ACCOUNTS
of chemical research

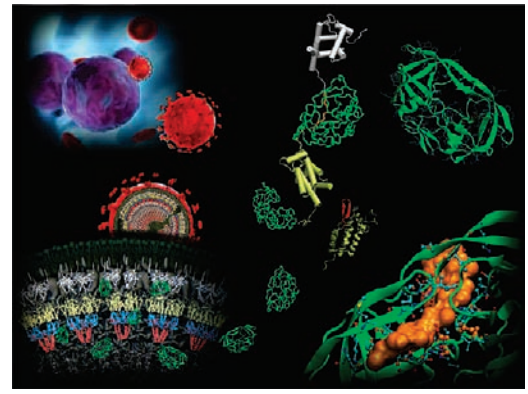
Multiscale Modeling of Proteins

VALENTINA TOZZINI*

NEST CNR-INFM, and Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

RECEIVED ON MAY 8, 2009

CONSPECTUS



The activity within a living cell is based on a complex network of interactions among biomolecules, exchanging information and energy through biochemical processes. These events occur on different scales, from the nano- to the macroscale, spanning about 10 orders of magnitude in the space domain and 15 orders of magnitude in the time domain. Consequently, many different modeling techniques, each proper for a particular time or space scale, are commonly used. In addition, a single process often spans more than a single time or space scale. Thus, the necessity arises for combining the modeling techniques in multiscale approaches.

In this Account, I first review the different modeling methods for bio-systems, from quantum mechanics to the coarse-grained and continuum-like descriptions, passing through the atomistic force field simulations. Special attention is devoted to their combination in different possible multiscale approaches and to the questions and problems related to their coherent matching in the space and time domains. These aspects are often considered secondary, but in fact, they have primary relevance when the aim is the coherent and complete description of bioprocesses.

Subsequently, applications are illustrated by means of two paradigmatic examples: (i) the green fluorescent protein (GFP) family and (ii) the proteins involved in the human immunodeficency virus (HIV) replication cycle. The GFPs are currently one of the most frequently used markers for monitoring protein trafficking within living cells; nanobiotechnology and cell biology strongly rely on their use in fluorescence microscopy techniques. A detailed knowledge of the actions of the virus-specific enzymes of HIV (specifically HIV protease and integrase) is necessary to study novel therapeutic strategies against this disease. Thus, the insight accumulated over years of intense study is an excellent framework for this Account.

The foremost relevance of these two biomolecular systems was recently confirmed by the assignment of two of the Nobel prizes in 2008: in chemistry for the discovery of GFP and in medicine for the discovery of HIV. Accordingly, these proteins were studied with essentially all of the available modeling techniques, making them ideal examples for studying the details of multiscale approaches in protein modeling.

Introduction

Biochemical processes occur on different scales of length and time[1] ranging from a few angstroms, the size of the active site of proteins, where the ultrafast triggering steps usually take place, up to the level of the cells and organs, where their macroscopic effects are detectable by the naked eye. Intermediate steps are the structural rearrangement of biomolecules (approximately nanometer and 10−100 ns), their aggregation/separation and folding/unfolding (∼10 nm to micrometer and greater than microsecond) and internal cell diffusion and dynamics (micrometers to millimeters and milliseconds to hours). This inherent hierarchical organization is responsible for the complexity of living matter: A single process involves a multiscale cascade of events whose description requires the combination of different methodologies in so-called multiscale approaches.[2,3]

Published on the Web 09/28/2009 www.pubs.acs.org/acr
10.1021/ar9001476

This Account is composed of two parts. The first reports a review of the modeling techniques typically used at each "scale" and of their multiscale combination, either implemented "in parallel", that is, treating different parts of the system with different resolutions in the same simulation,[4] or "in series", that is, performing subsequent simulations of the system at different resolutions and exchanging information among them[5] (see Figure 1). Rising up in the scale implies reducing the resolution of the system's description by eliminating a part of the internal degrees of freedom (DOF) and accounting for them implicitly. The strategies to achieve this goal are described especially focusing on the issue of coherently matching between scales in the space and time domains.

In the second part, applications are illustrated using two paradigmatic cases: the family of the green fluorescent proteins (GFPs) and the replication cycle of human immunodeficiency virus (HIV). Soon after its discovery,[6] the GFP from the jellifish *Aequorea victoria* and, more generally, all the members of the intrinsically fluorescent proteins (FP) family have proven a revolution in the field of molecular biology. They possess intrinsic fluorescence due to self-catalytic built-in chromophores and can be genetically fused to other proteins, and the protein−FP chimeric construct can be expressed in living cells. This allows one to follow the protein's trafficking by means of confocal fluorescence microscopy, making the FPs the most widely used probes or markers to monitor gene expression, protein localization, diffusion, and interactions *in vivo*.[7,8]

HIV replication relies on the action of some virus-specific enzymes, such as the integrase (HIV-1in), which integrates the viral genome into that of the infected cell, and the protease (HIV-1pr), which cleaves the Gag and Pol viral polyprotein into functional parts necessary for building and maturation of new virions.[9] Given their fundamental roles, those enzymes are targets for developing inhibitory drugs.[10] However, the virus' ability to rapidly mutate and reduce the affinity of drugs makes the therapeutic efficacy of the drugs limited in time. For this reason, efforts are currently directed to seek novel therapeutic strategies going beyond simple antagonism. This requires detailed knowledge of each step of the activity of those enzymes and of the subsequent events.

Methods

Quantum Mechanics Methods. At any resolution, the quality of a model depends on the accuracy with which the two following issues are addressed: the description of the interactions and the sampling of the configurations of the system. In this respect, there are a few concepts that iteratively occur. One is the potential energy surface (PES) or its finite temperature extension, the free energy surface (FES), that is, a multidimensional surface returning the potential (free) energy as a function of a set of internal coordinates. The method used to evaluate the P(F)ES strictly depends on the resolution level. For small molecules (up to a few tens of atoms), both the nuclear and electronic DOF can and must be explicitly treated in order to describe the chemical reactions. The concept of the PES is related to the Born−Oppenheimer approximation,[11] assuming that the much faster electrons adiabatically adjust their motion to that of the atomic nuclei. Thus, at any time, the Schrödinger equation for the electron system is to be solved in the external field generated by the atomic nuclei considered as frozen, and one is left with a nuclear-configuration-dependent set of energy eigenvalues $E_i(\{R_I\})$ that define the PESs of the ground and excited states (see Figure 1a). In turn, the PESs are effective electronic-structure-dependent potential energy functions that determine the dynamics of the nuclei.

The different methods used to solve the Schrödinger equation, called quantum mechanics (QM) approaches, basically differ by the way the electron−electron interactions are treated. The electron correlation can be accurately added as a perturbation of the exchange-only Hartree−Fock (HF) scheme at the expense of a large computational cost or *via* less expensive (and less predictive) semiempirical Hamiltonians.[12] Alternatively, in density functional theory (DFT) and, more specifically, in the Kohn−Sham scheme,[11] the many-electron problem is reduced to a single-electron Schrödinger problem in a self-consistent exchange−correlation potential depending on the electron density. DFT changed the way of approaching the QM calculations: its accuracy and predictive power are comparable to those of other *ab initio* methods but much cheaper computationally. Thus, DFT is conveniently used for molecular structure optimization or even for dynamic exploration of the PES, that is, *ab initio* molecular dynamics (MD). *Ab initio* MD underwent rapid development especially when Car and Parrinello (CP)[11] introduced their Lagrangian scheme to solve the coupled nuclei−electrons dynamical equation. The CP approach is nowadays the most used method to treat the *ab initio* MD of biosystems.[11] In addition, excited-state calculations are possible with the time-dependent extension of DFT (TDDFT[13]). Although TDDFT is known to suffer from large errors in the excitation energies, it is often used in biosystems thanks to its extremely low cost with respect to other excited-state methods.

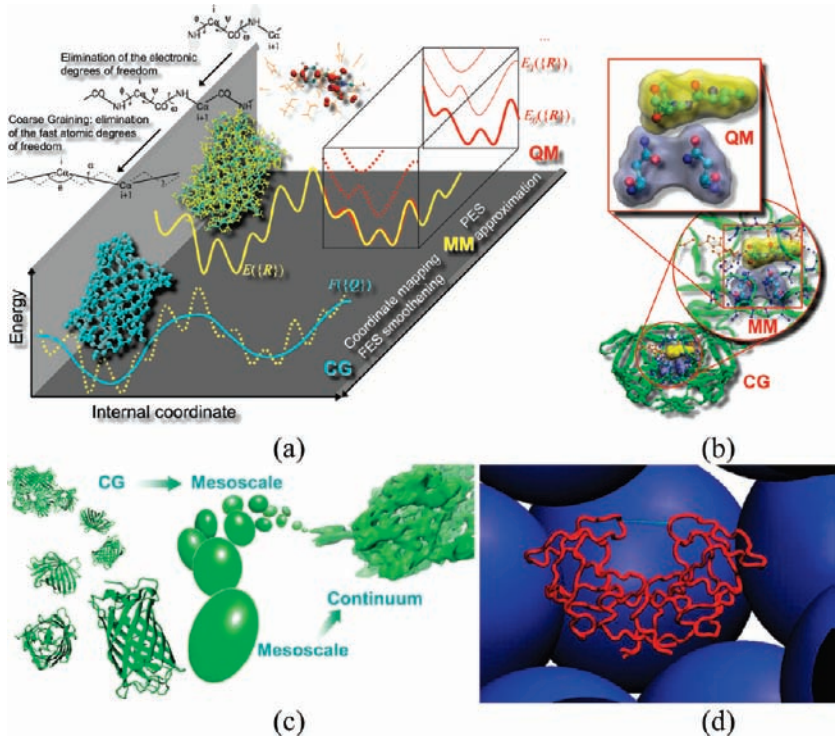


FIGURE 1. Schematic representations of the multiscale approaches. (a) In the QM methods (red, the atomic coordinates {*R*} and the electronic orbitals of the system are explicitly treated, and the PESs are evaluated quantum mechanically. In the MM (yellow) methods, the internal DOF are the coordinates of all the atoms {*R*}; the PES of an electronic state is approximated with empirical functions (the FF). In the CG models, groups of atoms are condensed in single interacting sites (whose coordinates are {*Q*}), and the PES is simplified (cyan). The GFP is represented at the three levels within a serial multiscale approach. (b) The same three methods applied in parallel to HIV-1pr: the two active aspartates and the residues of the substrate interacting with them are treated at the QM level; the active site residues (residues surrounding the aspartates and the substrate) are treated at the MM level, and all the rest are treated at the CG level. (c) Pictorial representation of the transition CG to mesoscale to continuum representations in GFPs, within a "serial" multiscale approach; (d) "parallel" multiscale model of HIV-1pr (CG representation) within the crowded cell environment: the crowding molecules are represented at the mesoscale level as spheres.

Force Field Based All-Atom Simulations and QM/MM Hybrid Methods. In order to enlarge the size of the addressable system, one must abandon the QM approach in favor of empirical descriptions of the interatomic forces, also called the molecular mechanics (MM) scheme. The elimination of the electronic DOF enables enormous savings of computational cost, allowing the simulation of fully hydrated proteins. In passing from QM to MM, coherence is naturally achieved by means of the PES: in MM, the chosen PES (usually the ground-state one) is approximated with a sum of analytical terms describing the chemical bonds and the van der Waals and the electrostatics interactions, globally called the force field (FF). The FF parameters are fitted on the PES evaluated with QM approaches in small molecules representing parts of the system (see Figure 1a). This procedure is often complemented by including structural and thermodynamic experimental data. The most commonly used FFs for proteins (Amber,[14] CHARMM,[15] Gromos[16]) differ by the relative amount of *ab initio* and experimental data used.

The hybrid ("parallel") QM/MM descriptions[17] exploit the fact that the chemical reaction usually concerns a small part of the system (the active site), allowing one to confine the QM treatment to it and describe the rest within an MM scheme (see Figure 1b). Termination of the electron density and matching the forces at the QM/MM boundary are achieved within a "boundary region" where the transition between the two representations occurs smoothly. QM/MM approaches enlarge the size of the system that can be simulated preserving accuracy, but they do not bear an analogous advantage in the time domain: the time length of a MD simulation is limited by the computational cost of the QM simulation, also related to the amplitude of the time step, Δt, for the integration of the equations of motion. Thus, while in the classical MD simulation, typically $\Delta t \approx 1$ fs, allowing one to reach 10−100 ns run lengths, in QM or QM/MM simulations within the CP approach, $\Delta t \approx 0.1$ fs and maximum run lengths of 100 ps to 1 ns can be reached. Typically, the chemical reactions, proton transfers or similarly fast events, are accessible,

TABLE 1. A Nonexhaustive List of CG Models for Proteins with Their Main Features

class	model	CG level	features	applications
structure based	network models elastic[24a] inelastic[24b] plastic[24c] and bimodal[24d] multiple resolution[24e] amino-acid specific[24f] phys chem properties[24g] Gō models standard[24h] solvation potential[24i] amino acid specific + phys chem properties[24j]	one bead to all atom	bias toward a single structure structural and topologic accuracy low transferability	principal modes anharmonic motions biased dynamics multiscale modeling minimally frustrated folding desolvation intermediates effects of mutations free energy landscape
empirical, knowledge-based	Head-Gordon[24k] Thirumalai[24l] Mukherjee[24m]	one bead two beads two beads	larger transferability	folding
statistical	Martini[24n] Unres[24o]	two to five beads three beads	thermodynamic consistency	aggregation slow transitions
free energy matching +local bias	Bahar[24p] Tozzini−McCammon[24q]	two beads one bead	structural accuracy	general dynamics
force matching	Voth[24r]	two to four beads	statistical mechanical consistency	general dynamics

but the subsequent structural rearrangement of the protein, which requires a fully classical MD simulation, is not.

Coarse-Grained Descriptions. Although microsecond all-atom simulations of entire proteins were performed on dedicated heavily parallel systems,[18] the simulation of macromolecular aggregates on the microsecond time scale requires further simplifications. Following the reductionist approach, the idea of coarse-graining (CG) the system, that is, condensing groups of atoms into single interacting sites, naturally arises. Very different levels of CG have been considered,[19] ranging from the "united atoms" approach, where only the hydrogen atoms are eliminated, to mesoscale models with interacting sites representing whole proteins. A critical issue is the combination of accuracy and predictive power in the FF, especially for the "minimalist models" such as the one bead per amino acid (OB-CG) model.[20] One can generally recognize three different parametrization methodologies: (i) structure based, (ii) statistical/empirical potentials, or (iii) force (or free energy) matching. In method i, the equilibrium values of the FF terms are based on a single (experimental) structure that is by construction an equilibrium one. The few free parameters are fitted on experimental data (amplitude of thermal fluctuations, specific heat of a transition, relative equilibrium concentrations, etc). In this class, characterized by low predictive power, a number of popular models fall, such as the network and the Gō models, used in the analysis of slow motions and folding. Transferability and predictability is improved in class ii, where a statistical set of structures is used to fit FF parameter with procedures based on the Boltzmann inversion,[21] often integrated including empirical or experimental data. This method generates models that are thermodynamically consistent with the data set (see Figure 1a), the diversity and nature of which also determine the transferability of the FF. Thus method ii is preferable when the CG simulation is performed within serial approaches or alone. Conversely, statistical mechanical consistency is inherent to method iii, which consists of fitting the CG forces on those obtained from sets of trajectories from all-atom MD simulations.[22,23] Thus method iii is more convenient in multiscale parallel approaches (Figure 1b). A nonexhaustive list of CG models classified according to their features is reported in Table 1.

The CG procedure cuts out the high frequencies of the system and smoothes the FES (see Figure 1a). Consequently, a further gain in computational cost stems from the possibility of choosing larger Δt and from the intrinsic acceleration of the system dynamics. However, the latter is a fictitious effect that must be corrected if one wants to have a realistic description and coherency also in the time domain. This can be done by using stochastic dynamics: the damping and random force in the generalized Langevin approach can reintroduce the dynamics of the eliminated DOF.[25] Finally, mesoscale models representing entire proteins or domains as spheres or ellipsoids, or even rigid bodies delimited by the solvent-accessible surfaces, are often considered even within multiscale approaches (see Figure 1c,d). Surface-based approaches are typically used for instance in docking and protein recognition methods.[26] A review of CG models can be found in ref 27.

The Continuum Representation. The continuum representations can be considered as an extreme case of coarse graining: the system is represented as a medium whose dynamics is described by a single functional variable depending on the space coordinates. Specific relevant cases are (i) the continuum representation of the solvent, (ii) the surface model

of the cell membrane, and (iii) the diffusional dynamics of proteins within the cell. In case i, the functional variable is the local ion or other cosolvent concentration, which can be treated within the Poisson−Boltzmann theory[28] or with generalized Born methods.[29] In case ii, the variable is the local membrane deformation, which can be treated within elastic or more sophisticated theories,[30] and in case iii, the variable is the local protein density, which can be treated, for instance, within a Liouville-like transport equation approach, describing the flux of a fluid (i.e., the protein ensemble) within given boundary conditions (e.g., the cell compartments)[31] (see Figure 1c). The transport equations describing the time evolution of the particle density can be rewritten within the time-dependent classical density functional formalism,[32] which offers a way to bridge the continuum with the particle representation: the information on the interactions among particles could be included within the classical energy density functional using data from the CG stochastic dynamics simulations. Curiously, the modeling approaches at two extremes of the scale (the QM and the continuum) can be formulated within the same framework, that is, DFT, although the density variables in the two cases assume very different meaning (i.e., the electron density and the protein density).

Applications

Colors and Shades of FPs. The self-catalytic chromophore building reactions slightly differ in the different FPs, leading to variously extended delocalized electron systems (see Figure 2[33]). The chromophores are relatively small (∼5 Å), and some of their optical properties can be observed even when they are isolated from the protein matrix.[34] Thus, they are ideal cases to be studied with QM approaches. The excitation energy of the neutral and anionic state of the GFP chromophore was evaluated with basically all available QM methods (see the Supporting Information). A range of over 2 eV is spanned by the different calculations. Compared with the *ab initio* and the semiempirical methods, the TDDFT turns out to be a good compromise between accuracy, predictive power, and computational cost. This circumstance allowed systematic studies on the whole series of the FP chromophores[33] in different protonation states, revealing an inverse proportionality relationship between the excitation energy and the charge displacement upon excitation (Figure 2). This roughly explains the variety of colors of the FPs: red-shifted chromophores are obtained from the green one by the progressive extensions of the electronic delocalization, while blue-shifted chromophores were obtained by artificially mutating the central tyrosine residue to effectively shorten the electron delocalization. This

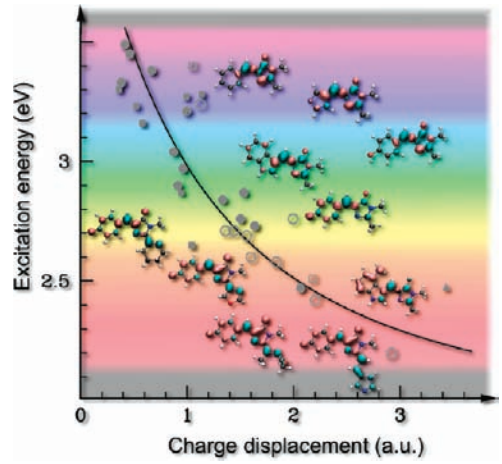


FIGURE 2. The relationship between color and chromophore structure. As the electronic delocalization increases, the charge displacement upon excitation increases and the excitation energy decreases. The different chromophore structures are reported with a surface representation of the distribution charge displacement and approximately placed in the color area corresponding to their emission wavelength.

study additionally allowed an evaluation of the TDDFT performances in this molecule's family, which performed particularly well for the blue chromophores (see Figure 3). A similar accuracy was found in the calculation of the two-photon absorption cross-section.[35]

The theoretical studies on the chromophore have helped to assign the two main absorption bands of the GFP (∼400 and ∼480 nm) emitting blue and green light to its neutral and anionic forms, respectively. Although the emitting color depends in first approximation on the chemical structure of the chromophore, this is not enough to explain the variety of available shades of GFP mutants. A systematic DFT theoretical study of the active site has clarified this complex landscape,[36] showing that the excitation wavelength is shifted by specific interactions occurring between the chromophore and the active site residues (Figure 4), also explaining an experimentally observed linear relationship between the absorption energy and a fingerprint Raman band. The CP MD approach was used to study the formation and properties of additional elusive states possibly forming by proton transfer from A or B. These are the intermediate I forming during the normal fluorescence cycle of the GFP (Figure 4[36]) and the hypothetical zwitterionic state (see Supporting Information), a possible candidate for the fast on−off blinking behavior of the fluorescence.

Dark State Dynamics in GFP and Cleavage Reaction of HIV-1pr. The (G)FPs enter a long-lived dark state after a certain number of fluorescence cycles. Some mutants, such as the E^2GFP, however, can be reconverted to the normal bright

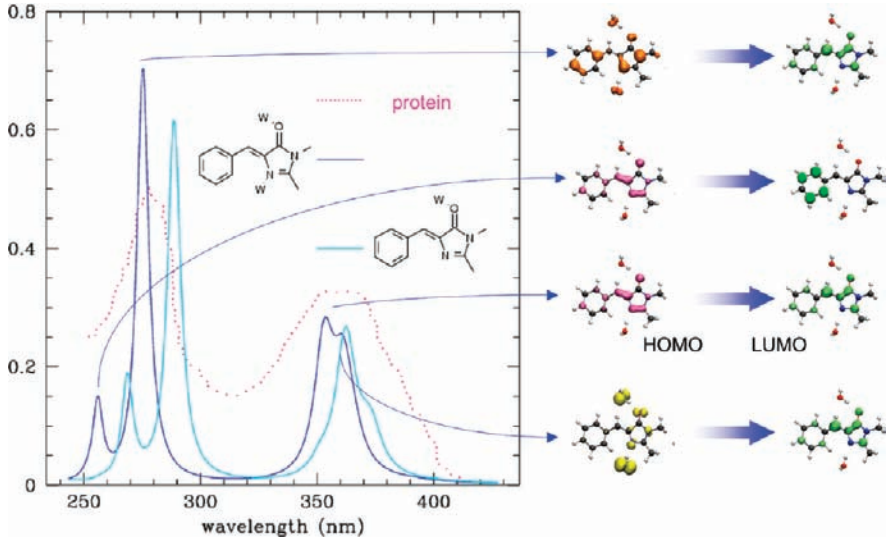


FIGURE 3. The TDDFT excitation spectrum of the BFPF chromophore in an environment mimicking the protein (blue and cyan) compared with the experimental protein spectrum (magenta). Different excitations seem to contribute to the large line shape, in addition to the vibronic structure. The chromophore excitation also appears to participate in the high energy peak, which was generally attributed to the protein aromatic amino acids.

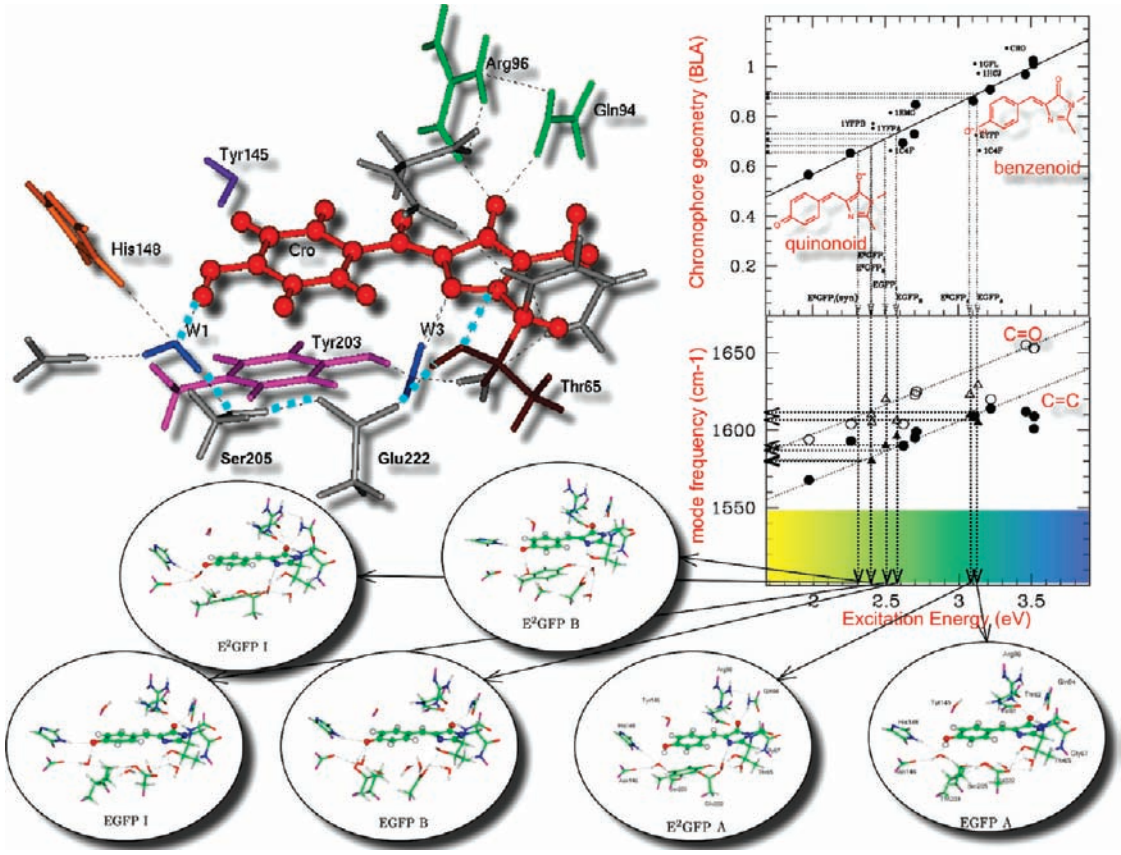


FIGURE 4. The chromophore environment and the relationship between shades, structure, and vibrational properties. The main panel represents the chromophore of the GFP in its neutral form and the residues in contact with it. The plots represent the linear relationship between the excitation energy and the chromophore geometry (represented by the bond-length alternation, upper graph) and the frequency of the C═C fingerprint Raman band (bottom graph). The emission color is reported as a colored band in the bottom graph. The arrows show the relationship between structure, excitation energy and vibrational frequency for the main states of two different mutants, EGFP and E^2GFP, whose structures are reported in the ovals.

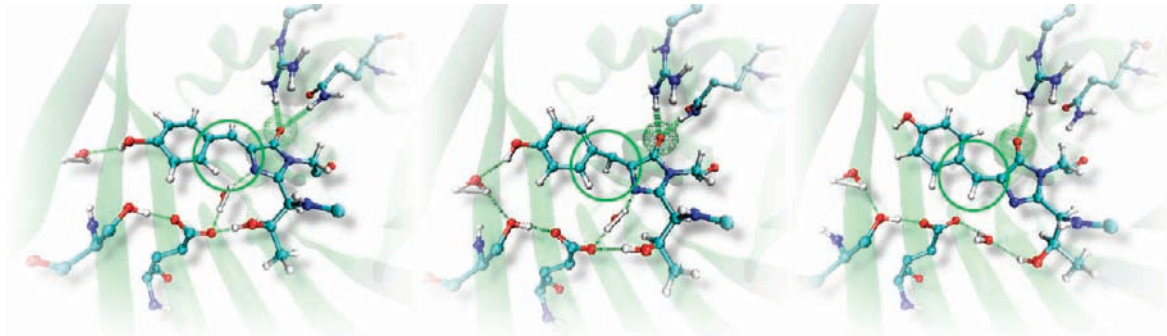


FIGURE 5. Three snapshots from the classical MD simulation of fully hydrated GFP, showing the cis−trans isomerization of the chromophore. The isomerization can take place within the small active site free volume because it occurs via the "hula-twist" rotation of the rings. The hydrogen-bonds are highlighted in green.

state by illumination in the UV (∼350 nm) where their dark state absorption spectrum displays an additional band (photoswitching). This phenomenology lead to the hypothesis of a cis−trans isomerization of the chromophore as the main mechanism for the photoswitching[37] (see also the Supporting Information). Such a transition is beyond the capability of the CP MD simulations, since it involves extended structural readjustments. On the other hand, a classical MD simulation requires a feedback with the QM calculations for the FF optimization: the cis−trans transition occurs through the excited state, whose parametrization required data from QM calculations.[38] The simulation showed that the cis−trans isomerization occurs despite the small volume available in the active site (see Figure 5). Furthermore, the structural changes of the active site induced by the transition can explain the observed blue shift of the absorption, giving strength to the cis−trans hypothesis. This is an example of "in series" combination of QM and MM calculations: the structure extracted from the classical MD simulation[39] was used as the basis for the QM calculations; in turn the QM calculations were used to optimize the FF parameters and gave the prediction of the absorption spectrum. A similar serial QM/MM approach was used for the proton transfer between the chromophore and the protein with QM calculations.[40] QM/MM parallel calculations of the absorption spectra were performed using QM regions of different sizes[41,42] and combinations of serial and parallel QM/MM approaches.[43]

The theoretical studies on HIV-1pr usually include the whole protein because the active moiety, the aspartic acid dyad that operates the hydrolysis reaction, cannot be separated from the protein matrix (see Figure 1c, blue in the squared inset). In the QM/MM studies of HIV-1pr, the aspartic dyad and model substrates are usually included in the QM representation.[44,45] These studies were successful in clarifying some fundamental aspects of the substrate cleaving reaction. However, the bottleneck steps of the process are the substrate recognition and binding. The large number of available classical MD simulations of HIV-1pr were mainly aimed at analyzing the differences in the dynamical behavior of the different mutants that the virus produces to survive the anti-AIDS drugs.[46] Some of these have shown that the flaps, two hairpin-like structures covering the active site, curl their tip on the nanosecond scale.[47] However, they generally fail in describing the opening of the flaps, which is a preliminary step for substrate binding. Only a few of them have succeeded in showing isolated or partial opening events, mostly by using "tricks" such as solvent with reduced viscosity to accelerate the dynamics.[48]

HIV-1 Protease and Integrase Dynamics. NMR data indicate that rearrangements in the structure of the HIV-1pr compatible with flap opening occur on the time scale of microseconds to milliseconds. Thus, the reason for the failure of the all-atom MD simulations in describing this event is its extreme rarity on the time scale that they can address. Conversely, a stochastic dynamic simulation with a OB-CG model of the HIV-1pr was able to reach that time scale, to model the complete flap opening dynamics, and to study a large statistic of events.[49,50] This model survived several stages of validation: it was able to correctly reproduce the effects of mutations on the opening tendency[51] and the location of an allosteric site for opening inhibition, to predict the open HIV-1pr structure subsequently experimentally confirmed (see Figure 6), and to reproduce the dependence of the flap opening frequency on the solvent viscosity.

The model was subsequently used to simulate the binding of a peptide modeling a part of the substrate and the product release after cleavage[52] (Figure 7). Conversely, the study of the binding of antiviral drugs required a serial CG/MM approach[53] (Figure 8). The slower phases of the substrate approaching and flap opening were simulated with the CG model. Subsequently, starting from selected snapshots of the almost formed complex, the final docking was simulated with

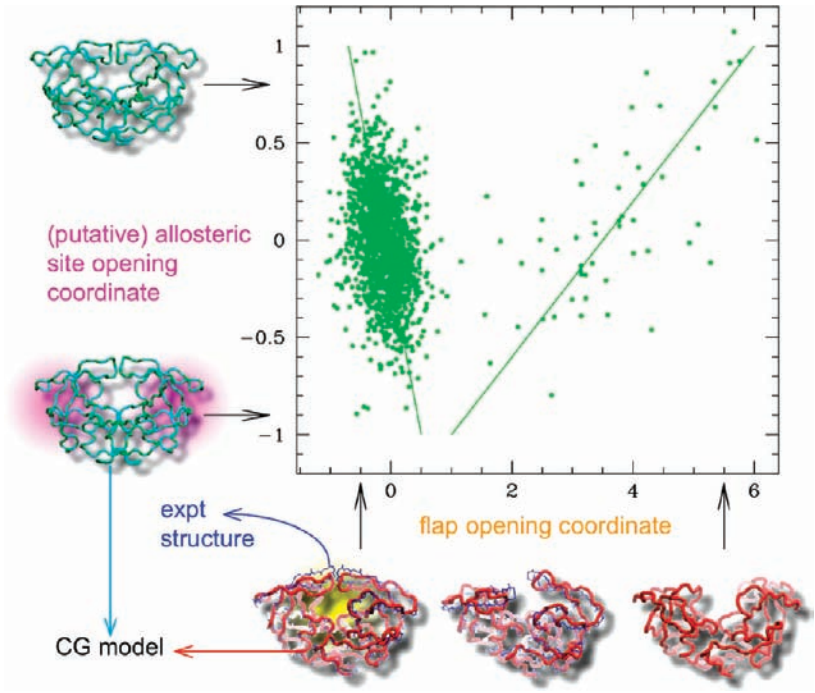


FIGURE 6. Red structures at the bottom provide a sequence showing the opening mode of the active site of HIV-1pr (highlighted in yellow). Superimposed to the closed and semiopen structures, in blue, are two experimental structures. Cyan structures on the left show two extremal conformations of the mode corresponding to the opening of a putative allosteric site (highlighted in magenta). The graph gives the correlation plot between the two modes, indicating that a drug constraining the allosteric site would probably inhibit the flap opening.

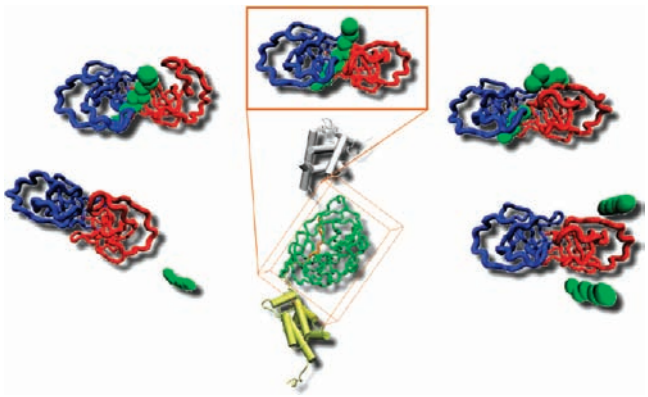


FIGURE 7. Five snapshots from the CG simulation of the substrate binding, cleavage, and release of HIV-1pr (in blue and red, substrate in green). From left to right, approaching, docking, flap closing and substrate adjustment, cleavage, and product release. The HIV-1pr bound to the Gag is also shown. The pr is in green, the Gag in white and yellow, and the linker peptide to which the pr binds in orange.

all-atom MD. This was necessary due to the more chemically various interactions between HIV-1pr and nonstandard ligands, such as the inhibitor drug. A parallel MM/CG model describing the active site at the all-atom level was also developed;[54] however the simulation length is limited to the nanosecond scale by the MM part.

A multiscale approach was also used to study anti-AIDS strategies targeting the HIV-1in. The acetylation of its specific lysine residues is thought to be involved in the viral DNA integration process.[55] Consequently, the inhibition of this process is a possible therapeutic strategy. A serial MM/CG/mesoscale approach was used to model the interaction of HIV-1in with p300 and GCN5 acetyl transferases (HATs):[56] a combination of standard docking methodologies with all-atom simulation was used to build a reliable complex of HIV-1in with HATs. These were subsequently used within a CG approach to study the dynamical properties of the complexes and to indicate possible inhibition strategies.

FPs and HIVpr in the Cell Environment: Facts and Perspectives. The cytoplasmic environment is crowded with macromolecular agents that are known to affect the dynamics and kinetics of many processes. This effect was studied in HIV-1pr with a parallel CG/mesoscale model where the protein was represented with the OB model and the crowding agents as large soft spheres (Figure 1d). It was shown that the flap opening frequency and the stability decreases at high concentrations of crowding agents due to steric effects.[57]

A similar approach applied to the FPs or a FP−protein construct would aid the interpretation of the fluorescence microscopy data. In fact the diffusion coefficient of those systems measured with special techniques such as fluorescent recovery after photobleaching (FRAP) depends on their dynamics and on the presence of crowding agents. Another application could be to the FRET technique (fluorescent resonant energy

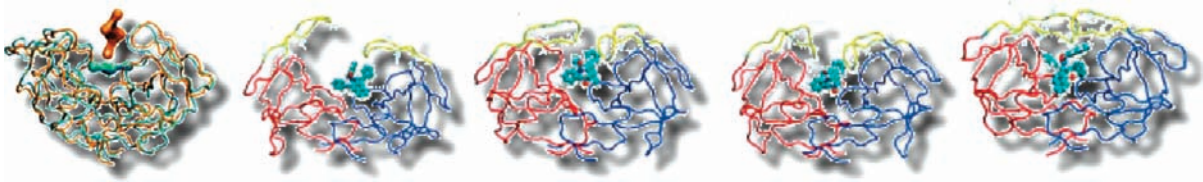


FIGURE 8. Illustration of the serial CG/MM simulation of the binding of an inhibitor drug: on the left, in orange and cyan two snapshots from the CG simulation of the drug approaching; on the right (HIV-1pr in red-green-yellow and drug in cyan), a sequence of snapshots from the MD simulation of the drug adjustment in the active site and flap closing.

transfer), which monitors the transfer of fluorescence between two FPs (donor D and acceptor A) whose spectra superimpose, indicating the interaction between the proteins tagged by D and A. The intensity of the effect depends both on the relative distance and on the orientation of D and A, which in turn depends on the D−protein−protein−A complex conformation. CG models complemented by docking techniques could allow extraction of fine quantitative information on the complex structures from FRET data.

The FRAP experiments were actually simulated at the "continuum" level.[58] In this case, a certain level of complexity resides in the definition of the boundary conditions, which must reproduce quite accurately the contours of the cell compartments (specifically, the endoplasmic reticulum). Clearly an interplay between this level of description and the CG/mesoscale level would give a more predictive picture and accurate contact with experimental data (Figure 1c).

Concluding Remarks

This paper has hopefully given the flavor of the variety of methods necessary to analyze and model the many different aspects of the biomolecular processes. Concerning the specific cases reported, while almost the all possible combinations of QM and MM techniques were used for the study of FPs and a few studies are available with the continuum models, very a little was done on them at the CG-mesoscale level. Considering the variety of different techniques based on FP-tagging that need a low-resolution level of structural and dynamic knowledge of the involved proteins, this gap is likely to be filled soon. On the other hand, for HIV replication the steps involving the HIV-1pr were well studied at all the levels (except the continuum). But again, the levels that probably would be worth elaboration are still the CG and mesoscale: the study of the cleavage of the entire Gag protein and of the subsequent rearrangements leading to the virus capsid formation is currently in progress (see Figure 7).

More generally, it is clear that studying isolated events, either in space or in time, often gives a partial, if not wrong, picture of the reality. Consequently multiscale simulations are necessary, and the different scale representations must be integrated coherently by means of exchange of information among them. The weak links of the chain are the CG and mesoscale models, which are still much less used than the more conventional FF-based and QM methods. This is partly because they are less standard and still in a development phase. Furthermore, while there are several software packages that integrate QM and MM levels, very a few packages implement the CG and coarser models, and even fewer integrate more levels in a fully multiscale software. An effort in this direction would be very useful to promote the use of multiscale approaches.

I thank Riccardo Nifosí, Joanna Trylska, Chia-En Chang, David Minh, and Dylan Morris for providing data to produce part of the graphics and Riccardo Nifosí and Fabio Beltram for useful discussions. This work was partially supported by IIT research unit Scuola Normale Superiore.

Supporting Information Available. A table reporting a summary of QM calculations of GFP chromophore excitation energies, with references, simulation of the formation of the zwitterionic state in S65T mutants of GFP, and theoretical on-resonance Raman spectra and vibrational properties of the isolated BFPF chromophore in cis and trans conformation. This information is available free of charge via the Internet at http://pubs.acs.org/.

BIOGRAPHICAL INFORMATION

Valentina Tozzini obtained her degree in Physics in 1993 at the University of Pisa and at Scuola Normale Superiore (SNS, Pisa, Italy). In 1997, she obtained her Ph.D. in Physics (at SNS) and in 1999 the Specialization in Medical Physics (at University of Pisa). In 1997−1999, she had a postdoctoral position at the National Institute of Condensed Matter Physics (INFM), and from 2000 she has held a scientist position at NEST-CNR-INFM, SNS. During these years, she spent some periods as visiting researcher at the University of Nijmegen (NL) and at the University of California, San Diego. Her research focuses on the modeling of complex, specifically biological, systems using a variety of theoretical techniques combined in multiscale approaches.

FOOTNOTES

*E-mail: tozzini@nest.sns.it.

REFERENCES

1 Russell, D.; Lasker, K.; Phillips, J.; Schneidman-Duhovny, D.; Velaszquez-Muriel, J. A.; Sali, A. The structural dynamics of macromolecular processes. *Curr. Opin. Cell Biol.* **2009**, *21*, 1–12.

2 Chu, J. W.; Ayton, G. S.; Izvekov, S.; Voth, G. A. Emerging methods for multiscale simulation of biomolecular systems. *Mol. Phys.* **2007**, *105*, 167–174.

3 Cascella, M.; Peraro, M. D. Challenges and perspectives in biomolecular simulations: From the atomistic picture to multiscale modeling. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.

4 Sherwood, P.; Brooks, B. R.; Sansom, M. S. Multiscale methods for macromolecular simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.

5 Ayton, G. S.; Noid, W. G.; Voth, G. A. Multiscale modeling of biomolecular systems: In serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.

6 Shimomura, O.; Johnson, F. H.; Saiga, Y. J. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239.

7 Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W.; Prasher, D. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802–5.

8 Tsien, R. Green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509—44.

9 Kohl, N.; Emini, E.; Schleif, W.; Davis, L.; Heimbach, J.; Dixon, R.; Scolnick, E.; Sigal, I. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *84*, 4686–90.

10 Kurup, A.; Mekapati, S.; Garg, R.; Hansch, C. HIV-1 protease inhibitors: A comparative QSAR analysis. *Curr. Med. Chem.* **2003**, *10*, 1679—1688.

11 Marx, D.; Hutter, J. *Ab initio molecular dynamics: Theory and Implementation*; Grotendorst, J., Ed.; NIC Series, Vol 1; John von Neumann Institute for Computing: Jülich, Germany, 2000.

12 Knowles, P.; Schütz, M.; Werner, H.-J. *Ab Initio Methods for Electron Correlation in Molecules; Modern Methods and Algorithms of Quantum Chemistry, Proceedings*, 2nd ed.; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000.

13 Marques, M.; Gross, E. Time-dependent density functional theory. *Annu. Rev. Phys. Chem.* **2004**, *55*, 427–455.

14 Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

15 Jr, A. D. M.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

16 Scott, W.; Hunenberger, P.; Tironi, I.; Mark, A.; Billeter, S.; Fennen, J.; Torda, A.; Huber, T.; Kruger, P.; van Gunsteren WF, W. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

17 Peraro, M. D.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. Investigating biological systems using first principles Car—Parrinello molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.

18 Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M.; Piana, S.; Shaw, D. E. Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4689–4694.

19 Tozzini, V. Coarse grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.

20 Tozzini, V.; Rocchia, W.; McCammon, J. A. Mapping all-atom models onto one-bead coarse grained models: General properties and applications to a minimal polypeptide model. *J. Chem. Theory Comput.* **2006**, *2*, 667–673.

21 Reith, D.; Puetz, M.; Mueller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1264–1275.

22 Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method I: A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114.

23 Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method II: Numerical implementation for coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244115.

24 (a) Yang, L.-W.; Chng, C.-P. Coarse-grained models reveal functional dynamics. I. Elastic network models—theories, comparisons and perspectives. *Bioinf. Biol. Insights* **2008**, *2*, 25–45. (b) Trylska, J.; Tozzini, V.; McCammon, J. A. Exploring global motions and correlations in the ribosome. *Biophys. J.* **2005**, *89*, 1455–1463. (c) Maragakis, P.; Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **2005**, *352*, 807–822. (d) Chu, J.-W.; Voth, G. A. Coarse-grained free energy functions for studying protein conformational changes: A Double-well network model. *Biophys. J.* **2007**, *93*, 3860–3871. (e) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys. J.* **2008**, *95*, 4183–4192. (f) Hamacher, K.; McCammon, J. A. Computing the amino acid specificity of fluctuations in biomolecular systems. *J. Chem. Theory Comput.* **2006**, *2*, 873–878. (g) Jeong, J. I.; Jang, Y.; Kim, M. K. A connection rule for α-carbon coarse-grained elastic network models using chemical bond information. *J Mol Graphics Modell.* **2005**, *24*, 296–306. (h) Hueda, Y.; Taketomi, H.; Go, N. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A three-dimensional lattice model of lysozyme. *Biopolymers* **1987**, *17*, 1531–1548. (i) Cheung, M.; Garcia, A.; Onuchic, J. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 685–690. (j) Das, P.; Matysiak, S.; Clementi, C. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10141–10146. (k) Yap, E.-H.; Fawzi, N. L.; Head-Gordon, T. A coarse-grained α-carbon protein model with anisotropic hydrogen-bonding. *Proteins: Struct, Funct., Bioinf.* **2008**, *70*, 626–638. (l) Klimov, D.; Thirumalai, D. Mechanisms and kinetics of β-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544–2549. (m) Mukherjee, A.; Bagchi, B. Correlation between rate of folding, energy landscape, and topology in the folding of a model protein HP-36. *J Chem Phys* **2003**, *118*, 4733–4747. (n) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834. (o) Liwo, A.; Khalili, M.; Sheraga, H. A. *Ab initio* simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367. (p) Jernigan, R.; Bahar, I. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195–109. (q) Tozzini, V.; McCammon, J. A coarse grained model for the dynamics of flap opening in HIV-1 protease. *Chem Phys Lett* **2005**, *413*, 123–128. (r) Thorpe, I.; Zhou, J. S.; Voth, G. A. Peptide folding using multiscale coarse-grained models. *J. Phys. Chem. B* **2008**, *112*, 13079–13090.

25 Izvekov, S.; Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys.* **2006**, *125*, 151101.

26 Vakser, I. A.; Kundrotas, P. Predicting 3D structures of protein-protein complexes. *Curr. Pharm. Biotechnol.* **2008**, *9*, 57–66.

27 Voth, G. A. *Coarse Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Washington DC, 2008.

28 Grochowski, P.; Trylska, J. Continuum molecular electrostatics, salt effects and counterion binding - A review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* **2008**, *89*, 93–113.

29 Tanizaki, S.; Clifford, J.; Connelly, B. D.; Feig, M. Conformational sampling of peptides in cellular environments. *Biophys. J.* **2008**, *94*, 747–759.

30 Ayton, G. S.; Voth, G. A. Systematic multiscale simulation of membrane protein systems. *Curr. Opin. Struct. Biol.* **2009**, *19*, 138–144.

31 Koumoutsakos, P. Multiscale flow simulation using particles. *Annu. Rev. Fluid Mech.* **2005**, *37*, 457–487.

32 Marini, U.; Marconi, B.; Tarazona, P. Dynamic density functional theory of fluids. *J. Phys.: Condens. Matter* **2000**, *12*, A413–A418.

33 Nifosí, R.; Amat, P.; Tozzini, V. Variation of spectral, structural and vibrational properties within the intrinsically fluorescent proteins family: A density functional study. *J. Comput. Chem.* **2007**, *28*, 2366–2377.

34 Luin, S.; Voliani, V.; Lanza, G.; Bizzarri, R.; Nifosí, R.; Amat, P.; Tozzini, V.; Serresi, M.; Beltram, F. Variation of spectral, structural and vibrational properties within the intrinsically fluorescent proteins family: A density functional study. *J. Am. Chem. Soc.* **2009**, *131*, 96–103.

35 Nifosí, R.; Luo, Y. Predictions of novel two-photon absorption bands in fluorescent proteins. *J. Phys. Chem. B* **2007**, *111*, 14043–14050.

36 Laino, T.; Nifosì, R.; Tozzini, V. Relationship between structure and optical properties in green fluorescent proteins: A quantum mechanical study of the chromophore environment. *Chem. Phys.* **2004**, *298*, 17–28.

37 Nifosí, R.; Ferrari, A.; Arcangeli, C.; Tozzini, V.; Pellegrini, V.; Beltram, F. Photoreversible dark state in a tristable green fluorescent protein variant. *J. Phys. Chem. B* **2003**, *107*, 1679–1684.

38 Nifosí; Tozzini, V. Cis-trans isomerization of the chromophore in the green fluorescent protein variant E2GFP: A molecular dynamics study. *Chem. Phys.* **2006**, *323*, 358–368.

39 Nifosí; Tozzini, V. Molecular dynamics simulations of enhanced green fluorescent proteins: Effects of F64L, S65T and T203Y mutations on the ground-state proton equilibria. *Proteins* **2003**, *51*, 378–389.

40 Vendrel, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. Potential energy landscape of the photoinduced multiple proton-transfer processes in the green fluorescent protein:

Classical molecular dynamics and multiconfigurational electronic structure calculations. *J. Am. Chem. Soc.* **2006**, *128*, 3564–3574.

41 Camilloni, C.; Provasi, D.; Tiana, G.; Broglia, R. A. Optical absorption of a green fluorescent protein variant: Environment effect in a density functional study. *J. Phys. Chem. B* **2007**, *111*, 10807–10812.

42 Patnaik, S.; Trohalaki, S.; Naik, R.; Stone, M. O.; Pachter, R. Computational study of the absorption spectra of green fluorescent protein mutants. *Biopolymers* **2006**, *28*, 253–263.

43 Demachy, I.; Ridard, J.; Laguitton-Pasquier, H.; Durnerin, E.; Vallverdu, G.; Archirel, P.; Levy, B. Cyan fluorescent protein: Molecular dynamics simulations and electronic absorption spectrum. *J. Phys. Chem. B* **2005**, *109*, 24121–24133.

44 Carnevale, V.; Raugei, S.; Micheletti, C.; Carloni, P. Large-scale motions and electrostatic properties of furin and HIV-1 protease. *J. Phys. Chem. A* **2007**, *111*, 12327–12332.

45 Trylska, J.; Bala, P.; Geller, M.; Grochowski, P. Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. *Biophys. J.* **2002**, *83*, 794–807.

46 Lauria, A.; Ippolito, M.; Almerico, A. Molecular dynamics studies on HIV-1 protease: A comparison of the flap motions between wild type protease and the M46I/G51D double mutant. *J. Mol. Modell.* **2007**, *13*, 1151–1156.

47 Perryman, A.; Lin, J.; McCammon, J. Restrained molecular dynamics simulations of HIV-1 protease: The first step in validating a new target for drug design. *Biopolymers* **2006**, *82*, 272–284.

48 Hornak, V.; Okur, A.; Rizzo, R.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.

49 Tozzini, V.; McCammon, J. A coarse grained model for the dynamics of flap opening in HIV-1 Protease. *Chem. Phys. Lett.* **2005**, *413*, 123–128.

50 Tozzini, V.; Trylska, J.; Chang, C.-E.; McCammon, J. A. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. *J. Stuct. Biol.* **2007**, *157*, 606–615.

51 Chang, C.-E.; Shen, T.; Trylska, J.; Tozzini, V.; McCammon, J. A. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. *Biophys. J.* **2006**, *90*, 3880–3885.

52 Trylska, J.; Tozzini, V.; Chang, C.-E.; McCammon, J. A. HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.* **2007**, *92*, 4179–4187.

53 Chang, C.-E.; Trylska, J.; Tozzini, V.; McCammon, J. A. Binding pathways of ligands to HIV-1 protease: Coarse-grained and atomistic simulations. *Chem. Biol. Drug Des.* **2007**, *69*, 5–13.

54 Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni1, P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.* **2005**, *95*, 218102.

55 Terreni, M.; Liverani, V.; Gutierrez, M.; Di Fenza, A.; Tozzini, V.; Albanese, A.; Arosio, D.; Giacca, M.; Cereseto, A. GCN5 dependent acetylation of HIV-1 integrase enhances viral integration. *Retrovirology*, submitted for publication.

56 Di Fenza, A.; Rocchia, W.; Tozzini, V. Proteins complexes of HIV-1 integrase with HAT proteins: Multi-scale models, dynamics and hypotheses on allosteric site of inhibition. *Proteins* **2009**, *76*, 946–958.

57 Minh, D.; Chang, C.-E.; Trylska, J.; Tozzini, V.; McCammon, J. The influence of macromolecular crowding on HIV-1 protease molecular dynamics. *J. Am. Chem. Soc.* **2006**, *128*, 6006–6007.

58 Sbalzarini, I. F.; Mezzacasa, A.; Helenius, A.; Koumoutsakos, P. Effects of organelle shape on fluorescence recovery after photobleaching. *Biophys. J.* **2005**, *89*, 1482–1492.