Coarse-grained representations in biology

Natural lump sizes:
- atoms
- residues
- base pairs
- duplexes, alpha helices, beta strands
- macromolecules
- molecular assemblies
- organelles

Cells, tissues, limbs, organs, organisms, societies, ecosystems, biospheres, intergalactic federations

(picometer to micron scale) (micron to mega-light-year scale)
Coarse-grained simulation of atom-based molecular models

• Full atom force field
• coarse-grained motion permits longer integration time steps
• Increasing coarseness
  – full Cartesian
  – torsion only (internal coordinate)
  – full rigid
Molecule internal coordinates

- Bond lengths, bond angles, dihedral angles
- Mandatory for Compound construction
- Optional for Compound simulation
Internal coordinates

• Instead of explicit Cartesian X, Y, Z coordinates for an atom
• Given three other atom positions, one distance, one bond angle, and one dihedral angle
Protein Internal Coordinates
Nucleic Acid Internal Coordinates

MOBILITY OF NUCLEIC ACID BACKBONE, SUGAR RING, AND BASE
Modelling full flexibility in torsion (internal) coordinates

NUCLEOTIDE UNIT
- Phosphate
- Sugar
- Base

SUGAR RINGS
- Ribose (RNA)
- Deoxyribose (DNA)

BASES
- Purines: Adenine (A), Guanine (G)
- Pyrimidines: Cytosine (C), Thymine (T), Uracil (U)

BACKBONE
- Chain direction
- 5' to 3'

RING MOBILITY (“PUCKER”)
- 0.09 nm
- 0.7 nm

Unit i-1
- O3’
- C’
- P
- 3’
- Base

Unit i
- O3’
- C’
- P
- 3’
- Base

Unit i+1
- O3’
- C’
- P
- 3’
- Base

See other side for proteins and legend.

Michael Sherman
May, 2007
Bioengineering 215 bugs/comments to msherman@stanford.edu
Mobility Exercises

- Simulate rigid protein
- Create second protein
- Convert to full Cartesian model
- Which way is faster?
- Simulate multi-grain AdenylateMobilitiesVTK example
VMD Molecular Graphics Viewer

http://www.ks.uiuc.edu/Research/vmd/
Viewing SimTK simulation in VMD

1. Start SimTK simulation using PeriodicVMDReporter
2. Load molecule in VMD
   1. VMD Main->File->New Molecule...
   2. Molecule File Browser->Browse
   3. Molecule File Browser->Load
3. Connect VMD to SimTK
   1. VMD Main->Extensions->Simulation->IMD Connect (NAMD)
   2. In IMD Connection window
      1. Hostname: localhost
      2. Port: 3000 (match argument to PeriodicVMDReporter)
      3. Timestep transfer rate: 1 (to see every reported step) PRESS ENTER
      4. Timestep keep rate: 1 (to remember every step for replay value) PRESS ENTER
      5. Press “Connect” button
RNA Mobility Exercises

• Simulate multi-grain AdenylateMobilities example program and connect to VMD
PDB: The Protein Data Bank

- [http://www.rcsb.org/pdb/](http://www.rcsb.org/pdb/)
- Repository of DNA, RNA, and protein atomic structures
- Contains experimental results, not perfect models
- Entries identified by 4 character ID, e.g. “1MRP”, “1GRZ”
- Use “advanced search” at PDB web site to find structures.
4-character ID
PDB File format

• Text format – humans can read it too
• Most of the information is atomic coordinates
• PDB files usually contain multiple molecules, including water molecules

...  
ATOM  1751  N  GLY C 250  32.286  1.882  43.206  1.00 22.00
ATOM  1752  CA  GLY C 250  32.365  1.086  41.969  1.00 21.39
ATOM  1753  C  GLY C 250  31.538  1.735  40.864  1.00 20.79
ATOM  1754  O  GLY C 250  30.621  2.527  41.152  1.00 21.58
...
LoadPdb Exercises

• Download 1AKG from Protein Data Bank
  http://www.rcsb.org/pdb
• Simulate 1AKG using Example
• Simulate 1AKG using stream constructor
Custom Molecule Construction

- BondCenters, Atoms, Bonds, Compounds
- Inboard bond center
- The first few atoms
- Ring closing bonds
Molecule Construction nomenclature

1: Atoms(3)
2: BondCenters(9)
3: Bonds(2)
4: Compound(1)
<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Protein example</th>
<th>RNA example</th>
<th>Argon example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>Molecule or part of molecule</td>
<td>hen egg-white lysozyme</td>
<td>yeast phenylalanine transfer RNA</td>
<td>argon</td>
</tr>
<tr>
<td>Atom</td>
<td>one particular atom in a molecule</td>
<td>serine 3 nitrogen</td>
<td>guanosine 3, carbon C3’</td>
<td>argon</td>
</tr>
<tr>
<td>Biotype</td>
<td>links Atoms to ChargedAtom Types</td>
<td>serine N</td>
<td>guanylate C3’</td>
<td>argon</td>
</tr>
<tr>
<td>ChargedAtom Type</td>
<td>Atom type with a particular charge</td>
<td>serine N: charge -0.42</td>
<td>guanylate C3’: charge +0.2022</td>
<td>argon: charge zero</td>
</tr>
<tr>
<td>AtomClass</td>
<td>Atom type</td>
<td>sp² amide nitrogen (N)</td>
<td>tetrahedral carbon (CT)</td>
<td>argon</td>
</tr>
</tbody>
</table>
Biotype

- Bridge between structural model and force field implementation
- Concept borrowed from TINKER molecular dynamics package
Argon

- Inert “noble” gas
- Has no chemistry
- van der Waals forces only
Lennard-Jones potential

Four ways to express length parameter:

<table>
<thead>
<tr>
<th></th>
<th>$\sigma (r_0)$</th>
<th>$r_{min}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>radius</td>
<td>= diameter/2</td>
<td></td>
</tr>
<tr>
<td>diameter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lennard-Jones potential function:

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

$$V(r) = \epsilon \left[ \left( \frac{r_{min}}{r} \right)^{12} - 2 \left( \frac{r_{min}}{r} \right)^6 \right]$$

$r_{min} = 2^{1/6} \sigma$
Argon Exercises

• Compile and run TwoArgons example program
• Add a third argon atom.
• What happens when atoms start too close together? Too far apart?
• Increase the attractive force between atoms ten-fold
• Why does this example lack temperature and enemy minimization stanzas?
Ethane

- Two carbons, six hydrogens
- One dihedral degree of freedom
- 3 kcal/mol energy barrier to rotation
Ethane Exercises

• Compile and run TwoEthanes example program

• Add a third ethane molecule
Submitting Bug Reports