Molecular Dynamics Simulations

Amedeo Caflisch
Department of Biochemistry
University of Zurich

www.biochem-caflisch.uzh.ch
Structure-based drug design identifies polythiophenes as antiprion compounds

with A. Aguzzi

Peptide Binding to a PDZ Domain by Electrostatic Steering via Nonnative Salt Bridges

Nicolas Blöchliger, Min Xu, and Amedeo Caflisch
1Department of Biochemistry, University of Zurich, Zurich, Switzerland

ABSTRACT

We have captured the binding of a peptide to a PDZ domain by unbiased molecular dynamics simulations. Analysis of the trajectories reveals on-pathway encounter complex formation, which is driven by electrostatic interactions between negatively charged carboxylate groups in the peptide and positively charged side chains surrounding the binding site. In contrast, the final stereospecific complex, which matches the crystal structure, features completely different interactions, namely the burial of the hydrophobic side chain of the peptide C-terminal residue and backbone hydrogen bonds. The simulations show that nonnative salt bridges stabilize kinetically the encounter complex during binding. Unbinding follows the inverse sequence of events with the same nonnative salt bridges in the encounter complex. Thus, in contrast to protein folding, which is driven by native interactions, the binding of charged peptides can be steered by nonnative interactions, which might be a general mechanism, e.g., in the recognition of histone tails by bromodomains.

\[
\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) = m_i \frac{d^2 x_i(t)}{dt^2}
\]
2013 Nobel Prize in Chemistry

Taking the Experiment to Cyberspace

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel “for the development of multiscale models for complex chemical systems”.

► Press release

Martin Karplus


► Have a look at Michael Levitt's

Michael Levitt


► Interviews with Chemistry

Arieh Warshel

Arieh Warshel, U.S. and Israeli citizen. Born 1940 in Kibbutz Sde-Nahum, Israel. Ph.D. 1969 from Weizmann Institute of Science, Rehovot, Israel. Distinguished Professor, University of Southern California, Los Angeles, CA, USA.
Introduction

To answer these questions (and many others) scientists use appropriate experimental techniques, e.g., single-molecule spectroscopy, NMR spectroscopy, site-directed mutagenesis etc.

In this lecture we will see how we can use a computer to analyze at atomic level of detail the structure, flexibility and function of biological (macro)molecules.
Introduction

• In Molecular Dynamics (MD) simulations the laws of classical physics are used to approximate the interactions and motion of (macro)molecules.

• MD simulations are a multidisciplinary field. The underlying laws and theories stem from mathematics, physics and chemistry; algorithms from various areas of applied mathematics and computer science are implemented in a wide variety of MD software packages.
Understanding biomolecular structure, dynamics and function

Development of (mathematical) models

Development of methods to explore models

Exploration of model phenomenology and properties

Understanding biomolecular structure, dynamics and function

Experimental information from:

- X-ray crystallography
- NMR spectroscopy
- Electron microscopy
- Atomic force microscopy
- X-ray, light, and neutron scattering
- Fluorescence techniques (FRET, FCS)
- Other spectroscopic techniques (CD, IR, ESR, UV)
- Calorimetry (ITC, DSC)
- Measurements of pH (titration), sedimentation (centrifugation), elution time (chromatography), ...

Development of new theories and models to rationalize and predict experimental observations

December 14, 2015
• Molecule(s) of interest (e.g. protein, DNA, protein/ligand complexes, protein/membrane).
• Solvent (water, ions, alcohol)
• Cuvette
Instead of expressing (e.g. in *E. coli*) the molecule of interest, we have to supply the computer with the 3D coordinates of the desired molecule(s).

For proteins and small organic compounds the atomic coordinates can be downloaded from the PROTEIN DATA BANK (www.rscb.org).

In the following slides the molecule of interest will always be protein, keep in mind though that for other types of (macro)molecules everything works analogously.
An example of a set of atomic coordinates which can be used to run molecular dynamics simulations. It is the PDZ3 domain of PSD-95/SAP90, PDB entry 1BFE.
Furthermore, we need a model to describe the geometry and the physical interactions of the atoms of our protein.

There are two main approaches:

- Molecular mechanics, where the atoms are rigid spheres connected by rigid and unbreakable bonds.
- Quantum mechanics, where also the dynamical evolution of the electrons is taken into account.

In this course we will discuss the model based on classical mechanics.
Structural Resolution in the Representation of Biomolecules

Ethene in electronic structure level representation (orbitals)

Butane in all-atom molecular mechanics representation (spheres)

Short linear carbohydrate in coarse-grained representation (single ellipsoidal bead)
Molecular Mechanics Representation

- Each atom is represented as a sphere.
- The biomolecular force field is what shapes the collection of spheres into things that look like molecules through so-called **bonded** interactions.
- The force field also describes the atom-atom interactions of distal parts within a flexible molecule as well as the interactions between molecules through **nonbonded** interactions.
- Sometimes all-atom representation is simplified to so-called united-atoms (e.g.: a methyl group is a single sphere).
- Electronic structure is in general coarse-grained out and expected to be captured by the force field.
An experiment on a computer

• The force field is an analytical function of the spatial coordinates of the atomic nuclei.

• Several different force fields are commonly available (CHARMM, AMBER, OPLS). Here we will focus on the CHARMM force field (Brooks et al., J. Comput. Chem. 30, 1545, 2009)
• In CHARMM the force field is given by:

\[ E = E_{\text{Bonds}} + E_{\text{Angles}} + E_{\text{Dihedrals}} + E_{\text{Nonbonded}} \]

• Let’s go through the single terms.
• $E_{Bonds}$ models the energy needed to stretch a covalent bond between two atoms. In CHARMM this term is represented by the following formula:

$$E_{Bonds}(r_{ij}) = k_{ij}^B (r_{ij} - r_{ij}^0)^2$$

where $i$ and $j$ are two covalently connected atoms and $r_{ij}$ is the distance between them. $r_{ij}^0$ is the equilibrium bond length and $k_{ij}^B$ is a constant representing the stiffness of the bond.
$E_{Angles}$ models the energy needed to bend the angle formed by two covalent bonds. In CHARMM this term is represented by the following formula:

$$E_{Angles}(\theta_{ijk}) = k_{ijk}^{\Theta} (\theta_{ijk} - \theta_{ijk}^0)^2$$

where $i$, $j$ and $k$ are three covalently connected atoms and $\theta_{ijk}$ is the angle between them. $\theta_{ijk}^0$ is the equilibrium angle width and $k_{ijk}^{\Theta}$ represents the rigidity of the angle.
• $E_{\text{Dihedrals}}$ models the energy needed to bend the dihedral angle formed by three covalent bonds. In CHARMM this term is represented by the following formula:

$$E_{\text{Dihedrals}}(\phi_{ijkl}) = k_{ijkl} [1 + \cos(n\phi_{ijkl} - \delta)]$$

where i, j, k and l are four covalently connected atoms and $\phi_{ijkl}$ is the dihedral angle between them.
• In the case of the dihedral angle energy term, the $k_{ijkl}^\Phi$ parameter stands for the rigidity of the dihedral, while the parameters $n$ and $\delta$ determine the value(s) where the energy has a minimum (for example for the $\omega$ dihedral angle of the protein backbone the two minima should always be close to 180° or 0°.)
• So far we have treated covalently bonded atoms. What about the interactions between non covalently bonded atoms, i.e., pairs of atoms that are separated by more than two covalent bonds in the same protein or pairs of atoms indifferent proteins?

• In the classical force field representation, these interactions are pairwise and modeled with the following formula:

$$E_{\text{Nonbonded}} = E_{\text{vdW}} + E_{\text{electrostatic}}$$

• In principle, $E_{\text{Nonbonded}}$ should be calculated for every pair of atoms in the system. In practice, a distance threshold is used, i.e., nonbonding interactions are actually calculated only for atom pairs whose distance is smaller than a threshold (usually about 1.2 nm).
- $E_{vdW}$ is modeled in CHARMM with the following energy function:

$$E_{vdW}(r_{ij}) = E_{\text{min}_{ij}} \left[ \left( \frac{r_{ij}}{r_{\text{min}}} \right)^{12} - 2 \left( \frac{r_{ij}}{r_{\text{min}}} \right)^6 \right]$$

- The repulsive term at short distances stems from the energy arising from the Pauli repulsion of overlapping electronic orbitals of different atoms;
- The attractive term at longer distances describes the attractive dispersion interactions.
• $E_{\text{electrostatic}}$ models the coulombic interactions between partial charges:

$$E_{\text{electrostatic}}(r_{ij}) = \frac{q_i q_j}{4\pi \varepsilon r_{ij}}$$

where $\varepsilon$ is the dielectric constant of the surrounding medium.

• What are $q_i$ and $q_j$?
• From chemistry we know that the electrons do not evenly distribute along a covalent bond (cfr. electronegativity values).

• In order to reproduce the polarization of a covalent bond, which leads for example to the hydrogen-bonding, a partial charge is assigned to each atom of our protein, depending on the bond partners.
All the parameters, like the equilibrium values and the spring constants in $E_{Bond}$ or the values of $r_{min}^{ij}$ and $E_{min}^{ij}$ in $E_{vdW}$ have to be calibrated for each desired molecule.

The parameters are determined by an optimization process, which consists of several steps. At each step, the calculated properties of a protein are compared to experimental results and then the set of parameters is iteratively changed in order to get as close as possible to the experimentally known properties of the examined protein.
An experiment on a computer

\[ E = E_{\text{Bonds}} + E_{\text{Angles}} + E_{\text{Dihedrals}} + E_{\text{Nonbonded}} \]

\[ = \sum \frac{1}{2} K_b (b - b_0)^2 + \sum \frac{1}{2} K_\theta (\theta - \theta_0)^2 \]

All Bonds  All Angles

Hooke 1635

\[ + \sum K_\phi [1 - \cos(n\phi + \delta)] \]

All Torsion Angles

Fourier 1768

\[ + \sum \varepsilon [(r_0/\varepsilon)^{12} - 2(r_0/\varepsilon)^{6}] \]

All Nonbonded pairs

Van der Waals 1837

\[ + \sum \frac{332q_i q_j}{r} \]

All partial charges

Coulomb 1736

Simple sum over many terms

©Michael Levitt 13
Cautionary Notes

- “Everything should be made as simple as it can be but not simpler” (A. Einstein).

- A force field (model) is never perfect. A model that is not quantitatively fully accurate does not preclude one from establishing results relevant to our understanding of biomolecules as long as the physics of the model are properly understood and explored.

- Calibration of the parameters is an ongoing and imperfect process. Questions and hypotheses should always be designed such that they do not depend crucially on the precise numbers used for the various parameters.

- A computational model is rarely universally wrong or right, it may be accurate in some regards, inaccurate in others. These subtleties can only be uncovered by comparing to all available experimental data.
• Most of the experiments in biochemistry are performed in aqueous solvent. Therefore, we need to model the interactions of our protein with the aqueous environment.

• There are two different ways to model the protein/solvent interactions:
  
  • Explicit solvent representation: Water molecules and ions are modeled considering all of their nuclei as interaction centers and degrees of freedom.

  • Implicit solvent representation: The interactions of the examined protein with the surrounding solvent are described by a mean field framework.
Explicit Water Representation

• The description of the water molecules is based on the same principles aforementioned for the protein.
• Water molecule’s geometry is described with spheres connected by rigid and unbreakable bonds.
• The interactions of the water molecules with each other and with the examined protein are given by the following term of the force field:

\[ E_{Nonbonded} = E_{vdW} + E_{electrostatic} \]
Explicit Water Representation

An experiment on a computer

December 14, 2015
Implicit Water Representation

• From the pictures in the previous slide it stands out that usually the protein is much “bigger” than a water molecule.
• Consider a boat on the Lake of Zurich: in order to study the motion of the boat on the water we do not need to consider the interactions of it with all the single water molecules surrounding it.
• Analogously, we can avoid representing each water molecule in our simulation, given that we have a function that models the mean interaction of the aqueous solvent with the examined protein. This is what is meant by Implicit Water Representation.
• We have seen how to model the protein and the solvent, now we have to “place” them somewhere.
• Where in experiments you would use a cuvette, in simulation we use a virtual box of a given geometry (cubic, orthorombic, truncated octahedron, etc.)
• To avoid finite-size effects and prevent molecules from leaving the box, Periodic Boundary Conditions (PBC) are usually applied.
• The actual simulation box is replicated so that it is fully surrounded by “image” boxes.
• Special algorithms deal with the interactions between the “real” molecules and their images.
• Once we have our protein properly solvated in a simulation box, we can start our experiment.

• In our approximation, the evolution in time of our system will be dictated by the Newton’s equation:

\[ \vec{F}_i(t) = m_i \vec{a}_i(t) = m_i \frac{d^2 \vec{x}_i(t)}{dt^2} \]

where for atom i, \( \vec{F}_i(t) \) is the force acting on it at time t, \( m_i \) is the mass, \( \vec{a}_i(t) \) is the acceleration at time t and \( \vec{x}_i(t) \) is the position at time t.
An experiment on a computer

- At t = 0, the positions of each atom are given from the PDB structure.
- To each atom we assign an initial velocity, taken from a Gaussian distribution.
- From the energy of the system, we can calculate the force acting on each atom at a certain time t:

\[ - \frac{\partial E(t)}{\partial x_i} = \vec{F}_i(t) = m_i \vec{a}_i(t) \]

Remember: to calculate the energy at time t, we just need to plug in the coordinates of all the atoms at time t in the force field function!
Knowing the acceleration, the velocity and the position at time $t$ for all the atoms, we can calculate the positions at time $(t + \Delta t)$:

$$\vec{x}_i(t + \Delta t) = \vec{x}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)\Delta t^2$$

while the velocities are given by the equation:

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \vec{a}_i(t)\Delta t$$

Usually, the time step is chosen to be $\Delta t = 2$ fs.
Changing the ensemble

- The system we have put together has constant number of particles, constant volume and constant energy. The simulation samples then what in statistical mechanics is called a **microcanonical ensemble (NVE)**.

- Often, it will be necessary to **match the conditions of a simulation to that of an experiment** or to something that is physically more suitable to the problem.

- **NVE → NVT**: Add a thermostat; in essence an auxiliary process that modifies particle velocities (common: Berendsen, Nose-Hoover, Andersen, Bussi-Parrinello)

- **NVE → NPT**: Add a manostat and a thermostat; in essence, a manostat is an auxiliary process that rescales the volume of the system in response to the difference between internal and external pressure (common: Berendsen, Parrinello-Rahman, Langevin piston)

- **NVE → μ_m VT**: Add a thermostat and a chemostat (particle bath); in essence, a chemostat is a reservoir of “bath” particles that allows the concentration of particles in the explicitly represented volume to fluctuate (remember aquaporins).
Langevin's Equation of Motion

- Alternatively to Newton’s equation, **Langevin**’s equation of motion can be used to reproduce the time evolution of our system. For each (classical) atom, we have a modified differential equation:

\[
m_i \vec{a}_i(t) = \vec{F}_i(t) + \sqrt{2\gamma_i k_B T m_i} \vec{R}(t) - m_i \gamma_i \vec{v}_i(t)\]

- Newton's equation is “extended” by two terms: a random force relying on a stationary and normalized random (Wiener) process, and a friction term that is velocity-dependent.

- The rate of kinetic energy dissipation (friction) and the magnitude of the random force resulting from equilibrium fluctuations are fundamentally linked because they are caused by the same underlying process: assumed collisions with “bath” particles.

- Because the system is coupled explicitly to an (assumed) bath, energy is not conserved and one naturally obtains a canonical ensemble (**NVT**) when using Langevin's equations of motion to propagate a dynamical system.
• Since modelling all atoms explicitly is computationally expensive, representations with lower resolution have been developed to simulate processes that take place on longer time scales:

<table>
<thead>
<tr>
<th>System</th>
<th>Time scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand unbinding of a weak binder from the binding pocket on a protein</td>
<td>~ 10 ns</td>
</tr>
<tr>
<td>Structural rearrangements of a domain of ~ 100 a.a. in the native state in explicit water</td>
<td>~ns to μs</td>
</tr>
<tr>
<td>Reversible folding of a miniprotein ~ 20 a.a. in implicit water</td>
<td>~μs</td>
</tr>
<tr>
<td>Aggregation of a model peptide in coarse-grained representation</td>
<td>~10 μs</td>
</tr>
</tbody>
</table>
**Typical applications:**

- Reversible **folding** of a miniprotein in implicit water

- Amyloid **aggregation** on the surface of a lipid vesicle

[http://www.biochem-caflisch.uzh.ch/movie/1/](http://www.biochem-caflisch.uzh.ch/movie/1/)  
[http://www.biochem-caflisch.uzh.ch/movie/9/](http://www.biochem-caflisch.uzh.ch/movie/9/)
Experimental approaches to the study of fragment binding to proteins have limitations in resolution. Molecular dynamics simulations of small molecule binding and unbinding provide femtosecond temporal resolution and full spatial resolution.

http://www.biochem-caflisch.uzh.ch/movie/7/
An experiment on a computer

- **Drug design:**

Unzue et al., J. Med. Chem. 2014, 57, 6834–6844
The simulation packages produce an output file called *trajectory*, that contains the coordinates of all the atoms at specific time intervals decided by the user, e.g. every 2-20 ps depending on the type of simulation and on the studied system. From the trajectory, we can calculate time evolution or averages over time of quantities like radius of gyration, internal energy, secondary structure propensity, interaction energy, root mean square deviation (RMSD) of the protein(s).
Simulations of PDZ3 of PSD-95/SAP90 in native state:

Over the complete MD trajectory we can compute the root mean square fluctuations (RMSF) for specific atoms.

The formula for the RMSF of an atom $i$ is the following:

$$ RMSF = \sqrt{\frac{1}{T} \sum_{t_j=1}^{T} (x_i(t_j) - \tilde{x}_i)^2} $$

where $T$ is the time over which one wants to average, $x_i(t_j)$ is the position of atom $i$ at time $t_j$ and $\tilde{x}_i$ is the reference position for atom $i$ (i.e. average position over time).

The plot of RMSF of the $\alpha$’s of each residue of PDZ3 shows that loop regions fluctuate much more than stable secondary structure elements. This alone might be not so surprising, more interesting it becomes when one compares with the crystallographic B-factor or a mutant.

An experiment on a computer

Since the MD trajectory allow us to follow the position of every single atom over time, we can also follow certain interactions over time.

If we plot the distances of the Donor and Acceptor of each hydrogen bond over time, we can see when and how often these bonds are formed during the simulation (remember that for H-bonds Donor-Acceptor distance is usually around 3 Å).

An experiment on MD simulations

From: Steiner et al., FEBS Letters 587, 2158 (2013)

Figure 2. (Top) Two representative snapshots extracted from the MD simulations of CREBBP show the swapping out of the side chain of the Tyr residue preceding the conserved Asn in the BC loop. The dashed line indicates the hydrogen bond between the Asn and the Tyr in the ZA loop. (Bottom) Surface representation of the crystal structure of CREBBP (PDB code 3P1C; left) and of a representative MD snapshot with a partially occluded, i.e., shallower, binding site (right). The colors of the boxes correspond to those in the two structures in the top panel while the color coding of the surface reflects atomic elements (carbon, nitrogen, and oxygen atoms in yellow, blue, and red, respectively). The location of the acetyl-lysine binding site is indicated with a black oval. Figure and

Figure 3. The two major binding modes of acetyl-lysine in the CREBBP bromodomain as observed in MD simulations of spontaneous binding. The N-binding mode (left) corresponds to the crystal structure, while the P-binding mode (right) emerges from the MD simulations and has not yet been reported in experimental studies.
Software Packages: CHARMM

Input files (force field, molecular topologies, coordinates)

Software setup of structural representation

Setup and use of force field

Implementation of sampling algorithms

Simulation data and their analysis

An experiment on a computer

From: Brooks et al., J. Comp. Chem. 30, 1545-1614 (2009)

December 14, 2015 MD simulations
Dynamics Software Packages

- CHARMM (co-development of ~20 research groups around the world over past few decades, originated at Harvard under Martin Karplus, 2013 Nobel laureate in Chemistry)
  - Developed over the last three decades with a primary focus on molecules of biological interest, including proteins, peptides, lipids, nucleic acids, carbohydrates and small molecule ligands.
  - Solution, crystals, and membrane environments.
  - Energy minimization, normal mode analysis, molecular dynamics, Monte Carlo sampling, umbrella sampling, free energy perturbation, large set of analysis techniques, and model-building capabilities.
  - Different energy functions and models including classical potential energy (force field), mixed quantum mechanical-molecular mechanical approach, and interface to ab initio quantum chemistry.
  - Explicit solvent and various boundary conditions, several implicit solvent models (surface based, generalized Born, finite-difference Poisson).
- AMBER (originated at UCSF)
- GROMACS (originated at University of Groningen)
- NAMD, DESMOND, and many more ...
Movies ....

http://www.biochem-caflisch.uzh.ch/movies/

Kinetic response of a photoperturbed allosteric protein

Brigitte Buchli\textsuperscript{a,1}, Steven A. Waldauer\textsuperscript{a,1}, Reto Walser\textsuperscript{a,1}, Mateusz L. Donen\textsuperscript{a}, Rolf Pfister\textsuperscript{a}, Nicolas Blöchliger\textsuperscript{b}, Sandra Steiner\textsuperscript{b}, Amedeo Caflisch\textsuperscript{b}, Oliver Zerbe\textsuperscript{a}, and Peter Hamm\textsuperscript{a,2}

Departments of \textsuperscript{a}Chemistry and \textsuperscript{b}Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

Edited by Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, NM, and approved June 3, 2013

Fig. 5. Change of water density as a function of simulation time, compared with that just before switching. Red depicts increased density and blue decreased density. The contour surfaces correspond to changes of ±0.01 water/Å\textsuperscript{3} (for comparison, the bulk water density is ∼0.033 water/Å\textsuperscript{3}). The protein is shown as a gray ribbon and the photoswitch (visible only in part) is shown in yellow. See also Movie S1.
Interested in a PhD position?