



ELSEVIER

# Multiscale methods for macromolecular simulations

Paul Sherwood<sup>1</sup>, Bernard R Brooks<sup>2</sup> and Mark SP Sansom<sup>3</sup>

In this article we review the key modeling tools available for simulating biomolecular systems. We consider recent developments and representative applications of mixed quantum mechanics/molecular mechanics (QM/MM), elastic network models (ENMs), coarse-grained molecular dynamics, and grid-based tools for calculating interactions between essentially rigid protein assemblies. We consider how the different length scales can be coupled, both in a sequential fashion (e.g. a coarse-grained or grid model using parameterization from MD simulations), and via concurrent approaches, where the calculations are performed together and together control the progression of the simulation. We suggest how the concurrent coupling approach familiar in the context of QM/MM calculations can be generalized, and describe how this has been done in the CHARMM macromolecular simulation package.

## Addresses

<sup>1</sup> STFC Daresbury Laboratory, Warrington WA4 4AD, UK<sup>2</sup> Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA<sup>3</sup> Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Corresponding author: Sherwood, Paul ([p.sherwood@dl.ac.uk](mailto:p.sherwood@dl.ac.uk)), Brooks, Bernard R ([brb@nih.gov](mailto:brb@nih.gov)) and Sansom, Mark SP ([mark.sansom@bioch.ox.ac.uk](mailto:mark.sansom@bioch.ox.ac.uk))

Current Opinion in Structural Biology 2008, 18:630–640

This review comes from a themed issue on  
Biophysical methods  
Edited by Samar Hasnain and Soichi Wakatsuki

Available online 17th September 2008

0959-440X/\$ – see front matter

© 2008 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2008.07.003

## Introduction

Recent years have seen a significant increase in the ability of computational approaches to help us understand the structure and function of complex biological systems. Alongside the advances in available computer power, we have seen a steady progress in the development and validation of modeling methods spanning a variety of length and time scales. Explicitly correlated electronic structure methods with linear scaling with system size now offer the prospect of chemical accuracy for the energetics of biological chemical reactions [1]. However, there is a massive disparity in the time and length scales on which the important cellular events occur and those that are accessible using these quantum-based tools.

Much larger length scales and longer timescales can be addressed by atomistic treatments based on much simpler, molecular mechanical expressions for the energy (important examples being molecular dynamics (MD) [2] and the elastic network models [3]), in addition to the methods based on larger primitive particles, for example, those in which each bead represents a small number of atoms. At longer length scales still, we will describe approaches, which treat whole macromolecules as the building blocks, allowing the study of intermolecular interactions. This being of particular interest in docking and in structural studies of macromolecular assemblies.

To make predictions of macromolecular behavior from first principles, it is clearly necessary to incorporate knowledge from studies at the small length scales into the more coarse-grained (CG) approaches. Since every method (except for first principles quantum chemistry) depends on parameters to describe the energy of interactions, one can foresee a process by which we can ascend the length scale ladder; each successive method deployed incorporating parameters taken from the previous one. This approach is already the key to atomistic simulation using MD techniques, as in the widely used parameter sets (force fields), which incorporate data from QM calculations, and other examples will be given below. This methodology has been referred to in a recent review in this journal [4] as the serial multiscale approach. Others have used the terms sequential [5], implicit, or message-passing methods to describe the approach. The serial approach has the weakness that any fitted parameter set, no matter how carefully constructed, is likely to have limits to its applicability or accuracy. For example, MD simulation using standard force fields will break down when a chemical bond is significantly stretched or an electronically excited state is present. For this reason, it is necessary to find other ways to integrate the calculations at the different length scales in a more closely coupled way. Invariably this requires the execution of multiple component calculations as part of a given simulation, leading to their description as parallel multiscale treatments (also known as concurrent [5] or explicit schemes). One obvious approach is to introduce the concept of a spatial decomposition, whereby most of the system treated using the simpler method and the part for which that method fails is treated by a higher level of theory. The most widely used example is the mixed quantum mechanics/molecular mechanics (QM/MM) approach [6], and one can imagine this description working at other length scales, but with more difficulty. While QM and MM share the same atomistic character, at the longer length scales mapping between the system descriptions

necessarily becomes more complex. In this article, we will follow [5] and use the terms sequential and concurrent, reserving the term parallel for the use of multiple compute processors.

It is possible to subdivide the sequential and concurrent approaches further, by looking into the nature of the information transfer between the different calculations [4]. This classification of methodologies is perhaps best developed in the field of process engineering [7], where the drive is to integrate detailed materials models into a predictive model of macroscopic behavior.

### QM/MM methods

As noted already, the close coupling of QM and classical MM approaches is an important example of a concurrent multiscale scheme. When applied to enzymic catalysis, a quantum mechanical method is used to study the reactive process (active site and some or all of the substrate) and the surroundings, which will include the remainder of the enzyme, solvent, counterions, etc., is treated by classical MM models. The QM/MM scheme, especially when used with a cost-effective QM method, such as a semiempirical theory, allows exploration of the potential energy surface (PES) and molecular dynamics to obtain free energies of reactions. It is now established as the method of choice for the study of enzymatic reactions. It is not necessary to discuss the history of QM/MM models, as there are many good reviews on the subject. In particular, we refer the reader to the comprehensive article by Senn and Thiel [6], in which many aspects of the method are described and a summary table of applications is provided, and to the forward-looking discussion of Lin and Truhlar [8].

### QM/MM models – the conventional and subtractive approaches

We start with a description of the established QM/MM methodology in which the QM and MM calculations are performed on different parts of the chemical system. The separate regions are coupled, of course, and the coupling terms can take a variety of forms depending on the particular model, usually on the basis of classical electrostatics and parameterized nonbonded interactions. Subsequently, Morokuma and his coworkers have introduced a subtractive [6] QM/MM model exemplified by the ONIOM scheme [9]. In this case, the domain chosen for the lower level (or more CG) calculation includes those parts of the system treated at a higher level. The multiple-counting is tackled by subtracting a third energy, that of the inner domain at the lower level of theory. In this way, the interactions between the domains appear within the lower level calculation and no coupling terms have to be devised. A similar approach has been developed for solid-state modeling by Sauer and Sierka [10], and the technique can simply be applied to other levels of theory. Mixing different QM models, for

example density functional theory (DFT) with coupled cluster models, is one important application. At longer length scales, we can now integrate atomistic and coarse-grain treatments, as discussed below.

### Energetics of reaction pathways

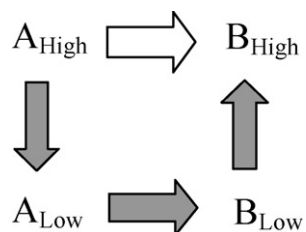
QM/MM calculations have been widely deployed in the study of enzymatic reactions. Many different QM methods have been used, ranging from semiempirical to high level; correlated [11] approaches and many of the techniques of small molecule computational chemistry can be used to explore the PES for macromolecules. Although there are serious limitations arising from the lack of sampling inherent in such an approach [12], the gross features of the PES can provide useful information in the discussion of reaction mechanisms.

Typical QM/MM studies in the past involved a single QM system embedded in a classical MM region. There has been a major effort to streamline QM/MM calculations, so that as many as 100 QM calculations are performed simultaneously in a single computer run. For example, a QM/MM potential energy function has been implemented between the CHARMM [13] and Q-Chem software packages [14]. This interface supports the minimization and dynamics via analytical first derivatives at the Hartree–Fock, DFT, and post-HF (RIMP2, MP2, and CCSD) levels of theory. In addition, the replica path (RPATH) [15], nudged elastic band (NEB) [16], and the combination of RPATH restrained distance (RPATH + RESD) [17] reaction pathway methodologies are fully supported. These are powerful techniques for studying reaction pathways in a multilevel parallel (or parallel/parallel) fashion, with each QM/MM pathway point being distributed to a different node of a Beowulf cluster. The software is designed to facilitate the optimization and analysis of a pathway in a single run.

### QM/MM approaches to free energies

Integrating QM into a macromolecular dynamical simulation provides a route to the free energy differences associated with biochemical processes. Direct computation of the free energy of a chemical system from the thermodynamical partition function is rarely practical and many creative approaches to extracting the relevant free energy changes for processes of biological interest from tractable MD simulations have been devised. The adequate sampling for conformational space remains a challenge, especially when higher-level QM calculations are used. The reader is referred to Ref. [18] for a review of the methods by which free energies based on *ab initio* free energy results may be obtained. In many cases, it is satisfactory to determine the zero-temperature reaction path for the QM system (with the MM surroundings relaxed at each point) and then use MD to sample the entropic contribution of the enzyme or solvent dynamics. Where the motion of the QM atoms needs to be included,

Scheme 1



a promising approach would appear to be that based on thermodynamic cycles [19], see Scheme 1.

The desired transformation (between states A and B at the high level computational technique) is shown by the white arrow. It is not usually directly accessible because of the cost of achieving adequate conformational sampling at the high level of theory. However, it can be computed by summing the three gray arrows, whereby the free energy difference between A and B is computed using a cheaper Hamiltonian (either MM or a semiempirical QM scheme such as EVB [19]). States A and B can be either distinct chemical states in a proposed reaction mechanism or different chemical systems, in which case the A–B transformations are nonphysical (known as ‘alchemical’ free energy perturbation). Recently Woods *et al.* [20] have suggested an approach to improving the accuracy of the two vertical integrals by using Metropolis–Hastings scheme (performing a Monte Carlo sampling at the MM level but using a QM/MM acceptance test).

In some cases, it is possible to estimate free energy changes without recourse to dynamical sampling. Within a harmonic approximation, the second derivatives of the energy with respect to the atomic coordinates (the Hessian) allow the computation of the vibrational entropy. As a result, there is also a renewed effort to utilize partial or full multiscale Hessians to calculate molecular flexibility and to estimate thermodynamic properties, such as reaction free energies. Expanding on the previous functionality, full QM/MM analytic second derivatives have recently been implemented into Q-Chem and interfaced with CHARMM’s VIBRan module [21]. Efficiency in calculating the Hessian can be achieved by limiting the number of environmental (i.e. MM) degrees of freedom and solving couple perturbed calculations on the allowed degrees of freedom. This is most easily done by treating distant molecular groups as rigid objects with only rotational and translational degrees of freedom.

A new vibrational subsystem analysis (VSA) method has been developed and implemented that couples global motion to a local subsystem while including the inertial effects of the environment [21]. The premise of the VSA method is partitioning of a system into a smaller region of interest and usually a larger part is referred to as environ-

ment. This method allows for researchers to address a vast array of interesting problems. Examples of these include, but are not limited to, more accurately estimating vibrational free energy contributions for parts of a large system; elimination of the ‘tip effect’ in elastic network model (ENMs) calculations; probing specific degrees of freedom that may contribute to free energy differences and estimating activation free energies in QM/MM reaction path calculations. The main use of this method will most probably be the estimation of pathway free energies as opposed to just enthalpies, without the need for extensive dynamic sampling.

### Adaptive schemes

We now consider the recent attempts to move beyond a static definition of the QM region. This has been a topic of research in the materials modeling community for some time [22] where it is required for problems such as crack propagation. A recent development of the Learn on the Fly (LOTF) method [23] has led to its application to biological systems, in which the AMBER QM/MM implementation [24] has been used (Gabor Csanyi, unpublished data). For biological systems, the most important application is solvation, where solvent molecules may move into the active site and thus require QM treatment for part of the dynamics trajectory. Heyden *et al.* have recently proposed the adaptive partitioning [25] approach, which addresses issues of discontinuities in the energy and gradients characteristic of some of the earlier methods.

### Electronic excitations

To finish this section of the review, it is worth mentioning the recent developments in two areas of protein science in which the QM/MM approach can be used to explore biologically important *electronic* transitions. We first consider the application of the QM/MM approach to photochemistry and the study of excited state PESs. Till date, the most popular QM methods for these studies are those based on multiconfigurational wavefunctions (e.g. CASSCF) and the availability of second order perturbation theory corrections for dynamical correlation (the CASPT2 scheme) now enables highly accurate energetics to be computed for excited state PESs. The CASPT2 scheme has recently been incorporated into a QM/MM approach [26] and used to study the fluorescence of the green fluorescent protein (GFP) used in the design a prototype light-driven molecular switch.

It is also possible to use QM/MM molecular dynamics to explore the influence of enzyme and solvent on redox-active transition metals. One approach, within the framework of Marcus theory, uses classical MD to sample the conformational space, combined with DFT calculations of the vertical ionization process at a number of snapshots. This approach was able to compute a difference in redox potential of 60 mV (between mesophilic and thermophilic

forms of Rubridoxin) to within 20 mV [27]. A scheme that incorporates dynamics on both the oxidized and reduced PES and switches between them in a grand-canonical MD scheme has been derived and applied to the redox potential of copper in azurin [28]. Very recently, a new approach to the simulation of redox processes has been suggested. The oxidation state is considered continuously variable, using fractional electron count [29] and can, therefore, be used as an order parameter in a thermodynamic integration free energy scheme.

### Elastic network models

One of the most important tools for the CG study of protein conformation changes is the ENM. Given the  $C_\alpha$  atomic coordinates for a protein's native structure, we build an ENM by using a harmonic potential with a single force constant  $C$  to account for pairwise interactions between all  $C_\alpha$  atoms that are within a cutoff distance ( $R_C = 10 \text{ \AA}$ ). The energy in the elastic network representation of a protein is

$$E_{\text{network}} = \frac{1}{2} \sum_{d_{ij}^0 < R_c} C(d_{ij} - d_{ij}^0)^2,$$

where  $d_{ij}$  is the distance between the dynamical coordinates of the  $C_\alpha$  atoms  $i$  and  $j$ , and  $d_{ij}^0$  is the distance between  $C_\alpha$  atoms  $i$  and  $j$ , as given in the reference structure. For recent reviews see [3,30]. It is especially useful for multiscale modeling in that it provides a simple CG model that allows protein flexibility while preserving a structure that is reasonably close to an experimentally observed structure. This model is purely harmonic, thus proteins cannot unfold. The low-frequency modes from an ENM model present an excellent basis for examining global motion. Limitations of ENM for examining processes with some local character have been observed [31], suggesting that more than a few lowest modes are often needed. However, the robustness of modes to ENM parameter perturbations was found to be useful in identifying functionally important modes [32], which may further reduce the subset of low-frequency modes required for focused conformational sampling.

ENMs can be used to facilitate structure refinement at low resolution [33] and a multibase generalization of ENM has been the basis for several recently proposed techniques for efficient generation of conformational transition pathways [34–36]. ENM-based methods for predicting dynamic coupling and ligand-binding-induced conformational changes have been successfully applied to HCV NS3 helicase [37].

Recent work with ENM models demonstrate that ENM modes can provide an efficient sampling of conformational space covered by the large number of available HIV-1 protease structures, in a manner similar to MD trajectory or NMR ensemble analysis [38] and that the existing

elastic models [ENM and Gaussian network model (GNM)] can be improved to achieve an optimal description of both protein conformational motions and thermal fluctuations [39].

A more complex ENM approach to conformational change is to combine information from multiple conformational states of the protein [40], for example to generate double-well models. This approach has been used to explore, for example, conformational changes in adenylate kinase [36], also see [41], a well-studied model system for such transitions [42]. A number of other models for conformational flexibility merit further exploration, including those which combine rigid 'blocks' within a protein with flexible connections [43].

### Coarse-grained MD simulations

Atomistic simulations of complex biomolecular systems provide a detailed picture of, for example, conformational dynamics and protein/ligand interactions. However, even with large-scale computational resources and well-scaling simulation codes, it is still challenging to reach long time scales ( $>1 \mu\text{s}$ ) for large and complex systems (e.g. [44]). Consequently, there has been considerable interest in the application of CG models to proteins and related biological systems.

A number of different levels of CG models are possible, varying in the number of atoms, which are represented by a single particle. For example, an amino acid residue within a protein could be represented by a single particle or by three to four particles where each particle represents two to five nonhydrogen atoms. In each case, both bonded and nonbonded interactions have to be parameterized to reproduce either thermodynamic properties of the constituent monomers (e.g. amino acids) or the overall dynamic properties of the resultant protein. Such coarse-graining reduces by an order of the magnitude the interactions that have to be evaluated for a given system and also enables the use of longer time steps (e.g. 40 fs for CG rather than 1 fs for atomistic). Together this may provide several orders of magnitude of speedup relative to atomistic simulations, thus enabling longer timescale processes for more complex systems to be explored.

Following the development and initial exploration of CG models for proteins (previously reviewed by e.g. [3]) and for membranes (previously reviewed by [45]), there has been considerable interest in exploring the potential for this approach and applying it to a number of key systems. In what follows this will be illustrated in terms of: firstly, modeling conformational changes of proteins; secondly, applications to membrane proteins; and thirdly, CG and related multiscale approaches to more complex assemblies. We will not discuss the history of CG simulations in the context of protein folding as this has recently been

reviewed in COSB [46] but will discuss some very recent developments later in this section.

### Conformational changes in proteins

A number of CG models have been used to explore mechanisms of conformational change in proteins. For all of these a major issue is that of how to parameterize. A promising general approach to this is a force-matching scheme in which the CG model is parameterized to match atomistic simulations for the same or a related system. This has been explored for, for example, ENMs [47] and for two particles per amino acid models [48] of proteins.

### Membranes and their proteins

There has recently been considerable interest in extending CG models previously used for lipid bilayers [49–51] and for simple models of membrane proteins [52–54] to more complex membrane proteins and peptides. This required the modification of existing CG models for lipids [55] to also include amino acids. It also required the issues of parameterization of amino acid/hydrocarbon interactions in CG models to be addressed, both by the comparison with experimental thermodynamic (partitioning) data and via comparison with atomistic-simulation based free energy profiles for amino acid sidechains versus location within a lipid bilayer [56].

Initial studies using CG models to simulate membrane/peptide and membrane/protein systems have been very encouraging, reproducing a range of experimental data on the location of model peptides and proteins within membranes [56,57] and reproducing patterns of protein/lipid interactions seen in extended atomistic simulations [58]. It should be noted that in general such models require either dihedral angle or elastic network restraints in order to retain the secondary and tertiary structure of the protein.

CG-MD may be used to selfassemble lipid bilayers around membrane proteins, thus enabling the prediction

of the localization and orientation of a protein within a lipid bilayer (see Figure 1). The smaller computational unit (CPU) demands of CG-MD mean that it is feasible to use this method to compare different families of membrane proteins and their interactions with lipids [59]. For example, the application to a nonredundant set of ~100 membrane proteins [60] enabled comparative analysis of protein/bilayer interactions, indicating how local bilayer deformations may be related to membrane protein class (e.g.  $\alpha$ -helix bundle versus  $\beta$ -barrel trans membrane (TM) domains). CG-MD has used to explore bilayers containing multiple (e.g. 16) rhodopsin molecules [61] to reveal how local bilayer deformation influences protein–protein association within a membrane.

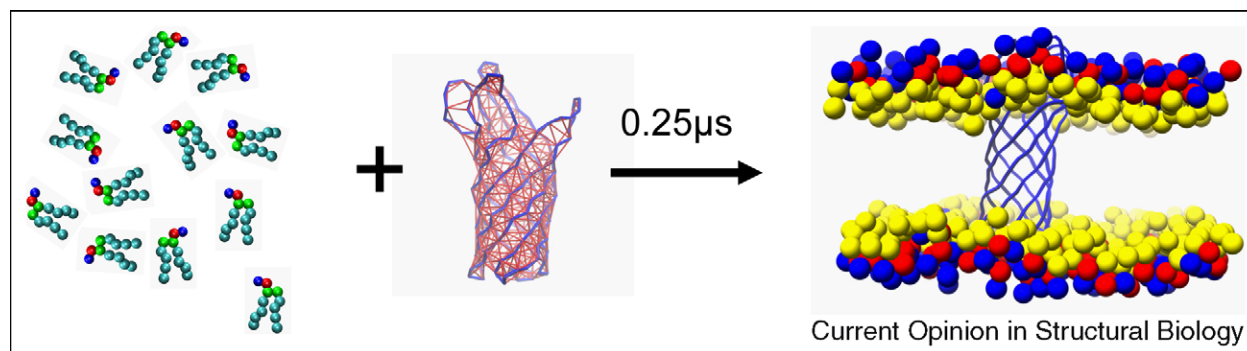
CG-MD simulations may also be used to characterize the interactions of proteins with the *surface* of a lipid bilayer, as has been shown both for small (~30-residue) toxins that bind to membrane surfaces and inhibit ion channels [62], and for more complex membrane-binding enzymes such as phospholipase A2 [63].

CG simulations have been used to explore functional aspects of a number of channel and related proteins in membranes. For example, CG simulations of lipid–protein interactions of isolated voltage sensor domains from  $K^+$  channels [64] suggested a degree of bilayer deformation as a result of interactions of lipid headgroups with the charged sidechains of the voltage-sensing S4 helix, as is also seen in (shorter) atomistic simulations [65,66]. CG simulations have been used to study possible mechanisms of gating of a number of channels and related proteins, including the mechanosensitive channel MscL [67], the voltage-gated channel Kv1.2 [68], and the translocon [69].

### CG and multiscale approaches to complex assemblies

One of the key areas for the future, which is only just beginning to be explored, is the application of CG and

Figure 1



Schematic representation of a CG-MD simulation of the selfassembly of a lipid bilayer/protein system from randomly positioned lipid (and water — not shown) molecules plus a known membrane protein structure. After ~0.25  $\mu$ s of simulation, the protein is stably inserted in a bilayer. The representation of the protein on the left illustrates the use of an elastic network model (in red) to restrain the structure of the protein.

multiscale approaches to more complex systems. A number of groups have successfully applied multiscale approaches to (reasonably complex) enzymes (e.g. acetylcholine esterase [70] and OmpT [71]) proving the utility of this approach. CG simulations have been used to self-assemble more complex systems such as lipoprotein particles [72]. Initial explorations of combining CG models of protein and of DNA in simulations of the nucleosome [73] are very promising. Highly CG models of lipids combined with radically simplified representations of proteins have been used to simulate protein-induced membrane vesiculation [74]. More recently, a multiscale approach has been used to explore bending and dynamics of membranes by multiple copies of a BAR domain [75], demonstrating that multiscale simulations have the potential to link structural and systems descriptions of cell biological processes.

### Protein folding in presence of osmolytes – the molecular transfer model

One limitation of many of these CG models is that the folding/unfolding reaction can only be initiated by a change in temperature, although many *in vitro* experiments use osmolytes (i.e. small organic molecules such as urea, guanidinium hydrochloride, trimethylamine-*N*-oxide, and sucrose) to modulate native state stability [76]. Recently, a novel method referred to as the molecular transfer model (MTM) has been developed to predict, rapidly and accurately, the effect of osmolytes on the thermodynamic properties of proteins [77]. It does this by combining information from CG molecular simulations using the BLN (hydrophoBic, hydrophiLic, Neutral) scheme, experimentally measured transfer free energies of individual amino acids, and the Tanford transfer model, all of which estimate the free energy of transferring a given protein conformation from water to aqueous osmolyte solution. With this information the MTM computes the partition function for any osmolyte type and osmolyte concentration of interest, thereby allowing almost any thermodynamic property to be computed. The MTM, being a postsimulation processing technique, is rapid. In a matter of minutes it can predict the effect of a large number of aqueous osmolyte solution conditions on the thermodynamics of protein folding and unfolding. The MTM has been validated against experimentally measured *m*-values (the rate of change of  $\Delta G$  as a function of denaturant concentration and a quantitative measure of the breadth of the unfolding transition) and single molecule fluorescence resonance energy transfer (FRET) measurements for protein L and a highly stable cold shock protein [77].

The MTM is a significant breakthrough in modeling because never before have molecular simulations been able to study protein folding/unfolding thermodynamics as a function of osmolyte concentration; it can accurately predict experimentally measurable quantities and offers

a molecular interpretation of experimental results. Although, in theory, the MTM can be applied to all-atom simulations, in practice it is only currently accurate for CG models, where adequately converged sampling can be achieved.

### Grid-based methods

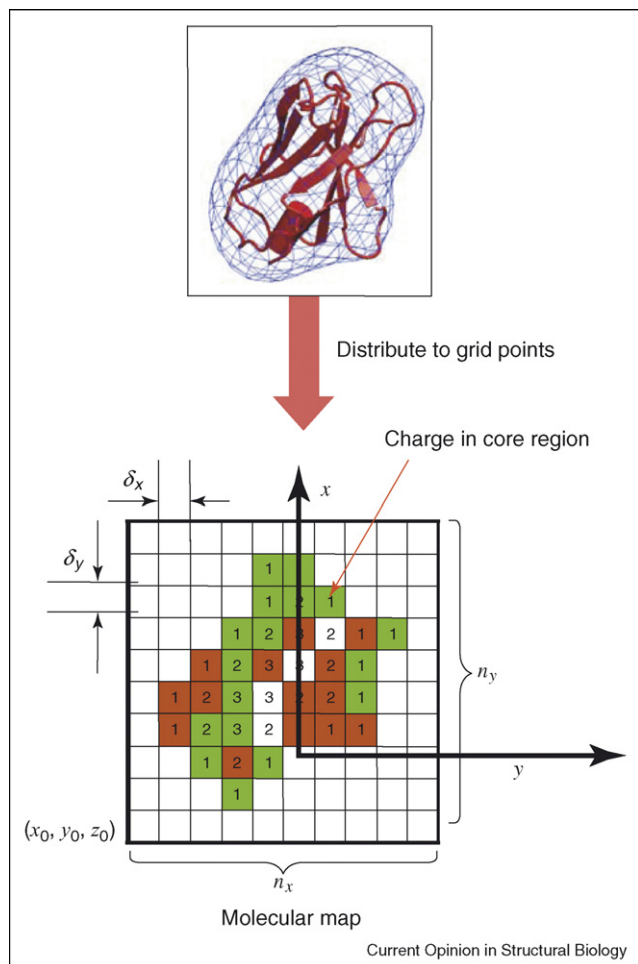
Molecular modeling has been a powerful approach to provide structural insights into biological procedures at the atomic level. New developments in experimental technologies, such as electron microscopy, provide an approach to obtain low-resolution structure information of large molecules and their assemblies. Extracting structure information from these low resolution maps and obtain atomic interpretation of the large biomolecular assemblies has become a powerful tool in modern structural biology. This requires molecular modeling to be conducted in conjunction with these low-resolution maps, as well as high-resolution atomic structures, to maximize the capability in structural biology studies.

Grid-based molecular modeling differs from the center-based models in that molecular properties are assigned to regions of space that are broken up into many smaller volume elements. This is a natural way to interact with electron microscopy data and X-ray data where electron densities are partitioned in the same manner. The main advantage of the grid approach is that operations, such as docking, can be performed rapidly and with a simplified energy function. One major drawback is that it is not easy to incorporate flexibility in a straightforward manner.

As with the development of structural biology, molecular modeling is now applied to larger and larger biomolecular machineries. As the systems become large, the atomic description becomes very inefficient and time-consuming and it is more efficient to treat large biomolecules as simplified shape objects, while ignoring their internal structures. Although molecular flexibility plays an important role in biological activity, in many cases molecular geometric shapes together with their surface properties are sufficient to describe many cellular processes such as molecule assembling and protein–protein binding.

As an example of the use of grid-based CG molecular modeling, we summarize the capabilities of CHARMM's [13] EMAP facility [78], which was originally developed to assist in structure determination with electron microscopy. Later, this facility was expanded to allow protein–protein docking analysis in the absence of experimental data. EMAP uses map objects (Figure 2) to represent space occupation of molecular structures. Unlike chemical descriptions of molecules that contains atoms and atoms are linked by chemical bonds, a map does not have internal chemical structures. Instead, a map represents a spatial distribution of certain properties, typically, electron density. This distribution is generally described by

Figure 2



A map object and its properties. A map object is defined as a grid with a given property mapped on the grid points. This could be either, for example, experimentally determined electron density or properties (e.g. electrostatic charges) generated from reference molecules.

scalar values at discrete grid points because of the irregularity of the distributions and because of limits in storage.

Map comparison provides a metric for fitting one map into another. Four types of cross correlation functions [78] are available for the comparison between map objects.

The energetics of molecular systems is the basis of molecular modeling. Calculation of molecular interaction using map objects is the crucial step to a successful modeling or simulation study. For atomic objects, interaction calculation is pairwise and is very time-consuming for large molecular assemblies. For map objects, we propose to use field interactions that can be calculated much more efficiently. Four types of energies are defined to describe the interaction between map objects: electric field interaction; surface charge–charge interaction; van

der Waals interaction; and desolvation interaction. The parameters in these functions are obtained by fitting into energies calculated with using standard force fields, making these another example of a serial multiscale algorithm.

The grid-threading Monte Carlo (GTMC) search algorithm [78] uses a combination of the grid search and Monte Carlo sampling to provide an efficient way for robustly fitting rigid domains to a target map. The conformational space is split into grid points and short Monte Carlo searches are performed to identify local maxima around the grid points. The global maximum is identified among the local minima.

The derivation of high-resolution molecular assembly structures from microscopy maps is a major application of the map approach. This method has been successfully applied into several experimental studies [79]. Figure 3 illustrates the steps to perform a fitting of high-resolution molecular structure into electron microscopy maps.

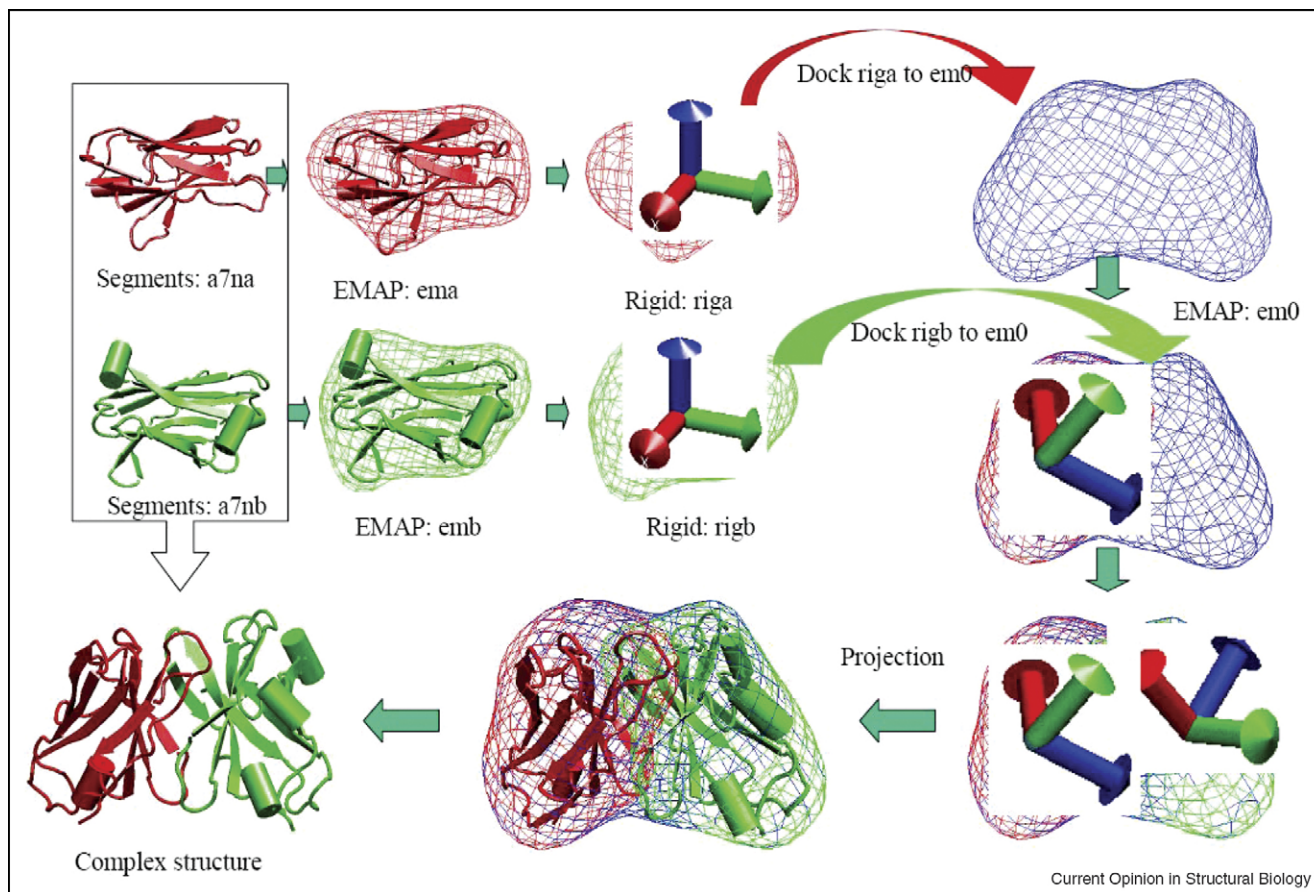
The structure obtained from map fitting generally is not optimized in atomic details. There are often atom overlaps or improper spacing between components. This structural mismatch can be removed by standard approaches such as energy minimization and simulated annealing as long as the fitting result is very close to the right structure. After the minimization, the root-mean-square (rms) deviation for the system illustrated in Figure 3 is 0.97 Å.

With map energy functions, it is possible to determine complex structures through minimizing the map interaction energy even in cases where the EM complex map is not available. It should be noted that the map object assumes a certain rigidity of a molecular object. Some flexibility of a loop region can be accommodated by the low-resolution character, while large flexibilities like domain movement should be dealt with multiple map objects. Recently, this method was successfully applied to modeling of the peroxiredoxin (Prx) complex [80].

### MSCALE – a generalized approach to multiscale simulation

With the exception of QM/MM, all of the multiscale methods reviewed in this article are typically deployed according to the sequential multiscale methodology, data from the high accuracy models being used to parameterize the coarser ones. However, it is interesting to consider the potential of a concurrent coupling of these disparate methods, by generalizing the ideas underpinning the additive and subtractive QM/MM schemes. Within a single framework, it is possible to support both additive and subtractive approaches (or a combination of these) and to allow a single energy and force calculation to be

Figure 3



Steps to derive molecular assembly structures by fitting molecular structures into electron microscopy maps. This is illustrated using a T-cell receptor (TCR) variable domain (PDB code: 1a7n) as an example complex to illustrate the modeling process with map objects. The TCR variable domain is a complex of two chains. The two chains are first blurred into maps of the same resolution (here 15 Å) as the EM map. Then each map is fitted into the EM map to get complex map. The complex map is projected back to atomic structures, which is the complex structure we are looking for. The root-mean-square (rms) deviations of the fitting result from X-ray complex is 3 Å.

done with diverse set of software and computer hardware. With such a tool, the multiscale combinations are endless.

CHARMM's recently implemented MSCALE command is such a general tool, supporting multiple, independent but connected calculations using defined molecular subsystems. Programs are run in parallel in a client-server mode, with basic communication. The server calculations can employ either CHARMM or several other supported programs (e.g. the quantum chemistry programs NWChem, Molpro, PSI3, and Gaussian 03) with a consistent interface.

This flexible implementation allows any number of subsystems, each running as a separate process, usually on a separate computer (or parallel cluster). The client process maps the forces and Hessian elements from the various atomic or CG centers, either by collocating them or by connecting them through constraints or restraints. For

example, the lone-pair facility in CHARMM can be employed to constrain one CG center to be at the center of mass of a collection of atoms. In constructing the total energy, a scale factor is applied to each subsystem; additive models will have all positive factors, with negative factors used for subtractive models. Scale factors can be tied to parameters ( $\lambda$ ), allowing free energy perturbation simulations with complex multiscale modeling.

With this capability, it is possible to mix models of different resolutions in a completely general manner. It is also possible to mix models at the same level of resolution, for example, combining the two residue-based CG schemes, ENM and BLN, to perform CG protein docking that allows protein flexibility but not protein unfolding. When used with all-atom modeling, it is possible to mix and match force field types in a single calculation. For example in simulating protein absorption



on a surface, the protein can be modeled with the CHARMM force field, while the surface could be modeled with the CFF force field.

## Conclusions

We have reported encouraging progress in the modeling of biomolecular systems at a range of different length scales. QM/MM techniques are now in routine use and the underlying quantum mechanical methods are becoming steadily more accurate. Significant methodological approaches to many chemically significant properties, such as redox potentials, free energy changes, and excited state reactivity have been made during the past few years. The coupling of QM and MM Hessian information has led to the development of new vibrational partitioning schemes with the application to harmonic free energies analyses for large molecular assemblies. We have reported on recent work with a number of different CG MD schemes and though there are still limitations to the types of processes that can be modeled with predictive accuracy, systematic studies of many complex biomolecules and assemblies are made possible by the computational efficiency of the methods. Likewise, ENMs are specialized for the study of local conformational changes, but recent developments allow the study of a broader range of processes and they also find the application in CG schemes where they can be used as soft restraints on protein structure. At the longest length scales, grid-based methods sacrifice flexibility of the biomolecule to allow rapid computation of interaction energies of multimillion atom systems, and fast interpretation of cryo electron micrograph (cryo-EM) images.

As the individual methods mature, a key challenge over the next few years will be the closer integration of the methods. The applicability of the CG schemes will be greatly enhanced if the detailed information required can be generated from accurate studies as part of a closely integrated computational scheme. Till date only QM/MM is really used this way, but the generalization of the ideas of additive and subtractive coupling to a wider range of multiscale models offers a route to explore many other concurrent coupling possibilities.

## Acknowledgements

Work in MSPS's group is supported by grants from the BBSRC, EPSRC, MRC and the Wellcome Trust. This research was supported in part by the Intramural Research Program of the NIH, NHLBI. We thank our coworkers for their contributions to some of the work described.

## References

- Mulholland AJ: **Chemical accuracy in QM/MM calculations on enzyme-catalysed reactions.** *Chem Cent J* 2007, **1**:19.
- Schulten K, Phillips JC, Kale LV, Bhatel A: **Biomolecular modeling in the era of petascale computing.** In *Petascale Computing: Algorithms and Applications*. Edited by Bader D. Chapman & Hall/CRC Press; 2008:165-181.
- Tozzini V: **Coarse-grained models for proteins.** *Curr Opin Struct Biol* 2005, **15**:144-150.
- Ayton GS, Noid WG, Voth GA: **Multiscale modeling of biomolecular systems: in serial and in parallel.** *Curr Opin Struct Biol* 2007, **17**:192-198.
- Lu G, Kaxiras E: **An overview of multiscale simulations of materials.** In *Handbook of Theoretical and Computational Nanotechnology*, vol. X. Edited by Rieth M, Schommers W. American Scientific Publishers; 2005:1-33.
- Senn H, Thiel W: **QM/MM methods for biological systems.** *Top Curr Chem* 2007, **268**:173-290.
- Ingram GD, Cameron IT, Hargos KM: **Classification and analysis of integrating frameworks in multiscale modelling.** *Chem Eng Sci* 2004, **59**:2171-2187.
- Lin H, Truhlar DG: **QM/MM: what have we learned, where are we, and where do we go from here?** *Theor Chem Accounts: Theory Comput Model (Theor Chim Acta)* 2007, **117**:185-199.
- Svensson M, Humbel S, Froese R, Matsubara T, Sieber S, Morokuma K: **ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(*t*-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> oxidative addition.** *J Phys Chem* 1996, **100**:19357-19363.
- Sauer J, Sierka M: **Combining quantum mechanics and interatomic potential functions in ab initio studies of extended systems.** *J Comput Chem* 2000, **21**:1470-1493.
- Claeysens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schütz M, Thiel S, Thiel W, Werner H: **High-accuracy computation of reaction barriers in enzymes.** *Angew Chem Int Ed Engl* 2006, **45**:6856-6859.
- Klähn M, Braun-Sand S, Rosta E, Warshel A: **On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions.** *J Phys Chem B* 2005, **109**:15645-15650.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **CHARMM: a program for macromolecular energy, minimization, and dynamics calculations.** *J Comput Chem* 1983, **4**:187-217.
- Woodcock HL, Hodošček M, Gilbert ATB, Gill PMW, Schaefer HF, Brooks BR: **Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations.** *J Comput Chem* 2007, **28**:1485-1502.
- Woodcock HL, Hodošček M, Sherwood P, Lee YS, Schaefer HF III, Brooks BR: **Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase.** *Theor Chem Accounts: Theory Comput Model (Theor Chim Acta)* 2003, **109**:140-148.
- Henkelman G, Uberuaga BP, Jónsson H: **A climbing image nudged elastic band method for finding saddle points and minimum energy paths.** *J Chem Phys* 2000, **113**:9901-9904.
- Woodcock H, Hodošček M, Brooks B: **Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms.** *J Phys Chem A* 2007, **111**:5720-5728.
- Hu H, Yang W: **Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods.** *Ann Rev Phys Chem* 2008, **59**:573-601.
- Trajbl M, Hong GY, Warshel A: **Ab initio QM/MM simulation with proper sampling: "First principle" calculations of the free energy of the autodissociation of water in aqueous solution.** *J Phys Chem B* 2002, **106**:13333-13343.
- Woods CJ, Manby FR, Mulholland AJ: **An efficient method for the calculation of quantum mechanics/molecular mechanics free energies.** *J Chem Phys* 2008, **128**:014109-14118.
- Woodcock, HL, Zheng, A, Ghysels, A, Shao, Y, Kong, J, Brooks, BR: **Vibrational subsystem analysis: a method for probing free energies and correlations in the harmonic limit.** *Journal of Chemical Physics* (submitted for publication).
- Shimojo F, Kalia RK, Nakano A, Nomura K, Vashishta P: **Metascale molecular dynamics simulation of nano-mechano-chemistry.** *J Phys: Condens Mat* 2008, **20**:294204.

23. Csanyi G, Albaret T, Payne MC, De Vita A: **Learn on the fly: a hybrid classical and quantum-mechanical molecular dynamics simulation.** *Phys Rev Lett* 2004, **93**:175503-175504.
24. Walker RC, Crowley MF, Case DA: **The implementation of a fast and accurate QM/MM potential method in Amber.** *J Comput Chem* 2008, **29**:1019-1031.
25. Heyden A, Lin H, Truhlar D: **Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations.** *J Phys Chem B* 2007, **111**:2231-2241.
26. Sinicropi A, Basosi R, Olivucci M: **Recent applications of a QM/MM scheme at the CASPT2//CASSCF/AMBER (or CHARMM) level of theory in photochemistry and photobiology.** *J Phys: Conf Ser* 2008, **101**:012001.
27. Sulpizi M, Raugei S, VandeVondele J, Carloni P, Sprik M: **Calculation of redox properties: understanding short- and long-range effects in rubredoxin.** *J Phys Chem B* 2007, **111**:3969-3976.
28. Cascella M, Magistrato A, Tavernelli I, Carloni P, Rothlisberger U: **Role of protein frame and solvent for the redox properties of azurin from *Pseudomonas aeruginosa*.** *Proc Natl Acad Sci U S A* 2006, **103**:19641-19646.
29. Zeng X, Hu H, Hu X, Cohen AJ, Yang W: **Ab initio quantum mechanical/molecular mechanical simulation of electron transfer process: fractional electron approach.** *J Chem Phys* 2008, **128**:124510-124510-10.
30. Bahar I, Rader AJ: **Coarse-grained normal mode analysis in structural biology.** *Curr Opin Struct Biol* 2005, **15**:586-592.
31. Yang L, Song G, Jernigan RL: **How well can we understand large-scale protein motions using normal modes of elastic network models?** *Biophys J* 2007, **93**:920-929.
32. Zheng W, Brooks BR, Thirumalai D: **Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations.** *Biophys J* 2007, **93**:2289-2299.
33. Schröder GF, Brunger AT, Levitt M: **Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution.** *Structure* 2007, **15**:1630-1641.
34. Zheng W, Brooks BR, Hummer G: **Protein conformational transitions explored by mixed elastic network models.** *Proteins* 2007, **69**:43-57.
35. Franklin J, Koehl P, Doniach S, Delarue M: **MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape.** *Nucleic Acids Res* 2007, **35**:W477-W482.
36. Chu J, Voth GA: **Coarse-grained free energy functions for studying protein conformational changes: a double-well network model.** *Biophys J* 2007, **93**:3860-3871.
37. Zheng W, Liao J, Brooks BR, Doniach S: **Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model.** *Proteins* 2007, **67**:886-896.
38. Yang L, Song G, Carriquiry A, Jernigan RL: **Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes.** *Structure* 2008, **16**:321-330.
39. Zheng W: **A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations.** *Biophys J* 2008, **94**:3853-3857.
40. Best RB, Chen Y, Hummer G: **Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor.** *Structure* 2005, **13**:1755-1763.
41. Lu Q, Wang J: **Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles.** *J Am Chem Soc* 2008, **130**:4772-4783.
42. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D: **A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.** *Nature* 2007, **450**:913-916.
43. Gohlke H, Thorpe MF: **A natural coarse graining for simulating large biomolecular motion.** *Biophys J* 2006, **91**:2115-2120.
44. Freddolino PL, Arkipov AS, Larson SB, McPherson A, Schulten K: **Molecular dynamics simulations of the complete satellite tobacco mosaic virus.** *Structure* 2006, **14**:437-449.
45. Nielsen SO, Lopez CF, Srinivas G, Klein ML: **Coarse grain models and the computer simulation of soft materials.** *J Phys Condens Matter* 2004, **16**:R481-R512.
46. Clementi C: **Coarse-grained models of protein folding: toy models or predictive tools?** *Curr Opin Struct Biol* 2008, **18**:10-15.
47. Moritsugu K, Smith JC: **Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian.** *Biophys J* 2007, **93**:3460-3469.
48. Zhou J, Thorpe IF, Izvekov S, Voth GA: **Coarse-grained peptide modeling using a systematic multiscale approach.** *Biophys J* 2007, **92**:4289-4303.
49. Marrink SJ, de Vries AH, Mark AE: **Coarse grained model for semiquantitative lipid simulations.** *J Phys Chem B* 2004, **108**:750-760.
50. Shelley JC, Shelley MY, Reeder RC, Bandyopadhyay S, Klein ML: **A coarse grain model for phospholipid simulations.** *J Phys Chem B* 2001, **105**:4464-4470.
51. Whitehead L, Edge CM, Essex JW: **Molecular dynamics simulation of the hydrocarbon region of a biomembrane using a reduced representation model.** *J Comput Chem* 2001, **22**:1622-1633.
52. Lopez CF, Nielsen SO, Moore PB, Klein ML: **Understanding nature's design for a nanosyringe.** *Proc Natl Acad Sci* 2004, **101**:4431-4434.
53. Nielsen SO, Ensing B, Ortiz V, Moore PB, Klein ML: **Lipid bilayer perturbations around a transmembrane nanotube: a coarse grain molecular dynamics study.** *Biophys J* 2005, **88**:3822-3828.
54. Venturoli M, Smit B, Sperotto MM: **Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins.** *Biophys J* 2005, **88**:1778-1798.
55. Marrink S, Risselada H, Yefimov S, Tieleman D, Vries A: **The MARTINI force field: coarse grained model for biomolecular simulations.** *J Phys Chem B* 2007, **111**:7812-7824.
56. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S: **The MARTINI coarse-grained force field: extension to proteins.** *J Chem Theory Comput* 2008, **4**:819-834.
57. Bond PJ, Sansom MSP: **Insertion and assembly of membrane proteins via simulation.** *J Am Chem Soc* 2006, **128**:2697-2704.
58. Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP: **Coarse-grained molecular dynamics simulations of membrane proteins and peptides.** *J Struct Biol* 2007, **157**:593-605.
59. Sansom MSP, Scott KA, Bond PJ: **Coarse-grained simulation: a high-throughput computational approach to membrane proteins.** *Biochem Soc Trans* 2008, **36**:27-32.
60. Scott KA, Bond PJ, Ivetac A, Chetwynd AP, Khalid S, Sansom MS: **Coarse-grained MD simulations of membrane protein-bilayer self-assembly.** *Structure* 2008, **16**:621-630.
61. Periole X, Huber T, Marrink S, Sakmar TP: **G protein-coupled receptors self-assemble in dynamics simulations of model bilayers.** *J Am Chem Soc* 2007, **129**:10126-10132.
62. Wee CL, Bemporad D, Sands ZA, Gavaghan D, Sansom MSP: **SGTx1, a Kv channel gating-modifier toxin, binds to the interfacial region of lipid bilayers.** *Biophys J* 2007, **92**:L07-L09.
63. Wee C, Balali-Mood K, Gavaghan D, Sansom MS: **The interaction of phospholipase A2 with a phospholipid bilayer: coarse-grained molecular dynamics simulations.** *Biophys J* 2008, **95**:107-123190.

64. Bond PJ, Sansom MSP: **Bilayer deformation by the Kv channel voltage sensor domain revealed by self-assembly simulations.** *Proc Natl Acad Sci* 2007, **104**:2631-2636.
65. Sands ZA, Sansom MSP: **How does a voltage sensor interact with a lipid bilayer? Simulations of a potassium channel domain.** *Structure* 2007, **15**:235-244.
66. Freitas JA, Tobias DJ, White SH: **A voltage-sensor water pore.** *Biophys J* 2006, **91**:L90-L92.
67. Yefimov S, van der Giessen E, Onck PR, Marrink SJ: **Mechanosensitive membrane channels in action.** *Biophys J* 2008, **94**:2994-3002.
68. Treptow W, Marrink S, Tarek M: **Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations.** *J Phys Chem B* 2008, **112**:3277-3282.
69. Gumbart J, Schulten K: **Structural determinants of lateral gate opening in the protein translocon.** *Biochemistry* 2007, **46**:11147-11157.
70. Gorfe AA, Chang CA, Ivanov I, McCammon JA: **Dynamics of the acetylcholinesterase tetramer.** *Biophys J* 2008, **94**:1144-1154.
71. Neri M, Baaden M, Carnevale V, Anselmi C, Maritan A, Carloni P: **Microseconds dynamics simulations of the outer-membrane protease T.** *Biophys J* 2008, **94**:71-78.
72. Shih AY, Arkhipov A, Freddolino PL, Schulten K: **Coarse grained protein-lipid model with application to lipoprotein particles.** *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys* 2006, **110**:3674-3684.
73. Voltz K, Trylska J, Tozzini V, Kurkal-Siebert V, Langowski J, Smith J: **Coarse-grained force field for the nucleosome from self-consistent multiscaling.** *J Comput Chem* 2008, **29**:1429-1439.
74. Reynwar BJ, Illya G, Harmandaris VA, Müller MM, Kremer K, Deserno M: **Aggregation and vesiculation of membrane proteins by curvature-mediated interactions.** *Nature* 2007, **447**:461-464.
75. Arkhipov A, Yin Y, Schulten K: **Four-scale description of membrane sculpting by BAR domains.** *Biophys J* 2008, **95**: (September 1, 2008 issue).
76. Santoro MM, Bolen DW: **Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants.** *Biochemistry* 1988, **27**:8063-8068.
77. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D: **Effects of denaturants and osmolytes on proteins are accurately predicted using the molecular transfer model.** *Proceedings of the National Academy of Science* (in press).
78. Wu X, Milne JLS, Borgnia MJ, Rostapshov AV, Subramaniam S, Brooks BR: **A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy.** *J Struct Biol* 2003, **141**:63-76.
79. Milne JLS, Wu X, Borgnia MJ, Lengyel JS, Brooks BR, Shi D, Perham RN, Subramaniam S: **Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy.** *J Biol Chem* 2006, **281**:4364-4370.
80. Lee D, Park SJ, Jeong W, Sung HJ, Oho T, Wu X, Rhee SG, Gruschus JM: **Mutagenesis and modeling of the peroxiredoxin (Prx) complex with the NMR structure of ATP-bound human sulfiredoxin implicate aspartate 187 of Prx I as the catalytic residue in ATP hydrolysis.** *Biochemistry* 2006, **45**:15301-15309.