## Article Title: Biomolecular simulations: from dynamics and mechanisms to computational assays of biological activity

## Article Type: Overview

## Authors:

*David J. Huggins and Adrian J. Mulholland are co-corresponding authors

| |
|---|
| **David J. Huggins*** |
| University of Cambridge, TCM Group, Cavendish Laboratory, 19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom |
| Unilever Centre, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW, United Kingdom |
| djh210@cam.ac.uk |
| This author declares no conflict of interest |
| **Philip C. Biggin** |
| Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom |
| philip.biggin@bioch.ox.ac.uk |
| This author declares no conflict of interest |
| **Marc A. Dämgen** |
| Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom |
| marc.daemgen@bioch.ox.ac.uk |
| This author declares no conflict of interest |
| **Jonathan W. Essex** |
| School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom |
| Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom |
| J.W.Essex@soton.ac.uk |
| This author declares no conflict of interest |
| **Sarah A. Harris** |
| School of Physics and Astronomy and Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom |
| S.A.Harris@leeds.ac.uk |
| This author declares no conflict of interest |
| **Richard H. Henchman** |
| Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom |
| School of Chemistry, The University of Manchester, Oxford Road, M13 9PL, United Kingdom |
| henchman@manchester.ac.uk |
| This author declares no conflict of interest |
| **Syma Khalid** |
| School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom |
| Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, United |

Kingdom
S.Khalid@soton.ac.uk
This author declares no conflict of interest

**Antonija Kuzmanic**
Department of Chemistry, University College London, London WC1E 6BT, United Kingdom.
a.kuzmanic@ucl.ac.uk
This author declares no conflict of interest

**Charles A. Laughton**
School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
Charles.laughton@nottingham.ac.uk
This author declares no conflict of interest

**Julien Michel**
School of Chemistry, Joseph Black Building, University of Edinburgh, David Brewster Road, Edinburgh, EH9 3FJ, United Kingdom
julien.michel@ed.ac.uk
This author declares no conflict of interest

**Adrian J. Mulholland***
Centre of Computational Chemistry, School of Chemistry, Cantock's Close, University of Bristol, BS8 1TS, UK.
adrian.mulholland@bristol.ac.uk
This author declares no conflict of interest

**Edina Rosta**
Department of Chemistry, King's College London, Room 113, Britannia House, 7 Trinity Street, London, SE1 1DB
edina.rosta@kcl.ac.uk
This author declares no conflict of interest

**Mark S.P. Sansom**
Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom
mark.sansom@bioch.ox.ac.uk
This author declares no conflict of interest

**Marc W. van der Kamp**
Centre of Computational Chemistry, School of Chemistry, Cantock's Close, University of Bristol, BS8 1TS, UK
School of Biochemistry, Biomedical Sciences Building, University Walk, University of Bristol, BS8 1TD, UK.
Marc.VanderKamp@bristol.ac.uk
This author declares no conflict of interest

## Abstract

Biomolecular simulation is increasingly central to understanding and designing biological molecules and their interactions. Detailed, physics-based simulation methods are demonstrating rapidly growing impact in areas as diverse as biocatalysis, drug delivery, biomaterials, biotechnology, and drug design. Simulations offer the potential of uniquely detailed, atomic-level insight into mechanisms, dynamics and processes, as well as increasingly accurate predictions of molecular properties. Simulations can now be used as computational assays of biological activity, e.g. in predictions of drug resistance. Methodological and algorithmic developments, combined with advances in computational hardware, are transforming the scope and range of calculations. Different types of methods are required for different types of problem. Accurate methods and extensive simulations promise quantitative comparison with experiments across biochemistry. Atomistic simulations can now access experimentally relevant timescales for large systems, leading to a fertile interplay of experiment and theory and offering unprecedented opportunities for validating and developing models. Coarse-grained methods allow studies on larger length- and timescales, and theoretical developments are bringing electronic structure calculations into new regimes. Multiscale methods are another key focus for development, combining different levels of theory to increase accuracy, aiming to connect chemical and molecular changes to macroscopic observables. In this review, we outline biomolecular simulation methods highlight examples of its application to investigate questions in biology.

## Graphical/Visual Abstract and Caption
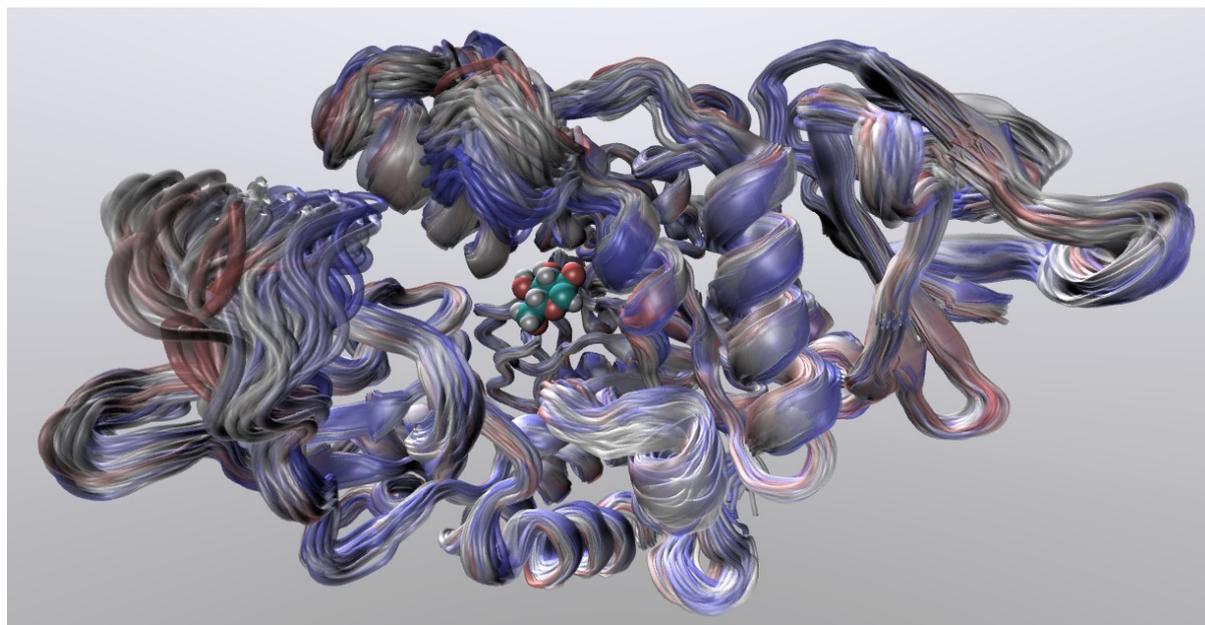


Biomolecular simulations reveal mechanisms, dynamics and interactions of biological molecules. Here, molecular dynamics simulation of the enzyme MalL reveal structural and dynamical changes in the protein during its catalytic cycle: these simulations allow calculation of the activation heat capacity which explains the optimum temperature of its catalytic activity

## Introduction

Biomolecular simulations are now making significant contributions to a wide variety of problems in drug discovery, drug development, biocatalysis, biotechnology, nanotechnology, chemical biology, and medicine. Biomolecular simulation is a rapidly growing field in scale and impact, increasingly demonstrating its worth in understanding mechanisms and analysing activities, and contributing to the design of drugs and biocatalysts. Physics-based simulations complement experiments in building a molecular level understanding of biology: they can test hypotheses and interpret and analyse experimental data in terms of interactions at the atomic level. Different types of simulation techniques have been developed, which are applicable to a range of different problems in biomolecular science. Simulations have already shown their worth in helping to analyse how enzymes catalyse biochemical reactions, and how proteins adopt their functional structures e.g. within cell membranes. They contribute to the design of drugs and catalysts, and in understanding the molecular basis of disease. Simulations have played a key role in developing the conceptual framework now at the heart of biomolecular science, that is, the understanding that the way that biological molecules move and flex - their dynamics - is central to their function. Developing methods from chemical physics and computational science will open exciting new opportunities in biomolecular science, including in drug design and development, biotechnology and biocatalysis. With high performance computing resources, atomistic simulations of large systems are possible, e.g. large-scale simulations of biological machines such as the ribosome, proton pumps and motors, membrane receptor complexes and even whole viruses. Useful simulations of smaller systems can be carried out with desktop resources, thanks to developments allowing e.g. graphics processing units (GPUs) to be used. A particular challenge across the field is the integration of simulations across length and timescales: different types of simulation method are required for different types of problems.(Amaro & Mulholland, 2018)

Biomolecular systems pose fundamental scientific challenges (e.g. protein folding, enzyme catalysis, gene regulation, disease mechanisms and antimicrobial resistance) and are at the heart of many advanced technological developments (drug discovery, biotechnology, biocatalysis, biomaterials and genetic engineering). Biomolecules, and biomolecular systems, are inherently complex and pose significant challenges in modelling. An essential underlying paradigm is the need to consider biomolecular ensembles, and their dynamics, rather than simply static biomolecular structures to understand and predict their behavior and properties. Simulations have been essential in developing this view, complementing experiments. X-ray crystallography is at the heart structural biology (Schröder, Levitt, & Brunger, 2010) but provides a static structure with limited dynamical information and usually in the non-physiological conditions of a low-temperature crystal.(Fraser et al., 2011) Nuclear magnetic resonance (NMR) reveals the heterogeneous structure of biomolecular systems and important details of dynamics,(Lindorff-Larsen, Best, DePristo, Dobson, & Vendruscolo, 2005) but structures may be incompletely determined, limited to moderate sizes, and at high, non-physiological concentrations. Cryo-electron microscopy is advancing rapidly, transforming the scope of structural biology by determining structures of larger systems at resolutions that approach atomic detail, and useful ensembles,(Bai, McMullan, & Scheres, 2015; Frank, 2002) without the need for crystals, but at cold, non-physiologically temperatures and typically lacking sufficient resolution to show dynamics. Biomolecular simulation complements these experimental techniques, e.g. by analysing the dynamics and conformational behaviour of proteins using experimentally determined structures as starting points, and analysing their interactions and mechanisms in atomic detail.

Simulation can probe the short timescale, short length-scale regime which experiment struggles to reach and (given sufficient sampling) can directly provide ensembles, which can in turn be used to calculate the changes in free energy and thus the equilibrium properties (see section on **Computation of Free Energy, Enthalpy and Entropy**). There are two main techniques to sample biomolecular ensembles: molecular dynamics (MD) and Monte Carlo (MC) simulations. MD simulations involve solving Newton's equations of motion (Alder & Wainwright, 1959; Martin Karplus & McCammon, 2002; McCammon, Gelin, & Karplus, 1977) whereas MC operates by accepting structures, generated by small moves, based on probability criteria.(Hastings, 1970; Li & Scheraga, 1987) A large number of biomolecular simulation software packages have been developed. Some widely used packages are presented in Table 1.

**Table 1** – Some examples of widely-used biomolecular simulation software packages

| Software Package | Technique | Website |
|---|---|---|
| CHARMM(Brooks et al., 2009) | MD | charmm.org |
| AMBER(Case et al., 2005) | MD | ambermd.org |
| GROMACS(Van Der Spoel et al., 2005) | MD | gromacs.org |
| MCPRO(Jorgensen & Tirado–Rives, 2005) | MC | cemcomco.com/BOSS_and_MCPRO_Distribution 125.html |
| NAMD(Phillips et al., 2005) | MD | ks.uiuc.edu/Research/namd |
| ProtoMS(Bodnarchuk et al., 2017) | MC | protoms.org |
| Desmond(Bowers et al., 2006) | MD | deshawresearch.com/resources_desmond.html |
| OpenMM(Eastman et al., 2012) | MD | openmm.org |
| GROMOS(Scott et al., 1999) | MD | gromos.net |
| DLPoly(Todorov, Smith, Trachenko, & Dove, 2006) | MD | scd.stfc.ac.uk/Pages/DL_POLY-FAQs.aspx |
| Sire(Woods, Mey, Calabro, & Michel, 2016) | MC/MD | siremol.org |

Owing to the vast energy landscapes of biomolecular systems and the often large energetic barriers between different states, the majority of these software packages have developed methods to expedite moving through configuration space (see section **Enhanced Sampling**). An accurate picture of the relevant configuration space also allows researchers to study transitions between different states and calculate kinetic properties (see section **Measuring Molecular Timescales**). MC and MD both require a potential function (force field) to model the interactions between atoms and calculate energy and forces from the atomic structure.(Levitt & Lifson, 1969; Lifson & Warshel, 1968) Typically, simple empirical 'molecular mechanics' (MM) forcefields are used. Electronic structure calculations on whole proteins are possible, but remain very computationally demanding: e.g. ab initio MD(Carloni, Rothlisberger, & Parrinello, 2002) will not be discussed here, as current methodology has not yet reached the lengthscales and timescales needed to routinely characterize large and complex biomolecular systems. The choice of an appropriate force field is one of the key ingredients in producing useful biomolecular simulations.(W. Wang, Donini, Reyes, & Kollman, 2001)

A number of modern well-tested forcefields are available. Despite similar functional forms, the parameters they contain are often very different, but nevertheless provide good descriptions of biomolecular structure and dynamics.(David E Shaw et al., 2010) For very large systems, for which atom-based simulations are not feasible, it is possible to use a coarse grained approach, where multiple atoms are represented as single particles.(Riniker, Allison, & van Gunsteren, 2012) Coarse graining has proven particularly useful in studying membrane bilayers and their associated proteins (see section **Coarse Graining and Membrane Simulation**). The majority of forcefields employed today come from a set of well-established forcefield families. These include GROMOS,(Oostenbrink, Villa, Mark, & Van Gunsteren, 2004) MARTINI,(Marrink, Risselada, Yefimov, Tieleman, & De Vries, 2007) CHARMM,(MacKerell Jr et al., 1998) AMBER,(J. Wang, Wolf, Caldwell, Kollman, & Case, 2004) and OPLS(Jorgensen, Maxwell, & Tirado-Rives, 1996). Up to date forcefields from each family can be found at the appropriate websites: GROMOS ([gromos.net](gromos.net)), MARTINI ([cgmartini.nl](cgmartini.nl)), CHARMM ([mackerell.umaryland.edu/charmm_ff.shtml](mackerell.umaryland.edu/charmm_ff.shtml)), AMBER ([ambermd.org/#ff](ambermd.org/#ff)), and OPLS ([zarbi.chem.yale.edu/oplsaam.html](zarbi.chem.yale.edu/oplsaam.html)). Note that it is important to distinguish between forcefields (which can be used with different biomolecular simulation programs) and simulation programs, which may have similar names (see Table 1). One of the limitations of classical force field based (MM) simulation is the inability to model electron rearrangements, for example in bond-breaking and bond–making, although force fields for reactions can be developed, and empirical valence bond methods also allow modelling of reactions). However, combining MM with quantum mechanics (QM) makes this possible (see section **QM/MM and Enzyme Catalysis**). Finally, simulation can reveal the important interactions between different types of biomolecules, such as proteins, DNA, RNA, and lipids (see section **Nucleic Acid Structure and Recognition**). This review will focus on some recent advances in biomolecular simulation and focus on the agreement between computed quantities and experimental observables.

## COMPUTATION OF FREE ENERGY, ENTHALPY AND ENTROPY

A central quantity of biomolecular systems is the Gibbs free energy, $G$, because changes in free energy govern many important processes such as molecular recognition, self-assembly and chemical reactions. This single number encapsulates all the possible molecular configurations of a particular state of the system at a given temperature and pressure through its connection to the partition function. Whilst the free energy itself is not measured, the difference in free energy, $\Delta G$, between two different states of a system relates to the equilibrium constant which expresses the relative probability of the two states. A state with lower free energy is more probable because it has configurations that are more numerous, more stable, or both.

There are two main classes of methods to calculate free energy from molecular simulation: one class computes $\Delta G$ and the other computes $G$; we refer to them, respectively, as *difference* and *state* methods. Difference methods evaluate $\Delta G$ from the probability ratio of two states. In principle, probability ratios can be evaluated directly from a molecular dynamics simulation. However, insufficient sampling using molecular dynamics alone usually necessitates enhanced-sampling methods (see section **Enhanced Sampling**) or unphysical short-cuts such as computational alchemy. (Michel & Essex, 2010) By perturbing atomic interactions, $\Delta G$ may be derived by measuring the overlap of configurations, as in the methods of exponential averaging,(Zwanzig, 1954) or the more efficient Bennett's acceptance ratio,(Bennett, 1976) by numerically integrating the energy gradient from one state to another, such as thermodynamic integration,(Kirkwood, 1935) or by taking the exponential average of the work.(C. Jarzynski, 1997) State methods evaluate $G$ directly as in the case of implicit solvent theories such as Poisson-Boltzmann,(Baker, 2004) Generalised Born,(Still, Tempczyk, Hawley, & Hendrickson, 1990) 3D-RISM,(Beglov & Roux, 1995) or many surface-area based methods or scoring functions. Other state methods compute $G$ from the enthalpy, $H$, and entropy, $S$. $H$ can be computed from long timescale simulations using the average force field energy for a set of Boltzmann-weighted configurations. $S$, on the other hand, which is a measure of the total number of configurations of a system, is difficult to evaluate because of the intricate complexity of biomolecular systems, making entropy calculation an active area of research. Examples of methods to calculate $S$ from molecular simulations include cell theory,(Henchman, 2007) inhomogeneous fluid solvation theory,(Lazaridis, 1998) mutual information expansion,(Killian, Kravitz, & Gilson, 2007) normal mode analysis,(Wilson, 1941) quasi-harmonic analysis,(M. Karplus & Kushick, 1981) and two-phase thermodynamic analysis.(Lin, Maiti, & Goddard III, 2010) A range of information may be used from simulation to calculate $S$, including coordinates, force constants, velocities or forces. These quantities may be considered independently or include correlations but at the cost of greater expense and complexity.

Free energy methods offer widely differing pros and cons. Difference methods are well established with detailed protocols and widely available software. They are in principle exact, although convergence and sampling issues limit either their accuracy or the sizes of the perturbations to which they can be applied. A drawback is that they typically require the simulation of many 'intermediate' states along a suitably chosen path connecting the two states of interest. Nevertheless, this demanding requirement of multiple simulations along the path is often amenable to parallelisation across multiple processors. Difference methods do not yield $G$, $H$ or $S$ directly, although $G$ can be extracted by perturbing to a reference state with an analytic value of $G$.(Tyka,

Clarke, & Sessions, 2006) Assuming that $\Delta H$ and $\Delta S$ are not temperature dependent, they can be calculated from suitable temperature derivatives of $\Delta G$ or ensemble averages of potential energies but the latter suffer from slow convergence.(Peter, Oostenbrink, van Dorp, & van Gunsteren, 2004)


By way of contrast, state methods only require a single simulation of the state of interest but still with convergence being a problem. Unlike experiment or difference methods, they can explain the value of $G$ directly not only in terms of $H$ and $S$, but also in relation to any specific part of the system. The main constraint on the widespread use of state methods is having theory with sufficient flexibility and accuracy to capture all the relevant degrees of freedom and correlations at every relevant length-scale. In the field of biomolecular simulations difference methods have been widely used to compute diverse physical quantities such as hydration free energies, partition coefficients, host-guest or protein-ligand binding free energies. Several recent potential energy function have been parameterised to reproduce the hydration free energies of small organic molecules with a mean unsigned error (MUE) under ca. 1 kcal mol$^{-1}$.(Harder et al., 2016) The decomposition of free energy into group contributions is often contentious because it is not entirely rigorous. If one is only concerned with the equilibrium of a particular process as governed by $\Delta G$, then nothing more would be required. However, evaluation of $\Delta H$ and $\Delta S$ (or heat capacities)(Kamp et al., 2018) can provide further physical insight; in some cases (e.g. for some enzyme-catalysed reactions), $\Delta H$ and $\Delta S$ are themselves temperature dependence, in which case the heat capacity should also be considered. In protein-ligand binding, considerable effort is often spent in ascribing structural interpretations to variations in the entropic and enthalpic components.(Klebe, 2015) Molecules that bind with very similar $\Delta G$ may have very different values of $\Delta H$ and $\Delta S$, depending on whether the binding is due to the formation of a few strong interactions, to a gain in flexibility or to water release) elsewhere in the system, and due to the phenomenon of enthalpy-entropy compensation. Binding may be said to be enthalpy or entropy-driven, but free energy is always the overall driver. The further breakdown into $H$ and $S$ for each system and even finer into molecular components provide further insights into thermodynamic processes. For example, spatially resolved methods such as WaterMap,(Abel, Young, Farid, Berne, & Friesner, 2008) Grid Inhomogeneous Solvation Theory,(Nguyen, Young, & Gilson, 2012) Solvaware,(David J. Huggins, 2014; D. J. Huggins, 2015, 2016), Grid Cell Theory,(Gerogiokas et al., 2014; Ross, Bruce Macdonald, Cave-Ayland, Cabedo Martinez, & Essex, 2017) and other techniques(Ross, Bodnarchuk, & Essex, 2015) can reveal the contributions of individual water molecules to the total Gibbs energy. However, the non-uniqueness of such decompositions means that care must be taken in their interpretation.

Such methods are now being routinely applied in small molecule drug design.(Cole et al., 2017) However, errors should be expected to increase for more complex polyfunctional drug-like molecules. The accuracy of protein-ligand binding free energies is highly variable, but favourable cases across congeneric series also feature MUEs of ca. 1 kcal mol$^{-1}$ for relative binding free energies, $\Delta \Delta G$.(L. Wang et al., 2015) Hydration and host-guest binding $\Delta H$ and $\Delta S$ are typically computed less accurately than $\Delta G$ and show a more pronounced dependence on force field parameters. (Gerogiokas et al., 2014; Henriksen & Gilson, 2017) Overall, the accurate computation of $\Delta G$ is an important predictive method in biomolecular simulation (and in molecular science generally), and decompositions can inform the design of tighter binding host-guest complexes or appropriate protein mutations, leading to practical applications in drug design(Irwin & Huggins, 2018; Vukovic, Brennan, & Huggins, 2016) or nanotechnologies.

## MEASURING MOLECULAR TIMESCALES

### The kinetics of binding and unbinding

Improvements in computer power and algorithmic advances over recent years (Aldeghi & Biggin, 2016) mean that there is now genuine overlap between the timescale of key biologically interesting events and those events that can be simulated in a reasonable amount of time. However, events that are at the longer end of the simulation timescale (a few milliseconds and beyond) require access to large amounts of computer time (either supercomputer or cloud-based solutions). Methods to study biomolecular kinetics have been developed and applied to a variety of different biomolecular problems ranging from protein folding through to the kinetics of ligand-binding. (David E Shaw et al., 2010) The latter has received a lot of attention in recent years, as it has become clear that under *in vivo* conditions, binding events are sometimes far from equilibrium and consequently thermodynamics may not offer the best prediction of activity. In other words, the most important factor in determining may be the lifetime of the *in-vivo* drug-receptor interaction. Kinetic parameters such as $k_{on}$ and $k_{off}$ (the on- and off-rate constants for ligand binding, which may be better predictors of biological activity than binding affinity) can often be measured experimentally but obtaining these via computational methods is challenging because of the long timescales involved. In principle, they can be computed via 'brute force' simulations, though the timescales required for simulations multiple (un)binding events for large systems are very long and in general extremely challenging. More practical is the use of enhanced sampling methods. There have been a number of recent reviews of such methods.(Dickson, Tiwary, & Vashisth, 2017; Pang & Zhou, 2017) In this section we briefly cover some of the recent highlights in simulations of ligand binding and unbinding and discuss where future opportunities and directions might lie.

### Direct calculation of kinetics

MD simulations can provide atomic resolution detail at a picosecond timescale and thus can provide physical insight into the way in which a ligand interacts with its receptor (Feixas, Lindert, Sinko, & McCammon, 2014). The timescale for many binding events is of the order of microseconds and this presents a problem for the timescale of typical unbiased MD simulations. A step-change occurred in 2008 with the development of ANTON(D.E. Shaw et al., 2008), a purpose-built MD simulation computer that could produce millisecond-long simulations in a few weeks. In 2011, Shan et al (Shan et al., 2011) reported a study using unbiased MD simulations that examined how the cancer drug dasatinib, and the kinase inhibitor PP1, can not only correctly find the binding site of Src kinase, but adopt a conformation indistinguishable from the crystallographic pose. Dror *et al.* simulated binding of ligands to the β-adrenergic receptor 1 and 2. From 82 simulations ranging from 1 to 19 ms, they found that ligands entered the binding site via a metastable extracellular vestibule. (Dror et al., 2011) They have also applied this long timescale MD approach to ligand binding to muscarinic receptors, where the simulations suggested that tiotropium binds transiently to an allosteric site *en route* to the orthosteric binding pocket(Kruse et al., 2012). However, routine use of such hardware is not possible. An alternative approach is to use a cloud-based community resources. The GPUGRID.net project is a good example of how this can be utilised and was first demonstrated to examine the binding of benzamidine to trypsin(Ignasi Buch, Giorgino, & De Fabritiis, 2011; I. Buch, Harvey,

Giorgino, Anderson, & De Fabritiis, 2010). This type of approach relies on being able to combine (typically) many hundreds of smaller simulations to build up enough statistics concerning the transitions between the many different states to reconstruct the full kinetic landscape. Fortunately, there is a well-developed framework for such reconstruction, namely Markov State Modelling (MSM).

**Markov-State Modelling**

Combining MD simulation with MSM has become a popular approach to shed light into molecular processes at timescales that are not accessible with single unbiased MD simulations. The use of this technique has found a broader use in the biomolecular simulation community thanks to software packages such as pyEMMA(Scherer et al., 2015) developed by the Noé group or MSMBuilder(Bowman, Beauchamp, Boxer, & Pande, 2009; Harrigan et al., 2017) developed by the Pande group. An MSM built from MD simulations is a model of the dynamics within a system and allows the analysis of key metastable states, thermodynamic and kinetic properties in a systematic manner. The key feature of an MSM is that it can recover the equilibrium distribution of states by simulating many short trajectories starting from different states, rather than having to simulate a single trajectory for such a long time that all states are sampled with equilibrium frequency. The approach has the key advantage that it is trivially parallelisable.

There have been many examples of the use of MSMs for biomolecular systems(John D. Chodera & Noé, 2014; Malmstrom, Lee, Van Wart, & Amaro, 2014) and they have been developed and applied to a variety of different biological problems ranging from the modulation of protein dynamics(Shukla, Hernández, Weber, & Pande, 2015) and protein folding and unfolding(Sirur, De Sancho, & Best, 2016) through to the kinetics of ligand- and protein-binding(Zhou, Pantelopulos, Mukherjee, & Voelz, 2017). To illustrate the process in the context of ligand binding events, we focus on a study from Plattner and Noé(Plattner & Noe, 2015). In the study, they constructed an MSM based on a total simulation time of 149.1 μs of the trypsin/benzamidine system. The MSM allows the calculation of on and off rates and also the overall free energy of binding. For the latter they obtained a value of −6.05 ± 1.00 kcal/mol, which is in excellent agreement with the experimental value of −6.2 kcal/mol(Mares-Guia & Shaw, 1965). The estimated association constant of $k_{on}$ = 6.4 ± 1.6 × 10$^7$ mol$^{-1}$s$^{-1}$ is higher than the experimental value of 2.9 × 10$^7$ mol$^{-1}$s$^{-1}$ and the dissociation constant is significantly overestimated with $k_{off}$ = 131 ± 109 × 10$^2$ s$^{-1}$ compared to experiment (6 × 10$^2$ s$^{-1}$). The large error in the latter is due to poor sampling of dissociation events, whereas systematic errors are likely due to the force field parameters, which have been fitted to thermodynamic rather than kinetic data. Using transition path theory, binding pathways and mechanisms can be obtained from the MSM, revealing that the ligand binding process in this case is more complex that either a simple induced fit or conformational selection model, which has also been reported in other cases(S. Gu, Silva, Meng, Yue, & Huang, 2014; Silva, Bowman, Sosa-Peinado, & Huang, 2011). In fact, in the apo state trypsin has 6 metastable conformations with varying affinities to benzamidine (Fig. 1), which fits into the conformational selection picture. However, upon benzamidine binding the populations of this ensemble are shifted (Fig. 2) and one additional metastable configuration emerges, which can be attributed to ligand induced conformational plasticity and thus shows that
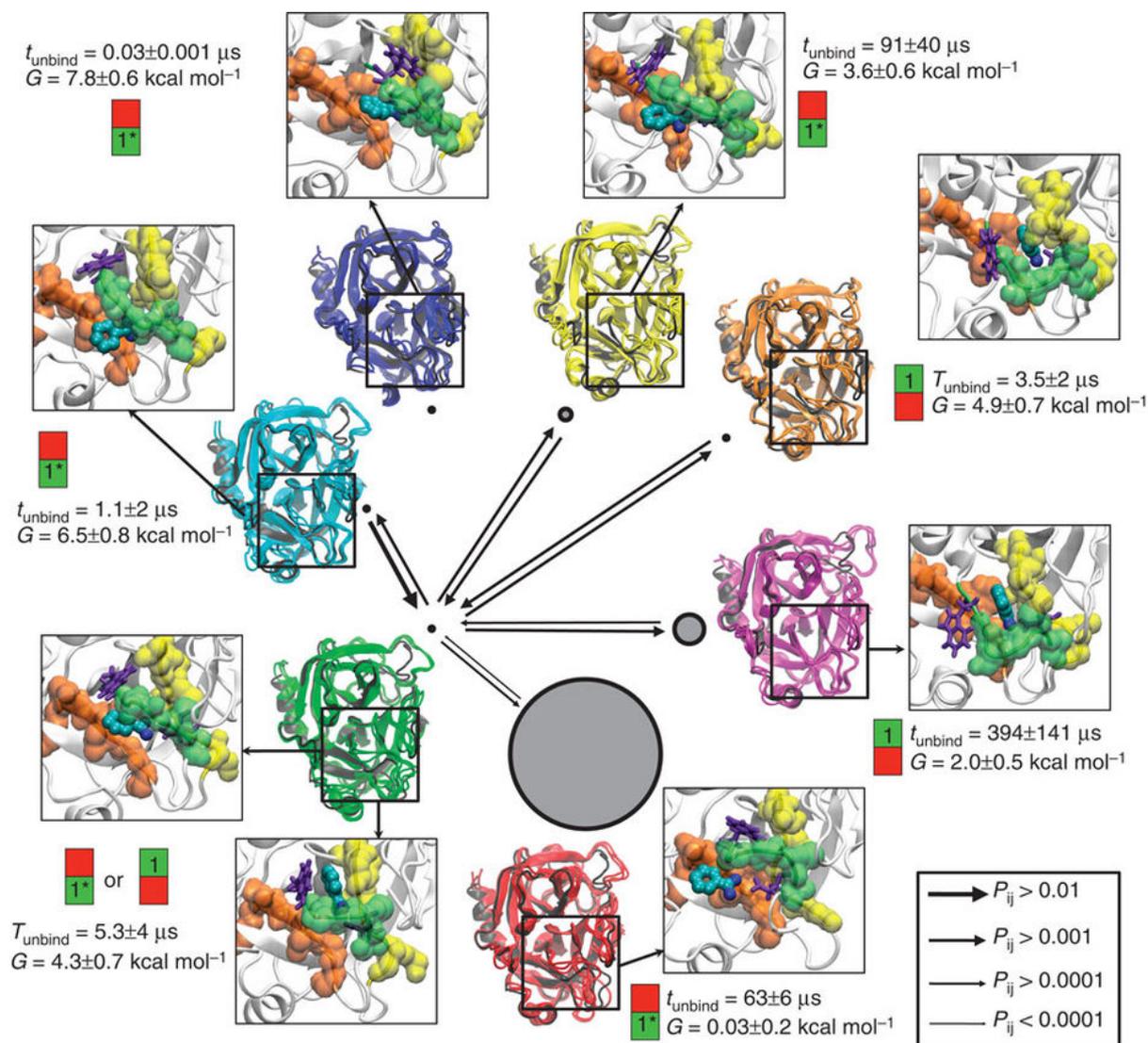
both induced fit as well as conformational selection are needed to fully describe the ligand binding process in this case.
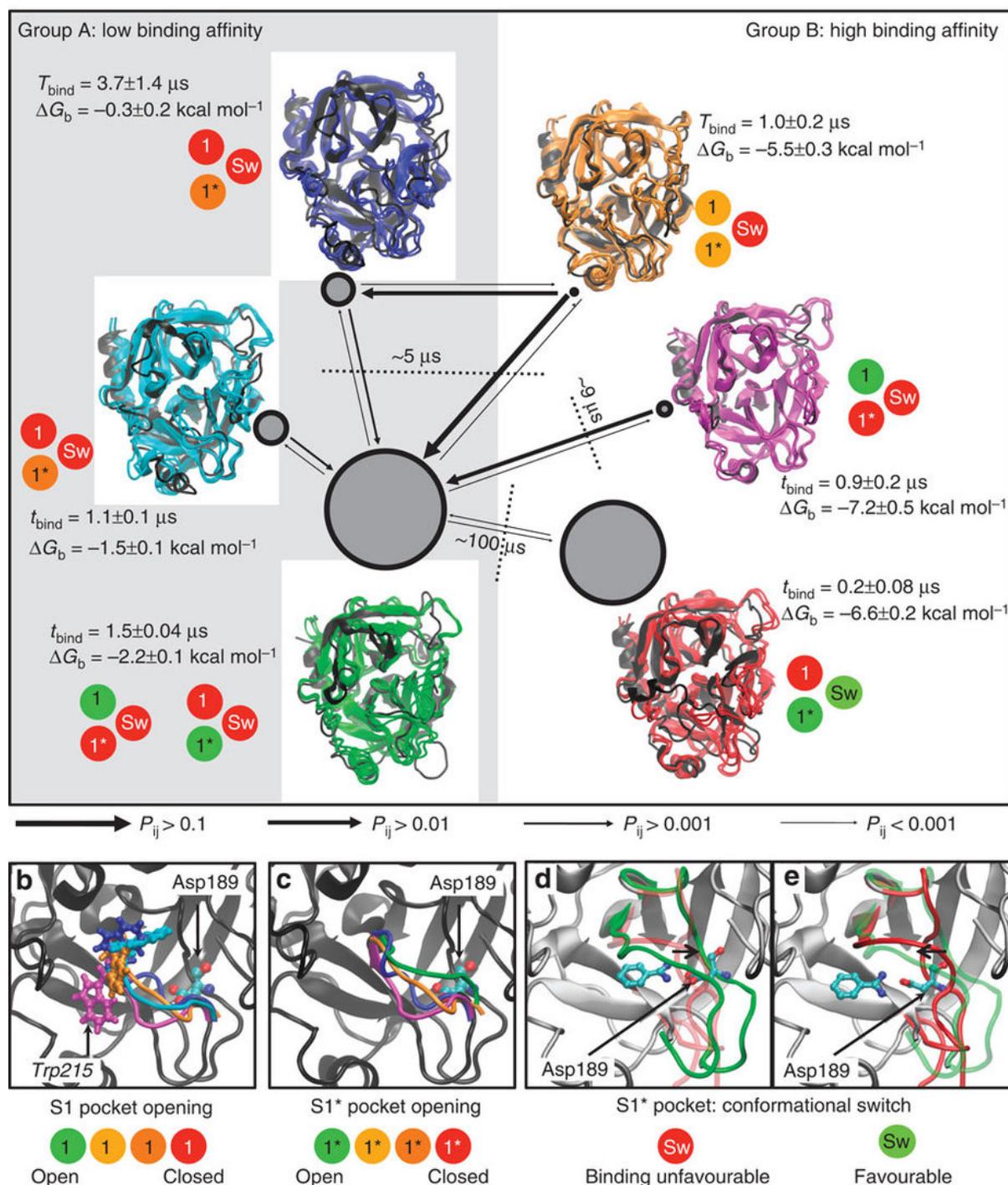


**Figure 1 -** Example of a typical MSM, in this case, describing the apo state of trypsin (a) Structural features, equilibrium distribution and kinetics of six unbound (apo) protein conformations. Transitions between them occur at timescales on order of tens of microseconds. The three slowest relaxation timescales and their corresponding transition process are indicated (dashed lines). The circles have an area proportional to the equilibrium probability $\pi_i$. Their respective free energy differences $\Delta G_b$ of binding a ligand to this conformation and the binding time $t_{bind}$ (mean first passage time to binding) are given. The arrows indicate the transition probabilities for direct transitions

between the different states (see legend). The most important structural differences concerning ligand binding are shown in b–e, and the structures are classified with respect to these features by green/orange/red bullets in a. The structures are classified by the state of S1 or S1*: open (green circle with '1' or '1*'), half-open (orange circle) or closed (red circle) and by the S1* pocket conformational switch: favourable for binding (green circle with 'Sw') or unfavourable for binding (red circle with 'Sw'). Reproduced from(Plattner & Noe, 2015).



$t_{unbind} = 0.03 \pm 0.001$ μs
$G = 7.8 \pm 0.6$ kcal mol$^{-1}$

$t_{unbind} = 91 \pm 40$ μs
$G = 3.6 \pm 0.6$ kcal mol$^{-1}$

$T_{unbind} = 3.5 \pm 2$ μs
$G = 4.9 \pm 0.7$ kcal mol$^{-1}$

$t_{unbind} = 1.1 \pm 2$ μs
$G = 6.5 \pm 0.8$ kcal mol$^{-1}$

$t_{unbind} = 394 \pm 141$ μs
$G = 2.0 \pm 0.5$ kcal mol$^{-1}$

$T_{unbind} = 5.3 \pm 4$ μs
$G = 4.3 \pm 0.7$ kcal mol$^{-1}$

$t_{unbind} = 63 \pm 6$ μs
$G = 0.03 \pm 0.2$ kcal mol$^{-1}$

$P_{ij} > 0.01$
$P_{ij} > 0.001$
$P_{ij} > 0.0001$
$P_{ij} < 0.0001$

**Figure 2 -** Trypsin conformations with benzamidine-bound and the binding mode of benzamidine. The seven conformational states shown are equal to the six apo states shown in Fig. 1, plus the yellow conformation that is only found with benzamidine-bound. The binding pocket conformation is defined by three loops: the yellow loop (residues 187–194) with Asp189, the green loop (residues 215–221) with Trp215 and the orange loop (residues 225–230). The circles have an area proportional to the equilibrium probability of the respective conformation, given that benzamidine is bound, $\pi_i$. Their respective relative free energies $G = -k_B T \ln \pi_i$ and the unbinding times $t_{unbind}$ (mean first passage time to unbinding) are given. The arrows indicate the transition probabilities for direct

transitions between the different states. The binding mode (pocket 1 or 1*) is indicated by the green square with '1' or '1. Reproduced from(Plattner & Noe, 2015).

The MSM model(Plattner & Noe, 2015) suggests that binding of benzamidine to trypsin cannot simply be described as a two state process, because the model reveals that the slowest transitions are not due to the binding/unbinding process. The model also reveals how different trypsin conformations are accessed and stabilised in different ways. MSM studies such as the one above are starting to provide unique insight into binding events and thus far tend to indicate not only that multiple routes are possible, but also that key metastable binding sites exist en-route to the main binding site. Another key aspect that is likely to be explored further as more systems are examined is the role of water in the binding/unbinding events. This is an aspect that is currently quite poorly understood and may have important ramifications for drug design in the future. MSM models (and indeed the brute force simulation approaches) also give reassurance that the force fields currently used are actually quite reasonable and are capable of delivering results in good agreement with experiment as far as thermodynamics averages are concerned, although there is always room for improvement. Another key advantage of the MSM approach is that it can make use of different levels of representation (including atomic and coarse-grained) to make experimentally testable (via site-directed mutagenesis) hypotheses (see section **Coarse Graining and Membrane Simulation**).(J. D. Chodera & Noe, 2014) Two interesting studies applied MSM to biologically relevant, yet computationally demanding cases: an activation mechanism of a kinase(Shukla, Meng, Roux, & Pande, 2014) and protein-protein association.(Plattner, Doerr, De Fabritiis, & Noé, 2017) The authors of the former examined Src kinase whose aberrant behavior ultimately leads to cancer. Extensive simulations provided information on the kinetics of the activation mechanism, which agreed with experimental observations, and revealed several metastable transition states which show promise as drug targets. In the case of protein-protein association, the authors reproduced a number of structural, kinetic, and energetic features of the process for ribonuclease barnase and its inhibitor barstar indicating a range of possibilities for future studies of macromolecular association.

In the future, models of biomolecular kinetics should be vastly improved by comparison with data from time-resolved crystallography.(Johansson, Stauch, Ishchenko, & Cherezov, 2017; Kupitz et al., 2014; Tenboer et al., 2014) This technique allows us to monitor bimolecular processes across biologically relevant timescales. Knowledge of kinetic timescales and pathways will lead to improved and transferable models for biomolecular processes.

**ENHANCED SAMPLING**

Notwithstanding the progress made in computer architectures, specialised hardware for MD,(David E. Shaw et al., 2014) and the parallelisation of MD codes, the typical timescales that can be accessed by atomistic MD simulations are shorter than what is required for many biomolecular systems. This has stimulated the development of many enhanced sampling methods. Reviewing the plethora of enhanced sampling algorithms and their many applications to biomolecular systems is beyond the scope of this article.(Bernardi, Melo, & Schulten, 2015) Here, we will only focus on recent examples

where enhanced sampling algorithms have been successfully used to explain complex biophysical phenomena and have been validated or combined with experiments.

Enhanced sampling algorithms can broadly be categorised according to their dependence on collective variables (CVs) over which sampling is enhanced. Methods such as metadynamics, (Barducci, Bonomi, & Parrinello, 2011; Laio & Gervasio, 2008) umbrella sampling,(Torrie & Valleau, 1977) Jarzynski's identity-based methods,(Christopher Jarzynski, 1997) steered MD,(Gullingsrud, Braun, & Schulten, 1999) and many other belong to the CV-based category. The CVs are functions of the system's coordinates that approximate the reaction coordinate. Choosing a set of relevant CVs for the enhanced sampling methods is not always straightforward. However, experimental observables can be helpful in the selection process and they can even be used directly as CVs, as in the study of the G protein beta subunit 3 folding mechanism where NMR chemical shifts guided the bias-exchange metadynamics.(Granata, Camilloni, Vendruscolo, & Laio, 2013) Recently, a CV-based algorithm (metadynamics) was successfully used to interpret the experiments and even point out some crystallographic artefacts in the activation mechanism of p38α MAP kinase.(Kuzmanic et al., 2017) Simulations showed that a large conformational change observed by X-ray crystallography is not triggered by the dual phosphorylation of the activation loop and can only be fully stabilised by binding of adenosine triphosphate (ATP) and a docking peptide. Further analyses confirmed that the X-ray structure is a product of an unfortunate binding of an atypically long His-tag to the peptide docking site.

Enhanced sampling methods have also been applied in computer-aided drug discovery to investigate the mechanism of action of a newly developed inhibitor (named SSR) of the fibroblast growth factor receptor (FGFR).(Bono et al., 2013; Herbert et al., 2013) FGFR, a receptor tyrosine kinase, is a promising target for anticancer drug development as it is involved in a number of fundamental cellular processes, including blood vessel formation. SSR was discovered serendipitously through a high-throughput screening. Experiments indicated that SSR binds extracellularly to the D3 Ig-like domain, but X-ray crystallography and NMR failed to identify its binding site due to the partially disordered nature of the domain. Using a combination of metadynamics and parallel-tempering metadynamics(Bussi, Gervasio, Laio, & Parrinello, 2006) together with CVs describing both the binding process and the folding of the domain, the authors were able to observe the reversible binding of SSR to the D3 domain of FGFR. Interestingly, a conformational change that elongates a small α-helix in the D3 domain was needed to open a hidden binding pocket for SSR. The predicted binding mode was subsequently validated by mutagenesis and used to design more potent derivatives of SSR.

Another success story comes from a study of L99A variant of bacteriophage T4 lysozyme where a range of metadynamics techniques was used not only to accurately capture key conformational and thermodynamic features, but also to uncover a transient tunnel which seems to allow ligands to unbind.(Y. Wang, Papaleo, & Lindorff-Larsen, 2016) Successful applications are not limited to metadynamics, as shown by a recent study that employed extensive umbrella sampling MD to

describe the molecular details of the energy conversion mechanism of the main rotary step in the synthesis cycle of $F_1$-ATPase.(Czub, Wieczór, Prokopowicz, & Grubmüller, 2017) To achieve high efficiency and timely substrate binding and product release, this molecular motor requires tight coupling of its subunits. By using the rotation of the γ-shaft as the order parameter in umbrella sampling, the authors were able to give structural support to single-molecule experiments, reproduce experimental rate constants, and capture a metastable state, which according to several of its features, corresponds to the elusive ATP-waiting state. Umbrella sampling has also been used in a multiscale scheme to elucidate the mechanism of activation of influenza A M2 proton channel which plays an important role in the viral life cycle.(Liang et al., 2016) The channel becomes active in the low-pH environment through an interplay between Trp and His tetrads, the latter of which becomes protonated. The C-terminal amphipathic helix, whose role in the mechanism was somewhat unclear, was shown not to affect the proton conduction which is in agreement with electrophysiological results.

When it comes to methods that are not directly based on geometric CVs, one of the approaches to enhance the sampling is to use a history-based potential that prevents the system from revisiting the previously explored conformations, such as local elevation technique (Huber, Torda, & van Gunsteren, 1994) or conformational flooding (Grubmuller, 1995) (the predecessors of metadynamics). Local elevation has been successfully combined with umbrella sampling (Hansen & Hünenberger, 2010) and used recently to build a conformational library for the construction of N- and O-glycan structures in glycoproteins (Turupcu & Oostenbrink, 2017). Temperature is also often used to enhance the sampling, as in the case of temperature replica exchange MD (T-REMD)(Sugita & Okamoto, 1999) and simulated tempering.(Marinari & Parisi, 1992) A recent application of REMD compared equilibrium populations of open and folded ring states for urotensin II and urotensin related protein, both of which are connected to various pathophysiologies.(Haensele et al., 2017) Obtained populations were in excellent agreement with experimental NMR chemical shifts and indicated that the different behaviour of the two proteins does not come from the ring conformation, but likely from their interaction with the G-protein coupled urotensin II receptor. Another notable example showed how to extract kinetic rates from REMD and demonstrated the method on alanine dipeptide and (un)folding of the neomycin RNA riboswitch.(Lukas S. Stelzl & Hummer, 2017) When it comes to simulated tempering, Pan et al.(Pan, Weinreich, Piana, & Shaw, 2016) have recently compared the efficiency of the method against extremely long (millisecond) MD simulations in sampling of the reversible folding of BPTI and ubiquitin and showed that simulated tempering can achieve a consistent and substantial sampling speedup of up to an order of magnitude or more relative to the conventional MD.

**QM/MM MODELLING OF ENZYME REACTIONS**

The combination of quantum mechanics and molecular mechanics, usually indicated with QM/MM, is a popular approach to allow a (relatively small) part of the system to be treated with a quantum mechanical method, while the rest is treated using molecular mechanics (Figure 3). While electronic structure calculations are now possible on quite large systems using low-scaling methods and efficient codes,(Lever et al., 2014) QM/MM methods offer a combination of flexibility and speed that

makes them attractive and practical. QM/MM methods allow calculation of electronic properties of (a small region in) biomolecular systems and are finding application in many areas of biomolecular science (e.g. in the calculation of spectroscopic properties,(Morzan et al., 2018) photochemistry,(Boulanger & Harvey, 2018) p$K_a$s, (Uddin, Choi, & Choi, 2013) and predictions of ligand binding affinities)(Steinmann, Olsson, & Ryde, 2018) and beyond (e.g. chemistry of solid-state materials),(Bernstein, Kermode, & Csanyi, 2009) but is particularly popular for the study of reactions (and interactions) in enzymes.(Senn & Thiel, 2009; Van der Kamp & Mulholland, 2013) A QM/MM method was first applied to an enzyme-catalysed reaction by Warshel & Levitt (1976) in their seminal study of the reaction mechanism of hen egg white lysozyme (Warshel & Levitt, 1976). The pioneers in this field, Warshel, Levitt and Karplus (Field, Bash, & Karplus, 1990), were awarded the Nobel Prize in Chemistry in 2013, recognising QM/MM methods as seminal in the development of multiscale modelling. Nowadays, over 100 research articles on QM/MM applied to enzymes are published each year,(Ranaghan & Mulholland, 2016) showing that QM/MM modelling of enzyme reactions has become and remains highly popular, which is helped by advances in readily available software (e.g. CHARMM, AMBER, ChemShell,(Sherwood, 2003) CPMD,(Hutter et al., 1995) Gaussian,(Frisch et al., 2016) Qsite(Friesner, 2004) and other packages).
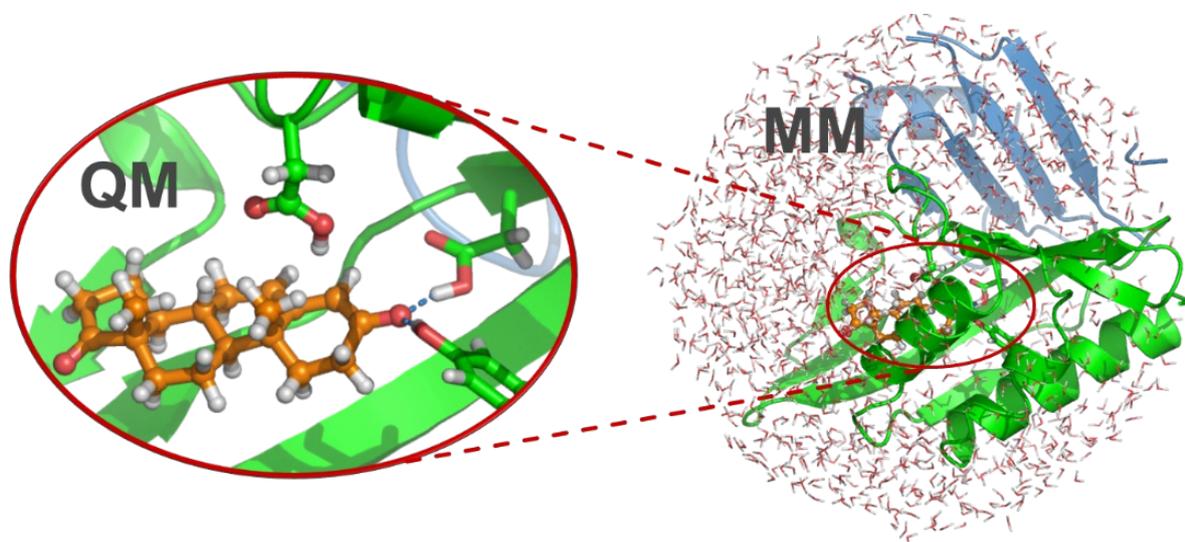


**Figure 3 -** Example of QM/MM modelling of an enzyme-catalysed reaction. The enzyme (ketosteroid isomerase) is divided into an MM region and a QM region. The transition state structure of the first proton transfer in the mechanism is shown.(van der Kamp, Chaudret, & Mulholland, 2013)

Different approaches to perform QM/MM simulations exist. One can distinguish the *additive* approach (whereby the energy of the system is the sum of the QM energy of the QM region, the MM energy of the MM region and an interaction term) and a *subtractive* approach (whereby the energy of the system is obtained by the MM energy of the total system plus the QM energy of the QM region and minus the MM energy of the QM region; this is used in the ONIOM method (Vreven et al., 2006)). In both cases, the QM region should ideally be polarised by the MM region (*electrostatic embedding*). A third popular multi-scale approach for enzyme reaction modelling is to

use a linear combination of the most important ionic and covalent resonance forms expected in the reaction, such as in the empirical valance bond (EVB) approach (Duarte, Amrein, Blaha-Nelson, & Kamerlin, 2015). EVB methods do not treat electrons explicitly, and so are relatively very fast; careful parameterization though is essential to develop each specific EVB model.

Here, we focus applications that combine quantum mechanics and molecular mechanics for modelling reactions. For QM/MM calculations, choices must be made regarding the QM method and MM potential to use, how the system is partitioned (including treatment of covalent bonds at the boundary), how the model is constructed and what type of simulation will be performed (e.g. energy minimisation or molecular dynamics simulation). The choice of QM method is particularly important, as this will typically be the most time-consuming factor in the calculation; there is generally a trade-off between accuracy and speed. In general, there are three main determinants of computational cost: (i) level of accuracy/complexity of the QM method, (ii) size of the QM region, and (iii) sampling of relevant conformations. For each enzyme system and problem of interest, the requirements for these three factors will differ. Far fewer QM energy evaluations are required for energy minimisation than molecular dynamics simulation. Potential energy profiles for enzyme reactions that result e.g. from a series of energy minimisations can therefore be performed with more demanding, more accurate QM methods; single-point *ab initio* energies from these structures can offer 'chemical accuracy', e.g. through local coupled-cluster ab initio QM methods(F. Claeyssens et al., 2006) or using coupled-cluster embedded in DFT for the QM region(Bennie et al., 2016). It must be noted, however, that for convergence of potential energy barriers, many different enzyme-substrate conformations may be required.(Oláh, Mulholland, & Harvey, 2011; Ryde, 2017) Potential energy profiles can also be used to estimate free energy profiles by additional sampling of the MM region. (Claeyssens et al., 2005; Hu, Lu, & Yang, 2007; Rosta, Haranczyk, Chu, & Warshel, 2008; Sodt et al., 2015)

Umbrella sampling molecular dynamics with lower-level QM/MM methods, using a few selected reaction coordinates, is widely used to calculate free energy surfaces for enzyme reactions (see Figure 4).
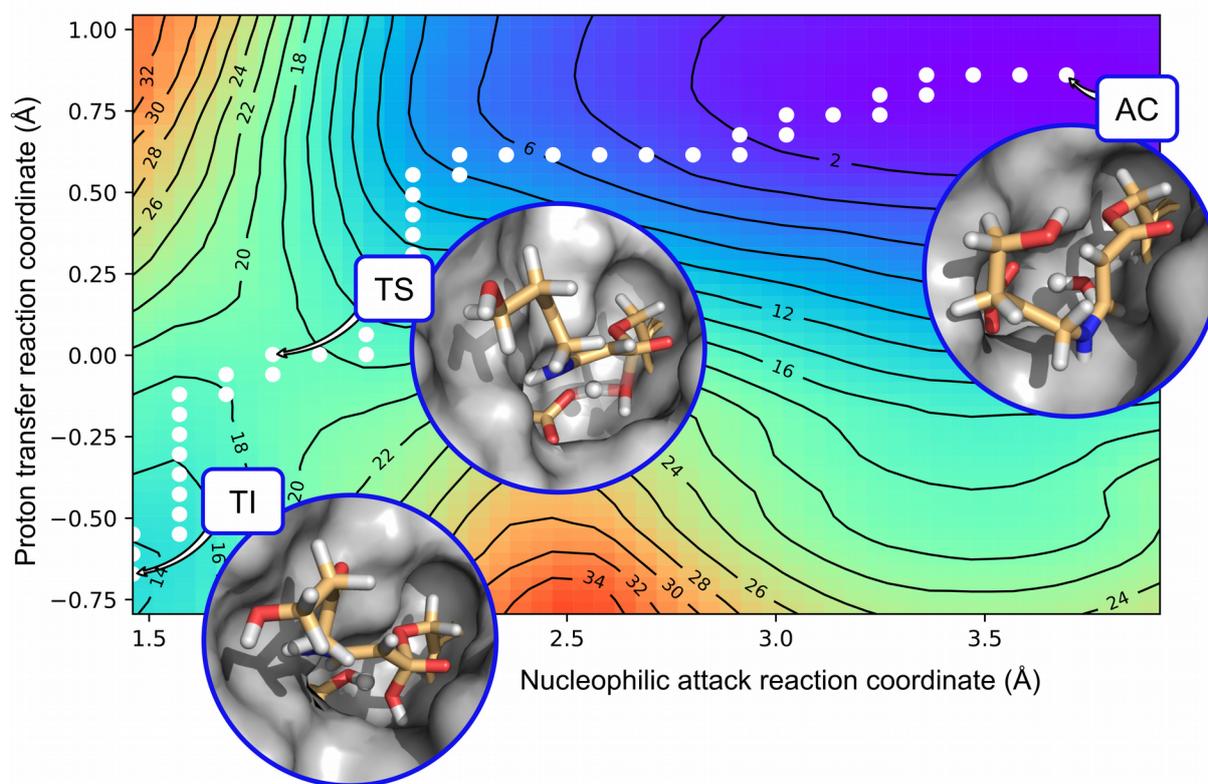
**Figure 4** - Illustration of a free energy surface obtained using QM/MM umbrella sampling along two reaction coordinates, with energy contours shown (values in kcal/mol). Here, the rate-limiting deacylation of the beta-lactamase inhibitor clavulanate by the beta-lactamase KPC-2 is shown, with QM/MM energies indicated on the surface plot.[cite] Active site structures of three states are depicted: the acyl-enzyme (AC), the transition state (TS) and the tetrahedral intermediate (TI).

In recent years, several developments have allowed simulations of reactions with without the need to prescribe reaction coordinates in detail in advance. These include 'string' type methods, such as the nudged elastic band or finite temperature string methods. While these string type methods usually rely on information for both the reactant and product states, methods related to umbrella sampling to adaptively explore the free energy profile have recently become very popular, and metadynamics is now used in QM/MM calculations frequently.(Raich, Nin-Hill, Ardèvol, & Rovira, 2016; Saitta & Saija, 2014) Both umbrella sampling and string type methods can be made more efficient(Zinovjev & Tuñón, 2017) by coupling them to replica exchange (see section **Enhanced Sampling**) within adjacent umbrella windows(Rosta, Nowotny, Yang, & Hummer, 2011) or metadynamics biases. Markov-chain based approaches have revolutionised how MD data is analysed (see section **Measuring Molecular Timescales**). These have now also been applied to QM/MM simulations, where dynamic information is now also available from the same data, both from biased and unbiased simulations.

A key experimental observable to compare to in enzyme reaction modelling is a rate constant, ideally for individual reaction steps, but more usually the overall enzymic reaction rate constant or

$k_{cat}$, which, can be converted using transition-state-theory into a free-energy barrier for the overall enzyme reaction, $\Delta G^{\ddagger}$ [Garcia-Viloca, Gao, Karplus, & Truhlar, 2004]. Sampling of conformations (e.g. by umbrella sampling) is required for the calculation of an activation free energy. Close agreement (approaching 'chemical accuracy') with experimentally derived barriers can be achieved for some enzymes when high level ab initio QM methods are used in QM/MM calculations of energetics. (Frederik Claeyssens et al., 2006) Lower-level QM methods (which are considerably less computationally demanding) do not provide quantitatively accurate barriers, but can be sufficient for identifying likely mechanism, and e.g. predicting relative reactivity of different substrates, or mutant enzymes. A strength of the QM/MM approach lies in the ability to compare different possible reaction mechanisms or, for example, different enzyme variants, without parameterization required for the specific system. QM/MM calculations are now widely applied to biomolecular systems, but it is important to warn the unwary user that despite increasingly routine setup, it is certainly possible to obtain artefacts through a poor choice of QM region, for example, so additional care is required in modelling with these hybrid techniques. Applied appropriately, QM/MM calculations can provide useful predictions of mechanism and activity. QM/MM calculations have been helpful to provide information on reaction mechanisms as well as the reactivity of mutant or modified systems compared to wild-type. (Lopata et al., 2015; Nagy et al., 2016; Senn & Thiel, 2009; Van der Kamp & Mulholland, 2013) Such applications were modest in terms of their predictive power, and often were used to explain experimental data and add key details regarding which reaction mechanism corresponds best to the observed kinetic measurements. In more recent applications, QM/MM calculations have been used to help design desired changes in enzyme reactivity, e.g. for selectivity or enhanced activity (e.g. with non-native substrates or reactions)(Hediger, De Vico, et al., 2013; Hediger, Steinmann, De Vico, & Jensen, 2013; Świderek, Tuñón, Moliner, & Bertran, 2015). In addition, QM/MM simulations have also been used as a computational assay, e.g. for breakdown of β-lactam drugs by β-lactamases, showing the ability to distinguish between enzymes that act as carbapenemases and those that do not.(Chudyk et al., 2014; Fritz, Alzate-Morales, Spencer, Mulholland, & van der Kamp, 2018) Time-resolved crystallography has the potential to determine structures of enzyme-bound reaction intermediates, which may be ideal starting points for QM/MM simulations, in developing atomically details models to connect with and explain biomolecular kinetics.(Olmos et al., 2018) QM/MM methods are also being usefully applied to study covalent inhibitors,(Callegari et al., 2018) which are of growing interest as drugs. QM/MM methods can also be coupled to other levels of simulation (e.g. atomistic and coarse-grained MD) to describe biochemical reactivity in multiscale schemes for tackling problems such as the prediction of drug metabolism.(Amaro & Mulholland, 2018; Lonsdale, Rouse, Sansom, & Mulholland, 2014)


**NUCLEIC ACID STRUCTURE AND RECOGNITION**

The DNA double helix is an iconic molecular structure that symbolises the scientific understanding of life. This widespread familiarity belies the difficulties of relating the sequence dependent structure and flexibility of the DNA biopolymer to biological function, which requires packing, regulation and control, as well as sequences that code for functional proteins. Here, we provide an overview of the most recent simulations of DNA that illustrate the multi-scale nature of molecular genetics. By focusing on connection to experiment, we highlight the difficulty of understanding DNA dynamics, as well as structure. We concentrate on DNA, rather than RNA, as for DNA the broadest range of length-scales are involved: from a few base pairs up to the scale of the entire genome (Figure 5).
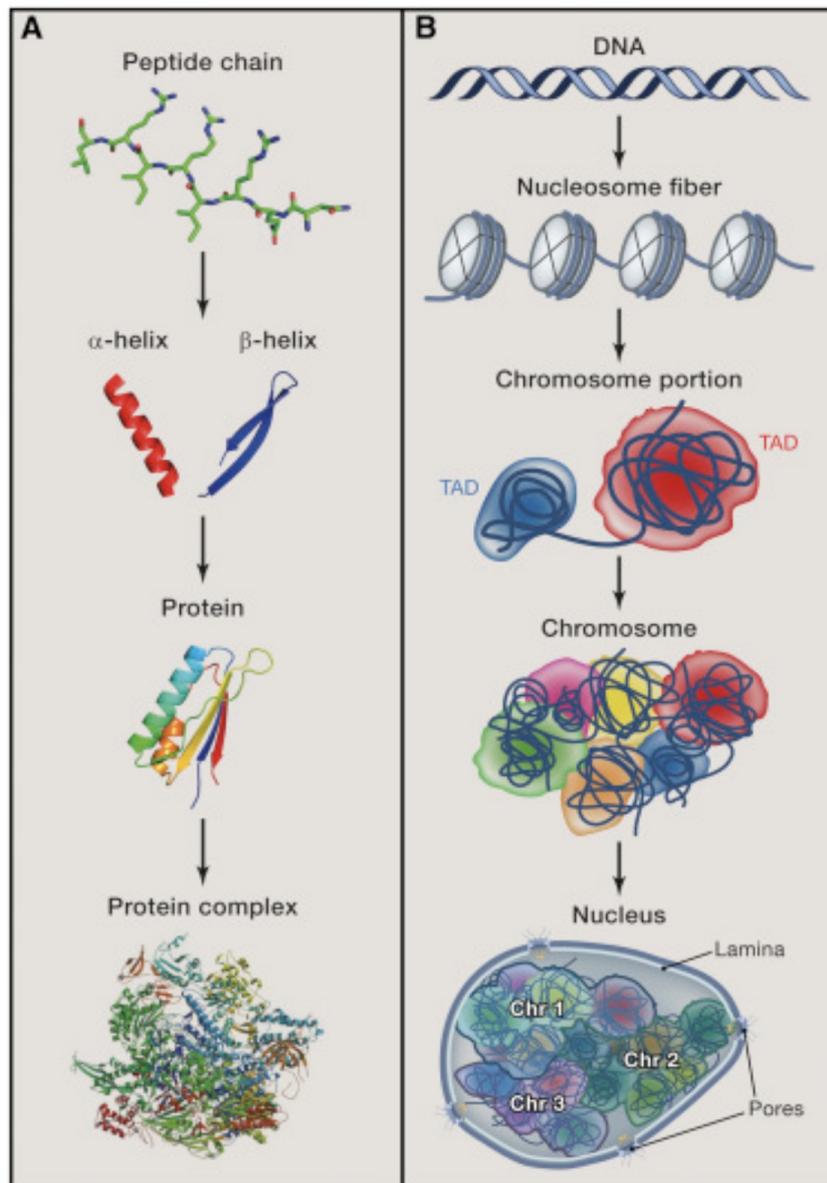
**Figure 5 -** The Hierarchical Structure of the Genome. Reproduced with permission from(Sexton & Cavalli)

Three examples of DNA structures that have been studied using biomolecular simulation are shown in Figure 6 and discussed below.
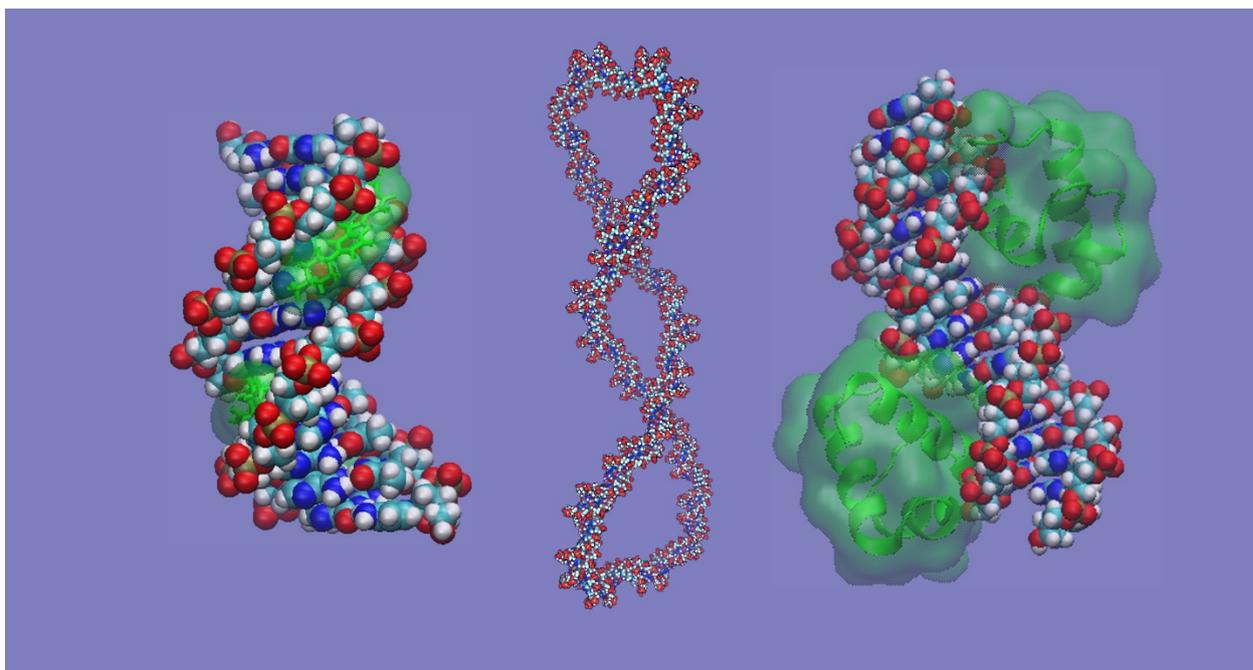
**Figure 6 –** Examples of the variety of DNA structures simulated atomistically. Left: A DNA-drug complex.(Harris, Gavathiotis, Searle, Orozco, & Laughton, 2001), centre: a supercoiled DNA micro-circle,(Irobalieva et al., 2015) right: A DNA-protein complex.(Garton & Laughton, 2013)

**Atomistic Simulations of Linear Duplex DNA**

The majority of the **s**imulations of the structure and dynamics of DNA at the atomistic level have been performed for short (<20 base pairs) DNA duplexes, due to the wealth of experimental information available in the Protein Data Base for such structures. For example, Pasi et al.,(Pasi et al., 2014) in a community-wide activity, have used simulations to catalogue the relationship between sequence and structure/flexibility for each of the 136 possible unique tetranucleotide base sequences. This reveals the remarkable richness, and heterogeneity, in the fine details of the structure and dynamics of double-stranded DNA – features which of course are vital to its biological roles. These community activities have benefitted the field of nucleic acid simulation by promoting a broad discussion of how such simulations can be validated and improved. The development of the new parmBSC1 force field for DNA(Ivani et al., 2016) produced ~140 µs MD simulations data covering the broad range of nucleic acid structures found in the PDB (including duplex, triplex, quadruplex and 3 and 4 way junctions). The BigNASim database(Hospital et al., 2016) provides access to the simulation trajectory and analysis data, and is designed to grow through interaction with the wider community. As well as saving time and valuable compute resources, such databases enable researchers to perform structural or dynamic analyses across multiple simulations, not just a single system (as is currently more common), which allows entirely new research questions to be addressed.

The validation of atomistic simulations against experiments remains challenging, in part because it is difficult by experiment to obtain atomic level resolution of the dynamics of biological macromolecules. For linear DNA sequences of between 20 and 50 base pairs, experiments using

pulsed electron-electron double resonance (PELDOR) have provided Angstrom level information about distances, orientations and dynamics, which have been compared with atomistic simulation. (Lukas S Stelzl, Erlenbach, Heinz, Prisner, & Hummer, 2017) Better quantitative agreement with the newer parmBSC1 and OL15 force fields were reported than for the older AMBER DNA force fields, reassuring the community that force fields are both accurate and improving. Nevertheless, the bending and twisting motions detected as being the principal modes of flexibility of DNA were observed in simulations as long as 18 years ago,(Sherer, Harris, Soliva, Orozco, & Laughton, 1999) and have later been shown to persist even when simulations are extended up to microsecond timescales on the ANTON supercomputer.(Galindo-Murillo, Roe, & Cheatham III, 2014) While the agreement between PELDOR measurements and MD simulations are encouraging, the experiments need to be performed at very low temperatures (40K), which adds complexity to the interpretation of the results (compared to simulations at room temperature in aqueous solution). More generally, biophysical measurements made in different experimental conditions, over disparate time or length-scales or on molecular fragments of a far larger macromolecular complex can be difficult to reconcile with each other, and in a cellular context. In these situations, simulations can assist to bridge the gap between experimentally accessible regimes, and provide valuable insight and understanding. For example, atomistic MD simulations have been used to reconcile the differences in experimental DNA persistence length measurements from Atomic Force Microscopy (AFM) stretching experiments, which probe the mechanical response over very long sequences, and from small angle X-ray scattering (SAXS), which detected stretching motions over distances of <40 base pairs. This analysis showed that the mechanical properties of DNA are different at the base pair level compared to the bulk, and that the cross-over occurs over around a helical turn.(Noy & Golestanian, 2012)


**Atomistic Simulations of Supercoiled DNA**

While short linear sequences of DNA are common in the PDB, in the cell they are non-existent, due to DNA degrading enzymes. Recent advances in high throughput sequencing have now made it possible to make an extensive genome wide profile of extrachromosomal circular DNAs (eccDNAs), which are an endogenous population of chromosome-derived circular DNA structures prevalent in repetitive DNA that codes for proteins with diverse isoforms, such as mucin and titin.(Shoura et al., 2017) However, the biological function (if any) of eccDNAs, the protein-DNA complexes they form within cells, and the molecular mechanisms that produce them remain poorly understood. Circular DNA is also challenging to simulate, because the relatively long persistence length of DNA (50 nm or 150 base pairs) requires DNA mini-circles that are over 200bp in size. Smaller circles are more difficult to synthesise biochemically, which places severe constraints on the biophysical tools that can be used to investigate them. An additional complexity of topologically closed DNA is that it can sustain supercoiling, where the biopolymer is over or under-twisted relative to the unrestrained linear sequence. This is highly biologically relevant, because DNA is usually maintained in a negatively supercoiled state by the action of DNA gyrase (in most prokaryotes), and subjected to large topological stresses during transcription and replication. Both the global nature of DNA supercoiling, and its transience make it impossible to study supercoiled DNA in the cell at atomistic resolution. However, the structure of supercoiled DNA mini-circles, which have a constrained structure relative to larger plasmids, has been characterised using cryo-electron tomography (cryo-ET) for a biologically relevant range of superhelical densities.(Irobalieva et al., 2015) The images

showed mini-circles adopting a surprisingly broad range of conformations, even within a single topoisomer. Complementary atomistic MD simulations also revealed supercoiled DNA to be highly dynamic, and provided structures sufficiently close in shape to the cryo-ET that they could be aligned to provide an atomistic interpretation of the measured electron density. For the most negatively supercoiled minicircles, good agreement was only obtained when the MD simulations had been run long enough for local complementary hydrogen bonds between the double strands to denature at the tips of loop apices, as has also been observed using coarse-grained simulations of far longer supercoiled DNA plectonemes (see section **Coarse Graining and Membrane Simulation**).(Matek, Ouldridge, Doye, & Louis, 2015)


**Simulations of DNA for Nanotechnology**

The last ten years has seen the rapid development of methods to build complex nanostructures by the self-assembly of carefully designed DNA oligonucleotides, so-called "DNA Origami". While these structures may be imaged – e.g. by AFM or cryo electron microscopy (cryoEM) – details of their atomistic structure cannot be resolved. This is a significant drawback because, while the sequence-based rules that are used in their design often work to the extent that the desired material is produced, in many cases the yield of the nanostructure is poor and purifying it from contaminants is difficult, if not impossible. A better understanding of the three-dimensional structure of the complex non-canonical multiway junctions that feature in these structures would lead to improved design methods. Recently Maffeo et al.(Maffeo, Yoo, & Aksimentiev, 2016) have reported the use of atomistic MD simulations to relax an idealised model of a 4.7 megadalton DNA Origami 'pointer' structure. The initial model, built using a simple rules-based approach which presumes, for example, that all DNA duplexes are completely regular and straight helical objects, was significantly at odds with the cryo-EM data on the same nanostructure. After 200 ns of carefully-controlled MD, the structure relaxed to a conformation that fitted the experimental data within experimental error. While optimising a pre-built model of such a nanostructure is now possible using standard atomistic simulation methods,(Maingi, Lelimousin, Howorka, & Sansom, 2015) these structures are too large for their *de novo* self-assembly to be simulated using the same approach. In response to this situation a number of 'coarse-grained' DNA models have developed. For example, Schreck et al. (Schreck, Romano, Zimmer, Louis, & Doye, 2016) have recently reported the simulation of the self-assembly of a triangular prism DNA nanostructure using the OxDNA force field. As well as reproducing key details of the structure that were evident in cryo-TEM images (e.g. the twist in the prism, such that the top and bottom triangular faces are offset), the simulations allow the study of the process of assembly, including intermediate structures that are formed. Coarse-grained models also permit longer timescale simulations of assembled nanostructures. For example, Maingi et al(Maingi et al., 2017) have used the Martini force field to study the structure and interactions of a DNA nanopore spanning a lipid bilayer. The simulations give insights into how the nanopore inserts into the membrane, and how it selectively transports cations across it. DNA sequencing with nanopores (protein and synthetic) has been an active area of research within academia and industry for the last 10 years or so. The idea being that DNA is driven through a nanopore by an electric field. As it enters the constriction within the nanopore, it causes a partial blockage of the baseline current that is present for the unoccluded pore. Each one of the four DNA nucleotides gives a slightly different partial blockage, allowing the sequence to be determined. Atomistic simulations have

provided vital details to improve the design of the nanopores used for sequencing. For example the energetic barriers to translocation for the different nucleotides have been calculated for different protein pores(Manara, Guy, Wallace, & Khalid, 2015) the mechanism of ssDNA translocation has been predicted(Guy, Piggot, & Khalid, 2012) and the importance of the directionality of the DNA has been determined.(Mathé, Aksimentiev, Nelson, Schulten, & Meller, 2005)

## Simulations of Protein-DNA Complexes

Most biological functions of DNA are mediated by its interaction with proteins. This interaction may be sequence-specific or not. Simulations can help to analyse the origins of selectivity. For example, Garton et al.(Garton & Laughton, 2013) have used extensive MD simulations to study the interaction between telomeric repeat-binding factor 1 (TRF1) and its cognate DNA sequence. They find that the recognition involves a dynamic, 'flickering' set of interactions, at any instant in time these are insufficient to guarantee selectivity for one DNA sequence over others, but averaged over a matter of nanoseconds they are. Through this analysis the simulations were able to explain known relative binding affinities of TRF1 for different DNA sequences. It is a characteristic of such interactions that they frequently involve significant changes to the structure of the DNA – e.g. bending or twisting. This has led to much investigation as to whether sequence-specific DNA-protein recognition comes about through the specificity of interactions made between the protein and particular bases in the DNA sequence, or indirectly through the particular ability of the correct DNA sequence to adopt the structure optimal for these interactions. In other words, analysing the process of induced fit is very important. Recently, Etheve et al.(Etheve, Martin, & Lavery, 2016) have studied three examples of protein-DNA complexes in which the proportion of 'direct' (via interaction) versus 'indirect' (via deformation) readout was predicted to vary. The authors find that this is indeed the case, and show good agreement with experimental data on sequence selectivity. Interestingly there are added subtleties – for example, some complexes may adopt a number of alterative low-energy states, each of which has a difference balance between direct and indirect readout. The simulations also suggest that in some cases, and perhaps not unexpectedly, ions can play a significant role in mediating the interactions between the highly charged DNA species and the protein.

## Chromatin and Beyond

The advent of new biophysical and chemical tools are switching the focus of DNA structural biology from individual protein-DNA complexes up to the nuclear level. The chromosome capture technologies, which use cross-linking and sequencing to map inter-chromosomal contacts, rely on simulation as an inherent part of the data analysis process, (Bascom & Schlick, 2017) in an analogous manner to the use of distance restraints in NMR structure refinement at the atomistic level. In these calculations, chromatin (or the prokaryotic equivalent) is represented as a simple polymer chain. Putative 3D structures of the chromosomes are then iteratively refined until sufficiently good agreement with the experimental data is achieved. This combined experimental/computational approach has shown that the nuclear material is decomposed into topologically associated domains (TADS), which can bring activators on distal sites into close proximity, and which appear to be vital to genome regulation and stability. For *E. coli*, it has been possible to gather sufficient experimental information to build a hierarchical structural model in which all 4.6 Mbp of DNA in the bacterial

chromosome were resolved at the single nucleotide level. Such models provide contact probabilities for distant promoters within the 3D chromosome, the physical environment surrounding a particular sequence of interest and the distribution and position of regions of empty space within the highly packed DNA, which has implications for the ability of large macromolecules such as ribosomes to access the interior of the chromosome.(Hacker, Li, & Elcock, 2017)

## MEMBRANE SIMULATIONS AND COARSE GRAINING

Simulation studies of biological membranes and membrane proteins have a long history of complementarity to ongoing experimental studies. The systems are challenging to study at the molecular level using experimental methods alone, and thus simulations have played a key role in facilitating interpretation of the experimental observables. For example, experimental studies of the *E. coli* outer membrane protein A (OmpA) gave some initially puzzling results; the X-ray structure of the protein revealed that the interior of OmpA did not have a continuous channel extending from one mouth of the protein to the other, yet electrophysiology experiments showed a conductance for the protein (see Figure 7). Molecular modelling and MD simulations revealed amino acid rearrangement within the lining of the protein, resulted in an 'open' pore with dimensions that closely matched those predicted by the conductance data.(Bond, Faraldo-Gómez, & Sansom, 2002) The simulation hypothesis of side-chain rearrangement was subsequently confirmed by NMR studies.(Hong, Szabo, & Tamm, 2006) This study showed the utility of simulations in rationalizing two apparently conflicting experimental results.
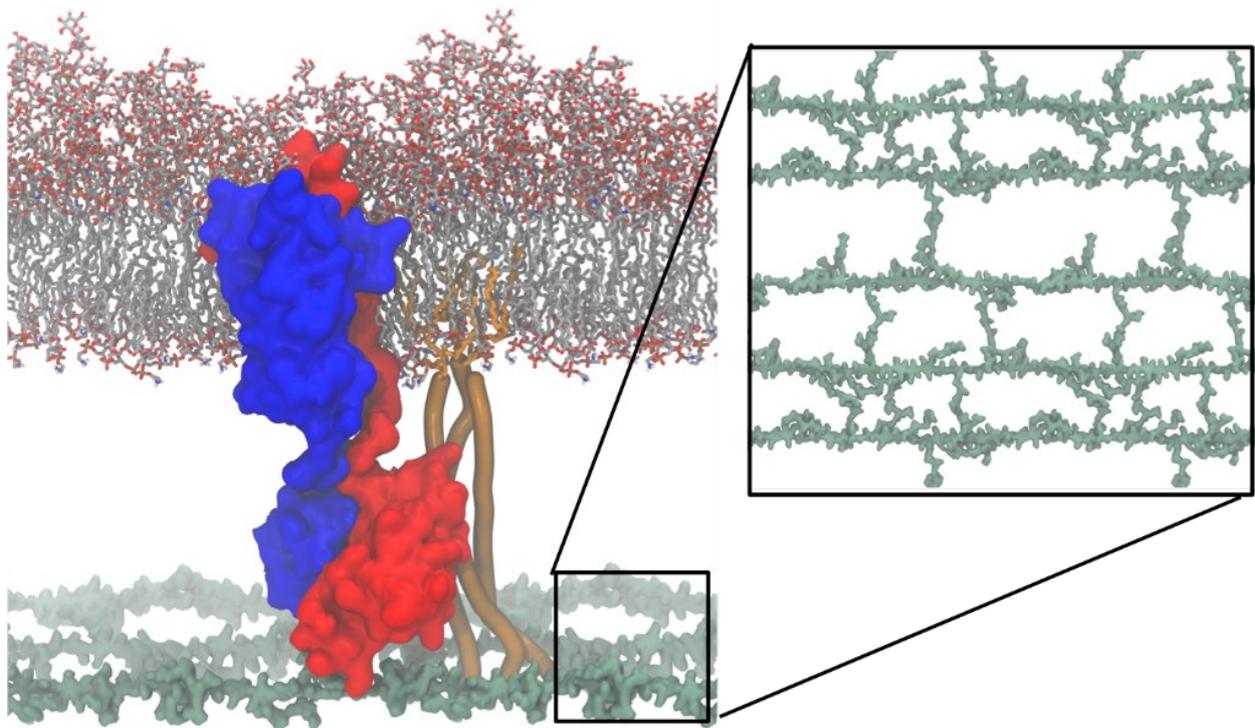


**Figure 7 -** The outer membrane protein OmpA is shown in its dimeric form in red and blue. Braun's lipoprotein, which is covalently bound to peptidoglycan is shown in orange. Peptidoglycan is shown

in green. The lipids of the outer membrane of <u>E. coli</u> are shown in grey and red.(Samsudin, Boags, Piggot, & Khalid, 2017)

In the following we discuss two areas in which membrane simulations have seen tremendous growth in the last five years or so, the membrane envelopes of bacteria and viruses. Bacteria can be divided into two categories according to the architectures of their cell envelopes. Gram-positive bacteria have a single membrane and a thick peptidoglycan cell wall, whereas Gram-negative bacteria have a thin cell wall separating two membranes. The two membranes of Gram-negative bacteria differ in their lipid compositions, the inner membrane is composed of phospholipids in both leaflets, in contrast the outer membrane is asymmetric; it contains phospholipids in the inner leaflet, but the larger, more complex lipopolysaccharide molecules in the outer leaflet. Experimental and computational studies have traditionally approximated both membranes as symmetric phospholipid bilayers. These studies have provided numerous insights into the structure-dynamics-function relationships of the proteins embedded within the membranes. For example a study of lipopolysaccharide (LPS) transport proteins D and E (LptD and LptE) from *S. typhimurium* and *S. flexneri* using mutagenesis, functional assays and molecular dynamics simulations proposed that a luminal gate composed of two luminal loops opens up to allow insertion of oligosaccharide and O-antigen regions of LPS into the outer leaflet.(Y. Gu et al., 2015) The simulations revealed the role of key cysteine residues in these loops; when in the reduced form and unable to form disulfide bridges to a third nearby loop, the gate was observed to close, in contrast in the presence of the disulfide bond, the gate remained open. In mutagenesis studies deletion of either of the loops resulted in cell death. A more recent study of the LptD/E complex from a different bacterial species; *P. aeruginosa*, in which the protein was simulated in its native LPS-containing membrane environment provided further support for the luminal gate mechanism.(Botos et al., 2016)

The first reported LPS-containing model of the outer membrane was developed with the Amber95 force field by Straatsma,(Lins & Straatsma, 2001) this was followed a few years later by models for CHARMM(Wu et al., 2013) and GROMOS.(Piggot, Holdbrook, & Khalid, 2011) One of the difficulties in developing parameters for membranes containing LPS is that experimental data for these systems is much scarcer than data for phospholipids, making validation of the molecular models difficult. Having said that, the three models mentioned show quantitative agreement with acyl chain deuterium order parameters obtained from NMR experiments and also reproduce the phase behaviour of these complex lipids.(Wu et al., 2013) A recent example of simulations and experiment in agreement with each in highlighting the importance of divalent cations for cross-linking LPS headgroups was reported by Clifton et al.(Clifton et al., 2015) Given the complexity of the LPS molecule and its slow rate of diffusion (an order of magnitude slower than phospholipids) even equilibration of systems containing these molecules is computationally very demanding and consequently it has only been in the last 6 or 7 years that simulations of such LPS-containing membranes have become widespread. These models are enabling the use of simulation to complement experimental data in identifying specific protein-lipid interactions.(J. Lee, Patel, Kucharska, Tamm, & Im, 2017) Progress in the development of atomistic models has made it possible to carry out computational studies of all components of the cell envelope of Gram-negative

bacteria and such studies are now beginning to emerge.(Ortiz-Suarez, Samsudin, Piggot, Bond, & Khalid, 2016; Samsudin, Ortiz-Suarez, Piggot, Bond, & Khalid, 2016) In the future these models and simulation studies are likely to play a major role in our understanding of the interaction of antibiotics with the cell envelopes that protect bacterial cells.


Large scale molecular dynamics simulations have also been used to explore the membranes of a number of enveloped viruses, providing both mechanistic insights into virus biology, and exemplars of very large scale simulations of biological membrane assemblies.(Reddy & Sansom, 2016a) One of the first simulation studies of the membrane envelope of a virus was of the immature HIV-1 virion by Ayton and Voth.(Ayton & Voth, 2010) In this landmark study, a combination of electron cryotomography data and both atomistic and coarse-grained simulations were used to identify potential mutations that may disrupt or modify the assembly process of the Gag polyprotein lattice in the immature HIV-1 virion. The simulations employed an iterative multiscale approach in which multiple coarse-grain parameters were explored in critical regions, with the aim of identifying the interactions that are critical to maintaining the experimentally observed structure of the virion. Subsequently the CG simulation data were used in an inverse fashion to guide all-atom MD simulations of select regions in order to refine the CG model. Simulation of a complete virion envelope has been possible for the influenza A virion,(Reddy et al., 2015) combining X-ray structures and transmembrane (TM) domain models for the hemagglutinin (HA) and neuraminidase (NA) proteins, an NMR structure for the TM domain of the M2 protein, and a reasonable approximation to the experimentally determined lipidome (Figure 8 below). The prevalence of glycans on the surface of the outer leaflet of the lipid bilayer of the influenza A model suggested that antibody or therapeutic compound access to the M2 proton channel may have to overcome steric barriers. The three species of influenza A envelope protein moved slowly within the cholesterol-rich membrane, with diffusion constants matching previous experimental measurements by solid-state NMR. (Polozov, Bezrukov, Gawrisch, & Zimmerberg, 2008) The spacing between membrane glycoprotein molecules on the influenza A surface was consistent with previous experimental measurements. (Wasilewski, Calder, Grant, & Rosenthal, 2012) These spacings were analysed in the context of multivalence, suggesting that polyvalent interactions between HA and/or NA on the viral surface and sialic acid residues on the host membrane are likely. This would enable strong virus-host association despite relatively weak (~2-3 mM affinity) viral HA-single host receptor interaction *in vitro*.(Sauter et al., 1989)
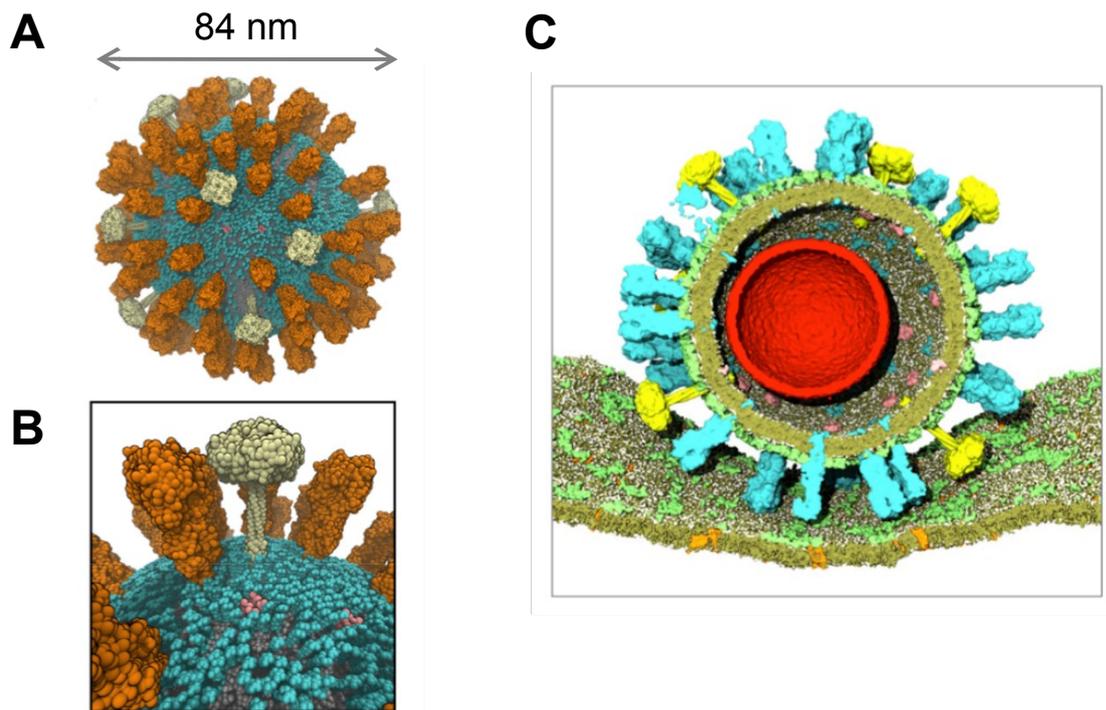
**Figure 8 -** The viral envelope membrane of influenza A modelled at coarse-grained resolution. The hemagglutinin, neuraminidase, and M2 proteins are orange, yellow and pink respectively. The glycolipid molecules are shown in cyan, and other lipids in grey. Overall (**A**) and zoomed-in (**B**) views are shown. **C** Model of an influenza A virion (with the red sphere indicating the approximate location of the genome within the virion, not currently modelled) docked against a simple model of a mammalian cell membrane(Koldsø & Sansom, 2015) with glycolipids in pale green, other lipids in darker green, and cell membrane proteins in orange.

The membrane envelope of the Dengue virus has been simulated in two independent studies (Reddy & Sansom, 2016b) using the Martini coarse-grain force field.(Marrink & Tieleman, 2013) Reddy & Sansom used a combination of CG modelling and simulation to 'add back' the lipid bilayer to the cryoEM structure of the Dengue virus envelope proteins(Reddy & Sansom, 2016b) (Figure 9 below). This allowed them to address diffusive properties of lipids within the outer envelope of the dengue virion. The dense crowding of protein TM domains and the enclosure of the outer leaflet of the lipid bilayer within a protein shell resulted in lipid diffusive properties similar to those in the raft-like influenza A virion membrane, namely reduced diffusion coefficients $D$ and exponents $\alpha$ less than 1, the latter indicative of anomalous diffusion. Simulation studies on crowded bacterial membrane models(Goose & Sansom, 2013) exhibited diffusive behaviour similar to that in the membrane of the enveloped viruses.

**A**

**B**

48 nm

protein  bilayer
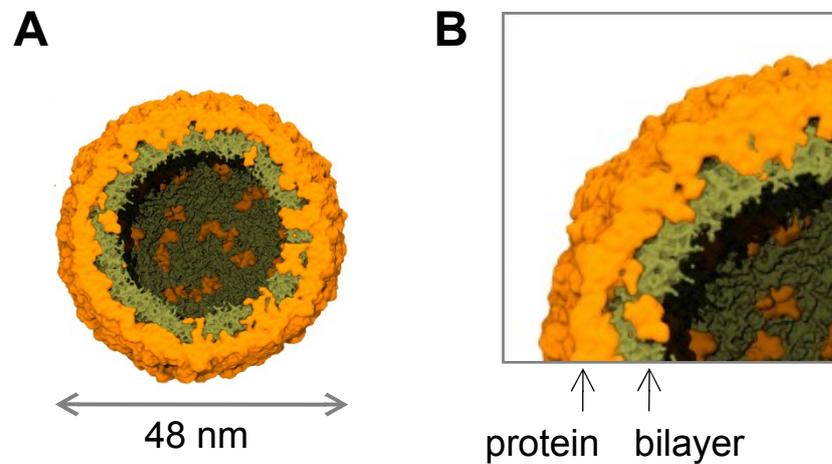
**Figure 9 -** A coarse-grained model of the envelope of dengue virus(Reddy & Sansom, 2016b) shown as (**A**) a cross-section of the entire virion envelope model is shown, along with (**B**) a zoomed-in view. Protein (orange) and lipid (dark green) components of the envelope are shown.

Bond and co-workers(Marzinek, Holdbrook, Huber, Verma, & Bond, 2016) developed a novel optimisation protocol to embed the cryo-electron microscopy structure of the envelope protein complexes of the DENV-2 icosahedral shell within a spherical lipid vesicle, the composition of which was guided by lipidomics data. Microsecond timescale simulations of the virion envelope enabled refinement of the lipid/protein complex, assessed by comparing density maps generated from simulations with those from cryo-electron microscopy. The refined structures revealed locally induced curvature resulting from specific electrostatic interactions with anionic phosphatidyl-serine phospholipids. These lipids were shown to stabilise the native architecture of the transmembrane domains, and may facilitate subsequent fusion of the viral envelope with the host membrane inside the endosome during infection. These and other studies demonstrate the potential of very large scale simulation of viral envelopes. Future challenges for such studies include a full model of glycosylation of viral surface proteins, which will enable more realistic and hence predictive modelling of virions binding to models of target cell membranes, and the effects of antibodies on such interactions. Models of virions and other complex mesoscale biomolecular assemblies can be built with tools such as cellPACK(Johnson et al., 2014) with LipidWrapper,(Durrant & Amaro, 2014) incorporating a variety of experimental data. These models can be built in atomic detail, allowing simulations from this level and potentially connecting to larger scales.(Amaro & Mulholland, 2018)

## Conclusion

Biomolecular simulation exploits and complements the vast amounts of structural data generated from techniques such as X-ray crystallography, NMR, and cryo-electron microscopy to model and thus understand interactions, mechanism and function. As experiments generate ever more structural, genetic and kinetic data, the need for simulation and model building becomes greater. Complementary tools, such as microscopy and cross-linking, often provide such different types of experimental information that it is impossible to reconcile them without computer modelling. The multi-scale nature of biology implies that no single experimental technique, or informatics dataset will be sufficient to inform reliable manipulation and modulation of biological systems. It is only by combining experiment with modelling and simulation that we will be able to connect together all of

the different types of biological information, and thus gain an understanding how information flow is controlled during genome activity, for example. Simulations have an essential role in interpreting, extending and adding value in structural investigations, complementing experiment. For example, Hub and co-workers have used atomistic models of macromolecules to extract information from SAXS data.(Chen & Hub, 2014) Simulations (e.g. with QM/MM methods) can exploit structures of enzyme reaction intermediates (Olmos et al., 2018) to model reactions in proteins and predict barriers.(Van der Kamp & Mulholland, 2013) Great opportunities (and challenges) for synergistic investigation come from cryo-EM e.g. in developing and tests models of biomolecular ensembles, from individual proteins to macromolecular complexes.

Simulation also allow the study of biomolecular ensembles, allowing calculations of the free energy differences which determine equilibria. Ongoing developments in biomolecular simulation will vastly increase their scope and utility. We are likely to see more progress on the computation of various aspects of ligand binding kinetics(Deganutti & Moro, 2017) particularly in a pharmaceutical context(Schuetz et al., 2017). One exciting example is using MSMs to characterise allosteric binding sites(Bowman, Bolin, Hart, Maguire, & Marqusee, 2015). One of the central challenges is how to use this information in a predictive fashion. It is of course important that, where possible, predictions from simulation are compared directly to experimental observables. Also, simulations should demonstrate significance and reproducibility (as with all experiments) in order to give reliable predictions. The effects of e.g. choice of forcefield and uncertainties in structural models should also be considered. When applied with care, simulations have shown that they can be computational assays for predicting important properties such as drug resistance.

Using methods such as constant pH MD,(M. S. Lee, Salsbury, & Brooks, 2004; Mongan, Case, & McCAMMON, 2004) polarisable force fields,(Albaugh et al., 2016; Patel, Mackerell, & Brooks, 2004) QM/MM (see section **QM/MM modelling of enzyme reactions)**, and ab initio MD(Marx & Hutter, 2009; Tuckerman, Laasonen, Sprik, & Parrinello, 1995) can yield more accurate ensembles and allow modelling of biomolecular reactions in addition to interactions. Simulations on larger scales will analyse the fundamental processes involved in complex biomolecular processes such as photosynthesis, ATP production, and ubiquitination.(Johnson et al., 2014) The increasing ability to reach long timescales brings into view detailed simulation of processes such as protein folding, translation, and transcription. It will also allow us to predict biomolecular kinetics with high accuracy. Finally, multiscale approaches(Amaro & Mulholland, 2018) promise to connect interactions and reactions modelled and revealed by biomolecular simulations to macroscopic observables at the levels of organelles and cells, and potentially tissues, and beyond.

## Acknowledgments

## References

Abel, R., Young, T., Farid, R., Berne, B. J., & Friesner, R. A. (2008). Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *Journal of the American Chemical Society, 130*(9), 2817-2831. doi:10.1021/ja0771033

Albaugh, A., Boateng, H. A., Bradshaw, R. T., Demerdash, O. N., Dziedzic, J., Mao, Y., . . . Case, D. A. (2016). Advanced potential energy surfaces for molecular simulation. *The Journal of Physical Chemistry B, 120*(37), 9811-9832.

Aldeghi, M., & Biggin, P. C. (2016). Advances in Molecular Simulation *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*: Elsevier.

Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics, 31*(2), 459-466.

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*, 0148. doi:10.1038/s41570-018-0148

Ayton, G. S., & Voth, G. A. (2010). Multiscale computer simulation of the immature HIV-1 virion. *Biophysical Journal, 99*(9), 2757-2765.

Bai, X.-C., McMullan, G., & Scheres, S. H. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences, 40*(1), 49-57.

Baker, N. A. (2004). Poisson-Boltzmann methods for biomolecular electrostatics. *Numerical Computer Methods, Pt D, 383*, 94-+.

Barducci, A., Bonomi, M., & Parrinello, M. (2011). Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 1*(5), 826-843. doi:10.1002/wcms.31

Bascom, G., & Schlick, T. (2017). Linking chromatin fibers to gene folding by hierarchical looping. *Biophysical Journal*.

Beglov, D., & Roux, B. (1995). NUMERICAL-SOLUTION OF THE HYPERNETTED-CHAIN EQUATION FOR A SOLUTE OF ARBITRARY GEOMETRY IN 3 DIMENSIONS. *Journal of Chemical Physics, 103*(1), 360-364. doi:10.1063/1.469602

Bennett, C. H. (1976). Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *Journal of Computational Physics, 22*(2), 245-268. doi:Doi 10.1016/0021-9991(76)90078-4

Bennie, S. J., van der Kamp, M. W., Pennifold, R. C. R., Stella, M., Manby, F. R., & Mulholland, A. J. (2016). A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *Journal of chemical theory and computation, 12*(6), 2689-2697.

Bernardi, R. C., Melo, M. C., & Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1850*(5), 872-877.

Bernstein, N., Kermode, J. R., & Csanyi, G. (2009). Hybrid atomistic simulation methods for materials systems. *Reports on Progress in Physics, 72*(2), 026501.

Bodnarchuk, M., Bradshaw, R., Cave-Ayland, C., Genheden, S., Cabedo-Martinez, A., Michel, J., . . . Woods, C. (2017). ProtoMS: University of Southampton: Highfield, Southampton.

Bond, P. J., Faraldo-Gómez, J. D., & Sansom, M. S. (2002). OmpA: a pore or not a pore? Simulation and modeling studies. *Biophysical Journal, 83*(2), 763-775.

Bono, F., De Smet, F., Herbert, C., De Bock, K., Georgiadou, M., Fons, P., . . . Carmeliet, P. (2013). Inhibition of tumor angiogenesis and growth by a small-molecule multi-FGF receptor blocker with allosteric properties. *Cancer Cell, 23*(4), 477-488. doi:10.1016/j.ccr.2013.02.019

Botos, I., Majdalani, N., Mayclin, S. J., McCarthy, J. G., Lundquist, K., Wojtowicz, D., . . . Buchanan, S. K. (2016). Structural and functional characterization of the LPS transporter LptDE from Gram-negative pathogens. *Structure, 24*(6), 965-976.

Boulanger, E., & Harvey, J. N. (2018). QM/MM methods for free energies and photochemistry. *Current Opinion in Structural Biology, 49*, 72-76.

Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., . . . Sacerdoti, F. D. (2006). *Scalable algorithms for molecular dynamics simulations on commodity clusters.* Paper presented at the Proceedings of the 2006 ACM/IEEE conference on Supercomputing.

Bowman, G. R., Beauchamp, K. A., Boxer, G., & Pande, V. S. (2009). Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys., 131*(12), 124101. doi:10.1063/1.3216567

Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C., & Marqusee, S. (2015). Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Nat. Acad. Sci. USA, 112*(9), 2734-2739.

Brooks, B. R., Brooks, C. L., MacKerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., . . . Boresch, S. (2009). CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry, 30*(10), 1545-1614.

Buch, I., Giorgino, T., & De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA, 108*(25), 10184-10189.

Buch, I., Harvey, M. J., Giorgino, T., Anderson, D. P., & De Fabritiis, G. (2010). High-throughput all-atom molecular dynamics simulations using distributed computing. *J. Chem. Inf. Model., 50*(3), 397-403. doi:10.1021/ci900455r

Bussi, G., Gervasio, F. L., Laio, A., & Parrinello, M. (2006). Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society, 128*(41), 13435-13441. doi:10.1021/ja062463w

Callegari, D., Ranaghan, K., Woods, C., Minari, R., Tiseo, M., Mor, M., . . . Lodola, A. (2018). L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chemical Science, 9*(10), 2740-2749.

Carloni, P., Rothlisberger, U., & Parrinello, M. (2002). The role and perspective of ab initio molecular dynamics in the study of biological systems. *Accounts of Chemical Research, 35*(6), 455-464.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., . . . Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry, 26*(16), 1668-1688.

Chen, P.-c., & Hub, J. S. (2014). Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophysical Journal, 107*(2), 435-447.

Chodera, J. D., & Noe, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol, 25*, 135-144. doi:10.1016/j.sbi.2014.04.002

Chodera, J. D., & Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struc. Biol., 25*, 135-144. doi:http://dx.doi.org/10.1016/j.sbi.2014.04.002

Chudyk, E. I., Limb, M. A., Jones, C., Spencer, J., van der Kamp, M. W., & Mulholland, A. J. (2014). QM/MM simulations as an assay for carbapenemase activity in class A β-lactamases. *Chemical Communications, 50*(94), 14736-14739.

Claeyssens, F., Freeman, C. L., Allan, N. L., Sun, Y., Ashfold, M. N., & Harding, J. H. (2005). Growth of ZnO thin films—experiment and theory. *Journal of Materials Chemistry, 15*(1), 139-148.

Claeyssens, F., Harvey, J. N., Manby, F. R., Mata, R. A., Mulholland, A. J., Ranaghan, K. E., . . . Werner, H. J. (2006). High-accuracy computation of reaction barriers in enzymes. *Angew. Chemie Int. Ed., 45*(41), 6856-6859.

Claeyssens, F., Harvey, J. N., Manby, F. R., Mata, R. A., Mulholland, A. J., Ranaghan, K. E., . . . Werner, H. J. (2006). High-accuracy computation of reaction barriers in enzymes. *Angewandte Chemie, 118*(41), 7010-7013.

Clifton, L. A., Holt, S. A., Hughes, A. V., Daulton, E. L., Arunmanee, W., Heinrich, F., . . . Webster, J. R. (2015). An accurate in vitro model of the E. coli envelope. *Angewandte Chemie International Edition, 54*(41), 11952-11955.

Cole, D. J., Janecek, M., Stokes, J. E., Rossmann, M., Faver, J. C., McKenzie, G. J., . . . Huggins, D. J. (2017). Computationally-guided optimization of small-molecule inhibitors of the Aurora A kinase–TPX2 protein–protein interaction. *Chemical Communications, 53*(67), 9372-9375.

Czub, J., Wieczór, M., Prokopowicz, B., & Grubmüller, H. (2017). Mechanochemical energy transduction during the main rotary step in the synthesis cycle of F1-ATPase. *Journal of the American Chemical Society, 139*(11), 4025-4034. doi:10.1021/jacs.6b11708

Deganutti, G., & Moro, S. (2017). Estimation of kinetic and thermodynamic ligand-binding parameters using computational strategies. *Future Med. Chem., 9*(5), 507-523. doi:10.4155/fmc-2016-0224

Dickson, A., Tiwary, P., & Vashisth, H. (2017). Kinetics of ligand binding through advanced computational approaches: A review. *Curr Top Med Chem., 17*, 2626-2641.

Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., . . . Shaw, D. E. (2011). Activation mechanism of the β2-adrenergic receptor. *Proc. Natl. Acad. Sci. USA, 108*(46), 18684-18689.

Duarte, F., Amrein, B. A., Blaha-Nelson, D., & Kamerlin, S. C. L. (2015). Recent advances in QM/MM free energy calculations using reference potentials. *Biochimica et Biophysica Acta (BBA) - General Subjects, 1850*(5), 954-965. doi:https://doi.org/10.1016/j.bbagen.2014.07.008

Durrant, J. D., & Amaro, R. E. (2014). LipidWrapper: an algorithm for generating large-scale membrane models of arbitrary geometry. *PLoS Computational Biology, 10*(7), e1003720.

Eastman, P., Friedrichs, M. S., Chodera, J. D., Radmer, R. J., Bruns, C. M., Ku, J. P., . . . Shukla, D. (2012). OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation, 9*(1), 461-469.

Etheve, L., Martin, J., & Lavery, R. (2016). Protein–DNA interfaces: a molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic Acids Research, 44*(20), 9990-10002.

Feixas, F., Lindert, S., Sinko, W., & McCammon, J. A. (2014). Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem., 186*(Supplement C), 31-45. doi:https://doi.org/10.1016/j.bpc.2013.10.007

Field, M. J., Bash, P. A., & Karplus, M. (1990). A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *Journal of computational chemistry, 11*(6), 700-733.

Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual Review of Biophysics and Biomolecular Structure, 31*(1), 303-319.

Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences, 108*(39), 16247-16252.

Friesner, R. A. (2004). Combined quantum and molecular mechanics (QM/MM). *Drug Discovery Today: Technologies, 1*(3), 253-260.

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., . . . Nakatsuji, H. (2016). Gaussian 16, revision A. 03. *Gaussian Inc., Wallingford CT*.

Fritz, R., Alzate-Morales, J. H., Spencer, J., Mulholland, A. J., & van der Kamp, M. W. (2018). Multiscale simulations of clavulanate inhibition identify the reactive complex in class A β-lactamases and predict efficiency of inhibition. *Biochemistry*.

Galindo-Murillo, R., Roe, D. R., & Cheatham III, T. E. (2014). On the absence of intra-helical DNA dynamics on the μs to ms timescale. *Nature Communications, 5*, 5152.

Garcia-Viloca, M., Gao, J., Karplus, M., & Truhlar, D. G. (2004). How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science, 303*(5655), 186.

Garton, M., & Laughton, C. (2013). A comprehensive model for the recognition of human telomeres by TRF1. *Journal of molecular biology, 425*(16), 2910-2921.

Gerogiokas, G., Calabro, G., Henchman, R. H., Southey, M. W. Y., Law, R. J., & Michel, J. (2014). Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *Journal of chemical theory and computation, 10*(1), 35-48. doi:10.1021/ct400783h

Goose, J. E., & Sansom, M. S. (2013). Reduced lateral mobility of lipids and proteins in crowded membranes. *PLoS Computational Biology, 9*(4), e1003033.

Granata, D., Camilloni, C., Vendruscolo, M., & Laio, A. (2013). Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proceedings of the National Academy of Sciences, 110*(17), 6817-6822. doi:10.1073/pnas.1218350110

Grubmuller, H. (1995). Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 52*(3), 2893-2906.

Gu, S., Silva, D.-A., Meng, L., Yue, A., & Huang, X. (2014). Quantitatively characterizing the ligand binding mechanisms of choline binding protein using markov state model analysis. *PLoS Comp. Biol., 10*(8), e1003767. doi:10.1371/journal.pcbi.1003767

Gu, Y., Stansfeld, P. J., Zeng, Y., Dong, H., Wang, W., & Dong, C. (2015). Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD. *Structure, 23*(3), 496-504.

Gullingsrud, J. R., Braun, R., & Schulten, K. (1999). Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *Journal of Computational Physics, 151*(1), 190-211. doi:https://doi.org/10.1006/jcph.1999.6218

Guy, A. T., Piggot, T. J., & Khalid, S. (2012). Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophysical Journal, 103*(5), 1028-1036.

Hacker, W. C., Li, S., & Elcock, A. H. (2017). Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. *Nucleic Acids Research, 45*(13), 7541-7554.

Haensele, E., Mele, N., Miljak, M., Read, C. M., Whitley, D. C., Banting, L., . . . Clark, T. (2017). Conformation and dynamics of human urotensin II and urotensin related peptide in aqueous solution. *Journal of Chemical Information and Modeling, 57*(2), 298-310. doi:10.1021/acs.jcim.6b00706

Hansen, H. S., & Hünenberger, P. H. (2010). Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *Journal of computational chemistry, 31*(1), 1-23. doi:doi:10.1002/jcc.21253

Harder, E., Damm, W., Maple, J., Wu, C. J., Reboul, M., Xiang, J. Y., . . . Friesner, R. A. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of chemical theory and computation, 12*(1), 281-296. doi:10.1021/acs.jctc.5b00864

Harrigan, M. P., Sultan, M. M., Hernández, C. X., Husic, B. E., Eastman, P., Schwantes, C. R., . . . Pande, V. S. (2017). MSMBuilder: Statistical models for biomolecular dynamics. *Biophys. J., 112*(1), 10-15. doi:https://doi.org/10.1016/j.bpj.2016.10.042

Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M., & Laughton, C. A. (2001). Cooperativity in drug– DNA recognition: a molecular dynamics study. *Journal of the American Chemical Society, 123*(50), 12658-12663.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97-109.

Hediger, M. R., De Vico, L., Rannes, J. B., Jäckel, C., Besenmatter, W., Svendsen, A., & Jensen, J. H. (2013). In silico screening of 393 mutants facilitates enzyme engineering of amidase activity in CalB. *PeerJ, 1*, e145. doi:10.7717/peerj.145

Hediger, M. R., Steinmann, C., De Vico, L., & Jensen, J. H. (2013). A computational method for the systematic screening of reaction barriers in enzymes: searching for Bacillus circulans xylanase mutants with greater activity towards a synthetic substrate. *PeerJ, 1*, e111. doi:10.7717/peerj.111

Henchman, R. H. (2007). Free energy of liquid water from a computer simulation via cell theory. *Journal of Chemical Physics, 126*(6). doi:064504

10.1063/1.2434964

Henriksen, N. M., & Gilson, M. K. (2017). Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host–Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *Journal of chemical theory and computation, ASAP*.

Herbert, C., Schieborr, U., Saxena, K., Juraszek, J., De Smet, F., Alcouffe, C., . . . Bono, F. (2013). Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. *Cancer Cell, 23*(4), 489-501. doi:10.1016/j.ccr.2013.02.018

Hong, H., Szabo, G., & Tamm, L. K. (2006). Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nature Chemical Biology, 2*(11), 627-635.

Hospital, A., Andrio, P., Cugnasco, C., Codo, L., Becerra, Y., Dans, P. D., . . . Gelpí, J. L. (2016). BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Research, 44*(D1), D272-D278. doi:10.1093/nar/gkv1301

Hu, H., Lu, Z., & Yang, W. (2007). QM/MM Minimum Free-Energy Path: Methodology and Application to Triosephosphate Isomerase. *Journal of chemical theory and computation, 3*(2), 390-406. doi:10.1021/ct600240y

Huber, T., Torda, A. E., & van Gunsteren, W. F. (1994). Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol Des, 8*(6), 695-708.

Huggins, D. J. (2014). Estimating Translational and Orientational Entropies Using the k-Nearest Neighbors Algorithm. *Journal of Chemical Theory and Computation, 10*(9), 3617–3625. doi:10.1021/ct500415g

Huggins, D. J. (2015). Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophysical Journal, 108*(4), 928-936. doi:10.1016/j.bpj.2014.12.035

Huggins, D. J. (2016). Studying the role of cooperative hydration in stabilizing folded protein states. *Journal of Structural Biology, 196*(3), 394-406. doi:10.1016/j.jsb.2016.09.003

Hutter, J., Alavi, A., Deutsch, T., Bernasconi, M., Goedecker, S., Marx, D., . . . Parrinello, M. (1995). cpmd. *MPI für Festkörperforschung and IBM Zurich Research Laboratory, Stuttgart, 1995*.

Irobalieva, R. N., Fogg, J. M., Catanese Jr, D. J., Sutthibutpong, T., Chen, M., Barker, A. K., . . . Chiu, W. (2015). Structural diversity of supercoiled DNA. *Nature Communications, 6*.

Irwin, B. W. J., & Huggins, D. J. (2018). Estimating Atomic Contributions to Hydration and Binding using Free Energy Perturbation. *Journal of Chemical Theory and Computation*.

Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., . . . Balaceanu, A. (2016). Parmbsc1: a refined force-field for DNA simulations. *Nature Methods, 13*(1), 55.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters, 78*(14), 2690.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters, 78*(14), 2690-2693. doi:10.1103/PhysRevLett.78.2690

Johansson, L. C., Stauch, B., Ishchenko, A., & Cherezov, V. (2017). A bright future for serial femtosecond crystallography with XFELs. *Trends in Biochemical Sciences, 42*(9), 749-762.

Johnson, G. T., Autin, L., Al-Alusi, M., Goodsell, D. S., Sanner, M. F., & Olson, A. J. (2014). cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nature Methods, 12*(1), 85.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc, 118*(45), 11225-11236.

Jorgensen, W. L., & Tirado–Rives, J. (2005). Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *Journal of Computational Chemistry, 26*(16), 1689-1700.

Kamp, M. W., Prentice, E. J., Kraakman, K. L., Connolly, M., Mulholland, A. J., & Arcus, V. L. (2018). Dynamical origins of heat capacity changes in enzyme-catalysed reactions. *Nature Communications, 9*(1), 1177.

Karplus, M., & Kushick, J. N. (1981). Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules, 14*(2), 325-332.

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology, 9*(9), 646-652.

Killian, B. J., Kravitz, J. Y., & Gilson, M. K. (2007). Extraction of configurational entropy from molecular simulations via an expansion approximation. *Journal of Chemical Physics, 127*(2). doi:024107

10.1063/1.2746329

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *Journal of Chemical Physics, 3*(5), 300-313. doi:10.1063/1.1749657

Klebe, G. (2015). Applying thermodynamic profiling in lead finding and optimization. *Nature Reviews Drug Discovery, 14*(2), 95-110. doi:10.1038/nrd4486

Koldsø, H., & Sansom, M. S. (2015). Organization and dynamics of receptor proteins in a plasma membrane. *Journal of the American Chemical Society, 137*(46), 14694-14704.

Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., . . . Kobilka, B. K. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature, 482*(7386), 552-556. doi:http://www.nature.com/nature/journal/v482/n7386/abs/nature10867.html#supplementary-information

Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N. A., Rendek, K. N., . . . Wang, D. (2014). Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature, 513*(7517), 261.

Kuzmanic, A., Sutto, L., Saladino, G., Nebreda, A. R., Gervasio, F. L., & Orozco, M. (2017). Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *eLife, 6*, e22175. doi:10.7554/eLife.22175

Laio, A., & Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics, 71*(12), 126601.

Lazaridis, T. (1998). Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *Journal of Physical Chemistry B, 102*(18), 3531-3541.

Lee, J., Patel, D. S., Kucharska, I., Tamm, L. K., & Im, W. (2017). Refinement of OprH-LPS interactions by molecular simulations. *Biophysical Journal, 112*(2), 346-355.

Lee, M. S., Salsbury, F. R., & Brooks, C. L. (2004). Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Structure, Function, and Bioinformatics, 56*(4), 738-752.

Lever, G., Cole, D. J., Lonsdale, R., Ranaghan, K. E., Wales, D. J., Mulholland, A. J., . . . Payne, M. C. (2014). Large-scale density functional theory transition state searching in enzymes. *The journal of physical chemistry letters, 5*(21), 3614-3619.

Levitt, M., & Lifson, S. (1969). Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology, 46*(2), 269-279.

Li, Z., & Scheraga, H. A. (1987). Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences, 84*(19), 6611-6615.

Liang, R., Swanson, J. M. J., Madsen, J. J., Hong, M., DeGrado, W. F., & Voth, G. A. (2016). Acid activation mechanism of the influenza A M2 proton channel. *Proceedings of the National Academy of Sciences, 113*(45), E6955-E6964. doi:10.1073/pnas.1615471113

Lifson, S., & Warshel, A. (1968). Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *The Journal of Chemical Physics, 49*(11), 5116-5129.

Lin, S.-T., Maiti, P. K., & Goddard III, W. A. (2010). Two-phase thermodynamic model for efficient and accurate absolute entropy of water from molecular dynamics simulations. *The Journal of Physical Chemistry B, 114*(24), 8191-8198.

Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M., & Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature, 433*(7022), 128.

Lins, R. D., & Straatsma, T. (2001). Computer simulation of the rough lipopolysaccharide membrane of Pseudomonas aeruginosa. *Biophysical Journal, 81*(2), 1037-1046.

Lonsdale, R., Rouse, S. L., Sansom, M. S., & Mulholland, A. J. (2014). A multiscale approach to modelling drug metabolism by membrane-bound cytochrome P450 enzymes. *PLoS Computational Biology, 10*(7), e1003714.

Lopata, A., Jambrina, P. G., Sharma, P. K., Brooks, B. R., Toth, J., Vertessy, B. G., & Rosta, E. (2015). Mutations decouple proton transfer from phosphate cleavage in the dutpase catalytic reaction. *ACS Catalysis, 5*(6), 3225-3237. doi:10.1021/cs502087f

MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., . . . Ha, S. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B, 102*(18), 3586-3616.

Maffeo, C., Yoo, J., & Aksimentiev, A. (2016). De novo reconstruction of DNA origami structures through atomistic molecular dynamics simulation. *Nucleic Acids Research, 44*(7), 3013-3019.

Maingi, V., Burns, J. R., Uusitalo, J. J., Howorka, S., Marrink, S. J., & Sansom, M. S. (2017). Stability and dynamics of membrane-spanning DNA nanopores. *Nature Communications, 8*, 14784.

Maingi, V., Lelimousin, M. l., Howorka, S., & Sansom, M. S. (2015). Gating-like motions and wall porosity in a DNA nanopore scaffold revealed by molecular simulations. *ACS nano, 9*(11), 11209-11217.

Malmstrom, R. D., Lee, C. T., Van Wart, A., & Amaro, R. E. (2014). On the application of molecular-dynamics based markov state models to functional proteins. *J. Chem. Theory Comput., 10*, 2648-2657.

Manara, R. M., Guy, A. T., Wallace, E. J., & Khalid, S. (2015). Free-Energy Calculations Reveal the Subtle Differences in the Interactions of DNA Bases with α-Hemolysin. *Journal of Chemical Theory and Computation, 11*(2), 810-816.

Mares-Guia, M., & Shaw, E. (1965). Studies on the active center of trypsin: The binding of amidines and guanidines as models of the substrate side chain. *J. Biol. Chem., 240*(4), 1579-1585.

Marinari, E., & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters, 19*(6), 451.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & De Vries, A. H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B, 111*(27), 7812-7824.

Marrink, S. J., & Tieleman, D. P. (2013). Perspective on the Martini model. *Chemical Society Reviews, 42*(16), 6801-6822.

Marx, D., & Hutter, J. (2009). Ab initio molecular dynamics. *Parallel computing, 309*(309), 327.

Marzinek, J. K., Holdbrook, D. A., Huber, R. G., Verma, C., & Bond, P. J. (2016). Pushing the envelope: dengue viral membrane coaxed into shape by molecular simulations. *Structure, 24*(8), 1410-1420.

Matek, C., Ouldridge, T. E., Doye, J. P., & Louis, A. A. (2015). Plectoneme tip bubbles: coupled denaturation and writhing in supercoiled DNA. *Scientific Reports, 5*.

Mathé, J., Aksimentiev, A., Nelson, D. R., Schulten, K., & Meller, A. (2005). Orientation discrimination of single-stranded DNA inside the α-hemolysin membrane channel. *Proceedings of the National Academy of Sciences of the United States of America, 102*(35), 12377-12382.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature, 267*(5612), 585-590.

Michel, J., & Essex, J. W. (2010). Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *Journal of Computer-Aided Molecular Design, 24*(8), 639-658. doi:10.1007/s10822-010-9363-3

Mongan, J., Case, D. A., & McCAMMON, J. A. (2004). Constant pH molecular dynamics in generalized Born implicit solvent. *Journal of Computational Chemistry, 25*(16), 2038-2048.

Morzan, U. N., Alonso de Armiño, D. J., Foglia, N. s. O., Ramírez, F., González Lebrero, M. C., Scherlis, D. n. A., & Estrin, D. A. (2018). Spectroscopy in Complex Environments from QM–MM Simulations. *Chemical Reviews, 118*(7), 4071-4113.

Nagy, G. N., Suardíaz, R., Lopata, A., Ozohanics, O., Vékey, K., Brooks, B. R., . . . Rosta, E. (2016). Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *Journal of the American Chemical Society, 138*(45), 15035-15045. doi:10.1021/jacs.6b09012

Nguyen, C. N., Young, T. K., & Gilson, M. K. (2012). Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit 7 uril. *Journal of Chemical Physics, 137*(4). doi:044101

10.1063/1.4733951

Noy, A., & Golestanian, R. (2012). Length scale dependence of DNA mechanical properties. *Physical Review Letters, 109*(22), 228101.

Oláh, J., Mulholland, A. J., & Harvey, J. N. (2011). Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. *Proceedings of the National Academy of Sciences, 108*(15), 6050-6055. doi:10.1073/pnas.1010194108

Olmos, J. L., Pandey, S., Martin-Garcia, J. M., Calvey, G., Katz, A., Knoska, J., . . . Oberthuer, D. (2018). Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biology, 16*(1), 59.

Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry, 25*(13), 1656-1676.

Ortiz-Suarez, M. L., Samsudin, F., Piggot, T. J., Bond, P. J., & Khalid, S. (2016). Full-length OmpA: structure, function, and membrane interactions predicted by molecular dynamics simulations. *Biophysical Journal, 111*(8), 1692-1702.

Pan, A. C., Weinreich, T. M., Piana, S., & Shaw, D. E. (2016). Demonstrating an order-of-magnitude sampling enhancement in molecular dynamics simulations of complex protein systems. *Journal of chemical theory and computation, 12*(3), 1360-1367. doi:10.1021/acs.jctc.5b00913

Pang, X., & Zhou, H. X. (2017). Rate constants and mechanisms of protein-ligand binding. *Annu. Rev. Biophys.. 46*, 105-130.

Pasi, M., Maddocks, J. H., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham III, T., . . . Laughton, C. (2014). µABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research, 42*(19), 12272-12283.

Patel, S., Mackerell, A. D., & Brooks, C. L. (2004). CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *Journal of Computational Chemistry, 25*(12), 1504-1514.

Peter, C., Oostenbrink, C., van Dorp, A., & van Gunsteren, W. F. (2004). Estimating entropies from molecular dynamics simulations. *Journal of Chemical Physics, 120*(6), 2652-2661. doi:Doi 10.1063/1.1636153

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*(16), 1781-1802.

Piggot, T. J., Holdbrook, D. A., & Khalid, S. (2011). Electroporation of the E. coli and S. aureus membranes: molecular dynamics simulations of complex bacterial membranes. *The Journal of Physical Chemistry B, 115*(45), 13381-13388.

Plattner, N., Doerr, S., De Fabritiis, G., & Noé, F. (2017). Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature Chemistry, 9*, 1005. doi:10.1038/nchem.2785

https://www.nature.com/articles/nchem.2785#supplementary-information

Plattner, N., & Noe, F. (2015). Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun., 6*. doi:10.1038/ncomms8653

Polozov, I. V., Bezrukov, L., Gawrisch, K., & Zimmerberg, J. (2008). Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nature Chemical Biology, 4*(4), 248-255.

Raich, L., Nin-Hill, A., Ardèvol, A., & Rovira, C. (2016). Chapter Seven - Enzymatic Cleavage of Glycosidic Bonds: Strategies on How to Set Up and Control a QM/MM Metadynamics Simulation. In G. A. Voth (Ed.), *Methods in Enzymology* (Vol. 577, pp. 159-183): Academic Press.

Ranaghan, K. E., & Mulholland, A. J. (2016). QM/MM Methods for Simulating Enzyme Reactions. *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*, 377.

Reddy, T., & Sansom, M. S. (2016a). Computational virology: From the inside out. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1858*(7), 1610-1618.

Reddy, T., & Sansom, M. S. (2016b). The role of the membrane in the structure and biophysical robustness of the dengue virion envelope. *Structure, 24*(3), 375-382.

Reddy, T., Shorthouse, D., Parton, D. L., Jefferys, E., Fowler, P. W., Chavent, M., . . . Sansom, M. S. (2015). Nothing to sneeze at: a dynamic and integrative computational model of an influenza A virion. *Structure, 23*(3), 584-597.

Riniker, S., Allison, J. R., & van Gunsteren, W. F. (2012). On developing coarse-grained models for biomolecular simulation: a review. *Physical Chemistry Chemical Physics, 14*(36), 12423-12430.

Ross, G. A., Bodnarchuk, M. S., & Essex, J. W. (2015). Water sites, networks, and free energies with grand canonical Monte Carlo. *Journal of the American Chemical Society, 137*(47), 14930-14943.

Ross, G. A., Bruce Macdonald, H. E., Cave-Ayland, C., Cabedo Martinez, A. I., & Essex, J. W. (2017). Replica-Exchange and Standard State Binding Free Energies with Grand Canonical Monte Carlo. *Journal of Chemical Theory and Computation, 13*(12), 6373-6381.

Rosta, E., Haranczyk, M., Chu, Z. T., & Warshel, A. (2008). Accelerating QM/MM free energy calculations: representing the surroundings by an updated mean charge distribution. *The journal of physical chemistry B, 112*(18), 5680-5692.

Rosta, E., Nowotny, M., Yang, W., & Hummer, G. (2011). Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc., 133*(23), 8934-8941.

Ryde, U. (2017). How Many Conformations Need To Be Sampled To Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *Journal of chemical theory and computation*. doi:10.1021/acs.jctc.7b00826

Saitta, A. M., & Saija, F. (2014). Miller experiments in atomistic computer simulations. *Proceedings of the National Academy of Sciences, 111*(38), 13768-13773. doi:10.1073/pnas.1402894111

Samsudin, F., Boags, A., Piggot, T. J., & Khalid, S. (2017). Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. *Biophysical Journal, 113*(7), 1496-1504.

Samsudin, F., Ortiz-Suarez, M. L., Piggot, T. J., Bond, P. J., & Khalid, S. (2016). OmpA: a flexible clamp for bacterial cell wall attachment. *Structure, 24*(12), 2227-2235.

Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J., & Wiley, D. C. (1989). Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry, 28*(21), 8388-8396.

Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., . . . Noé, F. (2015). PyEMMA 2: A software package for estimation, validation, and analysis of Markov Models.

. *J. Chem. Theory Comput., 11*, 5525-5542.

Schreck, J. S., Romano, F., Zimmer, M. H., Louis, A. A., & Doye, J. P. (2016). Characterizing DNA star-tile-based nanostructures using a coarse-grained model. *ACS Nano, 10*(4), 4236-4247.

Schröder, G. F., Levitt, M., & Brunger, A. T. (2010). Super-resolution biomolecular crystallography with low-resolution data. *Nature, 464*(7292), 1218.

Schuetz, D. A., de Witte, W. E. A., Wong, Y. C., Knasmueller, B., Richter, L., Kokh, D. B., . . . Ecker, G. F. (2017). Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discov. Today, 22*(6), 896-911. doi:https://doi.org/10.1016/j.drudis.2017.02.002

Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., . . . van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A, 103*(19), 3596-3607.

Senn, H. M., & Thiel, W. (2009). QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition, 48*(7), 1198-1229. doi:10.1002/anie.200802019

Sexton, T., & Cavalli, G. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell, 160*(6), 1049-1059. doi:10.1016/j.cell.2015.02.040

Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., & Shaw, D. E. (2011). How does a drug molecule find its target binding site? *J. Am. Chem. Soc., 133*(24), 9181-9183. doi:10.1021/ja202726y

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., . . . Ist, D. J. I. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Comm. ACM, 51*, 91-97.

Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., . . . Young, C. (2014). *Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer*. Paper presented at the Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, New Orleans, Louisana.

Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., . . . Shan, Y. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science, 330*(6002), 341-346.

Sherer, E. C., Harris, S. A., Soliva, R., Orozco, M., & Laughton, C. A. (1999). Molecular dynamics studies of DNA A-tract structure and flexibility. *Journal of the American Chemical Society, 121*(25), 5981-5991.

Sherwood, P. (2003). ChemShell for QM/MM Calculations. *Journal of Molecular Structure--Theochem, 623*, 1-28.

Shoura, M. J., Gabdank, I., Hansen, L., Merker, J., Gotlib, J., Levene, S. D., & Fire, A. Z. (2017). Intricate and Cell Type-Specific Populations of Endogenous Circular DNA (eccDNA) in Caenorhabditis elegans and Homo sapiens. *G3: Genes, Genomes, Genetics, 7*(10), 3295-3303.

Shukla, D., Hernández, C. X., Weber, J. K., & Pande, V. S. (2015). Markov state models provide insights into dynamic modulation of protein function. *Acc. Chem. Res., 48*(2), 414-422. doi:10.1021/ar5002999

Shukla, D., Meng, Y., Roux, B., & Pande, V. S. (2014). Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nature communications, 5*, 3397. doi:10.1038/ncomms4397

https://www.nature.com/articles/ncomms4397#supplementary-information

Silva, D.-A., Bowman, G. R., Sosa-Peinado, A., & Huang, X. (2011). A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput Biol, 7*(5), e1002054.

Sirur, A., De Sancho, D., & Best, R. B. (2016). Markov state models of protein misfolding. *J. Chem. Phys., 144*(7), 075101. doi:10.1063/1.4941579

Sodt, A. J., Mei, Y., König, G., Tao, P., Steele, R. P., Brooks, B. R., & Shao, Y. (2015). Multiple Environment Single System Quantum Mechanical/Molecular Mechanical (MESS-QM/MM) Calculations. 1. Estimation of Polarization Energies. *The Journal of Physical Chemistry A, 119*(9), 1511-1523. doi:10.1021/jp5072296

Steinmann, C., Olsson, M. A., & Ryde, U. (2018). Relative Ligand-Binding Free Energies Calculated from Multiple Short QM/MM MD Simulations. *Journal of Chemical Theory and Computation*.

Stelzl, L. S., Erlenbach, N., Heinz, M., Prisner, T. F., & Hummer, G. (2017). Resolving the Conformational Dynamics of DNA with Ångstrom Resolution by Pulsed Electron–Electron Double Resonance and Molecular Dynamics. *Journal of the American Chemical Society, 139*(34), 11674-11677.

Stelzl, L. S., & Hummer, G. (2017). Kinetics from replica exchange molecular dynamics simulations. *Journal of chemical theory and computation, 13*(8), 3927-3935. doi:10.1021/acs.jctc.7b00372

Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). SEMIANALYTICAL TREATMENT OF SOLVATION FOR MOLECULAR MECHANICS AND DYNAMICS. *Journal of the American Chemical Society, 112*(16), 6127-6129. doi:10.1021/ja00172a038

Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters, 314*(1), 141-151. doi:https://doi.org/10.1016/S0009-2614(99)01123-9

Świderek, K., Tuñón, I., Moliner, V., & Bertran, J. (2015). Computational strategies for the design of new enzymatic functions. *Archives of Biochemistry and Biophysics, 582*(Supplement C), 68-79. doi:https://doi.org/10.1016/j.abb.2015.03.013

Tenboer, J., Basu, S., Zatsepin, N., Pande, K., Milathianaki, D., Frank, M., . . . Koglin, J. E. (2014). Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science, 346*(6214), 1242-1246.

Todorov, I. T., Smith, W., Trachenko, K., & Dove, M. T. (2006). DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *Journal of Materials Chemistry, 16*(20), 1911-1918.

Torrie, G. M., & Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics, 23*(2), 187-199. doi:https://doi.org/10.1016/0021-9991(77)90121-8

Tuckerman, M., Laasonen, K., Sprik, M., & Parrinello, M. (1995). Ab initio molecular dynamics simulation of the solvation and transport of H3O+ and OH-ions in water. *The Journal of Physical Chemistry, 99*(16), 5749-5752.

Turupcu, A., & Oostenbrink, C. (2017). Modeling of Oligosaccharides within Glycoproteins from Free-Energy Landscapes. *Journal of Chemical Information and Modeling, 57*(9), 2222-2236. doi:10.1021/acs.jcim.7b00351

Tyka, M. D., Clarke, A. R., & Sessions, R. B. (2006). An efficient, path-independent method for free-energy calculations. *Journal of Physical Chemistry B, 110*(34), 17212-17220. doi:10.1021/jp060734j

Uddin, N., Choi, T. H., & Choi, C. H. (2013). Direct Absolute p K a Predictions and Proton Transfer Mechanisms of Small Molecules in Aqueous Solution by QM/MM-MD. *The Journal of Physical Chemistry B, 117*(20), 6269-6275.

van der Kamp, M. W., Chaudret, R., & Mulholland, A. J. (2013). QM/MM modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis. *Febs Journal, 280*(13), 3120-3131.

Van der Kamp, M. W., & Mulholland, A. J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry, 52*(16), 2708-2728. doi:Doi 10.1021/Bi400215w

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry, 26*(16), 1701-1718.

Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A., Morokuma, K., & Frisch, M. J. (2006). Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *Journal of chemical theory and computation, 2*(3), 815-826. doi:10.1021/ct050289g

Vukovic, S., Brennan, P. E., & Huggins, D. J. (2016). Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites. *Journal of Physics: Condensed Matter, 28*(34), 344007.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry, 25*(9), 1157-1174.

Wang, L., Wu, Y. J., Deng, Y. Q., Kim, B., Pierce, L., Krilov, G., . . . Abel, R. (2015). 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. *Journal of the American Chemical Society, 137*(7), 2695-2703. doi:10.1021/ja512751q

Wang, W., Donini, O., Reyes, C. M., & Kollman, P. A. (2001). Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annual Review of Biophysics and Biomolecular Structure, 30*(1), 211-243.

Wang, Y., Papaleo, E., & Lindorff-Larsen, K. (2016). Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *eLife, 5*, e17505. doi:10.7554/eLife.17505

Warshel, A., & Levitt, M. (1976). Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *Journal of molecular biology, 103*(2), 227-249.

Wasilewski, S., Calder, L. J., Grant, T., & Rosenthal, P. B. (2012). Distribution of surface glycoproteins on influenza A virus determined by electron cryotomography. *Vaccine, 30*(51), 7368-7373.

Wilson, E. B. (1941). Some mathematical methods for the study of molecular vibrations. *Journal of Chemical Physics, 9*(1), 76-84. doi:10.1063/1.1750829

Woods, C., Mey, A., Calabro, G., & Michel, J. (2016). Sire molecular simulations framework.

Wu, E. L., Engstrom, O., Jo, S., Stuhlsatz, D., Wildmalm, G., & Im, W. (2013). Molecular Dynamics Simulations of E. coli Lipopolysaccharide Bilayers. *Biophysical Journal, 104*(2), 586a.

Zhou, G., Pantelopulos, G. A., Mukherjee, S., & Voelz, V. A. (2017). Bridging microscopic and macroscopic mechanisms of p53-MDM2 binding with kinetic network models. *Biophys. J., 113*(4), 785-793. doi:https://doi.org/10.1016/j.bpj.2017.07.009

Zinovjev, K., & Tuñón, I. (2017). Adaptive Finite Temperature String Method in Collective Variables. *The Journal of Physical Chemistry A, 121*(51), 9764-9772.

Zwanzig, R. W. (1954). HIGH-TEMPERATURE EQUATION OF STATE BY A PERTURBATION METHOD .1. NONPOLAR GASES. *Journal of Chemical Physics, 22*(8), 1420-1426.