Molecular Simulations: Applications to biomolecular interactions and conformational flexibility

Martin Zacharias
International University Bremen
School of Engineering and Science
Outline

• Why performing computer simulations on complex molecular systems and what is possible?

• How does it work?

• Examples of own research

• Outlook
Structure and dynamics of biomolecules

- The structure, association and dynamics of biomolecules is important for its function.

- Why?
  - A biomolecule must adopt a defined three-dimensional (3D) structure to be active (e.g. to perform an enzymatic function).
  - A specific arrangement of biomolecules (or subunits) may be necessary to form a functional complex.
  - One static 3D structure may not be sufficient: conformational changes and transitions can be necessary for the function of biomolecules.
Limitations of experimental methods

- Most experimental methods to study the structure and dynamics of biomolecules have either (or neither) a high spatial or a high time time resolution.
  - X-ray crystallography allows to determine the structure of a biomolecule with a very high spatial resolution (the position of each atom in the molecule can be determined) but the time resolution is usually very low (hours).
  - Some spectroscopic methods allow to study conformational dynamics at high time resolution ($10^{-3}$ to $10^{-9}$ sec) but, at most, only the distance between chromophor groups can be followed.
Computer simulation of biomolecular structure and dynamics

• In a molecular dynamics (MD) simulation one tries to simulate the motion of each atom in one or several molecules as a function of time.

• Structure formation and association of molecules can be studied at high resolution in space and time.
  – time resolution \( \sim 10^{-15} \) sec.
  – Coordinates of each atom are calculated at every time during the simulation.
Purpose of molecular simulation studies

• Goals of MD simulation studies:
  – Understanding the driving forces for structure formation
  – Prediction of structure formation and association
  – Understanding the motions of a biomolecule and how this is coupled to its function.
Larger-scale motions

Protein folding

Local motions
Side chains, Loops

Collective motions
Helix subunits and domains

Current Computer Simulations capability
Development of computer performance

Doubling time = 1.5 yr.
# Extrapolated Future of Molecular Dynamics Simulations

<table>
<thead>
<tr>
<th>Year</th>
<th>Type</th>
<th>Length (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1964</td>
<td>atomic liquid (argon)</td>
<td>$10^{-11}$</td>
</tr>
<tr>
<td>1971</td>
<td>atomic liquid (water)</td>
<td>$10^{-11}$</td>
</tr>
<tr>
<td>1977</td>
<td>protein in vacuum</td>
<td>$2\times10^{-11}$</td>
</tr>
<tr>
<td>1983</td>
<td>protein in water</td>
<td>$2\times10^{-11}$</td>
</tr>
<tr>
<td>1989</td>
<td>protein-DNA complex</td>
<td>$10^{-10}$</td>
</tr>
<tr>
<td>1997</td>
<td>small peptide folding</td>
<td>$10^{-7}$</td>
</tr>
</tbody>
</table>

**Expectation:**
- 200x folding of small protein: $10^{-4}$
- 2029 biomolecules + water: $10^{-3}$
- 2034 E.coli bacteria: $10^{-9}$
- 2056 mammalian cell: $10^{-9}$
- 2080 biomolecule + water (as fast as nature): $10^{-6}$
- 2170 human body: 1

Source: van Gunsteren et al., Angew. Chem. 2006, 45, 4064.
Duan, Kollman and Harvey. Chemistry for the 21st Century (Wiley-VHC), 2000
Outline

• Why performing computer simulations on complex molecular systems and what is possible?
  • How does it work?
• Examples of own research
• Outlook
The basis of biomolecular simulations

- Structure formation of biomolecules is driven by interatomic interactions (this includes interactions within the molecule but also with surrounding solvent and ions).

![Diagram showing intermolecular and intramolecular interactions]

- How can one describe the interactions theoretically?
Quantum mechanics

- Quantum mechanics offers the principally most accurate description of intra- and intermolecular interactions.

- The electron coordinates are the principal variables in quantum mechanical calculations on molecules.

- In a quantum mechanical description a wave function for the electrons is calculated.

- The form of the electron wave function determines the number and geometry of chemical bonds of an atom, hence, the geometry and form of a molecule.

- The wave function contains all information necessary to calculate the physical properties of the molecule.
Limitations of quantum mechanical calculations on biomolecules

• Numerical problem:
  – Only an approximate electron wave function for a molecule can be calculated.
  – The numerical demand increases rapidly with the size of the molecule.
  – Many biomolecular systems that one wishes to tackle using molecular simulation approaches are too large to use QM methods.
Molecular mechanics force fields

- Atom-centered force field methods ignore the motions of electrons and calculate the energy only as a function of the atom (nuclear) positions.

- On the slow whole atom motion time scale atoms “feel” only an average electron distribution (Born-Oppenheimer approximation).

- Based on this (and other) approximations an energy function (force field) for a molecule can be introduced that depends only on the coordinates of atoms.

\[ E(r_1, r_2, r_3, \ldots, r_n) = E(x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, \ldots, x_n, y_n, z_n) \]
How does a force field look like?

Force field energy of a molecule:
$$E(r_1, r_2, \ldots, r_n) = \sum_{N\text{bonds}} \frac{1}{2}k_{bi} (b_i - b_{i,0})^2$$
Force field terms to control the bond length

• The energy function for "controlling" the bond lengths in a molecule:

\[ E(r_m, r_n) = \frac{1}{2} k_{bi} (b_i - b_{i,0})^2 , \]

\( b_{i,0} \) is the optimal bond length,
\( k_{bi} \) is an appropriate force constant
\( b_i = \vert r_m - r_n \vert \)
How does a force field look like?

Force field energy of a molecule:

\[ E(r_1, r_2, \ldots, r_n) = \]

\[ \sum_{N_{\text{bonds}}} \frac{1}{2} k_{b_i} (b_i - b_{i,0})^2 \]

\[ + \sum_{N_{\text{angles}}} \frac{1}{2} k_{\theta_i} (\theta_i - \theta_{i,0})^2 \]
Force field term to control the bond angle

- The function that describes the change in energy upon bond angle deformation:

\[ E(r_l, r_m, r_n) = \frac{1}{2} k_{\theta_i} (\theta_i - \theta_{i,0})^2, \]

subscript N indicates normalized vectors
\( \theta_{i,0} \) is the optimal bond angle
\( \theta_i \) : bond angle
How does a force field look like?

Force field energy of a molecule:

\[ E(r_1,r_2,\ldots,r_n) = \]

\[ \sum_{\text{N bonds}} \frac{1}{2} k_{bi} (b_i - b_{i,0})^2 \]

\[ + \sum_{\text{N angles}} \frac{1}{2} k_{\theta i} (\theta_i - \theta_{i,0})^2 \]

\[ + \sum_{\text{N torsions}} \sum_{n=1..N_i} k_{\tau ni} (1 + \cos [n_i \tau_i - \delta_i]) \]
Force field term to control the rotation around a bond

- Energy function for bond rotation:
  \[ E(r_k, r_l, r_m, r_n) = \sum_{j=1..N} k\tau_j (1 + \cos [j \tau - \delta]) \]
  \( \tau \): dihedral torsion angle around (l-m) bond.
  \( \delta \): shifts the minimum of the energy function.
How does a force field look like?

Force field energy of a molecule:

\[ E(r_1, r_2, \ldots, r_n) = \]

\[ \sum_{\text{Nbonds}} \frac{1}{2}k_{bi} (b_i - b_{i,0})^2 \]

\[ + \sum_{\text{Nangles}} \frac{1}{2}k_{\theta i} (\theta_i - \theta_{i,0})^2 \]

\[ + \sum_{\text{Ntorsions}} \sum_{n=1..N_i} k_{\tau ni} (1 + \cos [n_i \tau_i - \delta_i]) \]

\[ + \sum_{\text{nbpairs}} \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{d_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{d_{ij}} \right)^{6} \right] \]
Repulsion of non-bonded atoms at close distance

- Two atoms that are not connected by a bond are strongly repelled if their distance is below a certain minimum distance (sum of their atom van der Waals radii).
- This repulsion has a quantum mechanical origin → Pauli exclusion principle: two electrons are not allowed to have the same set of quantum numbers (same position and spin). This can happen if the electron cloud of one atom penetrates into the electron cloud of another atom.
- The repulsive interaction is modeled by:

\[ E(d_{ij}) = \varepsilon_{ij} \left(\frac{\sigma_{ij}}{d_{ij}}\right)^{12} \]

\[ d_{ij} = |r_i - r_j| : \text{distance between atoms } i \text{ and } j \]
\[ \varepsilon_{ij} \text{ and } \sigma_{ij} \text{ are parameters specific for atom types} \]
Short range attractive interactions between non-bonded atoms

- Fluctuations in the electron cloud of an atom can lead to a non-spherical electron distribution causing an instantaneous electric dipole.
- The electric field of this dipole can induce an electric dipole in a neighboring atom. The induced dipole interacts favorably with the inducing dipole.
- Since the electric field of the inducing dipole is $\sim 1/r^3$ (the magnitude of the induced dipole is proportional to the electric field) and the dipole-dipole interaction is $\sim 1/r^3$ the attractive interaction is $\sim 1/r^6$.
- This type of attractive interaction is called (London) dispersion interaction.
- The Lennard-Jones potential combines a $1/r^{12}$ repulsive and a $1/r^6$ attractive interaction to model the interaction between neutral atoms.

$$E(r_{ij}) = \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$
How does a force field look like?

Force field energy of a molecule:

\[ E(r_1, r_2, \ldots, r_n) = \]

\[ \sum_{\text{Nbonds}} \frac{1}{2} k_{bi} (b_i - b_{i,0})^2 \]

\[ + \sum_{\text{Nangles}} \frac{1}{2} k_{\theta i} (\theta_i - \theta_{i,0})^2 \]

\[ + \sum_{\text{Ntorsions}} \sum_{n=1..Ni} k_{\tau ni} (1 + \cos (n_i \tau_i - \delta_i)) \]

\[ + \sum_{\text{nbpairs}} \varepsilon_{ij} [(\sigma_{ij}/d_{ij})^{12} - (\sigma_{ij}/d_{ij})^6] + q_i q_j / (4\pi\varepsilon_0 d_{ij}) \]
Equations of motion for a set of particles

- Every atom of a molecule is considered as a point with a mass (which basically corresponds to the nuclear mass of the atom). Hence, the molecule is treated as a classical system of particles (mass points).

- The motion of such a classical system of atoms (molecule) is described by Newton's differential equation of motion (given for one selected particle i):

\[ F_i = m_i \ a_i = m_i \ \frac{dv_i}{dt} \]
= \[ m_i \ (\frac{d \nu_x}{dt}, \frac{d \nu_y}{dt}, \frac{d \nu_z}{dt}) \]
= \[ m_i \ (\frac{d^2 x}{dt^2}, \frac{d^2 y}{dt^2}, \frac{d^2 z}{dt^2}) \]

\( F_i \): force on atoms (from the energy function);
\( a_i \): acceleration of atoms:
- second derivative of position with respect to time
- first derivative of velocity \( \nu_i \) with respect to time
\( m_i \): mass of each atom
Solution of equations of motion

- Newton’s equation describes the change in particle coordinates and velocities upon changes in time.

- What do we get from an integration of Newton’s differential equation of motion for a system of atoms (point masses):
  - The solution provides the path each atom takes in space as a function of time.

![Diagram of an atom with initial velocity $v_0$ and final velocity $v_1$. Force at later time causes acceleration and change in velocity.]

Force at later time causes acceleration and change in velocity.
The motion of atoms

• Unfortunately, an analytic solution (integration) of Newton’s equation of motion (that is a function which we can write down for the path each atom takes in time) is not in general available for systems with more than two interacting particles.

• However, the equations of motion for a system of interacting particles can be integrated numerically in small time steps.

• The resulting set of (discrete) coordinates (trajectory) for each atom (particle) is an approximation to the “real” path the atom takes in time:
What needs to be done at every time step?

1. Calculate the forces on each atom in the system (due to all interactions with other atoms)
2. Calculate the acceleration of each atom
3. Calculate the change in velocities
4. Calculate new positions of each atom that are reached within the time step
5. Go to step 1 and continue

Atom at start position
What can be simulated?

• A molecular dynamics time step must be small enough such that the forces on atoms do not change.
  – time step $\sim 0.5 - 2$ fs ($10^{-15}$ s)

• Maximum simulation time for peptides and small proteins:
  – currently $\sim 0.1 - 1$ µs ($10^{-6}$ s)
  – fastest folding small proteins fold in about $\sim 10$-100 µs.
Visualization of molecular motions

- Example: Short MD simulation (0.2 ns) of a small peptide
Outline

• Why performing computer simulations on complex molecular systems and what is possible?

• How does it work?

• Examples of own research

• Outlook
Application example 1:
Dynamics of modified bases in RNA
Set up of molecular dynamics simulations

- Isolated DNA/RNA fragment
- DNA/RNA in a box + waters and ions used for MD
Fluorine-substituted bases in nucleic acids

- 2,4-difluorotoluene-nucleobases in DNA have been used to study fidelity in DNA replication*.
- In RNA 2,4-difluorophenyl (2,4diFP)-bases can act as universal bases**.
- Model system to study the influence of base modifications on the structure and dynamics of RNA.
- Questions:
  - Why universal bases?
  - What is the structural difference to „natural base pairs“?
  - Are there differences in the mobility/dynamics with respect to natural RNA?

**Parsch & Engels JACS, 2002, 124, 5664.
Molecular dynamics of dsRNA containing difluorophenyl nucleobases

- MD simulations (5.1 ns) were performed on six systems:
  - 5'-'CGCU'GCG/5'-'CGCA'GCG
  - 5'-'CGCF'GCG/5'-'CGCA'GCG
  - 5'-'CGCP'GCG/5'-'CGCA'GCG
  - 5'-'CGCF'GCG/5'-'CGCF'GCG
  - 5'-'CGUUA'CG/5'-'CGUA'ACG
  - 5'-'CGUFACG/5'-'CGUAACG

- MD simulation (Amber94* with PME):
  RNA + ~ 2500 waters + Na⁺ and Cl⁻ ions
  equilibration : 2 ns; data gathering: 3 ns

Conformational dynamics of 2,4diFP containing RNA

- Average conformation of all RNAs stays close to A-form (rmsd:~1.5 - 2 Å with respect to A-form start structure)
- Strongly enhanced conformational fluctuations at the central base pair in case of a modified base.
- Enhanced fluctuations not only of the modified base but also opposing base
- Small influence on neighboring base pairs
Character of the motion of the central base pair

- No stable base pairing geometry between central base analogue and opposite base
- Enhanced conformational fluctuations of the central (modified) base pair mainly in the plane perpendicular to the helical axis
- "flips" between "discrete" substates
Base-pair shear and opening motions

- Central base pair shows enhanced shear and opening motions in the presence of base analogue.
- Strongly localized effect (no/little increase of shear and opening motions of neighboring base pairs).
Correlation of helical and backbone dihedral torsion angle motions

- Mechanism of shear and opening motions in RNA?
  - Little correlation of any dihedral torsion angle motion and base pair shear and opening motions

- But: Shift and twist motions of single central bases correlate with backbone torsion angle changes $\varepsilon$ and $\zeta$.

~ helical axis

$\zeta$ dihedral angle
$\varepsilon$ dihedral angle

Simulation time (ns)

<table>
<thead>
<tr>
<th>2,4diPF base</th>
<th>opposing adenine</th>
</tr>
</thead>
<tbody>
<tr>
<td>base shift(Å)</td>
<td></td>
</tr>
<tr>
<td>0:central base</td>
<td></td>
</tr>
<tr>
<td>(0 vs. +1)</td>
<td></td>
</tr>
<tr>
<td>(0 vs. −1)</td>
<td></td>
</tr>
</tbody>
</table>

$\varepsilon$ (0)
$\varepsilon$ (-1)
$\zeta$ (0)
$\zeta$ (-1)
$\delta$(0)
$\alpha$ (0)
Conclusions

• Presence of 2,4-difluorophenyl containing base pairs preserves average A-form geometry during nanosecond MD simulations

• No stable base pairing geometry with opposing natural adenine base

• Strongly enhanced opening and shear motions of base analogue and opposing base (maybe similar in for other modified bases?)

• Conformational transitions to major and/or minor groove lead to greater overall accessibility of base analogues
  – May explain greater chemical reactivity of modified or mismatched bases with chemical probes
Application example 2:
Flexibility and deformability of DNA and its role for recognition by proteins
Protein-induced DNA deformation

- Protein binding frequently causes DNA deformation.
- Recognition influenced by:
  - protein-DNA interaction ("direct readout")
  - sequence-dependent DNA-deformability ("indirect readout")
- Examples:
  - Transcription repressors
  - TATA-Box binding protein (TBP)
  - DNA repair enzymes

TBP in complex with TATA-Box

Purin-Repressor-DNA complex
DNA deformation during molecular dynamics

Calculation of a potential of mean force (PMF) using umbrella sampling:

\[ V_{\text{umb}} = k_{\text{umb}} (d - d_{\text{ref}})^2 \]

d: reaction coordinate : distance between atom groups of two nucleotides on opposite strands (sugar heavy atoms)

\[ 5\text{'}-\text{GCGTATA}\text{TAGCGC} \]
\[ 3\text{'}-\text{CGCATATATGCG} \]
Potential of mean force for minor groove opening

- Calculated PMF for minor groove opening is larger in case of central AAATTTT sequence (~12-13 kcal/mol) than for TATATA case (~7-8 kcal/mol).

- Protein-DNA interaction needs to provide additional ~ 5-6 kcal/mol to allow opening of AAATTTT vs. TATATA sequence.

PMFs for first and last 0.5 ns windows of forward and backward simulations
Comparison with experimental DNA structures

• Structures along the deformation pathway compare well with experimental structures.
  
  – Backbone Rmsd of deformed TATATA-structure \( (d_{\text{ref}} = 18.0 \text{ Å}) \) with central segment of purR-DNA: 1.8 Å

  – Backbone Rmsd of deformed TATATA-structure \( (d_{\text{ref}} = 18.0 \text{ Å}) \) with central segment of TATA-box-DNA: 1.7 Å.
Conformational changes in deformed DNA

- Minor groove opening is accompanied by change in central twist and roll (only at large $d_{\text{ref}}$ in case of central AAATTT sequence)
- Partial unstacking at central base pair step at large induced deformation ($d_{\text{ref}} = 18.0 \text{ Å}$)
Application example 3:
Folding simulations on a DNA hairpin
Application to a DNA hairpin

- The DNA sequences 5′-GCGAAGC and 5′-GCGCAGC form very stable triloop hairpin structures (melting temp. ~65 °C).

\[
\begin{align*}
A \\
G--A \\
C--G \\
G--C
\end{align*}
\]

- Characteristic sheared G:A closing base pair

Zhu et al., Nat. Struct. Biol. 1995
Standard MD simulations at 310 K

- Extended stacked single stranded DNA structure:
  5′-CGAAGC
  + 6 K+, ~1200 waters
- Amber force field
- Temperature: 310 K
- Total simulation time: 10ns
Replica-exchange molecular dynamics

- Multi-temperature replica exchange MD:
  - Replicas of the system are run at N temperatures \( T_1, T_2, \ldots, T_N \)
  - Exchange between replicas \( i, j \) (at neighboring \( T \)), accepted according to:
    \[
    \exp\left[\frac{1}{k_B T_j} - \frac{1}{k_B T_i}\right] (E(j) - E(i)) \geq \text{rand}(0,1)
    \]
  - Momenta are adjusted according to:
    \[
    p[i] = \sqrt{\frac{T(i)}{T(j)}} p[j]
    \]

Replica Exchange MD starting from a stacked single stranded DNA chain

- The DNA sequences 5′-GCGAAGC
- Temperature: 310-420 K (16 replicas)
- Exchange every 1000 steps (22000 exchanges) =22 ns (1 fs t-step)
Final structure in good agreement with experiment

- Stem base pairs and sheared closing base pair has formed correctly
- Final structure shows an Rmsd of <1.5 Å from experiment

Superposition of one snapshot of the final stage of the RexMD on experimental structure (pdb1zhu)
Conclusions

• Effect of modifications on nucleic acid dynamics
  – Local conformational changes

• DNA minor groove opening
  – Large scale DNA deformation

• DNA hairpin formation
  – Example of a large scale conformational transition
Outline

• Why performing computer simulations on complex molecular systems and what is possible?

• How does it work?

• Examples of own research

• Outlook
Towards molecular dynamics simulations of nanomolecular structures

- Time scale of current simulations is still too small for many problems concerning nanomolecular structures.

- „Simulation tricks“ may help to overcome these limitations
  - Reduced models
  - Targeted molecular dynamics simulations to induce structural transitions during simulations

Pulling a molecule through a membran (Roccatano et al. submitted)

Induced rotation of protein domains (Boeckmann et al. 2003, and Schulten group)

Enforced protein unfolding (Schulten group, 2002)
Acknowledgements

Group members:
Andre Barthel (former postdoc)
Jeremy Curuksu (PhD student, VW-Stiftung)
Aleksandar Ivanov (grad. student, IUB)
Ragav Kannan (grad. student, IUB)
Andreas May (PhD student, DFG)
Nico Riemann (former PhD student)
Danilo Roccatano
(former Res. Associate, now IUB-lecturer)
Florian Sieker (PhD student, IUB)