



Amaro, R. E., & Mulholland, A. J. (2018). Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nature reviews. Chemistry*, 2(4), [0148].  
<https://doi.org/10.1038/s41570-018-0148>

Peer reviewed version

License (if available):  
Unspecified

Link to published version (if available):  
[10.1038/s41570-018-0148](https://doi.org/10.1038/s41570-018-0148)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer Nature at <https://www.nature.com/articles/s41570-018-0148>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Published in final edited form as:

*Nat Rev Chem.* 2018 April ; 2(4): . doi:10.1038/s41570-018-0148.

## Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures

Rommie E. Amaro<sup>1</sup> and Adrian J. Mulholland<sup>2</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0304

<sup>2</sup>Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

### Abstract

Drug action is inherently multiscale: it connects molecular interactions to emergent properties at cellular and larger scales. Simulation techniques at each of these different scales are already central to drug design and development, but methods capable of connecting across these scales will extend understanding of complex mechanisms and the ability to predict biological effects. Improved algorithms, ever-more-powerful computing architectures and the accelerating growth of rich datasets are driving advances in multiscale modeling methods capable of bridging chemical and biological complexity from the atom to the cell. Particularly exciting is the development of highly detailed, structure-based, physical simulations of biochemical systems, which are now able to access experimentally relevant timescales for large systems and, at the same time, achieve unprecedented accuracy. In this Perspective, we discuss how emerging data-rich, physics-based multiscale approaches are of the cusp of realizing long-promised impact in the discovery, design and development of novel therapeutics. We highlight emerging methods and applications in this growing field, and outline how different scales can be combined in practical modelling and simulation strategies.

### Introduction

Biomolecular simulations are essential tools for drug design and development, and for our understanding of the molecular bases of disease<sup>1,2</sup>. Simulations provide a ‘computational microscope’ to reveal biological mechanisms in atomic detail.<sup>3</sup> They can reveal cryptic drug binding sites<sup>4</sup> and predict important biological properties such as drug resistance<sup>5</sup>. Molecular dynamics (MD) simulations are the most widely used biomolecular simulation method: they apply empirical molecular mechanics (MM) force fields and can now be used to explore in atomic detail time-dependent phenomena at the scale of viral capsids<sup>6</sup> even over microseconds<sup>7</sup>, given sufficient computational power. MM methods are increasingly applied routinely in structure-based drug design, e.g. for free energy calculations to predict

Correspondence to: ramaro@ucsd.edu and adrian.mulholland@bristol.ac.uk.

Disclosures

REA is a co-founder, is on the Scientific Advisory Board, and has equity interest in Actavalon, Inc.

binding affinities of pharmaceutical leads to their targets, accelerating drug development<sup>8</sup>. Their importance and algorithmic efficiency (resting on years of development by many pioneers) have made atomistic molecular dynamics simulations one of the largest scientific consumers of computing time globally. These methods (and Monte Carlo simulations) can be applied in rigorous free energy calculations of relative binding affinities of small molecules to protein targets. However, computational demands in terms of the sizes of systems and timescales limit the use of MD methods, but at the same time, the relatively simple potential functions used to achieve computational efficiency somewhat limit their range of application and accuracy. Different types of simulation methods are therefore required for different types of problems. Each of these different simulation methods has strengths, weaknesses and practical limitations in terms of the size of system that can be simulated, length of simulation that can be achieved, and type of phenomena that can be modeled. For example, various types of coarse-grained methods allow simulations on large spatiotemporal scales phenomena, including protein-protein interactions, protein orientation in membranes and packaging of nucleic acids. Simple molecular docking approaches offer a limited level of detail of molecular interactions, conformational flexibility and solvation in favor of increased computational efficiency for the rapid identification of potential leads from large databases. At the other extreme of computational molecular science, quantum chemical methods can be used to model chemical reactions (*e.g.*, the mechanisms of enzyme catalysis) and calculate spectra. First principles (*ab initio*) electronic structure techniques for the structure optimisation of proteins<sup>9</sup>, and atomistic simulations to investigate dynamics of systems of appreciable size and complexity over nano- to microsecond<sup>10,11</sup> and even millisecond timescale<sup>12</sup> (for smaller systems) exemplify the upper limits of current capability for atomic and molecular methods<sup>13</sup>.

Multiscale modelling approaches are emerging in drug design with potentially enormous impact on human health. Drugs act at the molecular scale but obviously have macroscopic effects, so we must consider multiple length scales to understand how they exert their effects. The dynamic nature of drug targets and the breathtaking complexity of biological systems challenge our scientific understanding from the level of molecular structure all the way up to cellular organization (and beyond). Each level provides challenges and fascination in its own right, but a holistic approach requires an understanding of how changes at different levels are linked together and influence each other. No single simulation method can address all of the many questions involved, nor explain how phenomena at various spatiotemporal scales are coupled and linked. Multiscale simulation methods aim to model and analyse the connections across scales, *e.g.* how changes on one scale lead to changes at another. An obvious challenge is the integration of data and simulations across lengthscales and timescales. Current multiscale approaches are potentially capable of overcoming these limits by directly combining different levels of description, bringing a new perspective to drug discovery.

The award of the 2013 Nobel Prize in Chemistry to Karplus, Levitt, and Warshel for their seminal contributions in developing multiscale methods for modeling complex biochemical systems recognized the essential role of theoretical and computational methods as a direct and necessary complement to experiment, and the birth of multiscale molecular modelling in biochemistry<sup>14</sup>. Today, nearly half a century after the advent of these methods, multiscale

simulations offer the tantalizing possibility of understanding biology in intricate and exquisite detail. For example, multiscale modelling offers an understanding of how a chemical reaction occurring at an enzyme active site can affect other proteins and extend through the hierarchy of biological complexity - from subcellular neighborhoods, to cells, and tissues. The power of multiscale methods lies in the possibility to create a fluid knowledge landscape synthesizing disparate modelling and experimental data at different spatiotemporal scales. Practical insight will depend on combining diverse approaches, linking chemistry at the atomic scale with biological function at the cellular and higher levels to elucidate the mechanisms of emergent phenomena, and doing so in a way that circles back to drive drug design and development.

### [H3] Unleashing the potential of emerging experimental data sources

Technological innovations for the *in situ* acquisition of biological structural data, including advances in direct detector and phase plate technologies<sup>15</sup> for X-ray beamlines and electron microscopes (EMs), give access to new and vastly more detailed information across a range of previously inaccessible scales and, in some instances, time resolutions (Figure 1). Multiscale computational approaches are needed to fill in and connect datasets, which include data obtained from: serial block wide-field EM illumination of tissue and cellular ultrastructure to within tens of nanometers isotropic resolution for biologically real (endogenous, not cultured) samples<sup>16</sup>; cryoelectron tomography (cryoET) to localize supramolecular complexes and yield glimpses into cells with molecular resolution (~2–4 nm resolution in individual tomograms),<sup>17–19</sup>; soft x-ray tomography to image whole hydrated (not stained or frozen) cells in their near-native state<sup>20</sup>; near-atomic cryoelectron microscopy (cryoEM)<sup>19</sup>; small angle x-ray (SAXS) and neutron scattering (SANS); x-ray crystallography, diffuse scattering for an ensemble-based view of x-ray structures<sup>21</sup>; x-ray free electron lasers<sup>22</sup>, time resolved x-ray<sup>23</sup> and neutron diffraction<sup>24</sup>. In parallel, ongoing innovations in biophysical techniques such as NMR spectroscopy (*e.g.*, used for intein protein segmentation<sup>25</sup>) and hydrogen-deuterium exchange mass spectrometry (HDX-MS)<sup>26</sup> continue to enrich our understanding of the dynamics and interactions of molecular and macromolecular ensembles. Interpretation, refinement and understanding of the high-resolution data from all these techniques challenge current modelling approaches. In addition all these new data call for the development of models and provide tests for their validation.

### [H3] The coming fusion of simulation and data science

The convergence of improved and increased biophysical data, together with impressive algorithmic advances, occurs against a backdrop of an ever-expanding, increasingly diverse and more capable computing landscape. Porting simulation methods to the growing range of novel hardware architectures (*e.g.*, graphics processing units (GPUs), advanced RISC machine (ARM)-based high performance computing (HPC), cloud computing, petascale HPC machines, and the emerging horizon of exascale computing<sup>27</sup>) is extending the scope and range of simulations. The rapid growth of data science also offers transformative possibilities, not only in the manipulation of simulation data and linking across spatiotemporal scales, but also in its seamless integration with experimental data. Examples include the systematic development of Jupyter notebooks<sup>28</sup> and automated workflows<sup>29,30</sup>,

improved data sharing<sup>21</sup>, integration<sup>31</sup>, and analytics<sup>20</sup>. These developments are driving cultural shifts towards improved reproducibility, openness, sharing, robustness and, ultimately, predictive ability of the computational approaches discussed in this Perspective.

### [H3] Potential applications of multiscale methods in drug discovery

Multiscale methods span two or more spatial or temporal domains, or combine different types of treatment, aiming to give insight across scales. We note that the concept of multiscale modeling has been developed in several disciplines. Here, we focus on multiscale simulation methods that relate to the molecular level, that are based on fundamental physics, and are potentially relevant to drug discovery. Multiscale techniques can be constructed in different ways, depending on how different levels of description are combined or coupled, and how information is passed between the different levels<sup>34</sup>. Multiscale methods can combine different levels of theory or resolution, *e.g.*, combining MD with Brownian dynamics (BD) to access long timescales and larger lengthscales (MD/BD<sup>35</sup>) or combining quantum mechanics (QM) with molecular mechanics (MM) in (QM/MM<sup>36</sup>) to study electronic properties in a single simulation. Another class of multiscale methods comprises the hierarchical integration of sets of approaches carried out at different scales, which leads to one ultimate cohesive model. In this case, the final result is obtained through the interchange of key parameters across model scales<sup>37–40</sup>, even though the simulation platform itself may not directly interface two distinct physical regimes. A related form of multiscale modeling is based on the connection (and synthesis) of different types of biological, chemical, structural and biophysical data — a particularly exciting approach, given technological advances across a spectrum of experimental techniques. The development of increasingly accurate integrative models as an initial framework on which multiscale simulation methods subsequently operate presents a powerful emerging paradigm for drug discovery.

The study of multiscale methods for drug design is a wide and rapidly growing field; it is rich in potential but yet to realise its enormous promise. The diversity of approaches and applications means that we can only cover a few relevant examples in this Perspective to indicate the vast potential of this field. Here, we focus on some recent exciting methodological advances, as well as challenges for development and applications that highlight the promise of simulations to bridge scales from atoms to cells.

**[H1] Cellular to Subcellular**—The ongoing surge of high-resolution structural data is providing detailed views into many previously inaccessible biological compartments (*e.g.*, the cell nucleus) and enabling the development of correspondingly realistic molecular models of cells and subcellular organelles. New tools such as CellPack<sup>12</sup>, which interoperates with CellView<sup>41</sup> and LipidWrapper<sup>13</sup>, can be used to model complex biomolecular systems at the mesoscale, reaching atomic resolution. Combining different datasets across scales of resolution enables direct multiscaling from the standpoint of data integration (Box 1). The accessible spatial range now reaches the micron dimension, with essentially no limits on the complexity of the constituents of the system under investigation. A key advance is the ability to develop many ensembles of such models compatible with multiple sources of experimental data (*e.g.*, proteomics, structural data from x-ray through

cell tomography, *etc.*). This allows researchers to modify the construction of a system (for example, changing the expression level of a particular protein, introducing a structural perturbation to membrane or organelle shape, varying the molecular composition of viral strains (reassortment), or adding post-translational modifications) in order to test consistency against different types of data and predict the effects induced by these changes on the system.

Once structural models are assembled, researchers can explore biological heterogeneity at the molecular level, and statistical distributions of biological and chemical components within these complex environments. Monte Carlo based methods such as MCell,<sup>7</sup> and continuum-based methods such as SubCell,<sup>6</sup> enable researchers to investigate biological phenomena without explicitly accounting for molecular (particle) collision and interaction. Reaction-diffusion master equation (RDME) methods combine network- and particle-based approaches on discretized grids/lattice sites; in doing so, these approaches allow the development of whole-cell based models of drugs and their dynamical interactions with receptors. Two exemplary GPU-accelerated RDME programs are: Lattice Microbe<sup>46–48</sup>, which splits the reaction and diffusion operators to allow efficient models of *in vivo* crowding on particle diffusion; and ReaDDyMM<sup>49</sup>, which combines reactions at lattice sites with particle-particle interactions at off-lattice sites, as determined by MD. Another promising multiresolution method — lattice Boltzmann MD (LBMD) — employs a mixed approach in which (dynamic) proteins are represented as coarse-grained particles and the solution through which the proteins diffuse is represented probabilistically, such that multiple physical elements (including hydrodynamic and thermodynamic forces) can be included<sup>50</sup>.

Particle-based approaches range from coarse-grained (each particle represents a group of atoms, from part of an amino acid residue to a whole protein) to fully atomistic (every atom represented individually) representations of the molecular constituents. Although coarse-grained techniques offer the possibility of simulating the behavior of larger systems on longer time-scales compared to the fully atomistic approaches, the choice of particle representation (i.e., how atoms are grouped) and the related force field development still presents challenges<sup>51</sup>. Atomic details are neglected entirely in coarser models, such as fluctuating finite element analysis (FFEA). In this approach, the macromolecules are essentially treated as density maps, subject to thermal fluctuations within a continuum medium that encodes the material properties, such as shear and elastic. Lower resolution data, such as SAXS or cryoEM can be directly linked to FFEA, effectively bypassing the requirement of starting with a detailed molecular model<sup>52</sup>. Such approaches can be essential when, for example, higher resolution structural data for system components is not available, or not all of the molecular components in a particular system are known.

Particle-based simulations in which the molecular components are represented with atomic detail —rigid-body BD and fully-flexible atomistic MD simulations — are proving ever more capable<sup>6,53–55</sup>. Rigid-body BD uses fully atomistic representations of molecules, but neglects the molecular internal degrees of freedom. In this representation, rigid molecules (e.g. proteins) are free to tumble and rotate as they diffuse relative to each other in a viscous medium (with water represented as a continuum solvent) and subject to random motion

according to the fluctuation-dissipation theorem. The intermolecular forces that govern the interaction and collisions of the particles are electrostatic in nature and represented with a modified form of the Poisson–Boltzmann equation. In contrast, internal motions and conformational changes are considered in fully-flexible MD simulations. BD and MD can be used to explore the dynamics of molecules in crowded biological milieus, improving understanding of detailed biological and chemical interactions<sup>10,53</sup>. They can be combined in multiscale approaches such as SEEKR (Box 2). Recent efforts linking particle-based simulation to higher-level systems biology or network-based models exemplify how handoffs between such methods can be gainfully achieved. For example, association rates determined with BD, MD, and SEEKR in network-based Markov chain of states models have been combined in order to define the mechanisms underlying the cooperative nature of cAMP activation of protein kinase A<sup>37</sup>. Systems biology models and MD simulations have been combined e.g. to generate predictions of the effects of enzyme--substrate binding affinity changes due to genetic variation in human erythrocytes<sup>38</sup>. In both cases, linking molecular and cell-based models allowed mechanistic insights and predictions at the atomic-level to inform network-based whole-cell models of disease. Collectively, multiscale approaches spanning cellular to subcellular scales will help to address the challenge of better predictive models for off-target effects (*e.g.* binding of drugs to targets other than that desired). They also enable more complete understanding of chemical mechanisms of action and their effects at larger scales, especially for complex signalling pathways that require a broad view of molecular complexity. Hence, particle-based multiscale approaches at this scale may be particularly useful for understanding drug action.

Challenges at cellular and subcellular scales will continue to include the experimental determination of the constituents of cellular compartments (*e.g.*, protein counts, mRNA expression levels, etc.) with enough detail (spatial and temporal resolution) to enable the development of biologically accurate structural models. New tools for segmentation and refinement of tomographic data are needed, particularly when the data are highly complex. For example, recent ultrastructural 3D mapping of the cell nucleus shows how new imaging techniques, such as ChromEMT, can directly reveal details of the higher order structure of chromatin; at the same time, it underscores the need for new tools that are able to segment and refine the structural data at the level of individual chromatin fibers<sup>56</sup>. These data will transform our understanding of the relationship between chromatin structure and dynamics and the regulation of gene expression, which is an inherently multiscale challenge<sup>57</sup>. Additionally, once these structures are refined, new tools for developing numerically computable meshes from these reconstructions are required to create models that can be extended and interrogated with physics-based simulation. Innovations in simulation approaches are required to effectively handle hydrodynamic interactions<sup>58</sup>, as well as different concentration and diffusion regimes within a complex, crowded cell scene<sup>59</sup>. Finally, a major set of challenges relates to the ability of researchers to set up and execute complex models and simulations. Also, the substantial complexity of highly detailed large-scale models can make the interpretation of the simulated phenomena difficult. These challenges are intimately linked with data analysis and visualization, and will benefit from technological advances such as machine learning<sup>60</sup> and virtual reality<sup>61</sup>.

**[H1] Subcellular to molecular and atomistic**—The action (and metabolism) of many drugs is fundamentally based on the changes in and interactions among individual molecules. Multiscale methods are needed to connect molecular changes to changes induced in subcellular levels and beyond. Increasingly informed integrative models of macromolecular complexes<sup>62,63</sup> are providing atomically detailed views of complex drug targets and drug-target interaction<sup>64</sup>, thanks particularly to advances in cryoEM. The 2017 Nobel Prize in Chemistry recognised these developments<sup>65</sup>. However, moving beyond the traditional paradigm of studying single drug targets in isolation to tackle the dynamics of macromolecular complexes and, for example, their interactions with the genome<sup>64</sup> poses a challenge to atomic-scale modeling. The ability to model protein–protein complexes more accurately (e.g. allowing for changes in conformation driven by intermolecular interactions and chemical changes) will improve protein and antibody design for vaccine development.<sup>66–69</sup> Indeed, the resolution-revolution taking place in cryoEM and cryoET will drive methods development for multiscale simulation across the subcellular and molecular scales. Multiscale simulations will offer an expanded perspective of drug targets, elucidating their detailed and often critical interactions with realistically complex membranes and other proteins. The new structural understanding provided by diverse approaches has already helped industrial research teams reconcile seemingly divergent or otherwise inexplicable experimental assay results<sup>70</sup>.

There has been significant recent progress in the development of methods enabling the exploration and characterization of the dynamics of molecular scale systems. An example particularly relevant to drug discovery is provided by simulation-based approaches for the identification of so-called cryptic (hidden) pockets or sites of allosteric activation, which are not evident in x-ray crystallographic structures and can be novel drug-target sites<sup>4,71,72</sup>. Another promising class of techniques is based on Markov state models (MSMs), which enable the extension of temporal scales achievable in ensemble-based approaches through the extraction of long timescale dynamics from many short timescale simulations. MSMs take a more statistics-oriented view to trajectory analysis: individual states are defined or identified, and the dynamics between the interconnected states, assumed to be Markovian, are modelled as a transition probability matrix, populated from independent simulations. Dynamic information relating the states is typically obtained through many short timescale MD simulations that are integrated into one cohesive framework<sup>73–75</sup> MSMs have been used to predict the thermodynamic and kinetic landscapes for the activation of multiple kinases<sup>76,77</sup> and protein–protein association pathways<sup>78</sup>, characterizing biological processes that occur over timescales from microseconds to hours. Notably, MSMs make use of the statistical sampling necessary for the analysis of larger lengthscale (e.g., subcellular or cell-scale) simulations, which by their nature contain many independent copies of one particular drug target. One can use relatively short timescale (e.g., tens of nanoseconds) simulations of large and/or multicomponent systems, potentially comprising multiple hundreds of millions of atoms, in combination with MSMs to extract long timescale (e.g., kinetics on the order of milliseconds) information for the individual molecular components of a biological scene. These methods will facilitate an increasingly accurate understanding of how drugs act at a particular site and subsequently alter the dynamical landscape of their receptors, including, for example, for the highly dynamic G-protein coupled receptors<sup>79</sup>, one of the most



important classes of pharmaceutical targets. MD based methods also stand to benefit from and connect to advances in experimental structure characterization methods. These methods include diffuse x-ray scattering, which gives improved characterization of protein flexibility and ligand binding, and the detailed picture of biomolecular heterogeneity emerging from high resolution cryoEM, both of which can feed back into the development of more accurate MD force fields <sup>21</sup>.

The development of effective drugs is becoming increasingly reliant on our understanding and ability to quantify and optimize the kinetics of drug binding (associated with a rate constant  $k_{on}$ ) and unbinding ( $k_{off}$ ). Drugs must bind quickly enough to their target to avoid being cleared from the body before they can act, and must also remain bound long enough to exert an effect; such considerations are more important than the thermodynamic binding affinity in many cases. Drugs with slow rates of dissociation have longer interaction times (also referred to as residence times) with their targets and therefore are often found to be more efficacious<sup>80,81</sup>. The ability to quantify and predict binding and unbinding rates and residence times therefore represents a major and growing requirement in drug discovery and development programs. As such, the past few years have seen a dramatic increase in simulation and associated modeling methods to predict such quantities, and analyse the molecular and dynamical features that determine them<sup>81</sup>. These methods include direct quantification through enhanced sampling techniques<sup>82</sup> such as MSMs <sup>76,83</sup>, smoothed-potential MD<sup>84,85</sup>, and metadynamics <sup>86</sup> (e.g. with path collective variables and parallel tempering to calculate free energy profiles, and transition state-partial path transition interface sampling <sup>87</sup> for kinetics) or through multiscale simulation methods, such as SEEKR. SEEKR is a novel method that uses milestoning theory to combine atomic-scale rigid-body Brownian dynamics (if the two interacting molecular species are sufficiently far apart, with fully-flexible MD simulations when the particles are in close proximity (Box 2) <sup>35,88,89</sup>. The combined use of different dynamical propagators enables efficient computation of accurate binding kinetics and free energies of binding using a directly multiscale approach.

Predicting membrane permeability is another area of intense interest to drug discovery programs for which multiscale physics-based simulations promise significant advantages over phenomenological (or descriptor-based) models. Both the potential of mean force for a small molecule crossing the membrane and its position-dependent diffusion constant can be simulated in several ways and combined to predict the passive permeation of drugs through membranes (reviewed in <sup>90,91</sup>). Recent resurgence of interest in this area has included multiscale efforts to tackle system complexity, e.g. for notoriously challenging systems such as gram-negative bacterial membrane transport <sup>92–94</sup>; multiresolution methods that mix coarse-grained models of membranes with all-atom models of antibacterial compounds <sup>95</sup>; and methods that enable multiscaling in time through the integrated use of MSMs <sup>96</sup> or milestoning <sup>97,98</sup>.

Despite the developments described above, the ability of present methods to enable crossing of subcellular to molecular scales — ensuring biological and chemical realism — is not without technical and intellectual challenges. Sampling of slow conformational dynamics in all of these systems remains a key problem. Development of new approaches to accelerate

otherwise slow dynamics (e.g., multiensemble MSMs<sup>99</sup>) will be required. Innovative solutions to the challenges posed by slow dynamics are likely to require continued development of approaches for hierarchical coarse-graining<sup>57,100–104</sup>. As simulations over increasingly longer timescale become routine, research efforts must continue to shift from manual human-driven data curation (which remains the current *de facto* standard) to machine-learning-based methods that will enable the detection of new patterns, correlations, etc. within the vast amount of data being generated<sup>105</sup>.

**[H1] Atomistic to electronic**—Molecular structure, dynamics, and reactivity arise fundamentally from quantum mechanics. Calculations of the electronic structure of molecules are essential for the study of certain properties. In principle, as Dirac stated long ago, quantum mechanics provides the theoretical route to calculating all molecular properties, the “only” challenge being computational tractability. The utility of quantum mechanics has been amply demonstrated for small molecules (~tens of atoms), for which reaction barriers and spectra can be calculated from first principles ab initio electronic structure methods with accuracy often at least as good as experiment. The electronic structure of larger systems (~hundreds of atoms) can be calculated with density functional theory (DFT) methods. These are in general somewhat less accurate and are not systematically improvable, but have nevertheless revolutionized the role of computation in chemistry by providing useful insight (e.g., into reaction mechanisms) at a manageable computational cost. More approximate methods based on semiempirical molecular orbital theory or approximate DFT allow calculations on even larger systems (~thousands of atoms). Algorithmic developments (such as implementation on GPUs<sup>9</sup>) and methodological developments continue to extend the reach of quantum chemical calculations, both in terms of system size (for example, the thousands of atoms for modelling transition states in enzymes and properties of ion channels) and scope (for example, to MD and Monte Carlo simulations), bringing electronic structure calculations into new biological regimes<sup>106</sup>. All of these techniques address molecular electronic structure, which is inherently quantum mechanical, and are usually applied with a classical description of molecular dynamics/nuclear motion. Quantum dynamical effects such as quantum tunneling (which is significant in the determining the rate of transfer of hydrogen) can also be investigated, by methods that include the effects of quantum dynamics for nuclei.

In principle, quantum mechanical methods offer higher accuracy than empirical force fields, but in practice, it may often be more feasible — or indeed preferable — to apply a hybrid approaches, combining a quantum mechanical description of a small region (e.g., enzyme active site) with an empirical (MM) treatment of most of the system (e.g., protein, solvent, membrane, etc). Such QM/MM methods were a focus of the 2013 Nobel Prize in Chemistry and now provide an attractive combination of practicality and versatility for a range of problems. Applications in drug development include: informing inhibitor design from knowledge of interactions of transition states and intermediates<sup>107</sup>; understanding the reactivity and specificity of (and resistance to) covalent inhibitors<sup>108</sup>; analysing the coupling of chemical and conformational changes in biomolecules; predicting NMR and electronic spectra (e.g., to identify binding modes); developing of *in situ* structure-activity relationships, and predicting drug metabolism<sup>109–111</sup> (Figure 2). It is possible to extend

beyond DFT to highly accurate first principles *ab initio* methods, *e.g.*, by applying projector-based embedding techniques, which can be applied in a QM/MM framework to treat large systems such as proteins.<sup>112,113,114</sup> QM/MM MD simulations are possible with lower-level QM treatments and are increasingly practical with DFT methods. A combination of such methods can predict chemically accurate barriers for enzyme-catalysed reactions<sup>35,110</sup>. QM/MM methods can also now be applied in areas that were previously the domain of empirical force fields, such as in calculations of binding affinities and solvation energies using multilevel sampling approaches.<sup>8,117–121</sup> QM/MM methods can be applied in multiscale schemes coupling sampling at different levels, to combine the accuracy of the higher-level method with a more computationally efficient lower-level treatment (see Box 3). Potentially, this can overcome some limitations of empirical (MM) atomistic force fields (such as oversimplified descriptions of electrostatics and lack of electronic polarization), which may be particularly important for some target classes, such as metalloproteins<sup>122</sup>. For example, the binding affinity of water molecules to proteins is affected by changes in the polarization of water molecules<sup>123</sup>. This effect is larger for larger drug-like molecules and may affect predictions of drug binding/unbinding kinetics because of the important changes in solvation involved in these processes. Just as QM methods can generate data to inform development of atomistic force fields<sup>124</sup>, which are increasingly driven by machine learning approaches and integrated with experimental data, multiscale QM/MM schemes also offer the potential of testing and developing lower-level methods, *e.g.*, in on-the-fly (re)parameterization<sup>125</sup>.

Methods capable of modelling and predicting chemical reactivity in detail offer opportunities in a number of emerging challenging areas in drug discovery. A particularly important area of application is in the study of covalent inhibitors. There is renewed and growing interest in developing covalent binders and inhibitors (for enzymes and other targets), for increased affinity and *e.g.* altered pharmacokinetics<sup>126</sup>. There is a need to understand what governs the reactivity of covalent modifiers *in vivo* to maximise specificity and minimise off-target effects, and to understand resistance to covalent drugs<sup>127</sup>. The ideal covalent modifier is only activated at the specific target site. Just as for the prediction of drug metabolism, approaches based on the ligands alone cannot capture all the factors relevant for reactivity. Challenges include prediction of p*K*<sub>a</sub>s of target residues for modification (for cysteines, in particular), treatment of conformational effects and identification of unusual mechanisms. For modelling of reactivity, empirical valence bond (EVB) methods are a highly efficient alternative to QM or QM/MM calculations, being significantly less computationally demanding than electronic structure methods, but they require substantial effort in parameterization<sup>128</sup>. QM and QM/MM methods (and/or experimental data) can be used to parameterize EVB models<sup>115</sup>. Practical challenges in applying QM/MM methods<sup>129</sup> include the choice of the size of QM region for optimal efficiency and accuracy<sup>130</sup>. Adaptive schemes, in which the QM region changes during the simulation, are potentially useful for some applications, such as long-range electron transfer<sup>131</sup>. Consistency between the particular level of QM treatment and the MM force field is also important. Also, QM/MM simulations typically apply relatively low levels of QM theory, and connecting to higher levels of theory (for example, via embedding or perturbation approaches) can be important to achieve high accuracy. For large systems, models

combining QM/MM and coarse grained methods will be useful<sup>132</sup>. Generating reaction pathways and reactive configurations requires effective and efficient simulation methods and enhanced sampling techniques. A long-cherished goal is to use knowledge of transition states to design enzyme inhibitors, based on Pauling's proposal that transition state analogues should be high affinity ligands. Whilst it is naive to think that this is a universal approach to enzyme inhibition, there is real potential in using structurally detailed knowledge of the interactions of transition state structures in some enzymes, and reaction intermediates in others, to design and optimise binding interactions of drug leads.

Methods for understanding and predicting chemical reactivity in large biological systems also bring into view a range of other exciting possibilities relevant to drug design and development. Simulations of biochemical reactions will be important for understanding the modulation and control of reactivity by conformational effects and allosteric regulation. A fundamentally important theme is understanding how chemical and conformational changes are coupled in biomolecular systems. This is essentially a multiscale problem in itself: it requires knowledge of the role of macromolecular conformational changes in catalytic cycles, and how reactions such as the hydrolysis of ATP drive molecular motors and other biomolecular machines. Linking to larger scales will help in designed manipulation of metabolic cycles and signaling cascades, as noted above. More speculative practical challenges are also coming into sight: the control and manipulation of reactivity within biological systems promises entirely new types of therapy. Enzyme inhibitors are obviously important as pharmaceuticals and, in a few cases, enzymes are used as drugs, such as thrombolytics, and in enzyme replacement and enhancement therapies to correct genetic deficiencies, usually in rare conditions. The activation of prodrugs often depends on enzymes and therefore improved understanding of prodrug–enzyme reactivity will help the in the design and development of all types of directed enzyme prodrug therapy<sup>133,134</sup>. Potentially, engineered or evolved enzymes, catalytic antibodies or hybrid bio/chemocatalysts, could be used to control selective prodrug reactivity in cells. Photodynamic therapy is another area in which electronic structure calculations can potentially aid drug design, understanding and predicting photoactivation and contributing to improved selectivity of photoactivation<sup>135</sup>. More radical is the use of designed catalysts to remodel or destroy biological targets *in situ*<sup>136</sup>. One example is the possibility of gene editing offered by systems such as CRISPR/Cas 9. Application of catalysts in human patients will be accelerated by techniques for understanding and designing determinants of specificity and reactivity and their interactions *in vivo*. Multiscale simulation methods capable of modelling reactions and predicting their effects in complex biological systems will contribute to such developments.

**[H1] The changing role of computational science**—An expanding range of chemical, biological, biophysical, spectroscopic and structural techniques is increasingly integral to drug discovery and design programs. The scale and complexity of the data that they generate demand the concerted development of data-centric computational models to interpret and connect them. The combination of a range of data with multiscale models will provide detailed knowledge of drug targets, including their time dependence and dynamics, transforming our ability to understand and predict drug action. An exciting prospect is the

development of interconnected multiscale models spanning the full range of complexity, from chemical action at a target binding site through complex cellular interactions and beyond. Multiscale models of this scope will also help to identify and analyse adverse drug reactions, arising from drug–drug interactions and off-target effects. Developing reliable, integrated multiscale methods poses significant challenges, but they promise significant payoffs in the drug discovery arena. They will drive the generation of experimentally testable hypotheses and assist in experimental design. Understanding of biology will undoubtedly advance faster and more reliably through effective combination of experimental and multiscale computational science.

The role of biomolecular simulation in drug design and development (and in analysing mechanisms relevant to health and disease) is evolving rapidly. Whereas previously simulation provided simple models to help develop or illustrate hypotheses, increasingly simulations can be used as another form of experiment: a computational experiment or assay. As such, simulators must apply similar standards of statistical rigor in assessing the significance of their findings. Computational assays can be used to assess and predict biological properties, such as drug resistance in mutant systems. Simulations can also be used to explore and analyse processes that are otherwise inaccessible or unachievable with experiment. Ongoing increases in computer power, together with improvements in the reliability and scope of simulation methods (with detailed validation against experiment), mean that computational assays will become ever more important in drug development, offering speed and affordability, and complementing and managing the growing deluge of experimental data.

No single simulation technique can address all the many levels of challenge and understanding required for modelling of biological systems from the molecular to the cellular level: this is the essential driving motivation for the development of multiscale methods. As the examples outlined here highlight, the potential of multiscale modelling and simulation, and of its close integration with experiment, is only just starting to be realized in drug discovery. We expect that, within the next decade, multiscale methods are likely to be central in drug discovery and development programs. They will form the basis of, and inform, cohesive data-rich models for drug–target systems. Together, these will rationalize and synthesize experimental data, accelerate drug development and help discover effective therapeutics with novel mechanisms of action.

## Acknowledgements

Support from NIH DP2 OD007237 and NIH P41 GM103426 to REA is gratefully acknowledged. REA thanks Pek Jeong and Benjamin Jagger for assistance in figure preparation. AJM thanks EPSRC for support (EP/M022609/1 (for CCP-BioSim see [www.ccpbiosim.ac.uk](http://www.ccpbiosim.ac.uk)) and EP/M015378/1, and BBSRC (BB/M000354/1). AJM thanks co-workers including Christopher Woods, Christine Bathelt, Sarah Rouse, Richard Lonsdale and Mark Sansom for help in the preparation of figures.

## References

1. Jorgensen WL The many roles of computation in drug discovery. *Science* 303, 1813–8 (2004). [PubMed: 15031495]
2. Vivo M. De, Masetti M, Bottegoni G & Cavalli A Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem* 59, 4035–4061 (2016). [PubMed: 26807648]

3. Lee EH, Hsin J, Sotomayor M, Comellas G & Schulten K Discovery Through the Computational Microscope. *Structure* 17, 1295–1306 (2009). [PubMed: 19836330]
4. Wassman CD et al. Computational Identification of a Transiently Open L1/S3 Pocket for Reactivation of Mutant p53. *Nat. Commun* 4, 1407 (2013). [PubMed: 23360998]
5. Woods CJ, Malaisree M, Long B, McIntosh-Smith S & Mulholland AJ Computational assay of H7N9 influenza neuraminidase reveals R292K mutation reduces drug binding affinity. *Sci. Rep* 3, 3561 (2013). [PubMed: 24356381]
6. Zhao G et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 497, 643–646 (2013). [PubMed: 23719463]
7. Perilla JR & Schulten K Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat. Commun* 8, 1–10 (2017). [PubMed: 28232747]
8. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen ARJ Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc* 137, 2695–2703 (2015). [PubMed: 25625324]
9. Kulik HJ, Luehr N, Ufimtsev IS & Martínez TJ Ab initio quantum chemistry for protein structures. *J. Phys. Chem. B* 116, 12501–9 (2012). [PubMed: 22974088]
10. Yu I, Mori T, Ando T, Harada R & Jung J Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* 5, 1–22 (2016).
11. Dye A & Kramer W 2016 Annual Report Sustained Petascale in Action (2016).
12. Shaw DE Millisecond-Long Molecular Dynamics Simulations of Proteins on a Special-Purpose Machine. *Biophys. J* 104, 45a (2013).
13. Gray Alan, Harlen Oliver G., Harris Sarah A., Khalid Symba, Leung Yuk Ming, Lonsdale Richard, Mulholland Adrian J., Pearson Arwen R., Read Daniel J., and R. A. R. In pursuit of an accurate spatial and temporal models of biomolecules at the atomistic level: a perspective on computer simulation. *Acta Crystallogr. Sect. D Biol. Crystallogr* 71, 162–172 (2015). [PubMed: 25615870]
14. The Royal Academy of Swedish Sciences. Development of multiscale models for complex chemical systems. The Nobel Prize in Chemistry 2013 50005, (2013).
15. Danev R & Baumeister W Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol* 46, 87–94 (2017). [PubMed: 28675816]
16. Denk W & Horstmann H Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol* 2, (2004).
17. Villa E, Schaffer M, Plitzko JM & Baumeister W Opening windows into the cell: Focused-ion-beam milling for cryo-electron tomography. *Curr. Opin. Struct. Biol* 23, 771–777 (2013). [PubMed: 24090931]
18. Mahamid J et al. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* (80-. ) 351, 969–972 (2016).
19. Plitzko M, Schuler B & Selenko P Structural Biology outside the box — inside the cell. *Curr. Opin. Struct. Biol* 46, 110–121 (2017). [PubMed: 28735108]
20. Larabell CA & Nugent KA Imaging cellular architecture with X-rays. *Curr. Opin. Struct. Biol* 20, 623–631 (2010). [PubMed: 20869868]
21. Wall ME, Adams PD, Fraser JS & Sauter NK Diffuse x-ray scattering to model protein motions. *Structure* 22, 182–184 (2014). [PubMed: 24507780]
22. Neutze R, Branden G & Schertler GFX Membrane protein structural biology using X-ray free electron lasers. *Curr. Opin. Struct. Biol* 33, 115–125 (2015). [PubMed: 26342349]
23. Levantino M, Yorke BA, Monteiro DCF, Cammarata M & Pearson AR Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr. Opin. Struct. Biol* 35, 41–48 (2015). [PubMed: 26342489]
24. Casadei CM et al. Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase. *Science* (80-. ) 345, 193–197 (2014).

25. Shah NH & Muir TW Inteins: nature's gift to protein chemists. *Chem. Sci* 5, 446–461 (2014). [PubMed: 24634716]
26. Konermann L, Pan J & Liu Y-H Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev* 40, 1224–1234 (2011). [PubMed: 21173980]
27. Holdren JP National Strategic Computing Initiative Strategic Plan (2016).
28. Pérez F & Granger BE IPython: A System for Interactive Scientific Computing Python: An Open and General- Purpose Environment. *Comput. Sci. Eng* 9, 21–29 (2007).
29. Jeong PU et al. Progress towards Automated Kepler Scientific Workflows for Computer-aided Drug Discovery and Molecular Simulations. *Procedia Comput. Sci* 29, 1745–1755 (2014). [PubMed: 29399238]
30. Purawat S et al. A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics. *Biophys. J* 112, (2017).
31. Patwardhan A et al. Building bridges between cellular and molecular structural biology. *Elife* 6, e25835 (2017). [PubMed: 28682240]
32. Santos R et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov* 16, 19–34 (2017). [PubMed: 27910877]
33. Gathiaka S et al. D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. *J. Comput. Aided. Mol. Des* 30, 651–668 (2016). [PubMed: 27696240]
34. Woods CJ & Mulholland AJ Multiscale modelling of biological systems. *Chem. Model* 5, 13–50 (2008).
35. Votapka LW & Amaro RE Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLoS Comput. Biol* 11, 1–24 (2015).
36. Chudyk EI et al. QM/MM simulations as an assay for carbapenemase activity in class A  $\beta$ -lactamases. *Chem. Commun. (Camb)* 50, 14736–9 (2014). [PubMed: 25321894]
37. Boras BW et al. Bridging scales through multiscale modeling: A case study on protein kinase A. *Front. Physiol* 6, 1–15 (2015). [PubMed: 25688210]
38. Mih N, Brunk E, Bordbar A & Palsson BO A Multi-scale Computational Platform to Mechanistically Assess the Effect of Genetic Variation on Drug Responses in Human Erythrocyte Metabolism. *PLoS Comput. Biol* 12, 1–24 (2016).
39. Yu JS & Bagheri N Multi-class and multi-scale models of complex biological phenomena. *Curr. Opin. Biotechnol* 39, 167–173 (2016). [PubMed: 27115496]
40. Zhou G, Pantelopulos GA, Mukherjee S & Voelz VA Bridging Microscopic and Macroscopic Mechanisms of p53-MDM2 Binding with Kinetic Network Models. *Biophys. J* 113, 785–793 (2017). [PubMed: 28834715]
41. Muzic M. Le, Autin L, Parulek J & Viola I cellVIEW: a Tool for Illustrative and Multi-Scale Rendering of Large Biomolecular Datasets. in *Eurographics Workshop on Visual Computing for Biology and Medicine* (2015). doi:10.2312/vcbm.20151209
42. Johnson GT et al. cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nat. Methods* 12, 85–91 (2014). [PubMed: 25437435]
43. Durrant JD & Amaro RE LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry. *PLoS Comput. Biol* 10, (2014).
44. Hake J, Kekenes-Huskey PM & McCulloch AD Computational modeling of subcellular transport and signaling. *Curr. Opin. Struct. Biol* 25, 92–97 (2014). [PubMed: 24509246]
45. Kerr RA et al. Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces. *Siam J. Sci. Comput* 30, 3126–3149 (2009).
46. Roberts E Cellular and molecular structure as a unifying framework for whole-cell modeling. *Curr. Opin. Struct. Biol* 25, 86–91 (2014). [PubMed: 24509245]
47. Roberts E, Stone JE & Luthey-Schulten Z Lattice microbes: High-performance stochastic simulation method for the reaction-diffusion master equation. *J. Comput. Chem* 34, 245–255 (2013). [PubMed: 23007888]
48. Earnest TM et al. Challenges of Integrating Stochastic Dynamics and Cryo-Electron Tomograms in Whole-Cell Simulations. *J. Phys. Chem. B* 121, 3871–3881 (2017). [PubMed: 28291359]

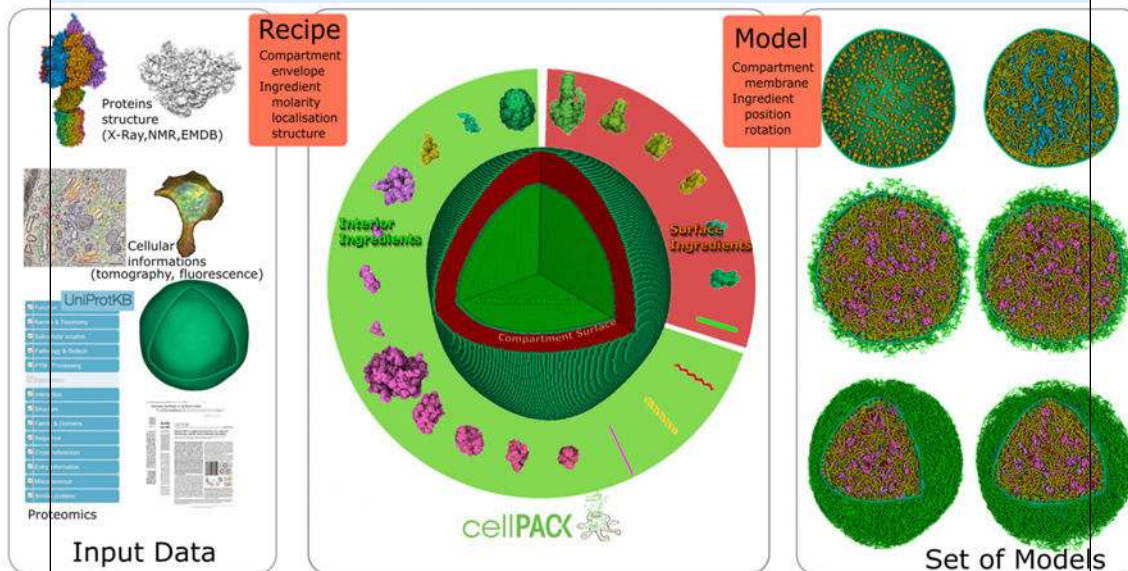
49. Noe F, Biedermann J, Ullrich A & Scho J ReaDDyMM: Fast Interacting Particle Reaction-Diffusion Simulations Using Graphical Processing Units. *Biophys. J* 108, 457–461 (2015). [PubMed: 25650912]
50. Chiricotto M, Sterpone F, Derreumaux P & Melchionna S Multiscale simulation of molecular processes in cellular environments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* 374, 20160225 (2016).
51. Tozzini V Coarse-grained models for proteins. *Curr. Opin. Struct. Biol* 15, 144–150 (2005). [PubMed: 15837171]
52. Oliver R, Read DJ, Harlen OG & Harris SA A Stochastic Finite Element Model for the Dynamics of Globular Macromolecules. *J. Comput. Phys* 239, 147–165 (2013).
53. McGuffee SR & Elcock AH Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol* 6, (2010).
54. Freddolino PL, Arkhipov AS, Larson SB, McPherson A & Schulten K Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* 14, 437–449 (2006). [PubMed: 16531228]
55. Phillips JC et al. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26, 1781–1802 (2005). [PubMed: 16222654]
56. Ou HD et al. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* (80-. ) 357, eaag0025 (2017).
57. Go H, Dans PD & Orozco M Multiscale simulation of DNA. *Curr. Opin. Struct. Biol* 37, 29–45 (2016). [PubMed: 26708341]
58. Skolnick J Perspective : On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *J. Chem. Phys* 145, 100901 (2016). [PubMed: 27634243]
59. Trovato F & Fumagalli G Molecular simulations of cellular processes. *Biophys. Rev* 9, 941–958 (2017). [PubMed: 29185136]
60. Rydzewski J & Nowak W Machine Learning Based Dimensionality Reduction Facilitates Ligand Diffusion Paths Assessment : A Case of Cytochrome P450cam. *J. Chem. Theory Comput* 12, 2110–2120 (2016). [PubMed: 26989997]
61. Stone E Chemistry to go boldly into virtual world. *Chemistry World* (2017).
62. Russel D et al. Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 10, 1–5 (2012).
63. Villa E & Lasker K Finding the right fit: Chiseling structures out of cryo-electron microscopy maps. *Curr. Opin. Struct. Biol* 25, 118–125 (2014). [PubMed: 24814094]
64. Demir O, Jeong PU & Amaro RE Full-length p53 tetramer bound to DNA and its quaternary dynamics. *Oncogene* 36, 1451–1460 (2017). [PubMed: 27641333]
65. The Royal Academy of Swedish Sciences. The development of cryo-electron microscopy 50005, (2017).
66. Willis JR et al. Long antibody HCDR3s from HIV-naïve donors presented on a PG9 neutralizing antibody background mediate HIV neutralization. *Proc. Natl. Acad. Sci. U. S. A* 113, 1518405113-(2016).
67. Bale JB et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* (80-. ) 353, 389 LP-394 (2016).
68. Fleishman SJ et al. Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin 979, (2010).
69. Hsia Y et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* 535, 1–12 (2016).
70. Jansen JM et al. Inhibition of prenylated KRAS in a lipid environment 1–21 (2017).
71. Mouchlis VD, Bucher D, McCammon JA & Dennis E a. Membranes serve as allosteric activators of phospholipase A2, enabling it to extract, bind, and hydrolyze phospholipid substrates. *Proc. Natl. Acad. Sci* 112, E516–E525 (2015). [PubMed: 25624474]
72. Hocker HJ et al. Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. *Proc. Natl. Acad. Sci. U. S. A* 110, 10201–6 (2013). [PubMed: 23737504]



73. Singhal N, Snow CD & Pande VS Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys* 121, 415–425 (2004). [PubMed: 15260562]
74. Swope WC et al. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a beta-Hairpin Peptide. *J. Phys. Chem. B* 108, 6582–6594 (2004).
75. Swope WC, Pitera JW & Suits F Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory †. *J. Phys. Chem. B* 108, 6571–6581 (2004).
76. Meng Y, Shukla D, Pande VS & Roux B Transition path theory analysis of c-Src kinase activation. *Proc. Natl. Acad. Sci* 113, 9193–9198 (2016). [PubMed: 27482115]
77. Malmstrom RD, Kornev AP, Taylor SS & Amaro RE Allostery through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun* 6, 7588 (2015). [PubMed: 26145448]
78. Plattner N, Doerr S, De Fabritiis G & Noé F Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem* 1–7 (2017). doi:10.1038/nchem.2785 [PubMed: 29256516]
79. Kohlhoff KJ et al. Cloud-Based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways. *Nat. Chem* 6, 15–21 (2013). [PubMed: 24345941]
80. Folmer RHA Drug target residence time: a misleading concept. *Drug Discov. Today* 0, 1–5 (2017).
81. Schuetz DA et al. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discov. Today* 22, 896–911 (2017). [PubMed: 28412474]
82. Zheng W, Gallicchio E, Deng N, Andrec M & Levy RM Kinetic network study of the diversity and temperature dependence of Trp-Cage folding pathways: Combining transition path theory with stochastic simulations. *J. Phys. Chem. B* 115, 1512–1523 (2011). [PubMed: 21254767]
83. Plattner N & Noé F Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun* 6, 7653 (2015). [PubMed: 26134632]
84. Segala E et al. Controlling the Dissociation of Ligands from the Adenosine A. *J. Med. Chem* 59, 7167–7176 (2016). [PubMed: 27391254]
85. Mollica L et al. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep* 5, 11539 (2015). [PubMed: 26103621]
86. Tiwary P, Limongelli V, Salvalaglio M & Parrinello M Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci* 112, E386–E391 (2015). [PubMed: 25605901]
87. Juraszek J, Saladino G, Van Erp TS & Gervasio FL Efficient numerical reconstruction of protein folding kinetics with partial path sampling and pathlike variables. *Phys. Rev. Lett* 110, 1–5 (2013).
88. Votapka LW, Jagger BR, Heyneman AL & Amaro RE SEEKR: Simulation Enabled Estimation of Kinetic Rates, a computational tool to estimate molecular kinetics and its application to trypsin-benzamidine binding. *J. Phys. Chem. B* 121, 3597–3606 (2017). [PubMed: 28191969]
89. Faradjian AK & Elber R Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys* 120, 10880–10889 (2004). [PubMed: 15268118]
90. Swift RV & Amaro RE Back to the Future: Can Physical Models of Passive Membrane Permeability Help Reduce Drug Candidate Attrition and Move Us Beyond QSPR? *Chem. Biol. Drug Des* 81, (2013).
91. Lee CT et al. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Model* 56, 721–733 (2016). [PubMed: 27043429]
92. Parkin J, Chavent M & Khalid S Molecular Simulations of Gram-Negative Bacterial Membranes: A Vignette of Some Recent Successes. *Biophys. J* 109, 461–8 (2015). [PubMed: 26244728]
93. Berglund NA et al. Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of *E. coli*: A Molecular Dynamics Study. *PLOS Comput. Biol* 11, e1004180 (2015). [PubMed: 25885324]
94. Pavlova A, Hwang H, Lundquist K, Balusek C & Gumbart J Living on the edge: Simulations of bacterial outer-membrane proteins. *Biochim. Biophys. Acta - Proteins Proteomics* 1858, 1753–1759 (2016).

95. Orsi M, Noro MG & Essex JW Dual-resolution molecular dynamics simulation of antimicrobials in biomembranes. *J. R. Soc. Interface* 8, 826–41 (2011). [PubMed: 21131331]
96. Dickson CJ, Hornak V, Pearlstein RA & Duca JS Structure–Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling. *J. Am. Chem. Soc* 139, 442–452 (2017). [PubMed: 27951634]
97. Cardenas AE et al. Unassisted transport of N -acetyl- l -tryptophanamide through membrane: Experiment and simulation of kinetics. *J. Phys. Chem. B* 116, 2739–2750 (2012). [PubMed: 22313494]
98. Votapka LW, Lee CT & Amaro RE Two Relations to Estimate Membrane Permeability Using Milestoning. *J. Phys. Chem. B* 120, (2016).
99. Wu H, Paul F, Wehmeyer C & Noé F Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci* 114, E3221–E3230 (2017).
100. Boninsegna L et al. A data-driven perspective on the hierarchical assembly of molecular structures. *J. Chem. Theory Comput* in press, (2017).
101. Dama JF et al. The Theory of Ultra-Coarse-Graining. 1. General Principles. *J. Chem. Theory Comput* 9, 2466–2480 (2013). [PubMed: 26583735]
102. Kmiecik S et al. Coarse-Grained Protein Models and Their Applications. *Chem. Rev* 116, 7898–7936 (2016). [PubMed: 27333362]
103. Graham JA, Essex JW & Khalid S PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories. *J. Chem. Inf. Model* 57, 650–656 (2017). [PubMed: 28345910]
104. Sharma S et al. A coarse-grained protein model in a water-like solvent. *Sci. Rep* 3, 3–8 (2013).
105. Altae-tran H, Ramsundar B, Pappu AS & Pande V Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci* 3, 283–293 (2017). [PubMed: 28470045]
106. Cui Q Perspective: Quantum mechanical methods in biochemistry and biophysics. *J. Chem. Phys* 145, 140901 (2016). [PubMed: 27782516]
107. Ranaghan KE et al. A catalytic role for methionine revealed by a combination of computation and experiments on phosphite dehydrogenase. *Chem. Sci* 5, 2191–2199 (2014).
108. Lodola A et al. Identification of productive inhibitor binding orientation in fatty acid amide hydrolase (FAAH) by QM/MM mechanistic modelling. *Chem. Commun. (Camb)* 214–6 (2008). doi:10.1039/b714136j [PubMed: 18092091]
109. Lonsdale R, Rouse SL, Sansom MSP & Mulholland AJ A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes. *PLoS Comput. Biol* 10, (2014).
110. Lonsdale R et al. Quantum mechanics/molecular mechanics modeling of regioselectivity of drug metabolism in cytochrome P450 2C9. *J. Am. Chem. Soc* 135, 8001–8015 (2013). [PubMed: 23641937]
111. Tyzack JD, Hunt PA & Segall MD Predicting regioselectivity and lability of cytochrome P450 metabolism using quantum mechanical simulations. *J. Chem. Inf. Model* 56, 2180–2193 (2016). [PubMed: 27753488]
112. Goodpaster JD, Barnes TA, Manby FR & Miller TF Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys* 140, (2014).
113. Barnes TA, Goodpaster JD, Manby FR & Miller TF Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys* 139, (2013).
114. Bennie SJ et al. Multiscale analysis of enantioselectivity in enzyme-catalysed ‘ lethal synthesis ‘ using projector-based embedding. *R. Soc. Open Sci* 5, 171390 (2018). [PubMed: 29515856]
115. Kazemi M, Himo F & Aqvist J Enzyme catalysis by entropy without Circe effect. *Proc Natl Acad Sci U S A* 113, 2406–2411 (2016). [PubMed: 26755610]
116. Bennie SJ et al. A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *J. Chem. Theory Comput* 12, 2689–2697 (2016). [PubMed: 27159381]
117. Ryde U & Soderhjelm P Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev* 116, 5520–5566 (2016). [PubMed: 27077817]

118. Woods CJ, Shaw KE & Mulholland AJ Combined quantum mechanics/molecular mechanics (QM/MM) simulations for protein-ligand complexes: Free energies of binding of water molecules in influenza neuraminidase. *J. Phys. Chem. B* 119, 997–1001 (2015). [PubMed: 25340313]
119. Genheden S, Cabedo Martinez AI, Criddle MP & Essex JW Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. *J. Comput. Aided. Mol. Des* 28, 187–200 (2014). [PubMed: 24488307]
120. König G, Pickard IV FC, Mei Y & Brooks BR Predicting hydration free energies with a hybrid QM/MM approach: An evaluation of implicit and explicit solvation models in SAMPL4. *J. Comput. Aided. Mol. Des* 28, 245–257 (2014). [PubMed: 24504703]
121. Ko G et al. Computation of Hydration Free Energies Using the Multiple Environment Single System Quantum Mechanical / Molecular Mechanical Method. *J. Chem. Theory Comput* 12, 332–344 (2016). [PubMed: 26613419]
122. Ho MH, De Vivo M, Peraro MD & Klein ML Unraveling the catalytic pathway of metalloenzyme farnesyltransferase through QM/MM computation. *J. Chem. Theory Comput* 5, 1657–1666 (2009). [PubMed: 26609858]
123. Woods CJ, Shaw KE, M. A. Combined quantum mechanics/molecular mechanics (QM/MM) simulations for protein-ligand complexes: free energies of binding of water molecules in influenza neuraminidase. *J. Phys. Chem. B* 119, 997–1001 (2015). [PubMed: 25340313]
124. Smith JS, Isayev O & Roitberg AE ANI-1 : an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci* 8, 3192–3203 (2017). [PubMed: 28507695]
125. Doemer M, Maurer P, Campomanes P, Tavernelli I & Rothlisberger U Generalized QM / MM Force Matching Approach Applied to the 11-cis Protonated Schiff Base Chromophore of Rhodopsin. *Journal* 10, 412–422 (2014).
126. Singh J, Petter RC, Baillie TA & Whitty A The resurgence of covalent drugs. *Nat. Rev. Drug Discov* 10, 307–317 (2011). [PubMed: 21455239]
127. Callegari D et al. L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chem. Sci* DOI: 10.10, 1–10 (2018).
128. Kamerlin SCL & Warshel A The empirical valence bond model: Theory and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci* 1, 30–45 (2011).
129. Ryde U QM/MM Calculations on Proteins. *Methods in Enzymology* 577, (Elsevier Inc, 2016).
130. Karelina M & Kulik HJ Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput* 13, 563–576 (2017). [PubMed: 28068092]
131. Mones L et al. The adaptive buffered force QM/MM method in the CP2K and amber software packages. *J. Comput. Chem* 36, 633–648 (2015). [PubMed: 25649827]
132. Sokkar P, Boulanger E, Thiel W & Sanchez-Garcia E Hybrid quantum mechanics/molecular mechanics/coarse grained modeling: A triple-resolution approach for biomolecular systems. *J. Chem. Theory Comput* 11, 1809–1818 (2015). [PubMed: 26574388]
133. Weidle UH, Tiefenthaler G & Georges G Proteases as activators for cytotoxic prodrugs in antitumor therapy. *Cancer Genomics and Proteomics* 11, 67–80 (2014). [PubMed: 24709544]
134. Schellmann N, Deckert PM, Bachran D, Fuchs H & Bachran C Targeted Enzyme Prodrug Therapies. *Mini Rev. Med. Chem* 10, (2010).
135. Bown SG Photodynamic therapy for photochemists. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* 371, 20120371–20120371 (2013).
136. Davis B et al. Selective protein degradation by ligand-targeted enzymes: towards the creation of catalytic antagonists. *Chembiochem* 4, 533–537 (2003). [PubMed: 12794865]
137. Lomize MA, Lomize AL, Pogozheva ID & Mosberg HI OPM: Orientations of proteins in membranes database. *Bioinformatics* 22, 623–625 (2006). [PubMed: 16397007]
138. Cavalli A, Spitaleri A, Saladino G & Gervasio FL Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Acc. Chem. Res* 48, 277–285 (2015). [PubMed: 25496113]

**Box 1:****Multiscale Data-Driven System Assembly**

CellPACK<sup>42</sup> is a system-construction framework allowing the integration of information from cellular ultrastructure, genomics and proteomics, and atomic structure to build integrated, data-centric 3D models of subcellular environments at molecular detail.

**Data Input Sources**

CellPack can connect to (ingest) many types of relevant datasets such as proteomics and genomics (for contents, stoichiometry of molecules in various compartments), X-ray crystallography (for atomic-resolution of molecular components), as well as various types of electron microscopies and tomographies (for macromolecular assemblies, subcellular and cellular ultrastructure).

**Compartment Construction Methods**

Once the compartments have been defined, polygonal models define membrane ultrastructure. If atomic membranes are desired, atomic structures of lipid membrane are modeled with LipidWrapper<sup>43</sup>. Membrane-embedded selective orientation and placement of membrane-bound macromolecules based on the Orientations of Proteins in Membranes (OPM) methodology<sup>137</sup>.

Solid voxelisation techniques are used to discretize the volume to be filled. Afterwards, crowded soluble environments are constructed with a size-priority voxel based method; where needed, schematic fibrous structures can be modeled with a random walk algorithm that “grows” the structures *in situ*.

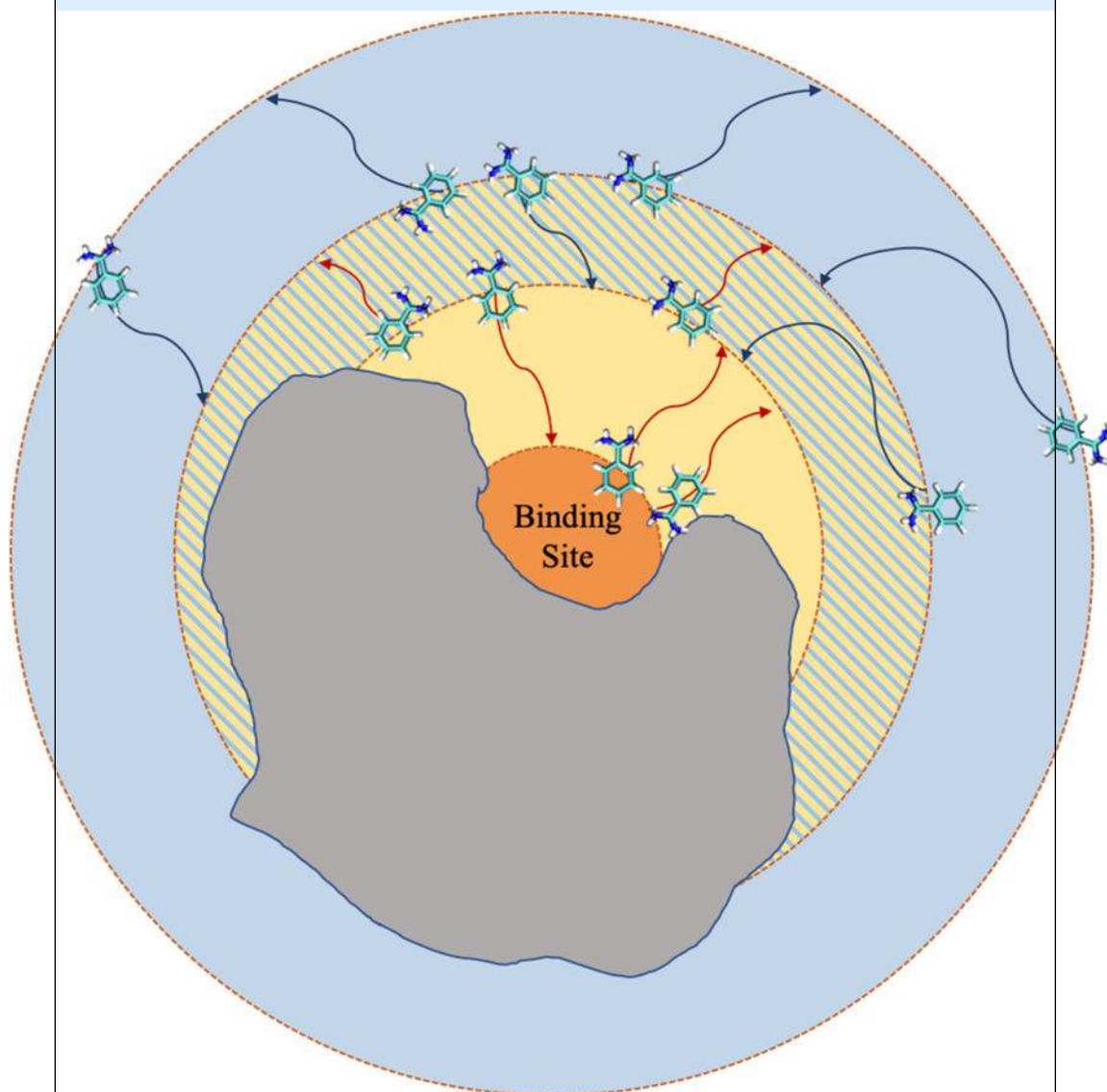
**Recipe Parameterization**

The various data input sources and methods are integrated into a “recipe” that describes how to build a model. Namely, the recipe defines the “what” (ingredient description

extracted from proteomics and structural biology), the “where” (cellular localization, *e.g.*, surface, interior, organelle, etc.), the “how” (packing methods and constraints) and “how many” (based on molarity or copy number extracted from literature or experimentation).

### **Model Generation**

Based on the recipe, and given a random seed, cellPACK generates a unique 3D model, which is simply a list of ingredients and their respective positions and orientations. By sampling different parameters (concentration, priority, order, packing methods) CellPack can easily generate hundreds of models to study statistical distributions, taking molecular models a step closer towards addressing biological heterogeneity, in hopes of understanding and, ultimately, accurately predicting emergent phenomena.

**Box 2:****A Multiscale Method for Drug Binding and Residence Time Prediction.**

SEEKR is a directly multiscale simulation approach that combines fully atomistic molecular dynamics simulations and rigid body Brownian dynamics simulations (i.e. two different dynamical propagators) with milestoning theory to calculate association and dissociation rates as well as binding free energies for protein ligand complexes. Fictitious ‘milestone’ surfaces (depicted with different colours) are placed at increasing distances from the determined ligand binding site. Short simulations are initiated from each milestone surface and are monitored until they touch an adjacent surface, where they are subsequently terminated. Each milestone surface is assigned a particular simulation type. These types can encompass a broad range of methods, as milestoning theory is agnostic to the simulation type implemented. Typical SEEKR simulations span the spectrum of

potential simulation detail, implementing fully atomistic MD regions and rigid body BD regions. This approach has demonstrated significant increases in both speed and accuracy for the calculation of kinetic parameters compared to “brute force” molecular dynamics simulations, which often require simulation timescales that are inaccessible by conventional computer simulations. The combination of a multiscale simulation approach with milestoning enables SEEKR to predict ligand dissociation events in a statistically robust and computationally efficient manner, facilitating the accurate computation of drug residence times, a particularly important property linked to the *in vivo* efficacy of drug molecules.

Yellow regions in the figure represent regions that are simulated using fully atomistic molecular dynamics simulations (red arrows represent these trajectories). These regions correspond to milestones close to the binding site, where an atomistic description of molecular interactions is required to model of the binding and unbinding processes. Blue regions instead are simulated using Brownian dynamics. These correspond to the outermost milestones of the figure, where atomistic detail is less important and rigid body dynamics with implicit solvent can rapidly sample long trajectories. The green region represents an area sampled by both molecular dynamics and Brownian dynamics simulations, which are brought together with milestoning theory.

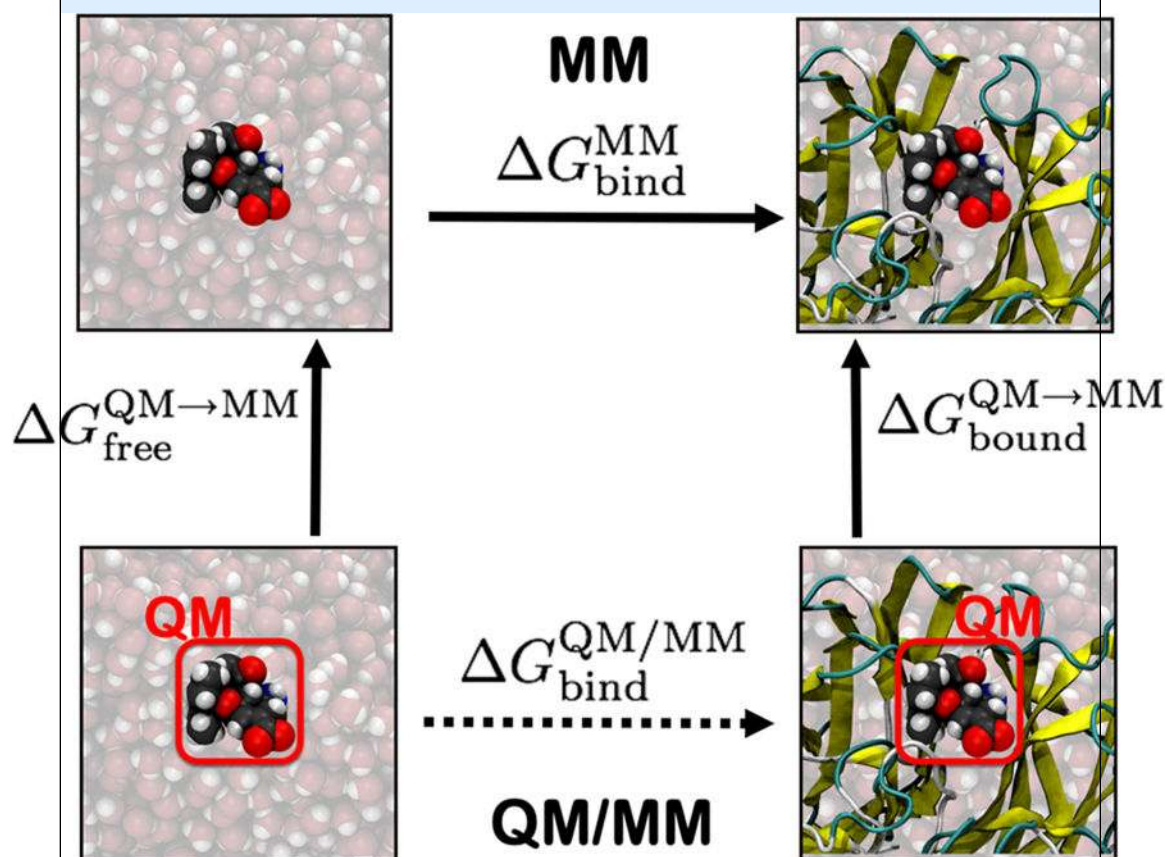
A typical milestoning procedure consists of multiple simulation steps. First an equilibrium distribution for each milestone surface is obtained, typically through umbrella sampling. The equilibrium distribution is then converted to a first hitting point distribution (FHPD) by initiating simulations from each point in the equilibrium distribution with reversed velocities to determine if they had been in a previous state. Any trajectories that cross the same milestone surface they were initiated from are excluded from the FHPD. Finally, trajectories are initiated from each point in the FHPD and allowed to propagate forward in time until it crosses another milestone surface. The trajectory is then terminated, and the transition, as well as the transition time, is recorded. This data can be used to compute a transition kernel,  $\mathbf{K}$ , where each element is computed by the formula:

$$K_{ij} = \frac{n_{i \rightarrow j}}{\sum_k n_{i \rightarrow k}}$$

The mean first passage time (MFPT),  $\langle \tau \rangle$ , can then be computed by the equation:

$$\langle \tau \rangle = p \cdot (I - \hat{K})^{-1} \langle \tau \rangle$$

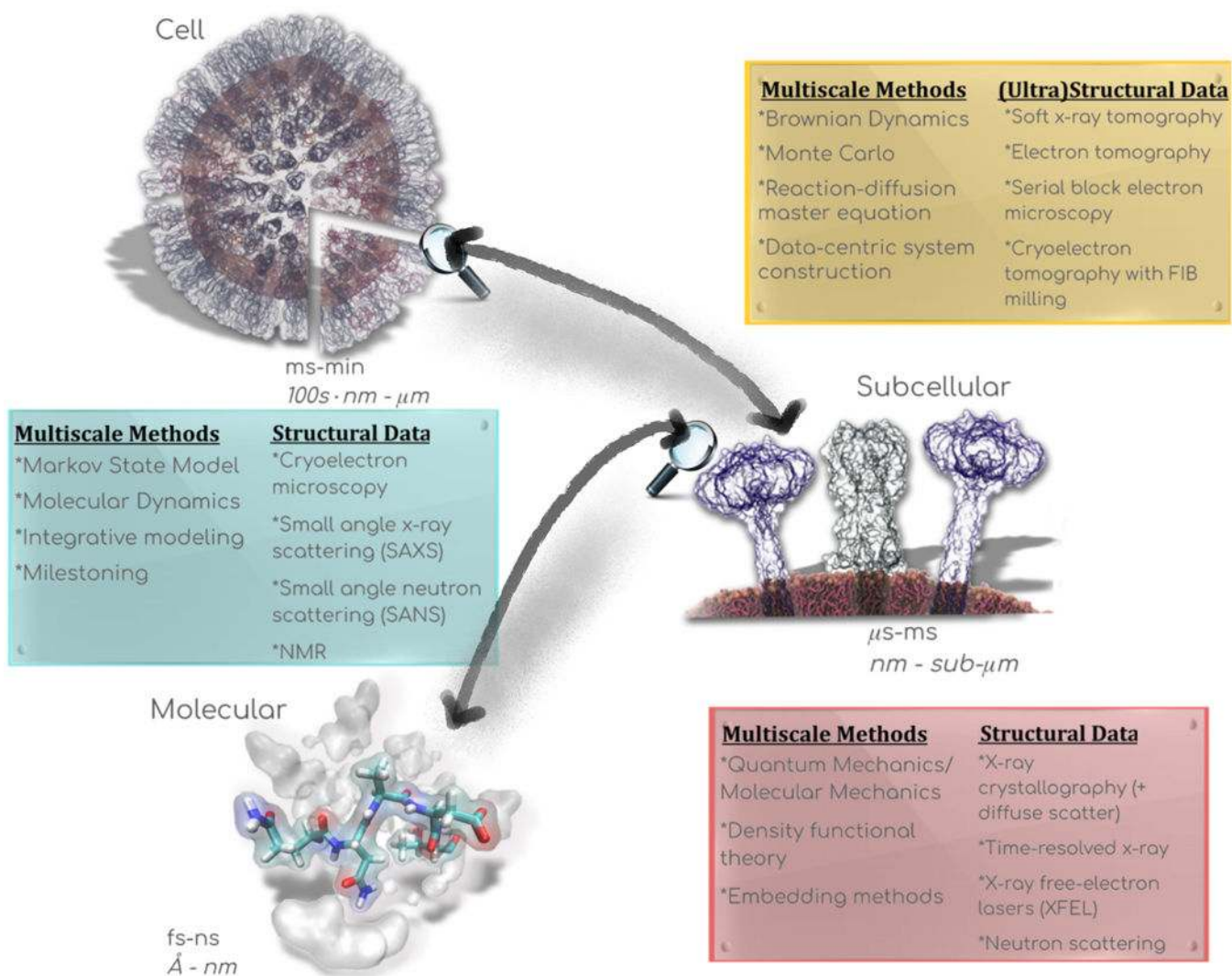
Where  $p$  is the initial probability distribution and  $I$  is the identity matrix. The unbinding rate constant,  $k_{\text{off}}$ , is then simply the inverse of the calculated MFPT.

**Box 3:****Combining QM/MM and MM simulations for calculations of Protein-Ligand Binding Affinity.**

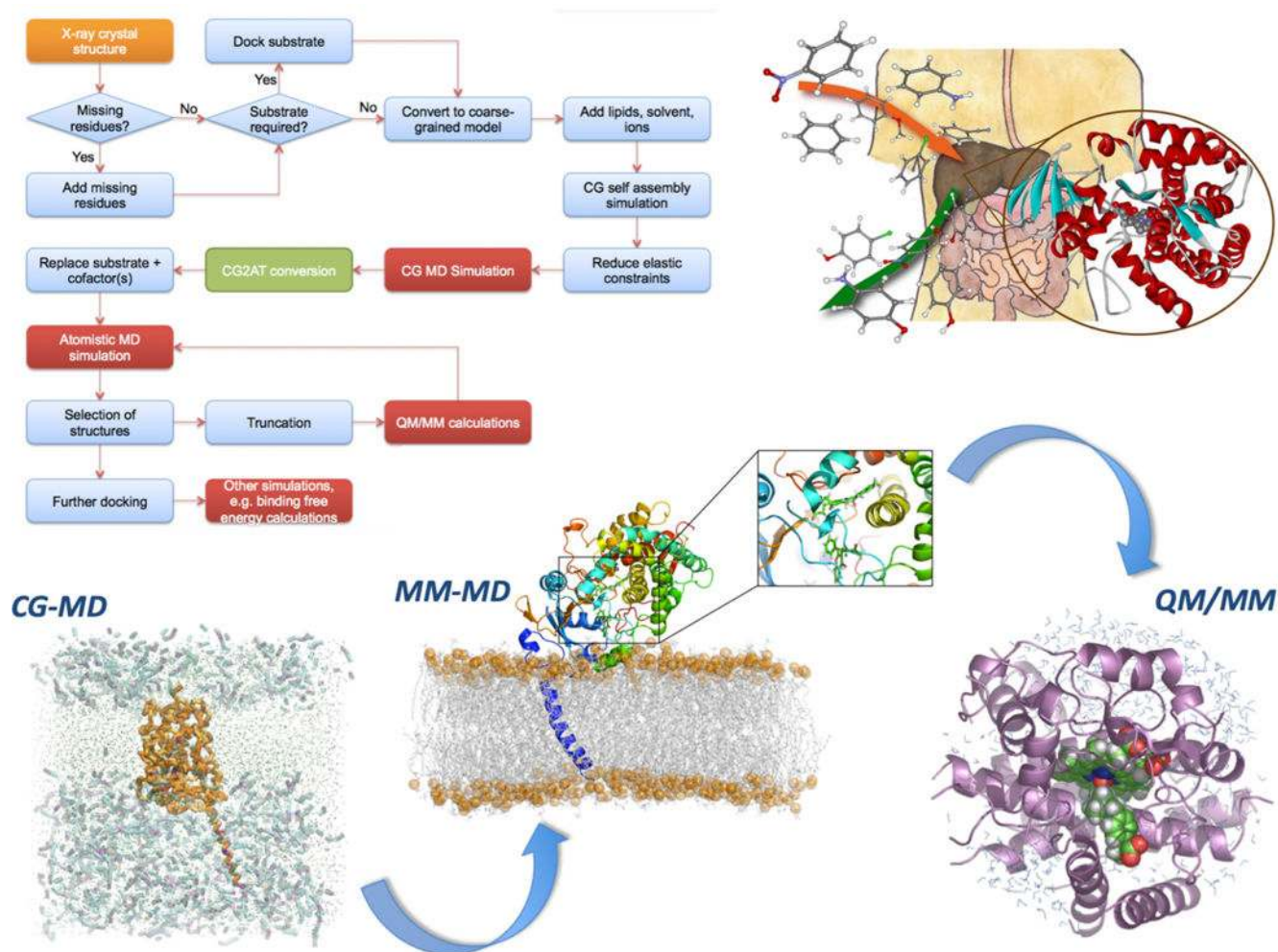
Multiscale sampling allows for efficient connection of high and low levels of simulation, combining a more accurate, high level method, with the greater computational efficiency of a lower-level method<sup>34, 115, 120</sup>. Free energy calculations are increasingly important in drug discovery and development. Typically, the relative binding affinity of different ligands is calculated (for example, via free energy perturbation simulations); alternatively, the absolute binding free energy of a ligand to its protein target,  $\Delta G_{\text{bind}}$ , can be calculated, as shown here, (for example using enhanced sampling molecular dynamics methods). Typically, such calculations apply empirical ‘molecular mechanics’ models, which are computationally efficient and provide atomic-level detail, but have limitations in their description of molecular interactions (due for example to the use of an invariant atomic charge model). Methods based on quantum mechanics (QM) are potentially more accurate because they treat the electronic structure of molecules, but they are much more computationally expensive, making direct application in free energy simulations impractical. A coupled multiscale approach, combining quantum mechanics (QM) calculations with MM simulations, allows protein-ligand binding affinities to be calculated at the combined QM/MM level, which may help overcome limitations of empirical MM force fields, potentially improving the accuracy of predictions, and



improving models. This is achieved by the thermodynamic cycle shown here, where the free energy change for going from a QM to a MM description of the ligand ( $\Delta G^{\text{QM} \rightarrow \text{MM}}$ ) is calculated, here for the ligand when bound to the protein, and when free in solution. Replica exchange thermodynamic integration can be used to calculate ( $\Delta G^{\text{MM} \rightarrow \text{QM}}$ ) efficiently. This transformation from a MM to a QM representation of the molecule takes advantage of efficient sampling at the lower (MM) level. The Metropolis-Hastings algorithm is used to accept a configuration into the higher level (QM/MM) ensemble. Replica exchange simulations across the QM $\leftrightarrow$ MM coordinate enhance sampling, and free energy differences can be calculated by thermodynamic integration, or other methods. Combined with the MM free energy of binding, these give the free energy of binding,  $\Delta G_{\text{bind}}$  at the QM/MM level. This can test (and correct) limitations of MM methods such as lack of changes in electronic polarization<sup>115–118</sup>. This type of scheme also allows rigorous connection between different levels in generalized multiscale simulation schemes.



**Figure 1: Multiscale structure- and physics-based methods bridging from atoms to cells.** Emerging multiscale computational methods coupled with increasingly accurate structural data on biological and chemical systems enables the development of highly detailed and predictive models of drug action across spatial scales ranging from angstroms to microns and temporal scales ranging from femtoseconds to minutes. Such approaches can be gainfully used to address a number of outstanding challenges in drug discovery and design (Table 1).



**Figure 2: Multiscale Simulation Methods to Predict Drug Metabolism by Cytochrome P450 Enzymes.**

Cytochrome P450 enzymes (CYPs) play a central role in metabolizing most drugs. Understanding their reactivity and selectivity is a central goal in predicting drug metabolism, and provides an example of a drug development challenge requiring multiscale simulation approaches. The flowchart shows a practical workflow for multiscale modeling of metabolic reactions of pharmaceuticals in CYPs<sup>109</sup>. Mammalian CYPs are membrane-bound enzymes, but typically only the structures of the soluble portions are determined experimentally, lacking the membrane-anchoring helix, and the membrane. The intact CYP, *in situ*, can be modeled by adding the transmembrane helix and assembling the membrane around the protein, which occurs spontaneously in coarse-grained (CG) molecular dynamics (MD) simulations. CG methods allow MD simulations on timescales of microseconds to milliseconds, showing how the protein is oriented in the membrane and how drug molecules such as warfarin move through the membrane and associate with the protein. Understanding how the drug accesses and binds within the active site requires more detailed, fully-flexible MD simulations with an atomistic molecular mechanics (MM) representation in which every atom in the simulation is represented explicitly, in contrast to the representation of amino acids by a small number of ‘beads’ that group atoms together at the CG level. The CG model

is converted into an atomistic (AT) model, which can be used for MD simulations of drug binding. Typical MM methods cannot be used to model chemical reactivity, therefore for potentially reactive poses of the enzyme/drug complex, the system is converted to a quantum mechanics/molecular mechanics (QM/MM) model, in which the reactive Compound I and the drug are included in the QM region for modeling of chemical reactions of the drug in the enzyme.

**Table 1:**

## Drug Discovery Challenges and Multiscale Computational Methods that Address Them

Drug Discovery Challenge	Relevant Multiscale Computational Method(s) and Examples
Mechanism of Action: Small molecule or protein-protein binding, transition or activation pathway analysis	Combined MD/BD (SEEKR) <sup>88</sup> , Markov state models (MSM) <sup>78</sup>
Mechanism of Action: Reaction mechanisms. Covalent inhibition/binding/modification	Combined quantum mechanics/molecular mechanics (QM/MM) <sup>5,36,115</sup> , including embedding methods <sup>116</sup> ; empirical valence bond methods
Mechanism of Action: Off-target effects	MD, BD of large-scale multi-component systems <sup>6,10,11</sup> , reaction diffusion master equation <sup>47,49</sup> , lattice Boltzmann MD <sup>50</sup>
Predicting Drug Resistance	QM/MM (e.g. for antimicrobial resistance due to beta-lactamases <sup>36</sup> , and for covalent binders <sup>124</sup> ); MD with free energy calculations (e.g. influenza neuraminidase) <sup>5</sup>
Drug Residence Time	SEEKR <sup>88</sup> , metadynamics <sup>138</sup> , MSM <sup>83</sup>
CYP P450 Drug Metabolism	Combined coarse grained MD, MD, QM/MM <sup>109,110</sup>
Small Molecule Membrane Permeability	Milestoning <sup>97,98</sup> , mixed coarse-grained/all-atom approaches <sup>95</sup> , combined PMF/MSM method <sup>96</sup>