From Small Molecules to Biological Molecules: Modelling Interactions

Dr. Antonio Chana
Milano, 22\textsuperscript{nd} April 2008
Computational Methods & Applicability

Simulation Methods for Soft Materials

- Electronic Structure - MO & DFT
- Ab initio MD
- Quantum MC
- Ab Initio
- Angstroms
- Molecular dynamics - Monte Carlo
- Molecular Simulation
- Mesoscale Simulation
- Macroscale Simulation
- CFD Mech
- Finite element
- Brownian Dynamics
- Lattice Boltzmann
- Cellular Automata
- DPD
- DDFT

*Diagram taken from S. C. Glotzer The University of Michigan.
Molecular Mechanics

Based on Newtonian Physics \((F = ma)\)

atoms = charged spheres (no electrons)
bonds = springs (no orbitals)

math of spring deformation:

- Bond Stretching
- Bending
- Twisting

\[
\text{Energy} = E_{\text{covalent}} + E_{\text{non covalent}}
\]

\[
E_{\text{covalent}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}
\]

\[
E_{\text{non covalent}} = E_{\text{van der Waals}} + E_{\text{electrostatic}}
\]
**Comp. Methods for Large Molecular Systems**

**Molecular Dynamics**

Application of statistical mechanics in order to include a new variable to the system: **TIME**

Depending on the conditions of the simulation we can find:
- Microcanonical ensemble (NVE, adiabatic)
- Canonical ensemble (NVT, thermostat)
- Isothermal-Isochoric ensemble (NPT)

It also uses Force Fields, including parameter sets:

- Atomic Mass
- van der Waals Radius
- Partial Charge
- Bond Length
- Equilibrium
- Spring Constant
- Bond angle
- Equilibrium
- Spring Constant
- Dihedral angle
- Equilibrium
- Spring Constant
Comp. Methods for Large Molecular Systems

MM or MD allows calculations over huge systems like in the examples

Pore formation
E. g.: PPC lipid bilayer

Nº Heavy atoms: 35,000
Simulation time: 1 ns
Total Real time of the simulation: 12 days

Coarse grain dynamics
Simulation time: 20 ns
Total Real time: 17 h

Diffusion and selectivity
E.g.: KcsA Channel
Handling Reversible Molecular Interactions with Biological Molecules
### Drug Design/Toxicity

<table>
<thead>
<tr>
<th>3D Structure</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Receptor</strong></td>
<td>Unknown, Known</td>
</tr>
<tr>
<td>Unknown</td>
<td>Combinatorial Chemistry</td>
</tr>
<tr>
<td></td>
<td>High-throughput screening</td>
</tr>
<tr>
<td></td>
<td>QSAR Pharmacophore identification on databases</td>
</tr>
<tr>
<td>Known</td>
<td><em>De novo</em> receptor-based structural search</td>
</tr>
<tr>
<td></td>
<td>Docking Virtual Screening</td>
</tr>
</tbody>
</table>

**Docking:**
Prediction of the interaction modes between two molecules

**Virtual Screening:**
Computational method consisting in searching the more promising compounds (vs. a biological target) within a database
Drug Discovery

Identification of Therapeutic Targets

Target’s 3D structure (free & ligand bound)
Crystallography, NMR, Homology Modeling

Chemical Databases

Virtual Screening (docking)

New molecules Design

Synthesis

Evaluation

Pre-clinical Studies

Drug

Candidate Optimization (Affinity, Toxicity, Especificity)

Clinical trials

Pre-clinical Studies

Drug
Docking: Software & References

...Dock AutoDock FlexE Grid FTDock EUDock Darwin Escher Flog ZDock Dot ConsDock FlexX DrugScore Bleep Rapid Monty CDock Hammerhead BiGGer SPDock SGDock Puzzle AGDock Gold Dockit QSDock CMIP PMF ChemScore Isostar Superstar PLP 3D-Dock Hex Gramm Ppd Merl MCSS Ludi Clix Adam EPDock QXP ProDock Pro_Leads Specitope CombiDock SanDock Glide ICM...

Citations per year  (Web of Science)
**Molecular Association: models**

\[
\Delta G_{\text{Binding}} = \Delta H - T\Delta S
\]

\[
K_a = \exp\left(-\frac{\Delta G_{\text{binding}}}{RT}\right)
\]

**Fischer (1894)**

Lock & Key model

**Koshland (1958)**

Ligand-Induced coupling model

Shape complementarity

Ligand-receptor adaptation
**Docking: concepts**

- Poses generation within the binding pocket
- Accurate Sampling Methods
  - Receptor Flexibility
  - Ligand Flexibility

**Quantitative interaction characterization**

\[
\Delta G_{bind} = \Delta G_{\text{int}} + \Delta G_{\text{solv}}^{R-L} - \Delta G_{\text{solv}}^R - \Delta G_{\text{solv}}^L = \Delta G_{\text{int}} + \Delta G_{\text{desol}} - T\Delta S_{\text{solv}} - T\Delta S_{\text{conf}}
\]

\[
E_{MM} = \sum_{i}^{\text{lig}} \sum_{j}^{\text{prot}} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{\varepsilon r_{ij}} \right]
\]

\[
\Delta G_{\text{desol}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} + \Delta G_{\text{elec}}
\]

\[
\Delta G_{\text{cav}} + \Delta G_{\text{vdW}} = a + \sum_{i}^{\text{atoms}} b \times \Delta(SASA_i)
\]
Docking: Structural Searches

- **Systematic**
  - Exhaustive variations
  - Over every degree of freedom
  - Combinatorial explosions
    - \[ N_{conf} = \prod_i \prod_j \frac{360}{\theta_{ij}} \]

- **Stochastic**
  - The starting state determines
  - The evolution of the system
  - Random changes over
  - The degrees of freedom

- **Deterministic**
Docking: Atomic mobility

Ligand effect upon binding pocket atoms

Euclidean coordinates define the position where exists the higher probability to find the atoms; atomic factors (B) define the extension of the atomic vibration as regards the equilibrium

\[ B = 8\pi^2 U^2 \quad \Rightarrow \quad U^2 \quad \Rightarrow \quad \text{(Mean quadratic fluctuations)} \]

\[ \Delta B = B_{bind} - B_{free} \]

< 0 movility lost

> 0 movility increase

≈ 70% of the atoms DB < 0

≈ 30% of the atoms DB > 0

Docking: Effect of the Ligands on the Receptor

Teague SJ., Nature Rev. Drug Discov. 2003, 2, 527-541
Docking: Effect of the Ligands on the Receptor

Teague SJ., Nature Rev. Drug Discov. 2003, 2, 527-541
Docking: Steric Contribution to Energy

Molecular Interactions: Lennard-Jones potential

\[ V(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right] \]

- Repulsion term
  - low \( r \)
  - Pauli Principle
- Attraction Term
  - High \( r \)
  - London dispersion forces
**Docking: Electrostatic contribution to Energy**

Molecular Interactions: Coulomb’s potential

\[ E_{\text{Coul}} (r) = \sum \sum \frac{q_i q_j}{4 \pi \varepsilon_0 r_{ij}} \]

Electrostatic Interaction

- Attraction/Repulsion
- Hydrogen Bond
- Solvation/Desolvation
- Hydrophobic effect
**Docking: Hydrogen Bonding**

Molecular Interactions: hydrogen bonding classes

Electrostatic interaction between one H atom bonded to an electronegative atom (N,O) and an additional electronegative atom or a π system.

![Chemical structures and interaction diagrams]
Molecular interactions: Hydrogen bonding parameters

<table>
<thead>
<tr>
<th></th>
<th>Main chain-Ligand</th>
<th>Side chain-Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-H---O</td>
<td>2.10</td>
<td>2.14</td>
</tr>
<tr>
<td>O-H---O</td>
<td>2.12</td>
<td>2.03</td>
</tr>
<tr>
<td>C-H---O</td>
<td>2.50</td>
<td>2.51</td>
</tr>
</tbody>
</table>

Sarkhel S. et al, Proteins 2004, 54, 247-259
Docking: Hydrogen Bonding Geometries

Molecular Interactions: Hydrogen Bonding geometries, relative strength and functions

\[
\begin{align*}
X = N, Y = C, N, O & \quad X = N, Y = C, N, O
\end{align*}
\]

Sarkhel S. et al, Proteins 2004, 54, 247-259

\[
\Delta G_{hbond} = \sum_{i,j} E(\Theta) \left[ \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]
\]

Autodock

GOLD
Molecular Interactions: solvation/desolvation

**Linear Response Theory**

\[ \Delta G_{elec} = \frac{1}{2} \left( \sum_{i,j} \frac{q_i q_j}{r_{ij}} \right) \]

**Poisson-Boltzmann Model**

\[ \Delta G_{elec} = \frac{1}{2} \int \rho(r) \phi(r) dv \]

\[ \nabla [\varepsilon(r) \nabla \phi(r)] = -4\pi \rho(r) \]

**Generalised Born Model**

\[ \Delta G_{elec} = -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{ij}} \]

\[ f_{ij} = \left( r_{ij}^2 + \alpha_i \alpha_j e^{-D_{ij}} \right)^{\frac{1}{2}} \]
Docking: Hydrophobic Interaction

Molecular Interactions: Hydrophobic interaction

Molecular Interactions: Other interactions

- Covalent ligand-receptor binding: irreversible inhibitors
- Water mediated Ligand-receptor interaction
- Metal proteins
Docking: Protocol

1. Binding pocket characterization

Receptor experimental 3D structure with Ligand
- Presence of cofactors
- Presence of $\text{H}_2\text{O}$ molecules
- Presence of metallic ions

Receptor experimental 3D structures without ligand
- Blind docking
- Cavities active site-like search

Without experimental 3D structure
- Homology modeling
Docking: Potential Maps

2. Interaction potentials calculation
Docking: Steric Potential Map
Docking: Electrostatic Potential Map
Docking: Poses Generation

3. Poses generation

- **N** grid points
- **M** orientations
- **P** conformations

25 Å → 0.5 30º → 172 x 10⁶ evaluations/conformer
4. How to classify the different solutions?

Scoring functions

    mathematical functions designed for the prediction of the biological activity by
    means the estimation of the ligand-receptor interaction energy

Three kinds of scoring functions

a) Based on Molecular Mechanics Force Fields

\[
E_{MM} = \sum_{i}^{lig} \sum_{j}^{prot} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + 332 \frac{q_i q_j}{\varepsilon r_{ij}} \right]
\]

b) Based on empirical data (affinity, physicochemical properties)

c) based on mean force potentials

Coordinates Mean Square Deviation

\[
RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2} \quad \text{(d atom pairs distance)}
\]
Virtual Screening

In vivo screening

In vitro screening

In silico screening

Chemical Databases

Bayer

Genomics

Proteomics

Molecular Interactions

Chemistry

Biopolymer Synthesis

Natural Product Synthesis

Chemical Databases

CrossFire
Virtual Screening: Data formats & Prefiltering

1D Formats:
- CAS Number
- SMILES
- InChI
- Other register formats

2D Formats:
- sd Files
- CF files
- Concord
- Connectivity Matrices

Conversion algorithms:
- Corina
- LigPrep
- Dundee PRODRG2 Server
- ZINC
- ...

3D Formats:
- PDB
- mol2
- XYZ
Virtual Screening: Building a Positive Control

Re-Dock the ligand manually
Accurate results?

No:
Search for a new set of Parameter/docking algorithms suitable for the problem

Yes:

Predicted Binding Free Energy (BFE)

BFE + SD = Threshold for new compounds

Run the docking for all the library
Running the docking: Virtual Screening

Chemical Database

1D 2D To 3D

Corina LigPrep ...

1D O=C1N([H])c2cccccc2C(c3cccccc3)=NC1
2D

Place chemical within docking area (receptor)

Dock

Autodock Glide DOCK 6 ...

Positive control: BFE ±SD

BFE PC >

Hits!!

<

No Hits
Virtual Screening Considerations

Search space

The searching spaces of all molecules must be comparable
Same active site definition
Same input for all the programs
Same searching parameters for all the compounds

Classification of solutions

Local optimization before calculating final BFE score

Other Factors

Reproducibility (different software versions)
Try different scoring functions (if available)
Statistical significance of results.
Docking: Conclusions

In general, docking software is able to generate different poses where it is possible to find the actual binding mode.

Docking with a single receptor structure and several families of compounds does not represent a problem for finding the native mode of interaction.

It is difficult to find results over 35% of native modes of interaction below 2 Å in all the cases. Such a problem is still to be resolved.

It is not common to find a statistical significance between scoring the results and affinity constants.

Does it means docking do not work?

It is not predictive

Give hints about the interaction and it is able to identify actual drug candidates.
Handling Irreversible Molecular Interactions with Biological Molecules
**Accurate Models for Studying Biological Systems**

Quantum Mechanics/Molecular Mechanics hybrid methods

**Partitioning** the system into

1. **chemical active** part treated by QM methods
2. Interface region
3. **large environment** that is modeled by a classical force field

\[ E^{\text{tot}} = \langle \Psi | \hat{H}^{\text{QM}} + \hat{H}_{el}^{\text{QM/MM}} \rangle + E^{\text{QM/MM}}_{\text{van}} + E^{\text{MM}} \]

**Active center**

The region of interest is treated by QM
Ab Initio
DFT
Semiempirical

\[ \hat{H}\Psi = E\Psi \]

We can vary geometries obtaining an energetic profile that will tell us under which conditions the reaction can happen.
The boundary term $\hat{H}_{QM/MM}$ is given by

$$\hat{H}_{QM/MM} = -\sum_{e,i} \frac{Q_i}{r_{e,i}} + \sum_{m,i} \frac{Z_m Q_i}{r_{m,i}} + \hat{H}_{vdw}$$

where $i$ is summed over all MM partial charges, $m$ over all QM nuclei, and $e$ over all QM electrons.

First term: 1-electron interaction between QM electron density and MM partial charges.

Second term: standard Coulomb interaction between QM nuclei and MM charges.
The final term is required because electron density (and hence dispersion) is explicitly treated in the QM region, but not in the MM region.

The HF energy of interaction between the QM system and a single MM partial charge $Q_i$ is then given by

$$E_{QM/MM}^i = \left\langle \Psi \left| \hat{H}_{QM/MM}^i \right| \Psi \right\rangle = \sum_\mu \sum_\nu P_{\mu\nu} I_{\mu\nu}^i + \sum_m \frac{Z_m Q_i}{r_{m,i}} E_{vdw}^i$$

where $I_{\mu\nu}^i$ is a one-electron integral, and $P_{\mu\nu}$ a density matrix element. Note that only one-electron terms are required.

How do we deal with bonds between the QM and MM regions?

- The valence of the QM region must be satisfied.
- MM bond, angle, dihedral terms need a partner atom to act on, in order to maintain the geometry of the system.

QM/MM is often used to simulate a solute quantum mechanically, with explicit solvent treated with MM — in this instance, the problem of QM-MM bonds is avoided.
‘link atoms’ (usually hydrogen atoms, but sometimes halogens or even methyl groups) are added along the bond $a$

- The link atom satisfies the valence of the QM region
- The QM atom is used for calculation of all MM bond terms
- For nonbond (electrostatic terms), originally the link atom did not interact with any MM atom (termed a ‘QQ’ link in CHARMM parlance)
- Better properties are usually obtained if the link atom interacts with the entire MM region (‘HQ’ link)
- Poor handling of electron density
Improved bond treatments

Local Self-Consistent Field (LSCF)\textsuperscript{a} uses a parameterized frozen orbital along the QM-MM bond, which is not optimized in the SCF.

Generalized Hybrid Orbital (GHO)\textsuperscript{b} includes the QM-MM orbitals in the SCF.

\textsuperscript{a}Warshel, A. and Levitt, M. J. Mol. Biol. 1976, 103, 227
\textsuperscript{b}Gao, J. et. al. J. Phys. Chem. A 1998, 102, 4714
Chemical reactions are often simulated by molecular dynamics, e.g. with umbrella sampling. Dynamics of a QM/MM system are almost identical to those of an MM system:

- Forces are calculated from first derivatives on each atom
- The QM nuclei are treated identically to the MM partial charges
- The system is propagated by standard Newtonian dynamics
Real Models Applied to Toxicology

Such methodology allows the calculation of the reactivity of compounds with natural enzymes.

So we are able to study the metabolic reactions that are going to happen due to a chemical substance.

We are able to study the whole system in an accurate way. Either from a chemical and structural point of view.
Some parameterization is still required for the boundary treatment
The choice of the size of the QM region is still something of an art
Although the QM region polarizes in response to the MM partial charges, the reverse is not also true (although fully polarizable QM/MM methods are being developed)
The free energy of a QM system can be determined via frequency calculation; however, this is rather inaccurate when applied to QM/MM systems (second derivatives are poorly determined, e.g. due to the harmonic approximation)
Require an extremely high computational power

**Drawbacks of QM/MM Methods**
Scoring Docking: RMSD validity

bad RMSD, bad interaction
bad RMSD, good interaction
good RMSD, good interaction
Molecules are represented as sets of spheres connected by springs.

The energy is defined as the sum of different contributions determined by the nature of the spheres and the springs.

\[ E = E_{str} + E_{bend} + E_{tor} + E_{vdW} + E_{elec} + E_{cross} \]

Each term represents the energetic cost to modify one structural characteristic as regards an equilibrium value.

Thus, the absolute value of the energy is meaningless, only increments of energy have full sense.
The $\Delta G_{\text{bind}}$ is estimated by means fitted functions with experimental parameters representing particularly relevant physicochemical properties within the Ligand-Receptor interaction.

It is an extension from the Hammet and Hansch theories
Free Energy linear relationships
QSAR

Scoring Functions: Empiric Functions

$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{h\text{-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic-\text{int}}} \sum f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipic}} A_{\text{lipic}} + \Delta G_{\text{rot}} NROT$

Hydrogen Bonds
Saline bridges
Lipophilic interactions
Ligand entropic term
Scoring Functions: Mean Force Potentials

Search (in structural databases) of the frequency of different kinds of interactions between atom pairs

\[ \Delta G_{bind} = \sum_{kl} A_{ij}(r) \]

all the interactions between atoms i and j

all the ligand-protein pairs

Mean Force Potentials

contact probability at distance r between atoms i and j

\[ A_{ij}(r) = -kT \ln \frac{g_{ij}(r)}{g(r)} \]

reference state probability (no interaction)
**Structural Searches: Monte Carlo**

Starting Coordinates ($X_o$)

Evaluation of the starting potential Energy: $E(X_o)$

Random generation of a new configuration ($X$)

Energy evaluation $E(X)$

¿$E(X) < E(X_o)$?  

NO  

ΔE Probability

NO  

Exclude configuration

Yes -> Add the configuration

yes
**Structural Searches: Molecular Dynamics**

- **Starting coordinates** $(X_o)$
- Evaluation of the starting energy: $E(X_o)$

**Starting velocities**

- $F = -\frac{dE_p}{dX}$
- $a = \frac{F}{m}$

- calculate $X$ and $V$ after $\Delta t$

- calculate $T$, is it constant?

  - yes
    - velocities scaling
      - yes
        - calculate $P$, is it constant?
          - yes
            - escalate positions
            - NO
          - NO
            - recalculate energy $E(X')$

- NO