

Computational Theory

Protein Structure and Function

The relationship between protein structure and function is central for understanding how the molecular world works. However, determining both can be very difficult and requires a combination of many different methods, as most methods are focused on determining either one or the other. X-ray crystallography, EM, and to some extent NMR spectroscopy are typical methods used to determine the structure of proteins [1–3]. Function is often coupled to the dynamical properties of the protein, which can be examined using e.g. Förster resonance energy transfer (FRET), variants of NMR spectroscopy, and infrared spectroscopy [4]. Each method has a timescale of which it can provide information ranging from vibrations of individual atoms on a sub-femtosecond timescale, to protein translation by the ribosome on a timescale of seconds [5]. Most of the available techniques provide ensemble properties, and do not report on the dynamics of single proteins.

MD simulation is a computational technique able to provide very high resolution in time and space which is unmatched by any experimental technique. It basically provides a three-dimensional movie of how the atoms in a system move relative to one another. MD is an important complementary technique to experiments, as it can help explain the mechanisms of molecular processes. In principle, simulation of any molecular system imaginable is possible, even if it would not be possible to set it up in practice.

The aim of MD simulations is to determine the motions of atoms in a system. The time-dependent Schrödinger equation describes the behavior of molecules; however, solving this is practically infeasible for large molecular assemblies. The motions of atoms and molecules heavier than the proton are essentially classical. Therefore, classical mechanics is used in MD to propagate the system based on a model, termed a FF, describing the energetics of the system [6].

Force Fields

The relationship between atomic coordinates and energy is defined in a FF. A FF is composed of a set of functions describing the energy of the interatomic interactions, e.g. bond stretching or angle bending, as well as an associated set of parameters determining the strength of these interactions. The parameters can be derived in various ways, e.g. by fitting to experimental data or quantum mechanical (QM) calculations. FFs rely on transferability, which means that parameters such as bond angles and force constants are assumed to be similar for the same chemical groups in different molecules. This allows fitting parameters to one set of molecules, and applying the parameters to other molecules of similar chemical structure. Transferability also leads to the concept of atom types. An atom type holds information about the element, hybridization, and neighboring atoms of a particle, and is used to define which parameters are used for a specific atom.

Most FFs use the same basic functional to describe the energy, which has not changed considerably for the commonly used FFs since they were first released [7]. The FF energy (E_{FF}) is described as a sum of bonded and non-bonded terms. The bonded contributions are from the stretch (or bond) energy (E_{str}), the bend (or angle) energy (E_{bend}), and the torsional energy (E_{tors}). The non-bonded contributions come from the van der Waals (E_{vdW}) and electrostatic energies (E_{elec}), and are usually included for atoms separated by more than two or three bonds. The vdW energy is usually described by the Lennard-Jones (LJ) potential, and the electrostatic energy is described by the Coulomb potential. The stretch and bend energies are both described by a harmonic function, whereas the torsional energy is described by a sum of cosine functions to account for periodicity.

$$\begin{aligned}
 E_{str}(R) &= k_{str}(R - R_0)^2 & E_{bend}(\theta) &= k_{bend}(\theta - \theta_0)^2 \\
 E_{tors}(\omega) &= \sum_{n=1} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \\
 E_{vdW}(R) &= \varepsilon_{LJ} \left[\left(\frac{R_{LJ}^0}{R} \right)^{12} - 2 \left(\frac{R_{LJ}^0}{R} \right)^6 \right] & E_{elec}(R) &= \frac{q^A q^B}{\varepsilon R}
 \end{aligned}$$

More terms can be added to the FF for special cases, such as cross terms to account for the coupling between different types of interactions.

The most commonly used FFs were originally fitted to QM calculations as well as experimental data from X-ray structures [7]. Recent advances in both experimental and computational ability have made it both feasible and necessary to reparameterize the FFs to include dynamical experimental data, e.g. from NMR experiments, in the fitting procedure [7]. Comparative studies have shown that the older FFs tend to overestimate the stability of α -helical structures [8–10]. The most recent FFs such as AMBER99SB*–ILDN [11], CHARMM22* [11], and CHARMM36 [12] have refitted torsional parameters providing a better balance between helix, sheet, and coil. A recent comparative study of multiple protein FFs

(not including CHARMM36) showed that the latest FFs are able to reproduce several structural and dynamical properties of small proteins; however, the temperature dependence of the FFs is still lacking, and the folding mechanisms varies from FF to FF [10]. It is unclear whether FF improvements are possible using the current functional, or if it will be necessary to include more sophisticated effects, such as polarizability, to improve the accuracy of the FFs [7].

Molecular Dynamics Simulation

In conventional MD simulations, each atom in the system is described by one particle. The energy of the system is described by the FF. It is possible to determine the force acting on each atom by differentiating the energy with respect to the atomic positions, and then using Newton's second law of motion to obtain the acceleration on the atoms. This allows propagation of the system in time using classical mechanics. Numerical integration is used to determine the trajectory of the system, requiring that the time-step of each integration step is very small. The fastest motions in a system are the bond vibrations, which have frequencies on the order of 10^{14} s^{-1} . It then follows that the time-step needs to be on the order of 1 fs to avoid atoms moving too close, which would produce very large forces that would eventually make the system explode and the simulation crash. This also means that one billion steps are needed to simulate a single μs .

The starting position of all the atoms in the system must be defined before the MD simulation can be initiated. For a protein, the coordinates are usually obtained from an X-ray crystal or NMR structure available from the RCSB Protein Data Bank (PDB) [13]. Starting velocities are usually assigned to each atom from a Maxwell-Boltzmann distribution at the simulation temperature (T).

Performing MD simulations on a system of finite size, e.g. in a cubic box, introduces boundaries to the system. The behavior of molecules close to the boundary is different from the bulk behavior. Periodic boundary conditions can help minimize boundary effects, and can be visualized as replicating the system an infinite number of times in all directions. Imposing the minimum image convention (i.e. no cutoff can be longer than half the length of the shortest box vector) ensures that a particle only interacts with the closest image of another particle, and never with itself. Periodic boundary conditions are also helpful in calculating the electrostatic interactions, as is described below.

MD simulations do not scale well computationally, due to the intrinsic sequential nature of the simulations, and the fact that all atoms in a simulation interact with one another through the non-bonded interactions. In principle, the computational time scales with N^2 , with N being the number of particles in the system; however, algorithmic advances in performing Fourier transformations have reduced the scaling with system size to $N \ln(N)$, as the long-range electrostatic interactions can be efficiently evaluated using the particle mesh Ewald method (PME) (see below) [6, 14].

The most computationally demanding part of an MD simulation is evaluation of the forces, and particularly the non-bonded forces. To minimize the number of evaluations, the vdW interactions can be truncated at a certain cutoff. This is a reasonable approximation since the vdW energy depends on R^{-6} , and therefore rapidly becomes negligible as the distance between atoms is increased. Truncation of the LJ potential can introduce discontinuities in the energy; therefore, a switching function can be introduced, which reduces the energy smoothly to zero between the switching distance and the cutoff distance. In itself, truncation does not save much time, since the distance between all pairs of atoms still has to be calculated. However, the introduction of a pair-list, which contains the pairs of atoms within a pair-list radius, can reduce the computation of all interatomic distances to every 10–20 steps. During each step, the distances and non-bonded forces are only calculated between the pairs of atoms included in the pair-list, which are within the cutoff distance.

The electrostatic potential depends on the distance as R^{-1} , which means that the distance where the energy becomes close to zero is large compared to the typical system size. Therefore, rather than using a cutoff, other means of reducing the computational time for evaluating the electrostatic forces have been developed, one of which is the use of Ewald sums [15]. Ewald summation divides the electrostatic contribution into two parts, the “near”-field and the “far”-field contributions [6]. The “near”-field contribution is the sum of the atomic partial charges and a screening charge distribution of Gaussian potentials. This screening potential exactly counters the atomic partial charges and is centered on the atomic positions. The sum converges quickly because of the screening field, and can therefore be truncated. The “far”-field contribution derives from a compensating Gaussian charge distribution which is exactly opposite the screening charge distribution. The compensating charge distribution converges rapidly in reciprocal space, and can be evaluated efficiently with Fourier transformation methods. The electrostatic potential at a point is the sum of the partial charges, the screening charge distribution, and the compensating charge distribution. The success of this method depends on efficient ways of evaluating the Fourier transformations; PME [14] uses one such method called Fast Fourier Transformation, which evaluates the transformation in a discrete number of grid points imposed on the system [6].

MD intrinsically samples the microcanonical (NVE) ensemble. Several methods have been developed to maintain the T and/or pressure (P) (near) constant, allowing sampling of other ensembles of interest. The isothermal-isobaric (NPT) ensemble is the most relevant for biomolecular simulations, since this is the condition that most experiments are performed under. Several algorithms exist for sampling constant T, and include scaling of the velocities, including coupling to an external heat bath (Berendsen thermostat); random reassignment of velocities from a Maxwell-Boltzmann distribution (Anderson thermostat); addition of frictional and random forces (Langevin dynamics); and extended system methods (Nosé-Hoover thermostat). It should be noted that scaling of the velocities does not strictly sample the

canonical (NVT) ensemble, as the fluctuations of T are not correct. The same algorithms can be used for constant P simulations by changing the positions of particles rather than the velocities.

Recent Advances in MD

Since all atoms in the system are interacting, the calculation cannot be accelerated by distributing the system to an infinite number of processors; at some point the communication between the different compute nodes will be slower than the actual calculation [7]. However, recent advances in hardware, other than an increase in the number of CPUs, have significantly accelerated the computation of MD trajectories. Graphics processing units (GPUs) originally developed to accelerate rendering of graphics for the gaming industry, have proven particularly useful in MD simulations [16]. The most popular MD engines, NAMD, GROMACS, and AMBER, have already implemented GPUs to calculate the time-consuming non-bonded forces, providing considerable speed-up of the calculations [17–20]. However, this approach is only viable up to a limited system size, as the communication between the GPUs is even worse than between the traditional processors. A more drastic approach has been taken at D.E. Shaw Research, where a special purpose computer chip named Anton has been developed specifically for performing MD simulations [21]. This has allowed simulations to extend into the ms timescale [22].

The current MD simulations being performed have come a long way from the first MD simulation of a folded protein, bovine pancreatic trypsin inhibitor, in 1977 by McCammon et al., which had a simulation length of 8.8 ps [23]. However, the amount of data being produced poses a great challenge in terms of handling, storage, and analysis. Currently, the largest entry in the PDB is an assembly of the HIV-1 capsid with more than 2 million atoms. It was modeled using all-atom MD Flexible Fitting to cryo-EM data [24]. The fully solvated system consisted of 64 million atoms, and was simulated for 100 ns, which is the largest all-atom MD simulation performed to date [25]. Development of new software was needed to handle both the analysis and the visualization of the large scale system, which illustrates that it is not just the desire to go bigger, but actual biological problems that drive the development of new methodologies [26, 27].

Enhanced Sampling Methods

Not everyone has access to a brute force computer such as Anton. It is therefore necessary to devise new methods to simulate rare events within the timescale that most scientists can obtain. Enhanced sampling methods have been developed to encourage the system to explore regions of phase space which are otherwise difficult to investigate. Some methods require knowledge of both the starting and

end-point of the process of interest. For example targeted MD explores the conformational change of a protein going from one state to another by applying a force to minimize the root mean square deviation (RMSD) of the initial structure compared to the final structure [28]. Such methods are not helpful when the end-point structure is not known, and may also move the system along a path, which is not a physically relevant transition path. Steered MD is an enhanced sampling method that does not require knowledge of the end-point structure [29]. A force is applied to one or more atoms while keeping other atoms fixed. This allows e.g. for simulation of the unbinding of ligands or the unfolding of proteins. While the final protein structure is not known before the simulation, a preconceived idea about the direction of change is imposed on the system by the user. This might impose a force on the system making it explore high energy regions of phase space along the path, which are not physically relevant. This is partly avoided in metadynamics, where a history-dependent bias potential is added to the system to discourage it from revisiting already explored areas of phase space [30]. When a potential energy well is filled by the history-dependent potential, the system is free to choose any path to move to another low-energy region of the phase space, which means that the system is not pushed or pulled towards high energy regions. Collective variables (CVs) are used to guide the system, and it is therefore important to choose CVs that cover the important regions of phase space. The advantage of metadynamics is that the phase space is efficiently explored simultaneously with the evaluation of the free energy surface. However, it is also more complicated to execute as the choice of CVs is essential to the success of the simulations, and again, prior knowledge about the system is often necessary. Accelerated MD (aMD) does not require prior knowledge about the system to enhance the sampling of the system. In aMD the potential of the low-energy regions of phase space is raised, which effectively lowers the barriers of transition between minima on the potential energy surface [31]. The amount that the potential energy surface should be raised needs to be optimized. A too aggressive approach leaves the risk of obtaining a flat potential energy surface, which would cause the system to spend a lot of time in regions of phase space which are not of interest.

CG MD simulation is a separate category of enhanced sampling methods in which the number of particles in the system is decreased by combining multiple atoms into a single particle. The main speed-up of these methods does not actually arise from the reduced number of particles, but from the possibility of longer time-steps. The MARTINI CG FF is a popular method for simulating biomolecular systems [32]. It was originally developed for simulating lipid systems, but has been further developed to include models for proteins and carbohydrates [33, 34]. The MARTINI FF combines approximately four heavy atoms into one CG bead, which allows the use of a time-step between 20 and 40 ns [35]. Conformational changes of proteins have proven to be challenging to simulate with MARTINI since hydrogen bonds are not included explicitly in the model, and the secondary structure has to be defined in the bead-type [35]. Gō models use the native interactions of the protein combined with a single particle for each residue to investigate the folding mechanisms of proteins [36]. However, this can thus only be used to

investigate folding into a conformation which is already known. The kinetics of amyloid aggregation has been studied using a phenomenological CG model for which the aggregation propensity can be tuned [37]. A short review and discussion regarding the use of CG models to investigate amyloid aggregation is given in the chapter “Coarse Grained Study of Amyloid Protofibril Aggregation”.

Protein-Ligand Binding

A large part of the study of proteins deals with the binding of ligands, e.g. substrates, signaling molecules, or inhibitors, to the protein. Therefore, a large part of the biomolecular modeling research is also devoted to the study of protein-ligand complexes. Software devoted to determining the structure of protein-ligand complexes is termed docking software, and usually contains a scoring function to estimate the strength of the complex [38].

Docking can be used for a variety of applications, from screening of a virtual library containing many thousands or millions of compounds to determining the structure and energy of binding of a single molecule. Most docking software includes flexibility of the ligand, either by generating multiple conformations of the ligand and then docking each as a rigid molecule, or by introducing flexibility into the ligand during the docking procedure. However, flexibility of the protein requires much more computational power, and is not a routine feature. It can be introduced e.g. by using several protein structures obtained from X-ray, NMR, or MD, or it can be a feature of the docking software. Another type of protein flexibility can be introduced using a rotamer library for the side chains in the binding pocket. It is especially important to incorporate protein flexibility when the target ligand does not resemble the ligand present in the binding pocket of the protein structure, or if the protein structure is a model generated based on a structure of a homologous protein [38].

Sometimes the binding site of a protein target is not known, or the goal is to develop a novel drug, which does not have to bind in the substrate binding pocket. Then it becomes necessary to use binding site prediction software, such as Sitemap [39]. This can be used to determine the location of potential binding pockets, and score them based on the likelihood of finding a good drug that will bind in the pocket. Furthermore, Sitemap can also give an idea of which functional groups should be placed in specific areas of the binding pocket to create a high-affinity drug.

However, binding site prediction and docking software fails when the binding site on the protein is not well-defined, such as in amyloid fibrils. First of all, the structural models of amyloid fibrils are usually based on a combination of NMR constraints and other biophysical methods, which do not necessarily give much information on e.g. the side chains of the protein [40]. Secondly, there are no well-defined binding pockets on the surface of the fibril, only shallow crevices which are highly exposed to the solvent. It can, therefore, be necessary to use MD

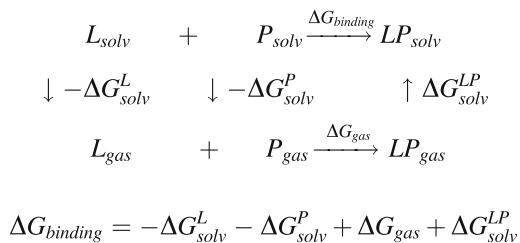
simulations to investigate possible binding sites of the ligand. A combination of multiple trajectories and free energy calculations of the resulting protein-ligand complexes can be highly effective in these cases [41–44].

Free Energy Calculations

Free energy calculations are of great importance in biomolecular modeling as it allows direct comparison with experiments. The Gibbs free energy is the relevant quantity when comparing simulations to experiments performed with a constant number of particles (N), pressure (P), and temperature (T), i.e. in the NPT ensemble. Unfortunately, the Gibbs free energy is difficult to calculate from MD simulations, as the higher energy regions of phase-space, which make important contributions to the free energy, are not sampled very much [45]. Often, the free energy difference of a reaction, such as the binding energy of a ligand to a protein or the relative energy of two conformational states, is the important property. The energy difference is a thermodynamic property, and therefore only depends on the difference between the two states [45]. This can be computed using rigorous and accurate methods such as free energy perturbation and thermodynamic integration. These methods sample a path between the two states, which make them computationally demanding, and consequently, they can only be performed for small perturbations [45, 46]. End-point models are alternatives to the more demanding methods, which only sample the two states of the reaction, and therefore can be used for a more diverse set of reactions [46].

MM-PBSA

Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) [47–49] is an approximate method to determine free energies. Binding free energies ($\Delta G_{\text{binding}}$) for the ligand-protein association ($L + P \rightarrow LP$) can be evaluated using a thermodynamic cycle [49].



ΔG_{gas} is the gas-phase interaction energy and ΔG_{solv} is the free energy of moving a molecule or complex from the gas-phase to the solvated state. The approximation in MM-PBSA lies in determining the energetic components. The gas-phase interaction energy is divided into enthalpic and entropic contributions.

$$\Delta G_{\text{gas}} = \Delta H_{\text{gas}} - T\Delta S \approx \Delta E_{\text{MM}} - T\Delta S_{\text{MM}}$$

The enthalpy can be approximated as the Molecular Mechanics (MM) energy (E_{MM}), and the entropy can be approximated using quasi harmonic analysis or normal-mode analysis [47, 49]. However, methods for calculating the entropy are computationally expensive and require extensive sampling to reach convergence. E_{MM} is extracted directly from the FF, and is therefore a sum of the bonded (stretch, bend, and torsion) and the non-bonded (vdW and electrostatic) contributions. The solvation energies are approximated using the PBSA method, which divides the solvation energy into an electrostatic (E_{PB}) and a non-polar contribution (E_{cav}) [47]. The electrostatic interaction between the solute and solvent is calculated using a numerical solution to the Poisson-Boltzmann (PB) equation [47], which is a continuum solvation approach using a combination of the Poisson equation and a Boltzmann distribution of ions [45]. The Poisson equation is a second-order differential equation that relates the change in electrostatic potential to the dielectric constant and the charge density. The Boltzmann equation accounts for the distribution of mobile ionic charges in the solvent [45]. The non-polar contribution is approximately proportional to the surface area (SA) of the solute and can be viewed as arising from two terms: The vdW interactions between the solute and solvent, and the cost of forming the solute cavity in the solvent [45].

The MM-PBSA method requires the determination of the MM and solvation energies of all species of the reaction (L, P, and LP), and the resulting binding free energy is the difference between the energies of the reactants and products.

$$\Delta G_{\text{binding}} = G_{\text{LP}} - G_{\text{P}} - G_{\text{L}}$$

Ensembles of structures for calculating the energy can be generated by explicit-solvent MD simulations, with subsequent removal of the solvent for the evaluation of the MM-PBSA energy [47]. The ideal situation uses different trajectories to calculate ensemble averages for the complex, protein, and ligand, respectively. However, using a single trajectory to evaluate all contributions can minimize the noise from sampling inconsistencies and the error in the FF and implicit solvation energies [46]. This procedure assumes that the change in structure and conformational freedom upon association of the complex is negligible [46]. It should be noted that using the single-trajectory approach eliminates the bonded contributions to the MM energy. If only relative binding energies of similar ligands are desired, it is common to neglect the entropic contribution, as it is assumed that the change in entropy is comparable for all ligands [49, 50].

References

1. Branden C, Tooze J (1998) In: Introduction to Protein Structure, 2nd edn. Garland
2. Rule GS, Hitchens TK (2006) In: Fundamentals of Protein NMR Spectroscopy: Chapter 17: Protein Structure Determination, Springer
3. Zhou ZH (2008) Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr Opin Struct Biol* 18:218–228
4. Cooper A (2004) In: Biophysical Chemistry, Royal Society of Chemistry
5. Dror RO, Jensen MØ, Borhani DW, Shaw DE (2010) Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J Gen Physiol* 135:555–562
6. Jensen F (2007) In: Introduction to Computational Chemistry, 2nd edn, John Wiley & Sons Ltd
7. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* 41:429–452
8. Best RB, Buchete N, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95:L07–L09
9. Beauchamp KA, Lin Y, Das R, Pande VS (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J Chem Theory Comput* 8: 1409–1414
10. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE (2012) Systematic Validation of Protein Force Fields Against Experimental Data. *PLoS One* 7:e32131
11. Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 100:L47–L49
12. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles, *J Chem Theor Comput* 8:3257–3273
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank, *Nucl Acids Res* 28:235–242
14. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
15. Ewald PP (1921) Die Berechnung Optischer Und Elektrostatisher Gitterpotentiale. *Ann Phys* 64:253–287
16. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K (2010) GPU-accelerated molecular modeling coming of age. *J Mol Graph Model* 29:116–125
17. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K (2007) Accelerating molecular modeling applications with graphics processors. *J Comput Chem* 28:2618–2640
18. Goetz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J Chem Theory Comput* 8:1542–1555
19. Salomon-Ferrer R, Goetz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J Chem Theory Comput* 9:3878–3888
20. Abraham M, Apol E, Apostolov R, Berendsen HJC, Buuren Av, Bjelkmar P, Drunen Rv, Feenstra A, Fritsch S, Groenhof G, Junghans C, Hub J, Kasson P, Kutzner C, Lambeth B, Larsson P, Lemkul JA, Marklund E, Meulenhoff P, Murtola T, Páll S, Pronk S, Schulz R, Shirts M, Sijbers A, Tieleman P, Wolf M, Hess B, van der Spoel D, Lindahl E (2010) Gromacs User Manual Version 4.6.5
21. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC (2008) Anton, a special-purpose machine for molecular dynamics simulation. *ACM* 51:91–97

22. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, Young C, Deneroff MM, Batson B, Bowers KJ, Chow E, Eastwood MP, Ierardi DJ, Klepeis JL, Kuskin JS, Larson RH, Lindorff-Larsen K, Maragakis P, Moraes MA, Piana S, Shan Y, Towles B (2009) Millisecond-scale molecular dynamics simulations on anton, proceedings of the conference on high performance computing networking, Storage and Analysis, 1–11
23. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267: 585–590
24. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683
25. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 497:643–646
26. Stone JE, Vandivort KL, Schulten K (2013) GPU-accelerated molecular visualization on petascale supercomputing platforms, UltraVis'13 Proceedings of the 8th international workshop on ultrascala visualization, p. 6
27. Stone JE, McGreevy R, Isralewitz B, Schulten K (2014) GPU-Accelerated Analysis and Visualization of Large Structures Solved by Molecular Dynamics Flexible Fitting, *Faraday Discuss* 169(12)
28. Schlitter J, Engels M, Krüger P (1994) Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* 12:84–89
29. Park S, Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys* 120:5946–5961
30. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99:12562–12566
31. Markwick PRL, McCammon JA (2011) Studying functional dynamics in bio-molecules using accelerated molecular dynamics, *Phys Chem Chem Phys* 13:20053–20065
32. Marrink S-J, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 111: 7812–7824
33. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J (2008) The MARTINI coarse-grained force field: extension to proteins. *J Chem Theor Comput* 4:819–834
34. López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink S-J (2009) Martini coarse-grained force field: extension to carbohydrates. *J Chem Theor Comput* 5:3195–3210
35. Periole X, Marrink S-J (2013) The Martini coarse-grained force field. *Methods Mol Biol* 924:533–565
36. Hills RD Jr, Brooks CL III (2009) Insights from coarse-grained gō models for protein folding and dynamics. *Int J Mol Sci* 10:889–905
37. Pellarin R, Caffisch A (2006) Interpreting the aggregation kinetics of amyloid peptides. *J Mol Biol* 360:882–892
38. Höltje HD, Sippl W, Rognan D (2008) In: *Molecular Modeling: Basic Principles and Applications: Virtual Screening and Docking*, 3rd edn, Wiley-VCH
39. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 49:377–389
40. Tycko R Solid (2011) Solid state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* 62:279–299
41. Wu C, Wang Z, Lei H, Zhang W, Duan Y (2007) Dual binding modes of congo red to amyloid protofibril surface observed in molecular dynamics simulations. *J Am Chem Soc* 129: 1225–1232
42. Wu C, Wang Z, Lei H, Duan Y, Bowers MT, Shea J (2008) The binding of thioflavin T and its neutral analog BTA-1 to protofibrils of the Alzheimer's disease A β_{16-22} peptide probed by molecular dynamics simulations, *J Mol Biol* 384:718–729

43. Wu C, Bowers MT, Shea J (2011) On the origin of the stronger binding of PIB over thioflavin T to protofibrils of the Alzheimer amyloid- β peptide: a molecular dynamics study, *Biophys J* 100:1316–1324
44. Skeby KK, Sorensen J, Schiott B (2013) Identification of a common binding mode for imaging agents to amyloid fibrils from molecular dynamics simulations. *J Am Chem Soc* 135: 15114–15128
45. Leach AR (2001) In: *Molecular modelling principles and applications*, 2nd edn, Prentice Hall
46. Swanson JMJ, Henchman RH, McCammon JA (2004) Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy, *Biophys J* 86:67–74
47. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TEI (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
48. Srinivasan J, Cheatham TEI, Cieplak P, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices, *J Am Chem Soc* 120:9401–9409
49. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of Efavirenz by docking and MM-PBSA. *J Am Chem Soc* 123:5221–5230
50. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations, *J Chem Inf Model* 51:69–82

Computational Modelling of the Human Islet Amyloid
Polypeptide

Skeby, K.K.

2016, XVI, 118 p. 58 illus., 47 illus. in color., Hardcover

ISBN: 978-3-319-20039-2