Theor Chem Acc (2006) 116: 51–59
DOI 10.1007/s00214-005-0022-z

REGULAR ARTICLE

Q. Cui

Theoretical and computational studies of vectorial processes in biomolecular systems

Received: 26 March 2005 / Accepted: 16 April 2005 / Published online: 3 January 2006

Abstract Efficient vectorial processes such as the transduction of bioenergy and signals are characteristics that strikingly distinguish living systems from inanimate materials. Recent developments in biophysical and biochemical techniques have provided new information about the structure, dynamics and interaction of biomolecules involved in vectorial life processes at multiple length and temporal scales. This wealth of data makes it possible to carry out theoretical and computational studied of key mechanistic questions associated with complex life processes at an unprecedented level. Using two "vectorial biomolecular machines", myosin and cytochrome c oxidase, as examples, we discuss the identification of interesting and biologically relevant questions that require thorough theoretical analysis. Technical challenges and recent progress related to these theoretical investigations are briefly summarized.

Keywords Vectorial processes · Biomolecular motor · Proton pump · Combined QM/MM · Multi-scale simulations

1 Introduction: vectorial processes in biology

Having envisioned far-reaching concepts such as nanotechnology and quantum computing long before their realization, Richard Feynman is often considered one of science's greatest visionsries. As early as 1963, he wrote down the famous statement,

> *"Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."*

As evidence of his exceptional wisdom, this statement remains proper today, if not more compelling than 40 years ago. Breath-taking developments in novel biophysical and biochemical techniques have made it possible to monitor [1] and manipulate [2,3] the behavior of biological systems at multiple length and time scales down to the level of conformational dynamics of a single biomolecule [4–7]. The ultimate dream of understanding life *"in terms of the jigglings and wigglings of atoms"* is evidently within reach. The long-term goal of my research group is to participate in such pursuit via pushing the limit of a broad range of theoretical and computational tools forward and applying those tools in the investigation of fundamentally important life processes.

Among the many fascinating mysteries presented by life, the specific class of problems we choose to focus on concerns how *vectorial* processes, which strikingly distinguish living from most inanimate entities, are implemented in biological systems. On the macromolecular scale, prominent examples include various biomolecular pumps that transport ions such as protons [8] and Ca^{2+} [9] across biological membranes *against* a concentration gradient, and biomolecular motors [10] that move *unidirectionally* along actin or microtubules. Both types of vectorial processes are driven by exothermic events such as O_2 reduction in cytochrome c oxidase and Adenosine Triphosphate (ATP) hydrolysis in many biomolecular motors and membrane transporters. The general challenge is to understand how such exothermic chemical reactions produce vectorial movements of ions or even a 10^5-Dalton protein with *high efficiency*. On the cellular scale, vectorial life processes involve the co-operative action of a large number of biomolecules; an excellent example concerns bacterial chemotaxis [11], in which the binding of a specific small molecule to the bacteria surface receptor(s) is translated into a sequence of vectorial signal cascades that results in a change in the rotational direction of the flagella, and consequently the swimming behavior of the bacterium. The challenge here concerns understanding not only the regulatory mechanism for the conformational properties of individual

Q. Cui
Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison 1101 University Avenue, Madison, WI 53706, USA
E-mail: cui@chem.wisc.edu

proteins involved in the signal transduction network, but also how these proteins work collaboratively to ensure that the vectorial response of the network to environmental stimuli is accurate, sensitive and robust [11,12].

Our fascination with these vectorial processes was triggered by exciting progress made in the experimental arena [3, 5], which made it possible, only in recent years, to conduct meaningful theoretical analyses that are able to establish coherent mechanistic pictures for complex life processes at a detailed level. For example, the issue of bioenergy transduction in the Ca^{2+} pump and the muscle motor myosin has been studied extensively in the 70s by Jencks [13] and Hill [14, 15], among others, based on kinetic and thermodynamic considerations. The lack of high-resolution protein structures, however, prevented a detailed understanding at the atomic level. With the recent developments in structural biology, site-directed mutagenesis and single molecule spectroscopy, innovative experiments [3,5] provided much more information at different spatial and temporal resolutions; e.g., single molecule measurements were crucial in determining the stepsize [16], movement pattern [17,18] and stoichiometry of ATP consumption [16,19] during the function of several molecular motors. Needless to say, we are in a far better position to combine all these data to establish a deeper and more precise understanding of bioenergy transduction. Similarly, phenomenological models based on reaction–diffusion equations for cellular-level vectorial processes such as chemotaxis have a long history in theoretical biology [20]. However, the explosive progress in genomics, chemical biology and various bio-imaging/analytical techniques [21] has generated much more precise information about the sequence and structure of all the proteins (including the flagella! [22]) involved in the signaling network as well as their spatial location in the cell [23] and the interaction pattern among them [24]. Clearly, a more sophisticated model for chemotaxis (and other signaling network) with appropriate structural and temporal details [25] can soon be developed.

2 The role of theoretical and computational chemistry

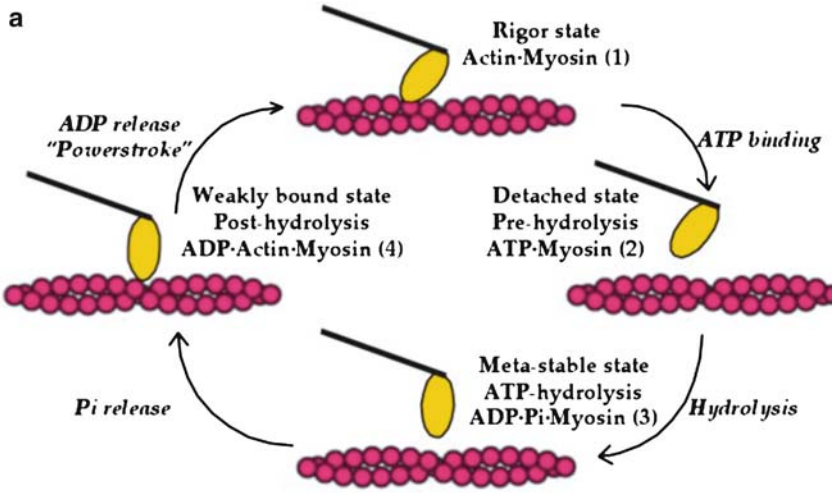
Considering the complexity of these vectorial life processes, there is no doubt that theoretical and computational studies will play an important role in terms of interpreting and unifying experimental observations as well as providing guidance to further experimental studies. The most serious challenge is to identify the most interesting problems that cannot be addressed using experiments alone. Here we briefly discuss two vectorial processes at the protein level that are currently under investigation in our group. The main objective is to illustrate that it is fruitful to consider the biological function of these "vectorial biomolecular machines" when defining the most interesting and *biologically relevant* questions for theoretical analysis. It is not our intention to discuss results from our studies, for which we refer the readers to published articles [26,27].

2.1 Mechanochemical coupling in conventional myosin

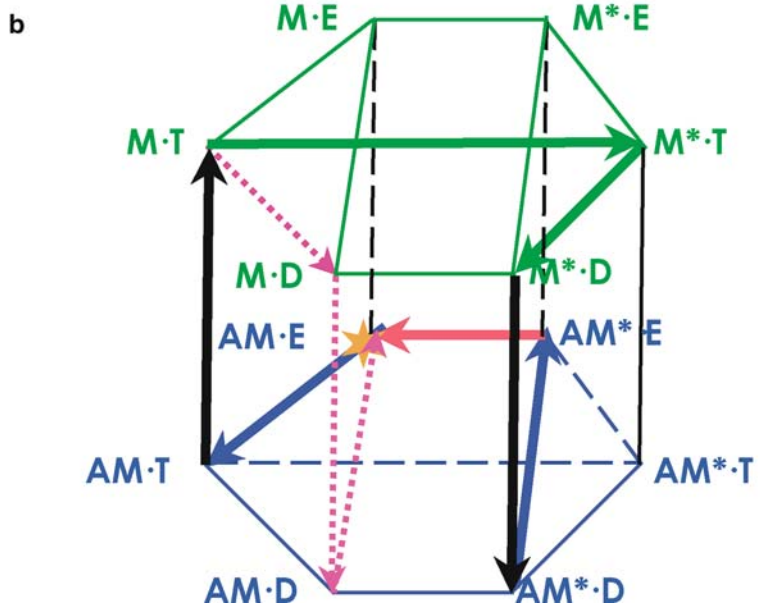
Conventional myosin (myosin II), which is involved in muscle contraction [28], is a typical molecular motor that utilizes chemical free energy in the form of ATP binding and/or hydrolysis to perform mechanical work; its efficiency has been estimated to be about 60% [29]. Myosin II was chosen over other members in the superfamily (e.g., myosin V and VI) due to the availability of multiple X-ray structures at high resolution and the rich biochemical and biophysical background associated with muscle research.

The most interesting question regarding motor proteins, such as myosin, concerns the mechanism, in structural and energetics terms, by which the ATP hydrolysis produces mechanical work with high efficiency [30,31]. The way to approach such a complex problem can be identified by considering Scheme I, in which the myosin–actin–ATP system is characterized in terms of (at the minimal) 12 states as different combinations of the myosin conformation (2), myosin–actin binding state (2) and the chemical state of ATP (3). The system makes stochastic transitions among those states with the probability governed by the rate (which depends on the rate constants and ATP, ADP, P_i concentrations) associated with the transitions. Based on such kinetic schemes, as pointed out by Hill [14,32], there are several conditions that the system has to satisfy to achieve high efficiency. First, the rates for different transitions have to adopt values such that only *one* major kinetic pathway contributes (shown in bold in Scheme I for myosin II [33]) because following multiple pathways may compromise the efficiency. For example, we note that force is generated (thus mechanical work is done) only when myosin rebinds to actin with a conformation ("cocked head") different from that in the beginning of the cycle, which involves tight binding between myosin and actin ("rigor" state). Therefore, the coupling between ATP hydrolysis and the conformational change in the *detached* myosin has to be tight, otherwise unproductive kinetic pathway (e.g., the dotted lines in Scheme I) may be followed, which causes futile ATP hydrolysis that reduces the motor efficiency. Moreover, the sum of free energy changes, ΔG_i, along any kinetic pathway for each cycle is rigorously equivalent to the hydrolysis free energy of ATP in *solution*, which is the ultimate thermodynamic driving force for the motor; this highlights the importance of keeping the ATP, ADP and P_i concentrations away from equilibrium for the motor to function. However, since force is generated only when myosin is bound to actin, the free energy changes for all other kinetic steps, including the actual hydrolysis in the myosin motor domain, should be small such that a large free energy drop is reserved for the force generation step. Large free energy drops and increases for non-force-generating steps would also produce a small cumulative free energy change, but large uphill transitions would compromise the overall rate (flux) of the cycle.

These kinetic and thermodynamic considerations have provided a framework for analyzing motor efficiency, but they do not provide a mechanism, just as stating "enzymes preferentially stabilize transition state over the ground state"



$$\sum_i \Delta G_i = \Delta G^{aq}_{ATP}$$



Scheme I **a** The Lymn–Taylor kinetic pathway for the functional cycle of conventional myosin. **b** A more complete, *minimal* kinetic model for the myosin–actin–ATP system, which includes two conformational state of the myosin (M), two binding states of myosin to actin(A) and three states for ATP: empty (E), ATP bound (T) and ADP/Pi bound (D). The Lymn–Taylor kinetic pathway is labeled as *bold lines*; a kinetic pathway involves futile ATP hydrolysis, which needs to be avoided for high efficiency, is shown as *dotted lines*

does not provide any specific mechanistic insights into enzyme catalysis. The challenge is to understand how these kinetic and thermodynamic constraints are *implemented* by the motor in terms of its sequence, three-dimensional structure and dynamical properties. Current experimental techniques do not yet have the resolution to fully meet this challenge; in fact, it is not uncommon to witness conflicting proposals based on different experiments [34–36]. Therefore, it is useful to complement experiments with detailed studies on the conformational dynamics of the myosin–actin system with the nucleotide in different chemical states. Based on the above discussions, the two questions that we found most relevant are: (1) What is the regulatory mechanism for the *tight* coupling between ATP hydrolysis and conformational property of myosin in the *detached* state? (2) What is the mechanism for ensuring a nearly thermal-neutral ATP hydrolysis in myosin, in contrast to the significantly exothermic reaction in solution? There are also other crucial questions regarding the myosin–actin interaction; e.g., it is possible that the rebinding to actin plays an important role in ensuring the tight

coupling between conformation of myosin and ATP hydrolysis; unfortunately, these issues are difficult for quantitative studies at this stage due to the lack of high-resolution X-ray structure of the actin–myosin complex [37], although the situation may change in the near future (I. Rayment, private communication).

2.2 Kinetic gating in cytochrome c oxidase

In the discussion of most biological systems, including molecular motors, it is often the case that the kinetics follow the same trend as the thermodynamics; i.e., fast (slow) rates are correlated with large (small) exothermicited. Violation of this trend may be of functional importance in certain bioenergy transduction processes; here we discuss this possibility using the example of cytochrome c oxidase.

Cytochrome c oxidase (ccO) is a crucial enzyme in bioenergetics [38]; it pumps (4) protons from the N(egative) to the P(ositive) side of the lipid membrane against concentration gradient using the exothermic process of O_2 reduction to water; the efficiency was estimated to be about 60%. Extensive experimental studies [8,39] led to the belief that each complete catalytic/pumping cycle can be separated into four steps, where each "sub-cycle" involves the pumping of one (physical) proton, consumption of one (chemical) proton and one electron in the O_2 reduction. In an outline form, the mechanism of the proton pumping does not seem to be profound: the oxidation–reduction reactions in the chemical centers alter the electrostatic potential inside the protein, which modulates the pK_a values of certain titratable residues that ultimately facilitate the translocation of protons across the membrane. The translocation can occur against a concentration gradient because it is coupled to an exothermic process (O_2 reduction).

More careful consideration, even at the thermodynamic and kinetic levels (Scheme II), however, reveals very interesting features. In each sub-cycle, the main controversy concerns the sequence of electron transfer, (physical) proton pumping and (chemical) proton consumption. It seems safe to assume that electron transfer occurs at least before the up-take of the chemical proton, because otherwise the chemical site would unlikely have the appropriate pK_a. We further assume that the electron transfer also proceeds the physical proton up-take, which seems consistent with available experimental information. Clearly, there are two possible pathways after the electron transfer, in which either the chemical or physical proton up-take occurs first. We note that the intermediate state following the chemical proton up-take as the first step *has to be lower* in free energy than the product of the sub-cycle, because the net difference between the two states is a proton on the N versus P side of the membrane; i.e., the extra stabilization of the intermediate, labeled as $[H^+(N) + PCH^{+\bullet}]$ in Scheme II, relative to the product state, $[PCH^{+\bullet} + H^+(P)]$, is exactly the chemical potential difference of proton between the two sides of the membrane, $\mu^{NP}_{H^+}$. Furthermore, we argue that the intermediate common to the two pathways, $[PH^+CH^{+\bullet}]$, *has to be higher* in free energy compared to the product state, since back-flow of protons from the P side toward N needs to be prevented to avoid futile cycling. As to the free energy of the first intermediate in the pathway initiated with the physical proton up-take, $[H^+(N) + PH^+C^{\bullet}]$, it is reasonable to speculate that it is only slightly higher than $[PH^+CH^{+\bullet}]$, assuming that the pK_a of the chemical site after electron transfer is suitable for proton up-take.

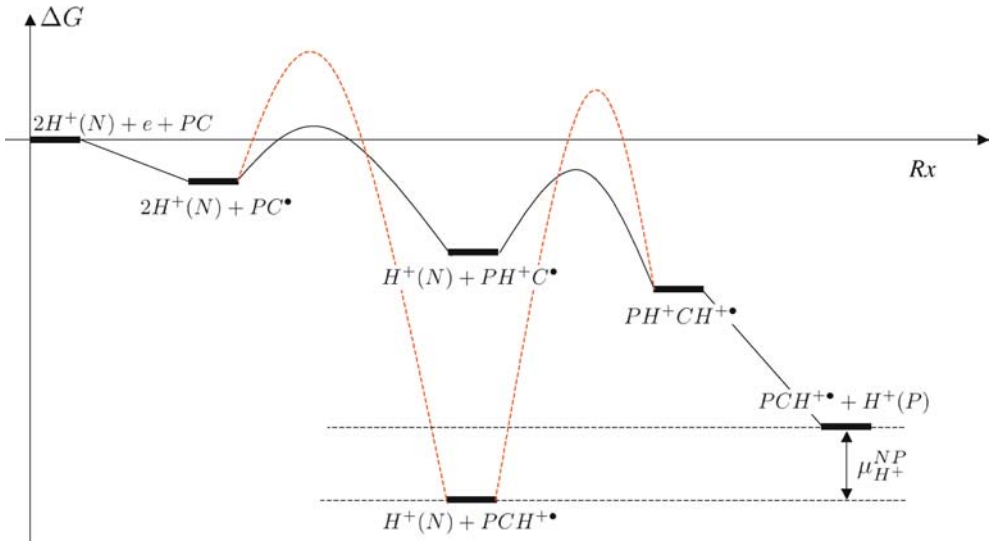
Therefore, based on *no actual calculations* but only basic functional properties of ccO, we can derive the qualitative free energy diagram in Scheme II for a step in the pumping cycle. Evidently, the pathway that initiates with chemical proton up-take has a thermodynamic trap, $[H^+(N) + PCH^{+\bullet}]$, while the one that starts with physical proton up-take is likely a thermodynamically downhill process; from the point of view of maximizing the pumping flux, which is the main function of ccO, the second pathway is clearly more preferable. In other words, for a pump that is not only efficient (which would imply that the free energy drop associated with the sub-cycle is close to be zero [14,32]) but also *fast*, the pathway involving the less stable intermediate has to be chosen. We term this *kinetic gating*, which was eluded to by Popovic and Stuchebrukhov [40] in their investigation of ccO, although they only considered thermodynamics in the electrostatics calculations without a detailed study of the kinetics. The major mystery, apparently, is about how ccO avoids the thermodynamic trap and accomplishes kinetic gating. The availability of several X-ray structures provided the starting point to answer this crucial question, although it is clear that energetics, especially barriers along different pathways, need to be determined to ultimately understand the mechanism. We further speculate that slipping into the thermodynamic trap state might be responsible for the decoupling of the chemical reaction and the proton pumping observed under certain conditions [41].

3 Technical challenges and recent progress

In the previous section, two "vectorial biomolecular machines" were discussed to illustrate key mechanistic questions that require *thorough* theoretical analyses. Considering the size and complexity of these systems, however, there are serious technical challenges that we need to overcome to conduct meaningful theoretical analysis. In the following, we briefly summarize some of these challenges and recent progress made by us and others in the field (to the limit of space). Since only protein-level vectorial processes are being studied in our group at this stage, we will leave the discussion of fascinating cellular vectorial problems to the future (see Sect. 4 for a few references and other articles in this special issue).

3.1 Chemical processes in complex environment

The above discussion made it clear that determining accurate energetics associated with the chemical processes in those "vectorial biomolecular machines" is crucial for answering



Scheme II Schematic free energy profile for a sub-cycle in cytochrome c oxidase. The process involves the transfer of one (physical) proton from the N(egative) side to the P(ositive) side of the membrane, consumption of one (chemical) proton and one electron (e). The pumping and chemical sites are labeled as " P" and "C" in the state notation, respectively; e.g., notation $PCH^{+\bullet}$ indicates loading of the chemical site with a proton and an electron, while $PH^{+}C^{\bullet}$ indicates loading of the physical site with a proton and chemical site with an electron. Electron transfer was assumed to occur first, after which there are two possible pathways (*solid and dashed*) that involves the up-take of physical and chemical proton, respectively. The chemical pathway is argued (see text) to have a thermodynamic trap, $[H^{+}(N) + PCH^{+\bullet}]$, which has to be avoided through *kinetic gating*

key mechanistic questions. Due to the quantum mechanical nature of chemical reactions and large conformational space available to biomolecules, this requires the seamless union between electronic structure and statistical mechanics methods. Although rapid progress has been made in linear-scaling electronic structure methods [42], the most effective approach with standard computational facility is to employ a hybrid quantum mechanical and molecular mechanical (QM/MM) potential function [43–46]. Although the basic idea of combining QM and MM potentials in condensed phase simulations is not profound and has been discussed for at least 30 years [47], it is a major technical challenge to develop QM/MM methods that are accurate, robust and efficient.

In the typical implementation [43], the QM/MM potential has the following components,

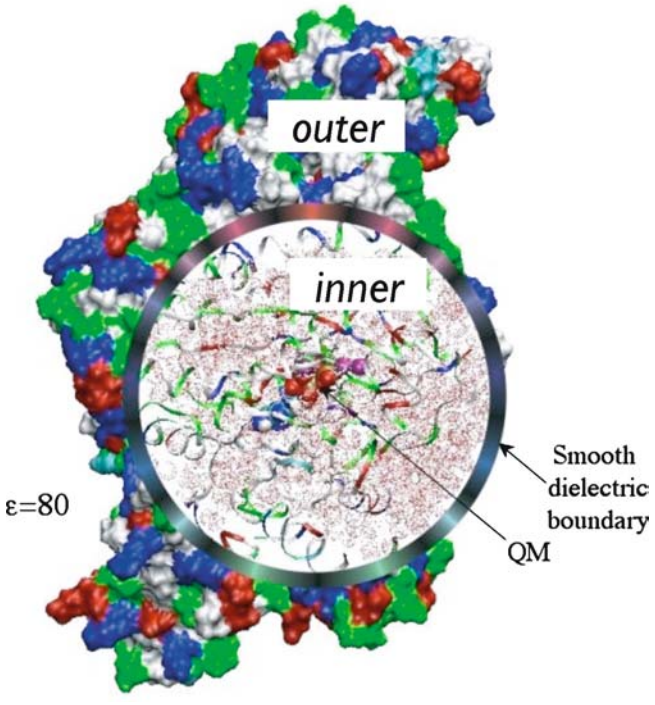
$$U^{\mathrm{tot}} = \left\langle \Psi | \hat{\mathrm{H}}^{\mathrm{QM}} + \hat{\mathrm{H}}^{\mathrm{QM/MM}} | \Psi \right\rangle + \mathrm{U}^{\mathrm{QM/MM}}_{vdW} + \mathrm{U}^{\mathrm{QM/MM}}_{\mathrm{bonded}} + \mathrm{U}^{\mathrm{MM}} \quad (1)$$

The accuracy of the method clearly depends on both the QM level and the MM force field. Most applications use non-polarizable force field due to computational efficiency and extensive calibration of such force fields. Including polarization in the MM part, at least in region close to the QM part, could be important for certain type of problems; this, however, has not been our major focus. As to the QM level, popular density functional theory and ab initio methods are powerful in minimum energy path (MEP) applications [48–51], but they remain impractical for simulations involving extensive conformation samplings. Therefore, there is an urgent need for developing a fast and sufficiently accurate QM method. A promising method, termed the self-consistent-density-functional-tight-binding (SCC-DFTB), has been proposed by Frauenheim, Seifert and co-workers, and is discussed in more details by Elstner [52] in this special issue. It is important to further extend the method to more elements (e.g., Phosphorus for ATP hydrolysis) and also open-shell systems (e.g., for oxygen chemistry), in which our group is actively involved.

In addition, our group has been focused on systematically investigating contributions from factors other than the QM and MM potentials, to the accuracy of QM/MM simulations as well as improving the reliability of such simulations. We believe this is important from a long-term perspective because the insights and methods established in these studies will apply to any QM and MM combinations. Recently, we showed that QM/MM free energy calculations are rather insensitive to the van der Waals parameters for QM atoms [53] as long as the parameters are reasonable, mainly due to error cancellation effects. Another issue often raised in the QM/MM community concerns the treatment of the QM/MM boundary, for which numerous investigations and developments have been made using link atoms [43,54–56], frozen orbitals [57], generalized hybrid orbitals [58], or pseudobonds [59]. Several studies [55,57,60] clearly showed that certain link-atom schemes (e.g., the single-link-atom approach) should be avoided in the calculation of quantities such as proton affinity, although other schemes often give similar results. For the energetics of reactions that conserve charge, the results are even less sensitive to the frontier treatments [56,60].

Considering the importance of electrostatic interactions in biomolecules [61], the treatment of electrostatics in QM/MM simulations has attracted most of our attention. Recently our group has implemented [62,63] the generalized

solvent boundary potential (GSBP)[64] and Ewald sums [65] for combined QM/MM simulations; these protocols have been quantitatively tested using reduction potential and pK_a calculations [63], which were chosen due to their sensitivity to electrostatics and the availability of reliable experimental data. Although both GSBP- and Ewald-based QM/MM protocols generated satisfying results in those quantitative tests, the GSBP scheme is more attractive for studying very large systems such as molecular motor and proton pump due to its computational efficiency. As illustrated in Scheme III, the system is partitioned into an inner region and an outer environment, where the dielectric property can vary (e.g., containing both bulk solvent and a slab of membrane). Atoms in the inner reagion are allowed to move during the simulation, whereas atoms in the outer region are fixed; part of the inner region can be treated with a QM potential. With such a partition, the GSBP approach allows efficient sampling of the region of interest while taking the contribution from the outer region atoms and bulk solvent into account. The advantage of the GSBP-based QM/MM protocol is particularly important for simulations that involve extensive conformational samplings, such as free energy calculations, because configuration distribution may depend sensitively on the treatment of electrostatics [62]. With further improvements of surface polarization effects [66] and better treatment of the dielectric response in the outer region, chemical events in complex biomolecular systems, such as those described here, can be studied at a quantitative level.



Scheme III Schematic representation of the generalized solvent boundary potential (GSBP) partitioning of a solvated biomolecule in the QM/MM framework. Atoms in the inner region are represented explicitly, part of which can be treated with quantum mechanical methods. In the outer region, the remaining biomolecular atoms are represented explicitly but their positions are fixed; the solvent (or lipid bilayer) in this region is replaced by a dielectric continuum

Once a reliable QM/MM potential function is chosen, it remains challenging to derive kinetic and thermodynamic properties associated with the chemical processes of interest. Although MEP analysis has been valuable for understanding qualitative contribution to the catalytic power of enzymes [48–51], it is less useful in the study of vectorial molecular machines because the conformational fluctuations of these intrinsically flexible systems and embedded water molecules would likely make significant contributions; in fact, for even a relatively short (8 Å) proton transfer in a small enzyme, carbonic anhydrase, MEP results were found to drastically differ from the potential of mean force [67]. Therefore, developing useful sampling techniques for determining reaction rates and energetics remains an important task. In this regard, the transition path sampling technique developed by Bolhuis et al. [68] is conceptually very powerful, although it remains computationally demanding and analysis of the results is not always straightforward. Computing a free energy profile along a carefully designed reaction coordinate remains an effective approach.

Finally, it is worth mentioning that QM/MM simulations can also contribute in a major way to the interpretation of various linear and non-linear spectra of biological molecules, which is tremendously useful for identifying the structure and dynamics of key species in the complex functional cycle of large systems [69,70]. The most widely used tools include one-dimensional infrared, Raman, Mossbauer, ESR and various NMR spectroscopies [71]. Rapid progress is being made in multi-dimensional infrared spectroscopy [72], which holds great promise for better resolved dynamical characterization of biomolecular structure; although the interpretation of such complex spectra would certainly benefit from careful theoretical analysis [73,74].

3.2 Conformational dynamics at multiple length and time scales

Another hallmark of those vectorial molecular machines is that their conformational dynamics span a wide range of length and time scales. This is most striking in molecular motors, which involve Å-level changes during the hydrolysis and $\sim 10^2$ Å scale movements of conformational domains; even larger-scale motions, such as flexing of the actin polymer and undulation of the biological membrane, may have important functional implications. These large-scale motions are typically very slow, in the range of μs–ms, which makes them difficult to study using standard simulation techniques. Since it is the coupling between events at different scales that makes vectorial processes effective, developing simulation techniques that can effectively cope with the presence of multiple length and time scales is an active area of research.

One class of approaches attempts to keep the atomic description of the system but speeds up the process of interest

by introducing an external bias; popular examples include targeted [75] and steered [76] molecular dynamics, in which holonomic constraints or harmonic restraints along specific degrees of freedom are used to drive the system from one conformational state to the other. Due to the presence of the bias, the time scale of transitions in such simulations is much faster (typically nanosecond) compared to the natural time-scale. Whether the sequence of events observed in such simulations reflect reality is always debatable; however, they do provide a framework for exploring interactions that are potentially important, and the results are useful for generating hypotheses that can eventually be tested experimentally. In a recent study [77], for example, TMD simulations were carried out to observe the conformational transitions involved in F_1-ATPase. The rupture of a series of salt-bridges during the transitions suggested the functional importance of a set of charged residues, which was subsequently confirmed by mutation studies [78]. The reliability of such biasing MD approaches clearly depends on the way that the bias is introduced; studies exploring coordinates that are more natural for large-scale conformational transitions, such as those based on internal variables or collective modes (see below) [79], are worthwhile and are being pursued. Another interesting avenue is to focus on the dynamics of the system in a smaller subspace spanned by several coordinates of interest using the mean force that is averaged over all other degrees of freedom [80]; the practical difficulty concerns the high cost of computing accurate mean force and, more importantly, identification of a subpace that captures the essential dynamical nature of complex systems.

The second class of methods attempts to extend the size and time-scale limits by reducing the resolution of the model and its dynamics. An area that has become popular involves the application of normal mode analysis (NMA) at various resolutions [26,81–88]. Although NMA invokes seemingly drastic approximations to the dynamics of biomolecules, increasing number of studies have shown strong correlation between the low-frequency modes and conformational transitions implicated in the function of large biomolecules [89,90] such as in several molecular motors [26,83,85,91]. It remains difficult to determine the amplitude and precise direction of large-scale motions based on NMA alone, but the technique provides a systematic framework for exploring *possible* transitions of functional importance, which can stimulate new experimental investigations. In the presence of additional information, such as low-resolution structural data from cyro-EM or small angle X-ray scattering, the character and amplitude of motion might be better resolved [88, 92]. Moreover, analysis of hinge locations in low-frequency modes is useful for revealing residues that are crucial to the flexibility, thus possibly, the function of the system [26]. Since low-frequency modes are often collective in nature, it is understandable (although not entirely expected) that fairly reliable results can be derived based on reduced-resolution models [84,93–95], especially when the system is globular in shape [84,96]. However, the fact that certain mutations can significantly perturb the mobility of motor systems [26] suggests that it is worthwhile pursuing coarse-grained models that take sequence into account. It is worth emphasizing that although NMA-related methods have been successful in capturing large-scale motions that are difficult to identify currently with other techniques, these low-resolution studies do not contain enough information to address detailed mechanistic questions, for which, as discussed in Sect. 2, require thorough analysis of coupling between different length and temporal scales.

Finally, a more ambitious direction that holds great promise involves developing truly multi-scale methods that treat different degrees of freedom with an appropriate level of details. An extreme case concerns the development of models that keep the atomic description for the biomolecule, while treating solvent or lipid membrane in an implicit manner [97]. A less drastic approach is to treat selected degrees of freedom such as solvent and lipid membrane far away from the region of interest with coarse-grained potential functions [98]; it is possible to describe regions even further away with a continuum level of description, using, e.g., a finite element scheme. For example, multi-scale methods that include QM, MM and continuum treatments of different regions have been successfully applied to problems such as material cracking [99]. However, the interface problem was much simpler for those solid systems that involve a rather regular distribution of atoms at most locations; it is more difficult to handle the interface issue (e.g., exchange of matter across the boundary) in heterogeneous soft-matters such as proteins embedded in fluidic lipid membrane and motor proteins attached to a network of actin filaments. Moreover, development of coarse-grained potentials that are *transferrable* for biomolecules is not straightforward at all [100], although progress is being made [101–104]. Finally, although such a hybrid approach is likely successful for describing equilibrium properties (e.g., structure), it is significantly more challenging to reproduce real-time dynamics [102,103], even for a localized region of interest. On the bright side, the rapid increase in the resolution and sensitivity of experimental techniques such as the optical tweezer and micropipette [2,3,105] made it possible to make direct connections between multi-scale simulations and solid experimental data, a valuable step for validating the simulation models that eventually will be applied in predictive type of applications.

4 Envoi: physical chemistry in an new era of biological science

Feynman is apparently correct in that *all* life processes are, ultimately, governed by the same physical and chemical principles that are equally valid to non-living systems. However, the fact that biological systems have been selected over billions of years of evolution makes them profoundly different from inanimate materials. The vectorial processes discussed above are just one type of staggering phenomena that we chose to focus on, there are certainly endless number of fascinating life processes [106,107] that illustrate the power of

combining physics, chemistry and well-motivated optimization. This is the reason, as highlighted in Sect. 2, that considering the biological function of the system under study is crucial in defining the most interesting mechanistic questions. Although only protein-scale phenomena has been discussed in some details, it is clear that ample opportunities exist for theoretical and computational studies at the larger scale that involves multiple macromolecules in the complex and noisy cellular or biomimetic [108] environments. Although systems biology is still in its infancy [109], carefully constructed models have already made an impact on the study of genetic circuits [110], metabolic pathways [111,112] and immunological synapse [113,114]. In this new era of biological science that is fueled with an explosively increasing amount of information, physical scientists, including theoretical and computational chemists, are poised to make essential contributions to in-depth mechanistic studies by bringing quantitative tools and physical insights.

Acknowledgements QC acknowledges generous supports from various funding agencies (Research Corporation, American Chemical Society, National Science Foundation, Alfred. P. Sloan Foundation and the National Institutes of Health) and group members as well as collaborators that made it possible to carry out the work discussed here.

References

1. Lakadamyali M, Rust MJ, Babcock HP, Zhuang XW (2003) Proc Nat Acad Sci 100:9280–9285
2. Bustamante C, Macosko, JC, Wuite GJL (2000) Nat Rev Mol Cell Biol 1:130–136
3. Bustamante C, Chemla YR, Forde NR, Izhaky D (2004) Annu Rev Biochem 73:705–748
4. Zhuang XW, Kim H, Pereira MJB, Babcock HP, Walter NG, Chu S (2002) Science 296:1473–1476
5. Shaevitz JW, Abbondanzieri EA, Landick R, Block SM (2003) Nature 426:684–687
6. Xie XS (2002) J Chem Phys 117:11024–11032
7. Yang H, Luo GB, Karnchanaphanurach P, Louie TM, Rech I, Cova S, Xun LY, Xie XS (2003) Science 302:262–266
8. Wikstrom M (2004) Biochim Biophys Acta Bioenergetics 1655:241–247
9. Stokes DL, Green NM (2003) Annu Rev Biophys Biomol Struct 32:445–468
10. Schliwa M (2003) Molecular motors. Wiley-VCH, New York
11. Bray D (1998) Annu Rev Biophys Biomol Struct 27:59–75
12. Bray D (1995) Nature 376:307–312
13. Jencks WP (1989) J Biol Chem 264:18855–18858
14. Hill TL (1977) Free energy transduction in biology. Academic, New York
15. Hill TL (1985) Cooperativity theory in biochemistry. Springer, Berlin Heidelberg New York
16. Schnitzer MJ, Block SM (1997) Nature 388:386–390
17. Yildiz A, Forkey JN, McKinney SA, Ha T, Goldman YE, Selvin PR (2003) Science 300:2061–2065
18. Yasuda R, Noji H, Yoshida M, Kinosita K, Itoh H (2001) Nature 410:898–904
19. Sako Y, Yanagida T (2003) Nat Cell Biol Suppl suppl:SS1–SS5
20. Murray JD (2003) Mathematical biology. Springer, Heidelberg New York
21. Sali A, Glaeser R, Earnest T, Baumeister W (2003) Nature 422:216–225
22. Yonekura K, Maki-Yonekura S, Namba K (2003) Nature 424:643–650
23. Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W (2002) Science 298:1209–1213
24. Xia Y, Yu HY, Jansen R, Seringhaus M, Baxter S, Greenbaum D, Zhao HY, Gerstein M (2004) Annu Rev Biochem 73:1051–1087
25. Rousseau F, Schymkowitz J (2005) Curr Opin Struct Biol 15: 23–30
26. Li G, Cui Q (2004) Biophys J 86:743–763
27. Li G, Cui Q (2004) J Phys Chem B 108:3342
28. Holmes KC, Geeves MA (1999) Annu Rev Biochem 68:687–728
29. Howard J (2001) Mechanics of motor proteins and the cytoskeleton. Sinauer Associates, Inc. Sunderland
30. Bustamante C, Keller D, Oster G (2001) Acc Chem Res 34:412
31. Karplus M, Gao YQ (2004) Curr Opin Struct Biol 14:250
32. Hill TL, Eisenberg E (1981) Q Rev Biophys 14:463–511
33. Lymn RW, Taylor EW (1971) Biochem 10:4617–4624
34. Shih WM, Gryczynski Z, Lakowicz JR, Spudich JA (2000) Cell 102:683–694
35. Highsmith S, Polosukhina K, Eden D (2000) Biochemistry 39:12330–12335
36. Malnasi-Csizmadia A, Pearson DS, Kovacs M, Woolley RJ, Geeves MA, Bagshaw CR (2001) Biochemistry 40:12727–12737
37. Holmes KC, Angert I, Kull FJ, Jahn W, Schroder RR (2003) Nature 425:423–427
38. Nicholls DG, Ferguson SJ (2002) Bioenergetics 3, Academic, New York
39. Brzezinski P (2004) Trends Biochem Sci 29:380
40. Popovic DM, Stuchebrukhov AA (2004) J Am Chem Soc 126:1858–1871
41. Bloch D, Belevich I, Jasaitis A, Ribacka C, Puustinen A, Verkhovsky MI, Wikstrom M (2004) Proc Nat Acad Sci 101:529–533
42. Yang W, Lee T-S (1995) J Chem Phys 103:5674
43. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11(6):700–733
44. Gao J (1996) Acc Chem Res 29:298
45. Shurki A, Warshel A (2003) Adv Prot Chem 66:249
46. Cui Q, Karplus M (2003) Adv Prot Chem 66:315
47. Warshel A, Levitt M (1976) J Mol Biol 1976 103:227
48. Guallar V, Friesner RA (2004) J Am Chem Soc 126:8501–8508
49. Lennartz C, Schaefer A, Terstegen F, Thiel W (2002) J Phys Chem B 106:1758
50. Cui Q, Karplus M (2001) J Am Chem Soc 123:2284
51. Liu H, Lu Z, Cisneros GA, Yang W (2004) J Chem Phys 121:697
52. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim Th, Suhai S, Seifert G (1998) Phys Rev B 58:7260
53. Riccardi D, Li G, Cui Q (2004) J Phys Chem B 108:6467–6478
54. Amara P, Field MJ (2003) Theor Chem Acc 109:43–52
55. Das D, Eurenius KP, Billings EM, Sherwood P, Chattfield DC, Hodošček M, Brooks BR (2002) J Chem Phys 117:10534–10547
56. Antes I, Thiel W (1999) J Phys Chem A 103:9290
57. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) J Phys Chem A 104:1720–1735
58. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714–4721
59. Zhang Y, Lee TS, Yang W (1999) J Chem Phys 110:46–54
60. Konig PH, Hoffmann M, Frauenheim Th, Cui Q (2005) J Phys Chem B 109, 9082–9095
61. Warshel A (2003) Annu Rev Biophys Biomol Struct 32:425–443
62. Schaefer P, Riccardi D, Cui Q (2005) J Chem Phys 123, 014905
63. Riccardi D, Schaefer PS, Cui Q (2005) J Phys Chem B 109, 17715–17733
64. Im W, Bernéche S, Roux B (2001) J Chem Phys 114(7):2924–2937
65. Frenkel D, Smit B (1996) Understanding molecular simulations: from algorithms to applications. Academic, San Diego
66. Warshel A, King G (1985) Chem Phys Lett 121:124–129
67. Riccardi D, Schaefer P, Yang Y, Yu H, Ghosh N, Prat-Resina X, Konig P, Li G, Xu D, Guo H, Elstner M, Cui Q, J Phys Chem B (submitted)
68. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Annu Rev Phys Chem 53:291–318

69. Cui Q, Karplus M (2000) J Chem Phys 112:1133
70. Cui Q, Karplus M (2000) J Phys Chem B 104:3721
71. Cantor CR, Schimmel PR (1980) Biophysical chemistry, part. II Techniques for the study of biological structure and function. W. H. Freeman and Company, San Francisco
72. Zanni MT, Hochstrasser RM (2001) Curr Opin Struct Biol 11:516–522
73. Mukamel S, Abramavicius D (2004) Chem Rev 104:2073–2098
74. Ham S, Hahn S, Lee C, Kim TK, Kwak K, Cho M (2004) J Phys Chem B 108:9333–9345
75. Schlitter J, Engels M, Kruger P, Jacoby E, Wollmer A (1993) Mol Sim 10:291
76. Isralewitz B, Gao M, Schulten K (2001) Curr Opin Struct Biol 11:224–230
77. Ma J, Flynn TC, Cui Q, Leslie AGW, Walker JE, Karplus M (2002) Structure 10:921–931
78. Bandyopadhyay S, Allison WS (2004) Biochemistry 43:2533–2540
79. Miyashita O, Onuchic JN, Wolynes PG (2003) Proc Nat Acad Sci 100:12570–12575
80. Hummer G, Kevrekidis IG (2003) J Chem Phys 118:10762–10773
81. Cui Q, Bahar I (2005) Normal mode analysis: theory and applications to biological and chemical systems. CRC, Bota Raton
82. Van Wynsberghe A, Li G, Cui Q (2004) Biochemistry 43:13083–13096
83. Cui Q, Li G, Ma J, Karplus M (2004) J Mol Biol 340:345–372
84. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2002) Biophys J 80:505–515
85. Tama F, Valle M, Frank J, Brooks CL (2003) Proc Nat Acad Sci 100:9319–9323
86. Tama F, Wriggers W, Brooks CL (2002) J Mol Biol 321:297–305
87. Ming D, Kong YF, Lambert MA, Huang Z, Ma JP (2002) Proc Nat Acad Sci 99:8620–8625
88. Ma JP (2004) Curr Prot Pept Sci 5:119–123
89. Tama F, Sanejouand Y (2001) Protein Eng 14:1–6
90. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu HY, Gerstein M (2002) Proteins: Struct Funct Genet 48:682–695
91. Ma J, Karplus M (1998) Proc Nat Acad Sci 95:8502–8507
92. Tama F, Miyashita O, Brooks CL (2004) J Mol Biol 337:985–999
93. Tama F, Gadea FX, Marques O, Sanejouand YH (2000) Proteins: Struct Funct Genet 41:1–7
94. Li G, Cui Q (2002) Biophys J 83:2457–2474
95. Tirion MM (1996) Phys Rev Lett 77:1905–1908
96. Li G, Van Wynsberghe A, Demerdash ONA, Cui Q (2005) Normal mode analysis: theory and application to biological and chemical systems. CRC Press, Boca Raton
97. Feig M, Brooks CL (2004) Curr Opin Struct Biol 14:217–224
98. Basdevant N, Borgis D, Ha-Duong T (2004) J Comp Chem 25:1015–1029
99. Broughton JQ, Abraham FF, Bernstein N, Kaxiras E (1999) Phys Rev B 60:2391–2403
100. Buchete NV, Straub JE, Thirumalai D (2004) Curr Opin Struct Biol 14:225–232
101. Ayton G, Voth GA (2002) Biophys J 83:3357–3370
102. Ayton G, Tepper HL, Mirijanian DT, Voth GA (2004) J Chem Phys 120:4074–4088
103. Nielsen SO, Lopez CF, Srinivas G, Klein ML (2004) J Phys Condens Matt 16:R481–R512
104. Marrink SJ, de Vries AH, Mark AE (2004) J Phys Chem B 108:750–760
105. Huang H, Kamm RD, Lee RT (2004) Am J Physiol Cell Physiol 287:C1–C11
106. Kirkwood TBL, Rosenberger RF, Galas DJ (1986) Accuracy in Molecular processes, Chapman and Hall, London
107. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) Molecular cell biology. W. H. Freeman and Company, San Francisco
108. Runyon MK, Johnson-Kerner BL, Ismagilov RF (2004) Angew Chim Int Edit 43:1531–1536
109. Kitano H (2002) Nature 420:206–210
110. McAdams HH, Arkin A (1998) Annu Rev Biophys Biomol Struct 275:199–224
111. Wolf DM, Arkin AP (2003) Curr Opin Micro 6:125–134
112. Rodionov DA, Dubchak I, Arkin A, Alm E, Gelfand MS (2004) Genome Biol 5:R90
113. Chakraborty AK, Dustin ML, Shaw AS (2003) Nat Immuno 4:933–936
114. Lee KH, Dinner AR et al. (2003) Science 302:1218–1222