Molecular Modeling 2018
Lecture 9

Automated homology modeling.
Manual re-alignment
Folding

Secondary Structure Elements (SSE): alpha helix or beta strand

Local

Secondary

Super-secondary

Tertiary

Quaternary

Initiation sites

like beta-alpha-beta units, hairpins
Super-Secondary Structure (SSS)

- SSS contains more than one SSE, interacting.
- Beta turns and helix caps are usually involved.
- Canonical SSS have names.

Coiled-coil

Helix hairpin

EF hand

alpha-alpha corner

www.cryst.bbk.ac.uk
Super-secondary structure.

"greek key"

hairpin

β

meander

β helix
Super-secondary structure.
Super-secondary structure. $\alpha\beta$

$\beta\alpha\beta$ supersecondary structure units are mostly right-handed

L-handed $\beta\alpha\beta$
1.5%

R-handed $\beta\alpha\beta$
98.5%
Theories for why βαβ units are right-handed.

Sternberg & Thornton: Twist of beta sheet makes right-handed crossover more of a straight line.
Theories for why βαβ units are right-handed.

Richardson, PNAS, 1976: Right-handed crossovers are trapped early in folding.
Theories for why $\beta\alpha\beta$ units are right-handed.

Phone Cord Effect: Northern versus Southern route to helix

Theories for why $\beta\alpha\beta$ units are right-handed.

Left-handed torque turns left-handed $\beta\alpha\beta$ to right-handed $\beta\alpha\beta$

Phone cord: Demonstrative Brownian Dynamics Simulations

http://www.youtube.com/watch?feature=player_embedded&v=6hQYjtmU6E0
$3$-helix bundles are also right-handed

$$p(R=0.50)<0.01$$
Methods for Predicting protein structure.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Algorithms used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary structure</td>
<td>stats, neural nets, HMMs,</td>
</tr>
<tr>
<td>Local structure</td>
<td>stats, HMMs</td>
</tr>
<tr>
<td>Supersecondary structure</td>
<td>MD, rules, HMMs</td>
</tr>
<tr>
<td>Inter-residue contacts</td>
<td>neural nets, rules, covariance</td>
</tr>
<tr>
<td>Tertiary structure</td>
<td>MD, homology</td>
</tr>
<tr>
<td>Sidechain conformation</td>
<td>MD, homology</td>
</tr>
<tr>
<td>Domain-domain interactions</td>
<td>MD, homology</td>
</tr>
<tr>
<td>Quaternary structure</td>
<td>MD, homology</td>
</tr>
</tbody>
</table>
Homology modeling, continued.

1. Database search

2. Alignment (automatic, manual)

3. Automated homology model script

4. Re-modeling, manual modeling, energy minimization.

5. Sidechain modeling, energy minimization.

6. etc....
Overrule automatic alignment when it...

- puts an indel in the **middle of a helix**
- puts an indel in the **middle of a strand**
- puts the anchor residues of a deletion **too far apart**
- puts an insertion at a buried position
- puts an insertion in the middle of a helix
- forces a database search of a loop of length zero.
- forces a database where the anchors are adjacent in the template.

**How to intervene:**

- Find a new spot for your indel
- Push the gap to the new spot before running (or re-running) Homology Model.
Deletion falls within a helix. What happens?

True indels within a contiguous alpha helix are rare. Alignment errors are pretty common.

Deletion in the middle of a helix

before deletion

after

chain crosses through middle of helix!?
An exception to the *no-insertions-in-helix* rule

Nature has selected an insertion of strongly anti-helix residues.
Not an exception to the *no-deletions-in-helix* rule

Deletion in the sequence leads to shorter helix, not a helix with a "deletion" in it.

This shows two actual structures. The one with the deletion is more extended, to span the distance.
Deletion falls within a strand. What happens if you accept it?

Before deletion. Residue in green will be deleted.


Move beta strand indel to turn

- Moving a deletion out of a helix or strand avoids disruption of the SSE caused by the energy minimization (which is incapable of saving a bad alignment).
- Moving a deletion move the disruption to where it can be tolerated.

**Alignment**

*Before*

XXXXXXX
YYY~YYYYYYYY

*After deletion*

after deletion

*After energy min.*

bad distorsion

alignment

XXXXXXX
YYY~YYYYYYYY

good H-bonding

alignment

XXXXXXX
YYY~YYYYYYYY
An exception to the \textit{no-insertions-in-strand} rule

\textbf{Beta-bulge}:

Two sidechains (usually polar) point to one side of the sheet, instead of just one. This causes a kink in an otherwise continuous pairing.

A beta-bulge is almost always a 1-residue insertion, here, usually a polar amino acid, most frequently D.

Occasionally, a many residue insertion occurs here. In that case it is called a \textit{beta-blowout}.
Hydrogen bonding pattern for beta bulge

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Bulge residue between two donor in a row.
Telling MOE how to anchor a loop search: the wrong way

Instead, carefully choose the “anchors” of the loop search. The anchors (or, pre-flex and post-flex) will define the loop search and will be restrained (tethered) to the template during energy minimization.
Telling MOE how to anchor a loop search: the right way

MOE will swap 2 residues into the space of 6. 2 for 6.

MOE will insert NPQ into the space of Q. 3 for 1.

The anchors for this loop search are DF and IY

The anchors for this loop search are IM and QS

Template positions to be treated as
Setup of loop search: **the wrong way**

This part of the template will be deleted. A long bond will connect the ends. Energy minimization will cause commotion.

Deleted cys is here, in the middle of a strand.
Results of Loop building: the wrong way

The resulting model has a short loop here.
Energy minimization has *pulled the sheet apart*.

No H-bonds in middle of sheet!

Deletion here results in straightening and shifting of the strand.
Setup of Loop building: the right way

Location of indel moved to here.

1 target residue unaligned, 2 template residues deleted

1 against 2

Old deletion
0 against 3

New deletions

2 against 5
2 against 3
Results of Loop building: the right way

Sheet has retained H-bonds.

New loops are at the ends of SSEs.

Take-home lesson: “Unaligning” makes loop search work better.
Demonstration of the power of re-alignment

1. Load "sequence 1"
2. Run the PDB search if necessary. Load 2IGD.
3. Align (use the defaults)
4. Your alignment should look like this:
5. Save as "before.moe"

<table>
<thead>
<tr>
<th>Tag</th>
<th>Chain</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| *   | strept 3| TKLVKGSGETKAVDVTEAEKSFQYANENEGKGYGEWSFD·TKLTFTVTAETYPKVLINGKTLKGETKAVD
| 2IGD| 2IGD.A | MTPAVTTYKLVINGKTLKGETTTEKAVDAETAEKAFQYANDNG--VDGWVTYDDATK--TFTVTE       |

6. **Protein > Homology Model**
   - Name **Current System**: automatic.moe
   - Name **Output Database**: automatic.mdb
   - Sequence: Chain #1
   - Template: Chain #2
   - Uncheck "disable C-terminal ..."
   - Set **Models**: 3
   - **Forcefield**: Amber12
   - Leave everything else as is.
   - **OK**
Demonstration of the power of re-alignment

7. Close
8. Open "before.moe"
9. Push the alignment to make it look like this

| Tag | Chain
|-----|------|
| 1:  | strept 
| 2IGD: | 2IGD.A |

-Protein > Homology Model-
Name **Current System:** realigned.moe
Name **Output Database:** realigned.mdb
Sequence: Chain #1
Template: Chain #2
Uncheck "disable C-terminal ..."
Set **Models:** 3
**Forcefield:** Amber12
Leave everything else as is.

**OK**
Demonstration of the power of re-alignment

11. Close
12. Open "automatic.mdb"
13. `Mouse-over right-mouse > send to MOE` each entry.
   When it asks whether to clear system, say No.
14. Protein | Geometry | Phi-Psi plot (Ramachandran plot)
15. **How many outliers? Where are they?**

16. **Look at the results**

17. Close
18. Open "realigned.mdb"
19. `Mouse-over right-mouse > send to MOE` each entry.
   When it asks whether to clear system, say No.
20. Protein | Geometry | Phi-Psi plot (Ramachandran plot)
21. **How many outliers? Where are they?**

22. **Look at the results**
Demonstration of the power of re-alignment

automatic alignment

Re-alignment

Lots of outliers.

a few outliers.
Exercise 6

Demonstration of the power of re-alignment

16. **Look at the results: automatic alignment**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Chain</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>1: strepto 3</td>
<td>TYKLI VKGNTFS GETTTKA VDVE TA EKS FKQY A NENG E K VYGEWSF D D KLTFTVTAE</td>
</tr>
<tr>
<td>21GD</td>
<td>2: 21GD.A</td>
<td>MTPAVTTYKL VING KTLK GETTTKAVDA ETAEKA FKQYAN DNG--VDGVW TYDDA TK--TFTVTE</td>
</tr>
</tbody>
</table>

Lots of outliers. Not coincidentally, in the indels
Demonstration of the power of re-alignment

22. Look at the results

<table>
<thead>
<tr>
<th>Tag</th>
<th>Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: strepto 3</td>
<td>TYKLIVKGN    TFSGETTTKAVDVE   TEAEKSF     KQYANENG   •• EKVYGE     WSD  ••  TKLTFTVTAE</td>
</tr>
<tr>
<td>2IGD 2: 2IGD.A</td>
<td>MTPAVTYKLVINGKTLGKETTTKAVDAETAEKAFKQYANNGVD---GVTYDDAT---KTFTVTE</td>
</tr>
</tbody>
</table>

Few outliers.  
Still in the indels
23. Find outliers in the **phi-psi plot**.
24. Select the sequence around one of the outliers.
   *Unfix* it. *Invert* selection. *Fix*.
25. `run 'gizcolorf.svl'
   Blue means little force. Red means high force, on an atom.
26. `run 'gizmin.svl'
27. Pull and push to make regular H-bonds, common secondary and local structures. Make sure it doesn't get too red!
28. Keep **Protein | Geometry | Phi-Psi plot** on. You are done when all of the outliers have moved to the allowed regions.
29. Unfix additional residues if necessary.
30. Save and upload the MOE file as "**Exercise 6**". Due Feb 17.
Review questions

• After Homology Model, are we done making a homology model?
• If you have a glycine and it is strictly conserved, where might it lie in the phi-psi (Ramachandran) plot?
• How do you as a protein modeler feel about a buried, unsatisfied, backbone hydrogen bond donor or acceptor?
• How can we guess which donor H-bonds to which acceptor?
• Are conserved sequence regions with no gaps always SCRs?
• Are conserved sequence regions with no gaps usually SCRs?
• Are side chain rotamers conserved in SCRs? (usually, never, always)