Molecular simulation and structure prediction using CHARMM and the MMTSB Tool Set

Coarse-grained Models

Charles L. Brooks III
MMTSB/CTBP
2006 Summer Workshop
Developing coarse-grained models in CHARMM

• Case studies
  – Cα-based Go models
    • Encode native interactions via Cα-Cα “contacts”, coarse-grain to level of one “bead” per aa
    • Useful as complement to protein folding mechanism studies
    • Helpful in understanding/informing single molecule “pulling” studies
    • MMTSB server available to provide “flavored” Go models for such studies
      – http://mmtsb.scripps.edu/webservices/gomodel.html
Developing coarse-grained models in CHARMM

- Case studies
  - Coarse-grained DNA models for sequence and salt effects on DNA melting
    - General coarse-graining of DNA to 3 “beads” per nucleotide (base, sugar, phosphate)
    - Developed by J. de Pablo and coworkers (Chem. Eng., U. Wisc.)
    - Helpful in understanding/informing thermodynamics of DNA melting
Essential Go Model Reductionism

From all atoms

To Cα only

All atom contacts are replaced by

Cα-Cα contacts
Essential Go Model Reductionism

• Native contact interactions encoded as
  – 1/0 (traditional Go model)
  – $\varepsilon_{ij}/0$ (scaled by empirical energy scale - flavored Go model)
  – All other pairs are repulsive

• Chain connectivity given by bonds, angles and dihedrals
  – Bonds and angle terms described by harmonic restoring forces centered at pseudo bond and pseudo angle separations from known structure
  – Torsions are treated either as
    • Simple cosine term centered at observed torsion (templated)
    • Information-based cosine series depending on pair of aa
Essential Go Model Reductionism

• Relevant references for Go-type models
  – Conventional Go models
Essential Go Model Reductionism

• Relevant references for Go-type models
  – Flavored Go models
Representing Go models in CHARMM

- Specifying topology and parameters

```plaintext
read rtf card
* Topology for Go model of 1bdc
*
20 1
MASS 1 G1       101.000000
MASS 2 G2       71.000000
MASS 3 G3       114.000000
MASS 4 G4       114.000000
MASS 5 G5       128.000000
MASS 6 G6       147.000000
MASS 7 G7       114.000000
MASS 8 G8       128.000000
MASS 9 G9       128.000000
MASS 10 G10     128.000000
MASS 11 G11     128.000000
MASS 12 G12     114.000000
MASS 13 G13     71.000000
.
.
MASS 57 G57     71.000000
MASS 58 G58     97.000000
MASS 59 G59     128.000000
MASS 60 G60     71.000000
```

Residue information

- DECL +CA
- AUTOGENERATE ANGLES
- DIHEDRAL
- RESI G1 0.0
- GROU
- Atom CA G1 0.0
- Bond CA +CA
Representing Go models in CHARMM

• Specifying topology and parameters

Bonds and angles

read param card
* Parameters for Go model of 1bdc
*

**BOND**

<table>
<thead>
<tr>
<th>G1</th>
<th>G2</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>378.000000</td>
<td>3.795046</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>378.000000</td>
<td>3.808982</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>378.000000</td>
<td>3.800045</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>378.000000</td>
<td>3.791182</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ANGLE**

<table>
<thead>
<tr>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>75.600000</td>
<td>108.672972</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75.600000</td>
<td>112.756549</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75.600000</td>
<td>124.755262</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75.600000</td>
<td>110.565786</td>
<td></td>
</tr>
</tbody>
</table>

**DIHEDRAL**

<table>
<thead>
<tr>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.382494</td>
<td>1 284.943180</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.026981</td>
<td>2 266.456266</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.017622</td>
<td>3 114.131745</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.195028</td>
<td>4 107.766228</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.434771</td>
<td>1 296.199841</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.524659</td>
<td>2 253.486984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.108980</td>
<td>3 25.409709</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.056961</td>
<td>4 96.428204</td>
</tr>
</tbody>
</table>

**NONBONDED**

<table>
<thead>
<tr>
<th>NBXMOD</th>
<th>ATOM</th>
<th>CDIEL</th>
<th>SHIFT</th>
<th>VATOM</th>
<th>VDISTANCE</th>
<th>VSWITCH</th>
<th>EPS</th>
<th>WMIN</th>
<th>CUTOFF</th>
<th>CTONNB</th>
<th>CTINBB</th>
<th>CTUNBB</th>
</tr>
</thead>
</table>

Non-specific non-bonded repulsion
Representing Go models in CHARMM

- Specifying topology and parameters
  - Residue pair specific (native contact) non-bonded parameters

<table>
<thead>
<tr>
<th>NBFIX</th>
<th>Residue 1</th>
<th>Residue 2</th>
<th>Parameter 1</th>
<th>Parameter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>G4</td>
<td>-0.043567</td>
<td>6.871368</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>G7</td>
<td>-0.043567</td>
<td>8.971603</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>G39</td>
<td>-0.047043</td>
<td>14.179823</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>G40</td>
<td>-0.046579</td>
<td>15.310104</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>G6</td>
<td>-0.080644</td>
<td>9.319967</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>G40</td>
<td>-0.037773</td>
<td>12.423546</td>
<td></td>
</tr>
</tbody>
</table>
Can we understand different mechanisms of folding in similar topologies?

Segment B1 of peptostreptococcal protein L (LB1) and segment B1 of streptococcal protein G (GB1) have very similar topologies but different folding mechanisms.

Sequence specific Go-like models yield two-state like folding for both proteins.

Similar heat capacities, cooperativity and folding free energy surfaces (versus q, fraction of native contacts)

Karanicolas & Brooks, Prot. Sci., 2002
Different sequences, analogous topology, yield different folding mechanism

Consistent with experimental findings

- For LB1 the N-terminal hairpin precedes folding of C-terminal hairpin
- In GB1 (as already seen from all-atom calculations) C-terminal hairpin forms earlier

Kinetics and mechanism of WW domains using Go-like models

- WW domains are simple β-sheet “proteins” that show a sequence dependent switch between 2-state and 3-state folding kinetics
- What is the folding mechanism?
- What is the origin of the switch?

Folding kinetics reproduce experimental observations

- PIN WW domain shows 2-state kinetics
- FBP WW domain follows 3-state kinetics
- FBP shows loop 2 folding dominates folding kinetics
- Parallel pathways for formation of loop 1 and loop 2
- Registration of loop 2 is rate determining in FBP

Karanicolas & Brooks, PNAS, 2003
Free energy landscapes indicate presence of intermediate in FBP WW domain

- Free energy landscapes calculated with detailed atomic models show intermediate “shoulder” in FBP WW domain
- Presence of meta-stable state consistent with Go model kinetics

Karanicolas & Brooks, *PNAS*, 2004
Multi-phase folding is a hallmark of functional substates - folding and function cooperate

- PIN and YAP domains bind different consensus sequences
- FBP binds two consensus sequence types

Karanicolas & Brooks, PNAS, 2004
Coarse-grained DNA model

- J de Pablo and coworkers
  - DNA reduced to three beads per nucleotide
  - Bond, angle and torsion potentials as in MM force fields
  - Non-nonded specific for specific base-stacking and pairing
  - Electrostatics via screened coulomb law

\[ u_{elec}(r) = \frac{e^{-\kappa r}}{\varepsilon r} \]
Coarse-grained DNA model

- J de Pablo and coworkers
  - Model reproduces salt-dependent DNA melting
Coarse-grained DNA model

- J de Pablo and coworkers
  - Model reproduces salt-dependent DNA melting
Coarse-grained model for virus assembly

Native and non-native associations possible

Model 1: Triangular Capsomers, $T=1_{20}$

Model 2: Quadrilateral Units, $T=1_{60}$
Probing viral assembly kinetics and thermodynamics
