❓

Biomolecular Modeling: Goals, Problems, Perspectives (2006)

▼ Table of contents

1.Introduction

- Quantum mechanics governs interaction between electrons of atoms and molecules.
- Nonbonded interactions can be well described by classical potential function or force field.

Four choices for modelling a biomolecular system

1. What level of degrees of freedom are explicitly considered in the model?
    a. mainly between 10^5 and 10^6 atoms or particles. wow so much less than Avogadro number.

2. Which force field is used to describe the energy of the system as a function of the cosen degrees of freedom?
3. How these degrees of freedom are to be sampled?
4. How spatial boundaries and external forces are modelled?

- What’s the point of periodic boundary condition?
    - Periodic boundary condition — box that contains the molecular system surrounded by an infinite number of copies of itself.

Methods	Degrees of freedom	Properties, processes	Timescale
Quantum Dynamics	Atoms,nuclei,electrons	Excited states, relaxation, reaction dynamics	picoseconds
Quantum Mechanics (ab initio, density functional, semiempirical, valence bond methods	Atoms, nuclei, electrons	Ground and excited states, reaction mechanism	no time scale
Classical statistical mechanics (MD, MC, Force Fields)	Atoms, solvent	Ensembles, averages, system properties, folding	nanoseconds
Statistical methods (database analysis)	Groups of atoms, amino acid residues, bases	Structural homology and similarity	no time scale
Continuum mechods (hydrodynamics and electrostatics)	Electrical continuum, velocity continuum etc.	Rheological properties	supramolecular
Kinetic equations	Populations of species	Population dynamics, signal transduction	macroscopic

Limitations of present day biomolecular modelling — Four problems in biomolecular modeling

Force field problem	1) very small (free) energy differences, many interactions 2) entropic effects 3) variety of atoms and molecules
search problem	1) convergence 2) alleviating factors 3) aggravating factors
ensemble problem	1) entropy 2) averaging 3) nonlinear averaging
experimental problem	1) averaging 2) insufficient number of data 3) insufficient accuracy of data

Why computer simulation is used in science?

1. experiment is impossible	collision of stars or galaxies / weather forecast
2. experiment is danger	flight simulation / explosion simulation
3. experiment is expensive	high pressure simulation / wind channel simulation
4. experiment is blind	Many properties cannot be observed in very short time scales and very small space scales

2. Force field problem

- Force field equation consist of potential energy terms for covalent interactions between atoms
    - bonds
    - angles
    - dihedrals
    - impropers
- and nonbonded (electrostatics + van der waals): between atoms in different molecules and between atoms in a molecule separated by more than two or three covalent bonds.
    - this is the critical interest for force field parametrization because these interactions govern the thermodynamics process and equilibria in most biological process (e.g. protein folding, micelle formation)

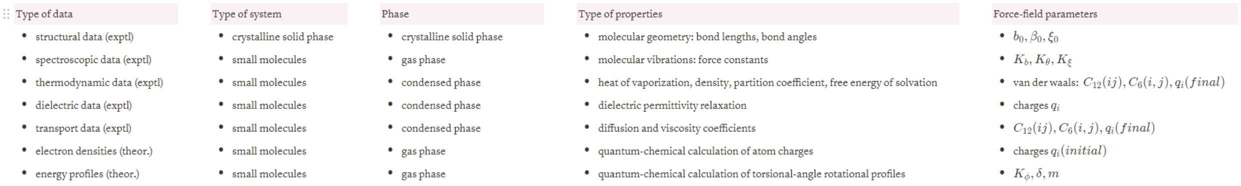
Problems

- very small (free) energy differences, many interactions
    - result from summation over many atom pairs contributing to non-bonded interaction.
    - accuracy ↓ when # of atom-pairs ↑
- entropic effects
    - only at 0 K that there is no entropy and we are not interested in that.
    - entropy is the measure of the extent of conformational space accessible to the molecular system at a given temperature T.
    - finding global minimum is meaningless when its entropy accounts for a sizeable fraction of its free energy.
    - parameters should be derive to be consistent w/ entropic effects.
    - contribution of entropy to free energy $F = U - TS$
- variety of atoms and molecules
    - if force field parameters are transferable between atom groups or atoms, then this problem may be alleviated.

2.2. Calibration of Force-Field Parameters

- type of data
- type of systems
- thermodynamic phase
- properties to be used as calibration set for specific force-field parameters
- pdb bank has it in different properties of the folded molecule

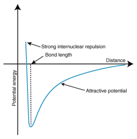
📊 how force field is derived (pic of table)



Type of data	Type of system	Phase	Type of properties	Force-field parameters
structural data (exptl)	crystalline solid phase	crystalline solid phase	molecular geometry: bond lengths, bond angles	b_0, β_0, ξ_0
spectroscopic data (exptl)	small molecules	gas phase	molecular vibrations: force constants	K_b, K_θ, K_ξ
thermodynamic data (exptl)	small molecules	condensed phase	heat of vaporization, density, partition coefficient, free energy of solvation	van der waals: $C_{12}(ij), C_6(i,j), q_i(final)$
dielectric data (exptl)	small molecules	condensed phase	dielectric permittivity relaxation	charges q_i
transport data (exptl)	small molecules	condensed phase	diffusion and viscosity coefficients	$C_{12}(ij), C_6(i,j), q_i(final)$
electron densities (theor.)	small molecules	gas phase	quantum-chemical calculation of atom charges	charges $q_i(initial)$
energy profiles (theor.)	small molecules	gas phase	quantum-chemical calculation of torsional-angle rotational profiles	K_ϕ, δ, m

- energy profiles (theor.)
- charges $q_i(initial)$ is used as initial estimate
- changing a subset of parameters by taking them from other force fields or models may introduce inconsistencies and inaccuracies.
- consistent between solute and solvent molecules
- free energy of solvation
- GROMOS good for studying protein folding bc of its own kJ/mol difference →more accurate

2.3. Long range forces



- electrostatics are long range bc ↓ radius → ↑ interaction energy
- two charged molecules — interaction energy proportional to r^{-1} and corresponding force proportional to r^{-2}
- two neutral molecules — interaction energy proportional to r^{-2} and corresponding force proportional to r^{-3}
- two neutral molecules w/ dipole moments — interaction energy proportional to r^{-3} and corresponding force proportional to r^{-4}
  - will increase with increasing dipoles (e.g. quadrupole moment r^{-5}
- $\int_0^\infty V^{el} 4\pi r^2 dr$
- Two techniques are used to evaluate long-range (electrostatic) interactions

1. lattice sum method to put it into particularly shaped boxes (cubic,rectangular, triclinic, truncated octahedral) and surrounding it by infinite number of identical copies of itself.
   - so boundary is moved to infinite but not really.
   - artificial periodicity, treats interaction beyond the box size as **periodic**
   - lattice-sum methods
     - Ewald summation
     - Particle-particle mesh

2. approximate the medium beyond a given cut-off distance from a specific atom by dielectric continuum of uniform permittivity and ionic strength
   - does not introduce artificial periodicity

    - Particle-mesh-Ewald method
  - does not involve average

2.4. Testing Biomolecular Force Field

- many ways to compare simulated conformational distribution
  - x-ray diffraction, NOE, J-coupling constant, chemical shift values calculated from simulations of solutes in solution may be compared with NMR.
- force field should reproduce the conformational distribution of the solute in a particular thermodynamic condition.

2.5. Perspectives in Force-Field Development

1. van der waals parameters and partial charge distributions of charged moieties should be based on free energies of solvation.
   a. a heavy task from both simulation and experiments.
2. properties of solvent mixtures should be evaluated as a function of their composition
   a. thermodynamic properties of energy and density of mixing, are important when free energy of solvation and folding of solute is calculate
3. limited accuracy when systems is under varying dielectric conditions
   a. biomolecule FF take a average of varying dielectric condition and this will sacrifice the accuracy of the system for faster time.

- the advent of polarizable force fields
  - varying dielectric condition
  - need completely new reparameterization
- to simulate larger biomolecular systems and slow process (e.g. membrane and micelle formation)
  - coarse graining is preferred.
  - number of covalently bound atoms are treated as a single particle or bead
  - faster with expense of losing atomic detail.

3. The search (sampling) problem

- The search for global energy minimum of a high-dimensional function with large number of degrees of freedoms or the search for those regions of the surface that contribute most to the free energy of the system — daunting and impossible task.
- biomolecular system cannot be described by a single global min energy configuration or structure, but only by a stat mech ensemble of configurations, in which the weight of the configuration is given by the Boltzmann factor

$$P(x) \sim \exp(-V(x)/k_B T)$$

- exponential weighting implies that high energy regions of the energy hypersurface will not contribute configurations that are relevant to the state of the system, unless they are very numerous (entropy).
- equilibrium properties of the system is dominated by configurations where V(x) is low.
- One challenge is to develop methodology to efficiently search the vast biomolecular energy surface for regions of low energy.

3.1. Methods to Search and Sample Configuration Space

- **Two basic types of search methods: systematic and heuristic**
- Systematic scans through complete or significant fraction of the config space biomolecular system
    - can only be applied for molecules w/ few degrees of freedom. grow and exponential computing time.
- Heuristic — visit only tiny fraction of the configuration space, aim at generating a representative set of system configuration.
    - nonstep methods — config generated independent of one another
    - step methods that build a complete molecular or system config from config of fragments of the molecule or system in step-wise method
    - step methods that generate new config from a previous config

Reason for change	EM (energy minimization)	MC (monte carlo)	MD (molecular dynamics)	SD (stochastic dynamics)	PEACS (potential energy annealing conformational search
energy	yes	yes	no	no	yes
energy gradient	yes	no	yes	yes	yes
second derivative of the energy	yes	no	no	no	no
memory	no	no	yes	yes	yes
randomness	yes	yes	no	yes	no

-
    - For MD,
        - energy min is chosen based on energy gradient values (steepest descent and conjugate-gradient methods)
        - step is determined by force $\partial V / \partial x$
    - For MC, the step direction is taken at random, limited by Boltzmann acceptance
        - when $\triangle V < 0$, the step is accepted
        - when $\triangle V > 0$, the probability of being accepted is $\exp(\triangle V / k_b T)$

- Efficacy of search is hindered by the dominant action of energy surface to find lower potential energies
  - The high-energy barriers between local minima means that radius of convergence of step methods is very small.

3.1.1 Deformation or Smoothening of the Potential-Energy Hypersurface to Reduce Barriers

1. smoothening of the potential energy function V(x) allows for faster search of minimas
   a. smoothening enhances radius of convergence of structure refinement.
2. electron density is smoothed by omission of high-resolution diffraction intensities when back calculating the electron density from Fourier transforms.
3. softening hard-core atoms by removing repulsive short-range nonbonding interaction → smooth energy surface.
4.
5. avoiding the repeated sampling of an energy well through local potential-energy elevation or conformational flooding
   a. Once a local minima is found, it is removed from the potential energy surface by a suitable potential energy function. **metadynamics**
6. softening of geometric restraints derived from experimental data from time averaging
7. to overcome barrier in energy hypersurface of 3D cartesian coordinate by performing MD in 4D and free energy differences can be calculated.
8. Freeze the highest frequency degrees of freedom

3.1.2. Scaling of System Parameters To Enhance Sampling

1. simulating at higher temperature, to surmount energy barriers, and then gradually cooling it down — annealing.
2. scaling masses — Because mass does not appear in the partition function, configurational integral, it can be exploited to enhance sampling
   a. By ↑ the mass of part of molecule, the inertia ↑ → ↓ energy barriers → ↑ timesteps
   b. hydrogen mass partition
3. mean force approaches — goal is to reduce computational cost by treating a multi-body problem as a one-body problem.
   a. separating into 2 parts and N_A identical copies of part A and N_B identical copies of part B.
   b. AA and BB: force = 0.
   c. AB: force factor of N_B^{-1}; force A exert on B
   d. BA: force factor of N_A^{-1}; force B exert on A

3.1.3. Multi-copy Simulation with a Given Relationship between the Copies.

1. genetic algorithm
2. replica-exchange algorithm
   1. multiple copies each simulated at distinct temperature
   2. close enough temperatures, copies are exchanged through exchange probability based on Boltzmann factor
3. SWARM-type MD — combining a collection of copies of the system each with its own traj into a cooperative multicopy system that searches configurational space.
   1. less attracted by local minima and is more likely to follow an overall energy gradient toward the global energy minimum.
   2. drives it to the average of the trajectories

3.2. Convergence of Simulated Properties

- ranges from femtoseconds to seconds or even longer.
- timescale of the change or relaxation of a system will depend on
  1. type of system
  2. thermodynamic starting point
  3. particular quantity or property
- potential energy relaxes faster than rmsd
- different ways to analysis the relaxation and dynamics
  1. equilibrium simulation, monitor time-series, average value, fluctuations, or autocorrelation function

3.3. Alleviation of the Search and Sample Problems

- Not the end of the world because it contains way fewer conformations at equilibrium than all the possible conformations.
- This is due to the hydrogen bonding that restrict a lot of conformational changes
- Therefore, # of conformation does not increase exponentially with # of polypeptide length.

3.4. Aggravation of Search and Sampling Problems

- Free energy of a system of N particles in volume V
- limitations
  1. solvent degrees of freedom also contribute to the free energy of folding, not just solute degrees of freedoms.
  2. in protein-ligand algorithm, in docking
     - the inclusion of protein degrees of freedom should also be addressed but this will aggravate the search and sampling of docking.

3. dependence on the magnitude of hydrophobic effect on the hydrophobic cluster and the composition of the solvent.

3.5. Perspectives Regarding the Search and Sampling Problem

- single strep perturbation methodology allows ligand-binding free energies or solvation free energies to be obtained for a great many compounds.

4. The ensemble (sampling) problem

- govern by stat mech
- The state of a system is characterized by a Boltzmann ensemble of configurations or structures.

4.1. Free Energy, Energy, and Entropy of Solvation

ΔG_s broken down to energetic contribution, change in solute solvent energy, and entropic contribution

4.2. Temperature Dependence of Folding Equilibria

- entropy changes - temperature dependent

4.4 Perspectives in Calculating Entropies

- entropy is the key property for understanding phenomena such as hydrophobic interactions, solvation, ligand binding, etc.
  - requires complete sampling of phase space
- to obtain free-energy difference between 2 states of a system or between two systems, it is sufficient to extensively sample the relevant parts of phase or configuration space where the two states or systems differ

4 ways to compute entropy difference

coupling parameter λ approach

1. Entropy difference from energy difference and thermodynamic integration (TI) of the free energy
2. Entropy difference directly from TI
3. Entropy difference from finite temperature difference
4. Solvation entropy difference from solute-solvent entropy difference (using TI) and end state solvent-solvent energy differences.

5. The experimental problem

- we need experimental data to facilitate with force field developments. Quantum-alone does not suffice.

- 3 problems with experimental data in biomolecular modeling
    1. Every experiment involves an averaging over time and the space or molecule → does not yield direct information on all configurations constituting a simulation trajectory.
    2. Experimental data for biomolecular systems are scarce relative to the number of degrees of freedom involved.
        a. difficult to get conformational ensemble from experimental data
    3. Experimental data may be insufficient accuracy used to (in)validate simulation predictions.

5.1. Averaging Limitation

- Not a problem just to NMR experiments, but also to CD, NOE,

6. Perspectives in Biomolecular Modeling

- Driving Forces of Biomolecular Modeling
    1. computing power is steadily and rapidly increasing by a factor of 10 in every 5 years
        a. advent of parallel computing
        b. most time-consuming calculating is force or interaction calculation, needing parallel computing.
    2. advancement of modeling techniques
        a. computing long-range electrostatic forces
        b. methods for extended and enhanced sampling
        c. refined force fields.

How can biomolecular models be extended, improved, or simplied?

- It requires the inclusion of electronic degrees of freedoms (QM/MM)
    - for appropriately describing enzyme reactions
    - To simulate proton-transfer reaction
- In terms of classical FF,
    - improvements will come from introduction of polarizability in biomolecular FF.
        - useful for co-solvent effects through explicit simulation
        - extend sampling power of simulation
- Simplified, in terms of coarse grained models
    - averaging over atomic degrees of freedom will allow for simulation of slower processes

> 🧠 Why simulation and modeling?
> 1. To provide a microscopic picture of unrivaled resolution in time, space, and energy that compliments a limited set of properties to be observed experimentally.
> 2. System parameters can be changed easily to study cause-effect relationships → enhanced understanding of biomolecular systems.

The choices made about the biomolecular model needs to consider the criteria

1. properties interested / size of configurational space / estimated time scale needed to be searched
2. required accuracy of the properties
3. available computing power

Choice made:

- molecular model
- force field
- sample / time scale