2536–40
24. Shakhnovich E, Abkevich V, Ptitsyn O. 1996. Conserved residues and the mechanism of protein folding. *Nature* 379:96–98
25. Heath AP, Kavraki LE, Clementi C. 2007. From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. *Proteins* 68:646–61
26. Nymeyer H, García AE, Onuchic JN. 1998. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* 95:5921–28
27. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, et al. 2003. Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68:91–109
28. Sheinerman FB, Brooks CL III. 1998. Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* 278:439–56

1. Introduced simulations of protein folding, as well as CG models.

2. Presents an alternative approach for simulating protein folding, and appeared later in the same year as Ref. 1. This approach has proven to be highly popular; however, it is less physical than the model introduced in Ref. 1.

16. Introduced QM/MM multilevel approaches.

21. Successfully exploited the CG idea in order to introduce and articulate the concept of a folding landscape.

29. Roca M, Messer B, Hilvert D, Warshel A. 2008. On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* 105:13877–82

30. Introduced a general approach for moving from CG to explicit models in folding studies.

30. Fan ZZ, Hwang J-K, Warshel A. 1999. Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. *Theor. Chem. Acc.* 103:77–80

31. Brandt A. 2008. Principles of systematic upscaling. In *Bridging the Scales in Science and Engineering*, ed. J Fish, pp. 193–205. New York: Oxford Univ. Press

32. Matysiak S, Clementi C. 2006. Minimalistic protein model as a diagnostic tool for misfolding and aggregation. *J. Mol. Biol.* 363:297–308

33. Kamerlin SCL, Haranczyk M, Warshel A. 2009. 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. *J. Phys. Chem. B* 113:1253–72

34. Luzhkov V, Warshel A. 1992. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J. Comput. Chem.* 13:199–213

35. Rosta E, Haranczyk M, Chu ZT, Warshel A. 2008. Accelerating QM/MM free energy calculations: representing the surroundings by an updated mean charge distribution. *J. Phys. Chem. B* 112:5680–92

36. Rosta E, Klähn M, Warshel A. 2006. Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *J. Phys. Chem. B* 110:2934–41

37. Introduced the DFT embedding strategy, which presents what is currently the most effective multilevel QM approach.

37. Wesolowski TA, Warshel A. 1993. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* 97:8050–53

38. Hwang J-K, Warshel A. 1993. A quantized classical path approach for calculations of quantum mechanical rate constants. *J. Phys. Chem.* 97:10053–58

39. Hwang J-K, Warshel A. 1996. How important are quantum mechanical nuclear motions in enzyme catalysis? *J. Am. Chem. Soc.* 118:11745–51

40. Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain A, Warshel A. 2010. Multiscale simulations of protein landscapes: using coarse grained models as reference potentials to full explicit models. *Proteins* 78:1212–27

41. Roca M, Vardi-Kilshtain A, Warshel A. 2009. Toward accurate screening in computer-aided enzyme design. *Biochemistry* 48:3046–56

42. Shurki A, Warshel A. 2003. Structure/function correlations of proteins using MM, QM/MM, and related approaches: methods, concepts, pitfalls, and current progress. *Adv. Protein Chem.* 66:249–313

43. Riccardi D, Schaefer P, Cui Q. 2005. pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J. Phys. Chem. B* 109:17715–33

44. Rothlisberger U, Carloni P, Doclo K, Parrinello M. 2000. A comparative study of galactose oxidase and active site analogs based on QM/MM Car-Parrinello simulations. *J. Biol. Inorg. Chem.* 5:236–50

45. Zheng J, Wang D, Thiel W, Shaik S. 2006. QM/MM study of mechanisms for compound I formation in the catalytic cycle of cytochrome P450cam. *J. Am. Chem. Soc.* 128:13204–15

46. Ranaghan KE, Mulholland AJ. 2004. Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chem. Commun.* 2004:1238–39

47. Olsson MHM, Hong G, Warshel A. 2003. Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin. *J. Am. Chem. Soc.* 125:5025–39

48. Zhang Y, Liu H, Yang W. 2000. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* 112:3483–92

49. Rod TH, Ryde U. 2005. Quantum mechanical free energy barrier for an enzymatic reaction. *Phys. Rev. Lett.* 94:138302

50. Liu WB, Wood RH, Doren DJ. 2003. Density and temperature dependences of hydration free energy of Na^+ and Cl^- at supercritical conditions predicted by ab initio/classical free energy perturbation. *J. Phys. Chem. B* 107:9505–13

51. Iftimie R, Salahub D, Schofield J. 2003. An efficient Monte Carlo method for calculating ab initio transition state theory reaction rates in solution. *J. Chem. Phys.* 119:11285–97

52. Crespo A, Marti MA, Estrin DA, Roitberg AE. 2005. Multiple-steering QM/MM calculation of the free energy profile in chorismate mutase. *J. Am. Chem. Soc.* 127:6940–41
53. Pradipta B. 2005. Accelerating quantum mechanical/molecular mechanical sampling using pure molecular mechanical potential as an importance function: the case of effective fragment potential. *J. Chem. Phys.* 122:091102
54. Wesolowski T, Muller RP, Warshel A. 1996. Ab initio frozen density functional calculations of proton transfer reactions in solution. *J. Phys. Chem.* 100:15444–49
55. Warshel A. 1991. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York: Wiley & Sons
56. Kamerlin SCL, Warshel A. 2010. The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discuss.* 145:71–106
57. Wesolowski T, Warshel A. 1994. Ab initio free energy perturbation calculations of solvation free energy using the frozen density functional approach. *J. Phys. Chem.* 98:5183–87
58. Olsson MHM, Warshel A. 2004. Solute solvent dynamics and energetics in enzyme catalysis: the S_N2 reaction of dehalogenase as a general benchmark. *J. Am. Chem. Soc.* 126:15167–79
59. Warshel A. 1982. Dynamics of reactions in polar solvents: semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *J. Phys. Chem.* 86:2218–24
60. Hwang J-K, King G, Creighton S, Warshel A. 1988. Simulation of free energy relationships and dynamics of S_N2 reactions in aqueous solution. *J. Am. Chem. Soc.* 110:5297–311
61. Štrajbl M, Shurki A, Warshel A. 2003. Converting conformational changes to electrostatic energy in molecular motors: the energetics of ATP synthase. *Proc. Natl. Acad. Sci. USA* 100:14834–39
62. Pisliakov AV, Cao J, Kamerlin SCL, Warshel A. 2009. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci. USA* 106:17359–64
63. Hatcher E, Soudackov AV, Hammes-Schiffer S. 2007. Proton-coupled electron transfer in soybean lipoxygenase: dynamical behavior and temperature dependence of kinetic isotope effects. *J. Am. Chem. Soc.* 129:187–96
64. Watney JB, Agarwal PK, Hammes-Schiffer S. 2003. Effect of mutation on enzyme motion in dihydrofolate reductase. *J. Am. Chem. Soc.* 125:3745–50
65. Mones L, Kulhánek P, Simon I, Laio A, Fuxreiter M. 2009. The energy gap as a universal reaction coordinate for the study of chemical reactions. *J. Phys. Chem. B* 113:7867–73
66. Hong G, Rosta E, Warshel A. 2006. Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *J. Phys. Chem. B* 110:19570–74
67. Wesolowski TA, Weber J. 1996. Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules. *Chem. Phys. Lett.* 248:71–76
68. Wu Q, Kaduk B, Van Voorhis T. 2009. Constrained density functional theory based configuration interaction improved the prediction of reaction barrier heights. *J. Chem. Phys.* 130:034109
69. Jackson TA, Karapetian A, Miller A-F, Brunold TC. 2005. Probing the geometric and electronic structures of the low-temperature azide adduct and the product-inhibited form of oxidized manganese superoxide dismutase. *Biochemistry* 44:1504–20
70. Wu Q, Van Voorhis T. 2006. Direct calculation of electron transfer parameters through constrained density functional theory. *J. Phys. Chem. A* 110:9212–18
71. Wesolowski TA. 2006. One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modeling of complex polyatomic systems. In *Computational Chemistry: Reviews of Current Trends*, ed. J Leszczynski, pp. 1–82. Singapore: World Sci.
72. Xiang Y, Warshel A. 2008. Quantifying free energy profiles of proton transfer reactions in solution and proteins by using a diabatic FDFT mapping. *J. Phys. Chem. B* 112:1007–15
73. Wu Q, Van Voorhis T. 2005. Direct optimization method to study constrained systems within density-functional theory. *Phys. Rev. A* 72:024502
74. Oberhofer H, Blumberger J. 2009. Charge constrained density functional molecular dynamics for simulation of condensed phase electron transfer reactions. *J. Chem. Phys.* 131:064101
75. Schmidt JR, Shenvi N, Tully JC. 2008. Controlling spin contamination using constrained density functional theory. *J. Chem. Phys.* 129:114110

62. Expanded the use of the renormalization approach in the first realistic study that actually explored the coupling between the conformational and chemical motions in enzymes.

76. Cembran A, Song L, Mo Y, Gao J. 2009. Block-localized density functional theory (BLDFT), diabatic coupling and their use in valence bond theory for representing reactive potential energy surfaces. *J. Chem. Theory Comput.* 5:2702–16
77. de la Lande A, Salahub DR. 2010. Derivation of interpretative models for long range electron transfer from constrained density functional theory. *J. Mol. Struct.* 943:115–20
78. Kamerlin SCL, Cao J, Rosta E, Warshel A. 2009. On unjustifiably misrepresenting the EVB approach while simultaneously adopting it. *J. Phys. Chem. B* 113:10905–15
79. Laio A, Parrinello M. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* 99:12562–66
80. Muller RP, Warshel A. 1995. Ab initio calculations of free energy barriers for chemical reactions in solution. *J. Phys. Chem.* 99:17516–24
81. Huber T, Torda AE, van Gunsteren WF. 1994. Local elevation: a method for improving the searching properties of molecular dynamics simulations. *J. Comput. Aided Mol. Des.* 8:695–708
82. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. 2006. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* 39:73–81
83. Chandler D. 1998. *Finding Transition Pathways: Throwing Ropes over Rough Mountain Passes*. Singapore: World Sci.
84. Bolhuis PG, Dellago C, Chandler D, Geissler P. 2002. Transition path sampling, throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318
85. Dellago C, Bolhuis PG, Geissler P. 2003. Transition path sampling. *Adv. Chem. Phys.* 123:1–78
86. Plotnikov N, Kamerlin SC, Warshel A. 2010. Paradynamics: an effective and reliable model for ab initio QM/MM free energy calculations and related tasks. Manuscript under review
87. Saven JG, Wang JG, Wolynes PG. 1994. Kinetics of protein folding: the dynamics of globally connected rough energy landscapes with biases. *J. Chem. Phys.* 101:11037–43
88. McQuarrie DA. 1976. *Statistical Mechanics*. New York: Harper & Row
89. Wang MC, Uhlenbeck GE. 1945. On the theory of the Brownian motion II. *Rev. Mod. Phys.* 17:323–42
90. Wong YT, Clark TW, Shen J, McCammon JA. 1993. Molecular-dynamics simulation of substrate-enzyme interactions in the active-site channel of superoxide-dismutase. *Mol. Simul.* 10:277–89
91. Adelman SA. 1976. Fokker-Plank equations for simple non-Markovian systems. *J. Chem. Phys.* 64:124–30
92. Grote RF, Hynes JT. 1980. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.* 73:2715–32
93. Kurnikova MG, Waldeck DH, Coalson RD. 1995. A molecular dynamics study of dielectric friction. *J. Chem. Phys.* 105:628–38
94. Dryga A, Warshel A. 2010. Renormalizing SMD: The renormalization approach and its use in long time simulations and accelerated PMF calculations of macromolecules. *J. Phys. Chem. B* 114:12720–28
95. Jarzynski C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–93
96. Gullingsrud JR, Braun R, Schulten K. 1999. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *J. Comp. Chem.* 151:190–211
97. Kamerlin SCL, Warshel A. 2010. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* 78:1339–75
98. Chen J, Dima RI, Thirumalai D. 2007. Allosteric communcation in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J. Mol. Biol.* 374:250–66
99. Karplus M. 2010. The role of conformation transitions in adenylate kinase. *Proc. Natl. Acad. Sci. USA* 107:E71
100. Hummer G. 2010. Catching a protein in the act. *Proc. Natl. Acad. Sci. USA* 107:2381–82
101. Kamerlin SCL, Warshel A. 2010. Reply to Karplus: Conformational dynamics have no role in the chemical step. *Proc. Natl. Acad. Sci. USA* 107:E72
102. Khandogin J, Brooks CL III. 2005. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* 89:141–57
103. Baptista AM, Martel PJ, Soares CM. 1999. Simulation of electron-proton coupling with a Monte Carlo method: application to cytochrome c3 using continuum electrostatics. *Biophys. J.* 6:2978–98
104. Olsson MHM, Warshel A. 2006. Monte Carlo simulations of proton pumps: on the working principles of the biological valve that controls proton pumping in cytochrome c oxidase. *Proc. Natl. Acad. Sci. USA* 103:6500–5

105. Braun-Sand S, Burykin A, Chu ZT, Warshel A. 2005. Realistic simulations of proton transport along the gramicidin channel: demonstrating the importance of solvation effects. *J. Phys. Chem. B* 109:583–92
106. Braun-Sand S, Strajbl M, Warshel A. 2004. Studies of proton translocations in biological systems: simulating proton transport in carbonic anhydrase by EVB based models. *Biophys. J.* 87:2221–39
107. Kato M, Pisliakov AV, Warshel A. 2006. The barrier for proton transport in aquaporins as a challenge for electrostatic models: the role of protein relaxation in mutational calculations. *Proteins* 64:829–44
108. Warshel A, Sharma PK, Kato M, Parson WW. 2006. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta* 1764:1647–76
109. Kurnikova MG, Coalson RD, Graf P, Nitzan A. 1999. A lattice relaxation algorithm for three-dimensional Poisson-Nernst-Planck theory with application to ion transport through the gramicidin A channel. *Biophys. J.* 76:642–56
110. Liu H, Shi Y, Chen XS, Warshel A. 2009. Simulating the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases. *Proc. Natl. Acad. Sci. USA* 106:7449–54
111. Warshel A, Chu ZT. 2001. Nature of the surface crossing process in bacteriorhodopsin: computer simulations of the quantum dynamics of the primary photochemical event. *J. Phys. Chem. B* 105:9857–71
112. Warshel A, Parson WW. 2001. Dynamics of biochemical and biophysical reactions: insight from computer simulations. *Q. Rev. Biophys.* 34:563–670
113. Braun-Sand S, Sharma PK, Chu ZT, Pisliakov AV, Warshel A. 2008. The energetics of the primary proton transfer in bacteriorhodopsin revisited: It is a sequential light induced charge separation after all. *Biochim. Biophys. Acta* 1777:441–52
114. Lee FS, Chu ZT, Bolger MB, Warshel A. 1992. Calculations of antibody antigen interactions: microscopic and semimicroscopic evaluation of the free energies of binding of phosphorylcholine analogs to Mcpc603. *Protein Eng.* 5:215–28
115. Warshel A, Sussman F, King G. 1986. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* 25:8368–72
116. Tomasi J, Persico M. 1994. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* 94:2027–94
117. Muegge I, Tao H, Warshel A. 1997. A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Protein Eng.* 10:1363–72
118. Sham YY, Chu ZT, Tao H, Warshel A. 2000. Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins* 39:393–407
119. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* 111:7812–24
120. Essiz SG, Coalson RD. 2009. Dynamic linear response theory for conformational relaxation of proteins. *J. Phys. Chem. B* 113:10859–69
121. Parson WW, Warshel A. 2004. Dependence of photosynthetic electron-transfer kinetics on temperature and energy in a density-matrix model. *J. Phys. Chem. B*: 10474–83
122. Calderon CP, Janosi L, Kosztin I. 2009. Using stochastic models calibrated from nanosecond nonequilibrium simulations to approximate mesoscale information. *J. Chem. Phys.* 130:144908
123. Calderon CP. 2007. On the use of local diffusion models for path ensemble averaging in potential of mean force computations. *J. Chem. Phys.* 126:084106
124. Jimenez JC, Ozaki T. 2005. An approximate innovation method for the estimation of diffusion processes from discrete data. *J. Time Ser. Anal.* 27:77–97
125. Savelyev A, Papoian GA. 2009. Molecular renormalization group coarse-graining of polymer chains: application to double-stranded DNA. *Biophys. J.* 96:4044–52
126. Savelyev A, Papoian GA. 2009. Molecular renormalization group coarse-graining of electrolyte solutions: applications to aqueous NaCl and KCL. *J. Phys. Chem. B* 113:7785–93
127. Hwang J-K, Chu ZT, Yadav A, Warshel A. 1991. Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *J. Phys. Chem.* 95:8445–48
128. Wang Q, Hammes-Schiffer S. 2006. Hybrid quantum/classical path integral approach for simulation of hydrogen transfer reactions in enzymes. *J. Chem. Phys.* 126:184102

110. Used the renormalization approach outlined in the text to successfully simulate the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases.

119. Adopted the CG protein folding idea for studies of membranes and has helped make the CG idea highly popular.

122. Presents the interesting surrogate process approximation method, which summarizes the state of complex systems by the time series of low-dimensional system observables, such as nonequilibrium trajectories, in order to examine long-timescale dynamics.

129. Major DT, Gao JL. 2005. Implementation of the bisection sampling method in path integral simulations. *J. Mol. Graph. Model.* 24:121–27
130. Wang M, Lu Z, Yang W. 2006. Nuclear quantum effects on an enzyme-catalyzed reaction with reaction path potential: proton transfer in triosephosphate isomerase. *J. Chem. Phys.* 124:124516
131. Major DT, Heroux A, Orville AM, Valley MP, Fitzpatrick PF, Gao J. 2010. Differential quantum tunneling contributions in nitroalkane oxidase catalyzed and the uncatalyzed proton transfer reaction. *Proc. Natl. Acad. Sci. USA.* In press; doi: 10.1073/pnas.0911416106
132. Craig I, Manolopoulos DE. 2004. Quantum statistics and classical mechanics: real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* 121:3368–73
133. Åqvist J, Warshel A. 1989. Energetics of ion permeation through membrane channels: solvation of Na^+ by Gramicidin A. *Biophys. J.* 56:171–82
134. Alden RG, Parson WW, Chu ZT, Warshel A. 1996. Macroscopic and microscopic estimates of the energetics of charge separation in bacterial reaction centers. In *Reaction Centers of Photosynthetic Bacteria Structure and Dynamics*, ed. ME Beyerle, pp. 105–16. Berlin: Springer Verlag
135. Faraldo-Goméz J, Smith GR, Sansom MSP. 2002. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* 31:217–27
136. Feller SE, Pastor RW. 1999. Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities. *J. Chem. Phys.* 111:1281–87
137. Essmann U, Berkowitz ML. 1999. Dynamical properties of phospholipid bilayers from computer simulation. *Biophys. J.* 76:2081–99
138. Armen RS, Uitto OD, Feller SE. 1998. Phospholipid component volumes: determination and application to bilayer structure calculations. *Biophys. J.* 75:734–44
139. Tobias DJ, Tu KC, Klein ML. 1997. Atomic-scale molecular dynamics simulations of lipid membranes. *Curr. Opin. Colloid Interface Sci.* 2:15–26

RELATED RESOURCE

Warshel Group home page: **http://futura.usc.edu**

Annual Review of
Physical Chemistry
Volume 62, 2011

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at http://physchem.annualreviews.org/errata.shtml