

Bioinformatics 2 -- Lecture 7

Homology modeling:

Aligning sequence to template.

Adding loops manually.

Automated loop search.

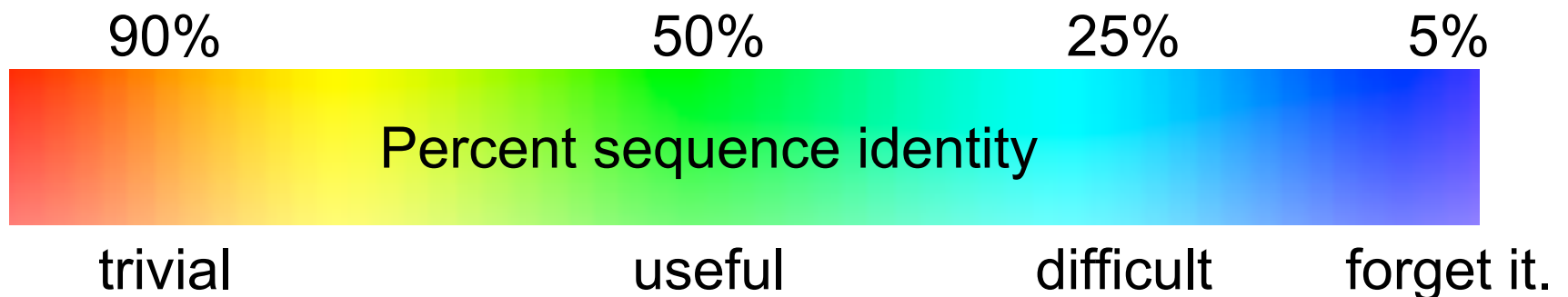
7.1 Homology modeling

definitions

Homology modeling

also called “comparative modeling”

- Sequences that have similar sequence have similar structure.
- Therefore we can model a sequence of unknown structure based on a homolog of known structure.
- Differences between homolog functions can be inferred.

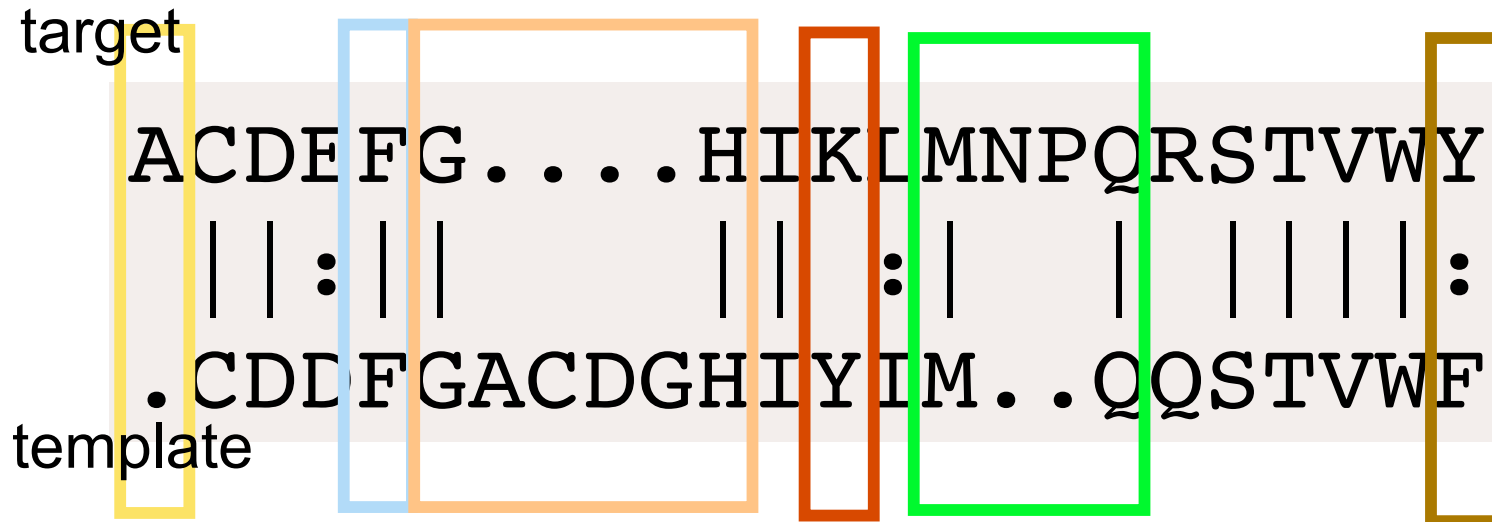


Why do homology modeling?

- Predict functional differences.
 - Predict binding sites.
 - Predict mutational effects.
 - Predict drug interactions.
 - Design all of the above.
-
- It's easier than growing crystals.
 - We have ~50K crystal structures, but millions of sequences.

What an alignment means to a biologist

Given this alignment...



Biologist infers...

- The gene was extended by one residue at the N-terminus.
- The Phe is conserved.
- Four residue deletion occurred between G to H.
- A non-similar mutation Y->K occurred.
- A two-residue insertion occurred between M and Q.
- A similar mutation F->Y occurred.

Aligned positions share a common ancestral position.

In what order do I model?

Modeling philosophy: MAXIMIZE USE OF THE INFORMATION IN THE TEMPLATE

confidence

HIGH



LOW

This means...

- Assign coordinates based on homology first.
- Identity** first, then **similarity**.
- Then make **deletions**.
- Then add **insertions**.
- Then build **extensions**.

Specifically, How?

What actions do I take to build a model?

- identity** == keep coordinates. no change.
- similarity** == mutate sidechain
- deletions** == remove residues, make new peptide bond, energy minimize.
- insertions** == predict loop conformation, position a loop. Make two new peptide bonds, energy minimize.
- extensions** == predict extension conformation, position an extension. Make new peptide bond, energy minimize.

Jargon word: SCR

- Structurally Conserved Regions (SCR) are assumed to be structurally invariant.
- SCRs should generally be 'fixed' during energy minimization. (Initially fix all atoms, and finally just fix the backbone atoms.)



From a multiple structure alignment like this you can identify SCRs versus designated/variable Loops. SCRs can include multiple SSEs and the loops between them.

Jargon word : Loop

Three types of loops:

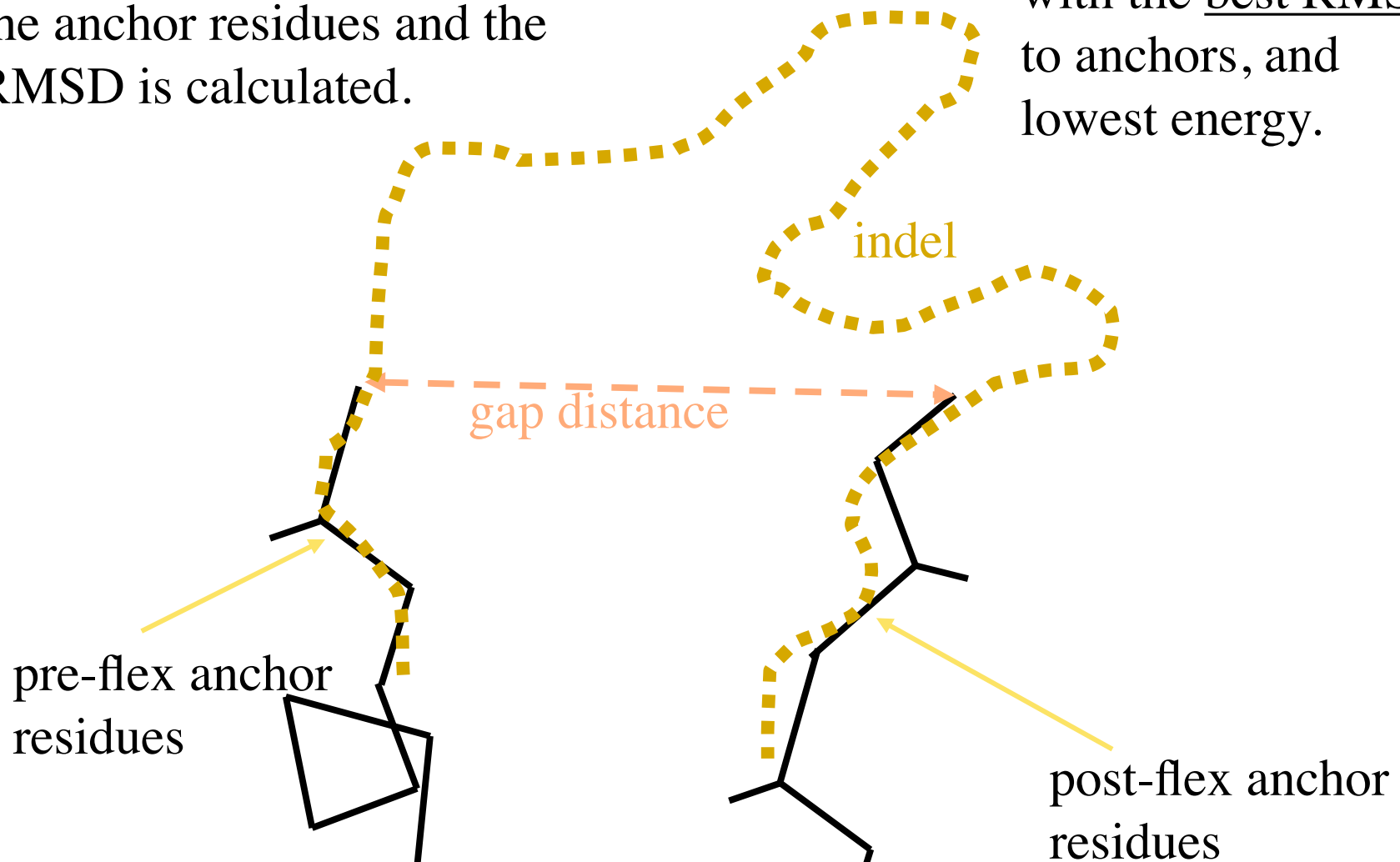
- **Designated Loop:** coordinates derived from a *template*. Not variable from model to model. But *not fixed* during energy minimization.
- **Variable Loop:** coordinates derived from a *database search* or simply constructed. May be variable from model to model. Not fixed during energy minimization.
- **Outgap:** a variable loop at the end of a chain. May be derived from a secondary structure prediction or experimental data.

7.2 Automated loop search

Automated Loop Search

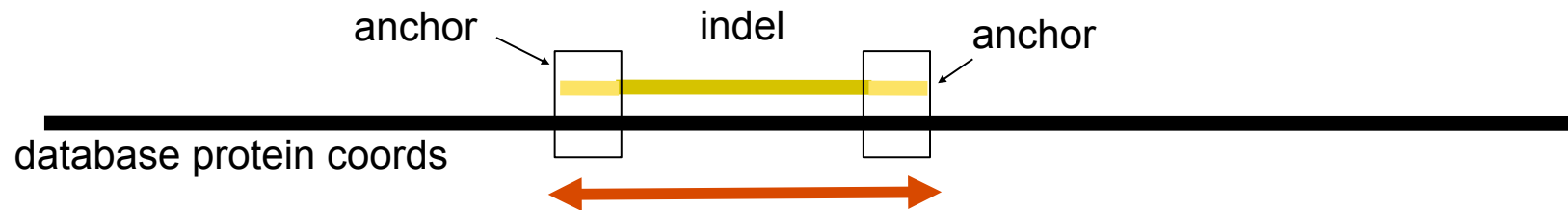
Loops of the right length in the database are superimposed on the anchor residues and the RMSD is calculated.

MOE keeps the loops with the best RMSDs to anchors, and lowest energy.

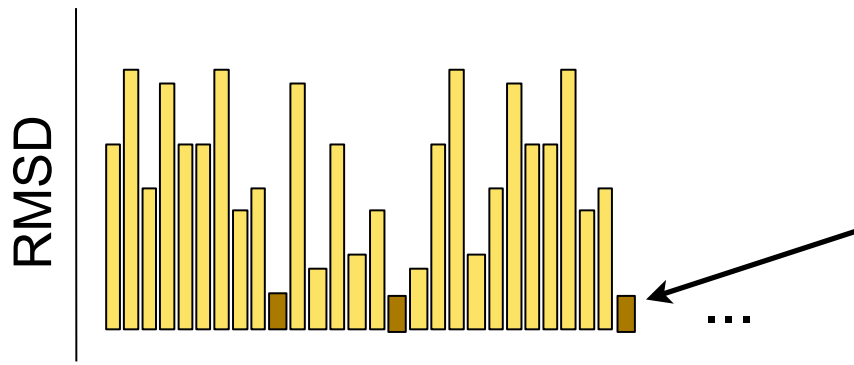


Searching the database for good loops

PDB structures are scanned with a window size = indel + prefix + postfix

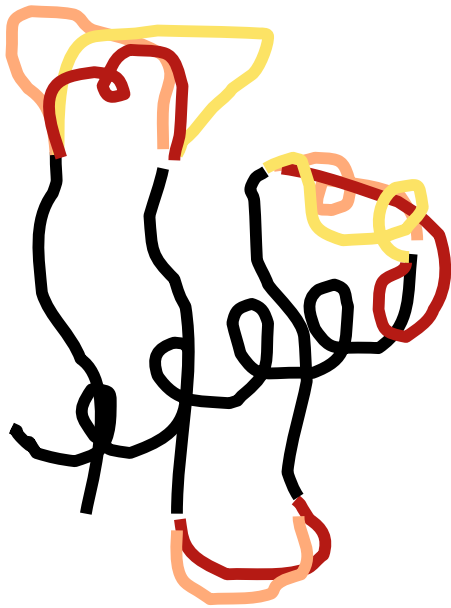


Anchor coords are superimposed on the database protein, and the RMSD is calculated. Low RMSD fragments are superimposed on the model to check the energy, after quick-n-dirty energy minimization.



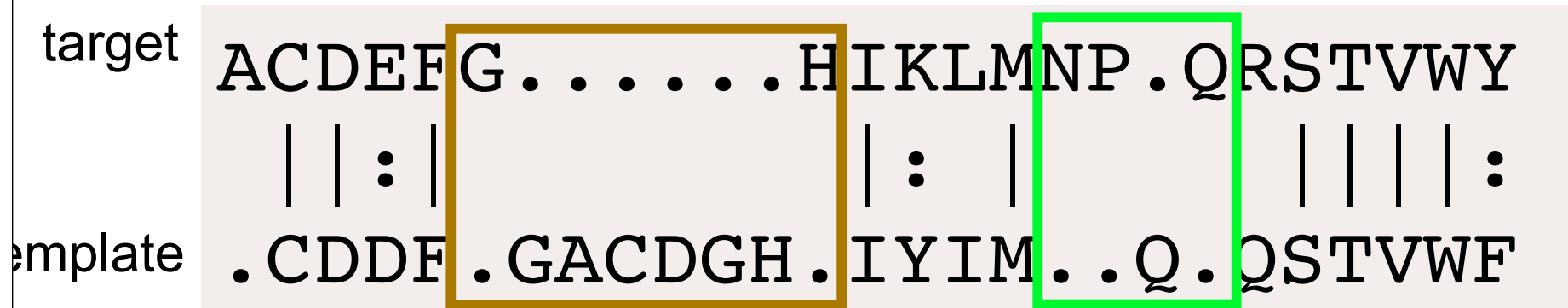
Low-RMSD, low-energy loops are saved for later use in loop-swapping step if modeling more than one loop.

Loop swapping for finding multiple compatible loops.



Each indel has a set of conforming loops (low-RMSD and low-energy). Indels are chosen in random order and loops are chosen from the loop set. When all indels have loops, the energy is (optionally minimized and then..) calculated. If it is one of the best energies in the list, the model is saved.

Telling MOE how to anchor a loop search: the right way



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

The anchors for this loop search are DF and IY

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

The anchors for this loop search are IM and QS

Overruling automated alignment

Automatic alignment might not always give you the right answer. Can we tell the program to give us the best alignment given a constraint? yes. *Alignment | constraints | constrain residue*

When to manually intervene:

When automatic alignment...

- 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

Also whenever there is experimentally known information is available, such as...

- *location of a disulfide bridge*
- *location of a metal binding site*
- *partial NMR data, or other evidence of interaction.*

In class exercise 7.2: re-aligning with constraints.

Suppose your sequence alignment places an indel in a helix. You can shift the indel out of the helix by constraining the alignment.

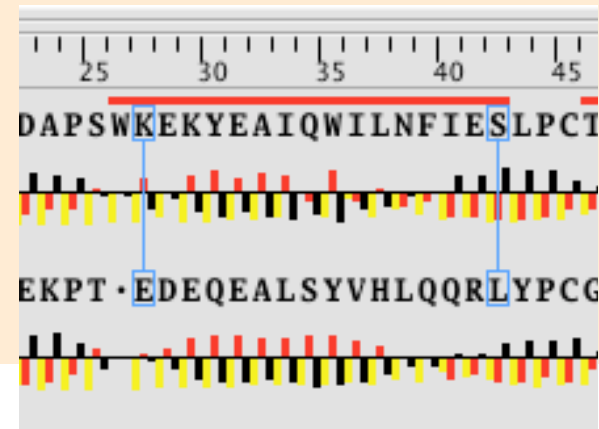
On course website, download moe file linked to “realignment exercise”

Alignment | Align (revise settings to Blosum30, gap opening penalty 3, repeat Align). Note location of deletion. Now, fix it.

Alignment | constraints | constrain residue

Select residues to align so that there are no deletions in the helix. Hit “New constraint” <esc> to exit.

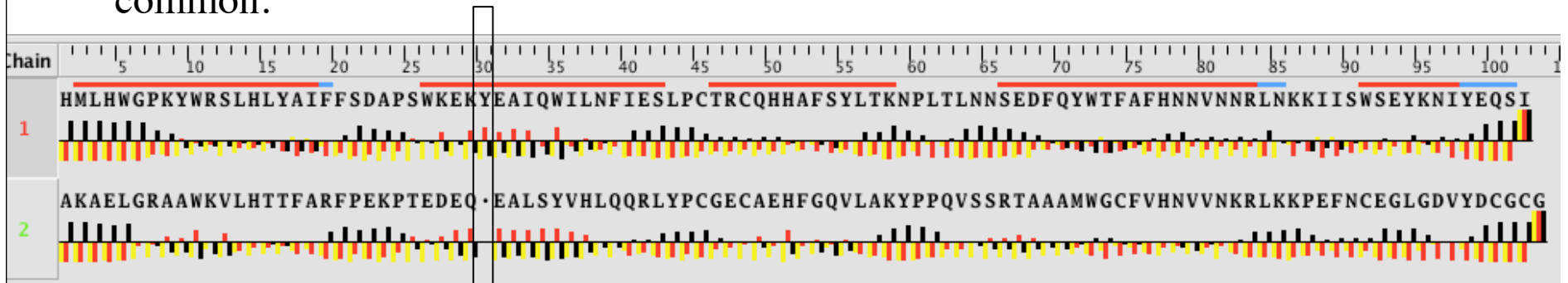
Run **Alignment | Align**



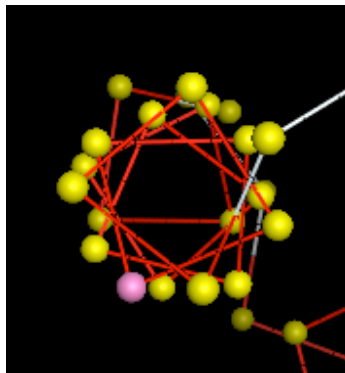
7.3 Deletions

Deletion falls within a helix. What happens if you accept it?

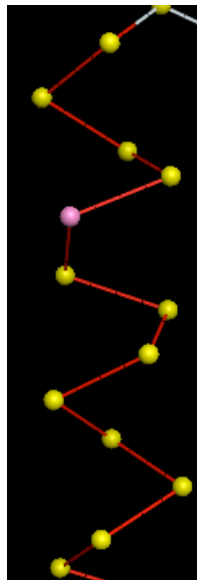
True indels within a contiguous alpha helix are rare. Alignment errors are pretty common.



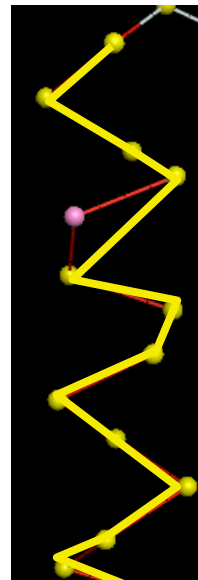
Deletion in the middle of a helix



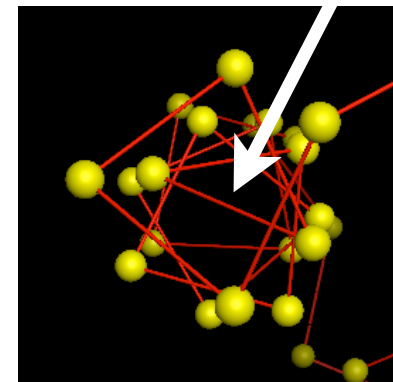
before deletion



after

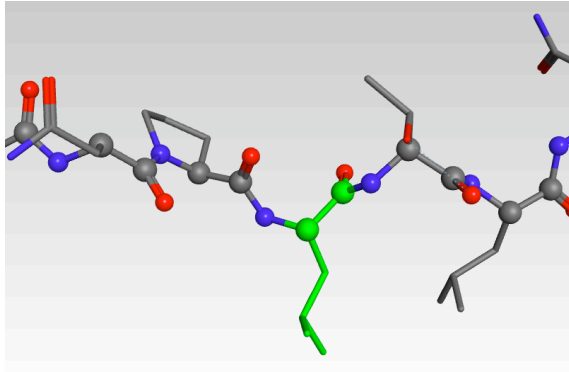


chain crosses through middle of helix!?

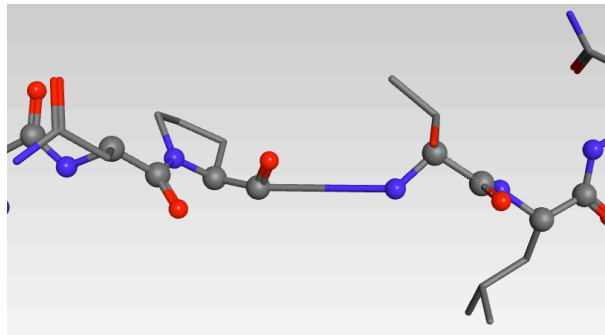


20

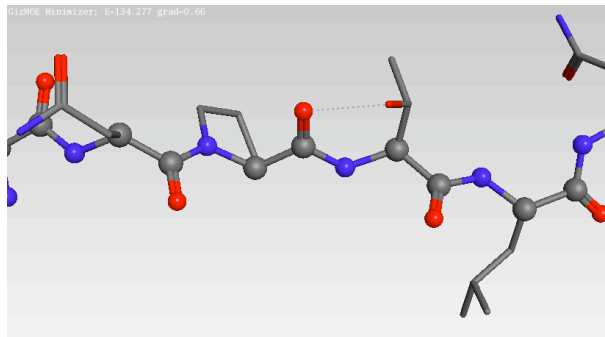
Deletion falls within a strand. What happens if you accept it?



Before deletion. Residue in green will be deleted.



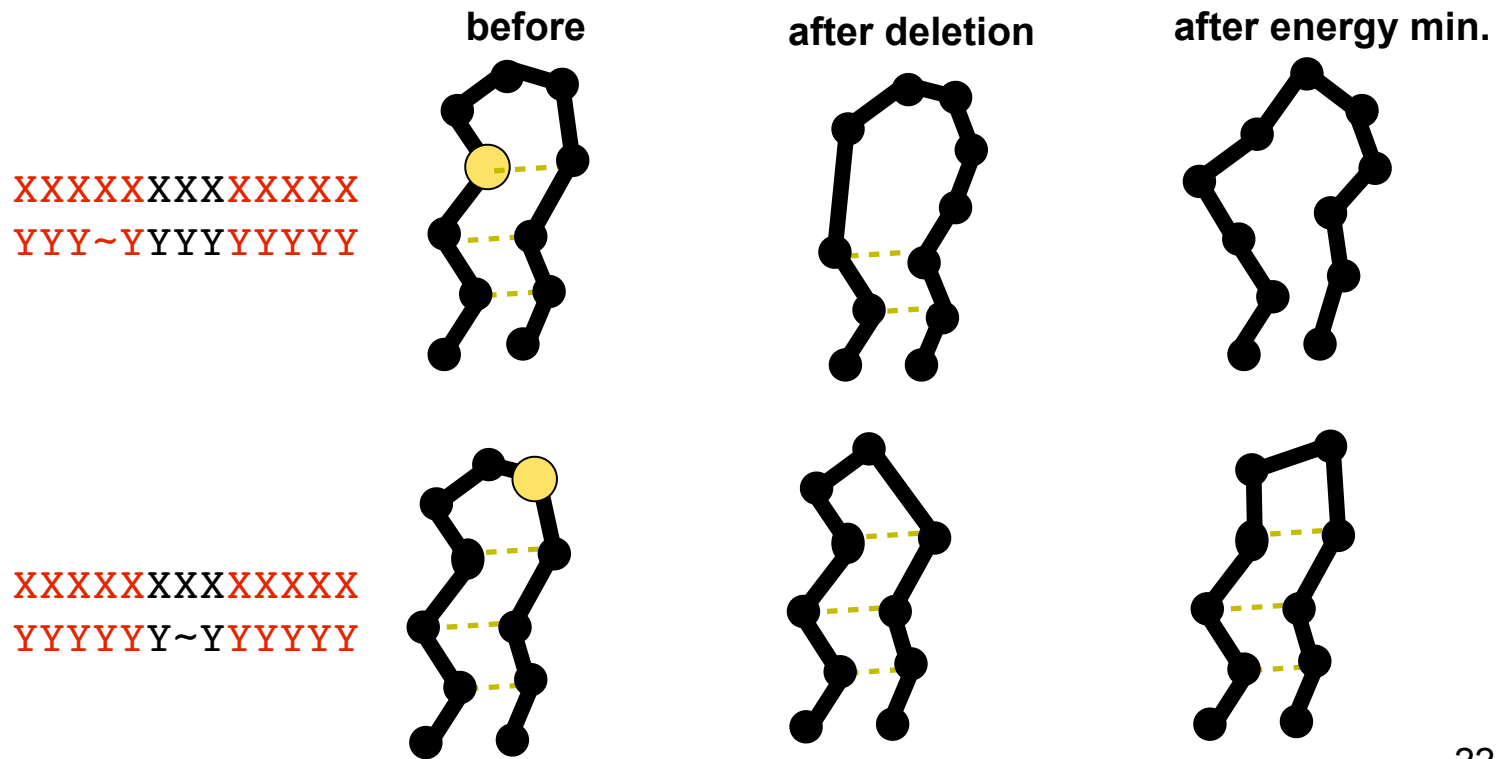
After deletion. Before energy minimization. Long peptide bond.



After energy minimization. Long peptide bond. Bad energy. Hyper-extended. H-bonds pointing the wrong way.

Moving the indels

- Moving a deletion out of a helix or strand avoids disruption of the SSE caused by the energy minimization, which is incapable of unfolding and refolding a protein.
- Moving a deletion is like unfolding a protein and allowing it to refold.



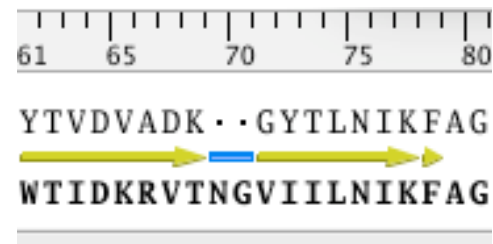
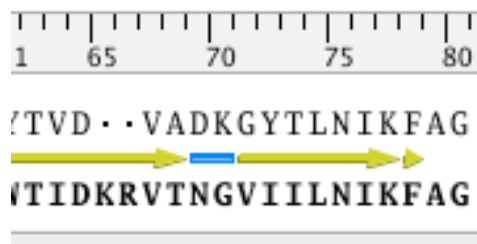
In class exercise 7.3: making a deletion.

Suppose your sequence alignment places a deletion in the middle of a beta strand. Move it to the end of the strand, then make the deletion.

On course website, download moe file linked to “deletion exercise”

Find the deletion in the SEQ window.

Put the deletion over the two turn residues by **middle-mouse dragging** over the sequence.



Delete the two template residues.
Create a new peptide bond.
Energy minimize.

Work on Homework 3

- Homology modeling by hand....
- Due Feb 17.

Also do Homology Model Tutorial before next class.

Help/Tutorials/Homology Modeling of Proteins

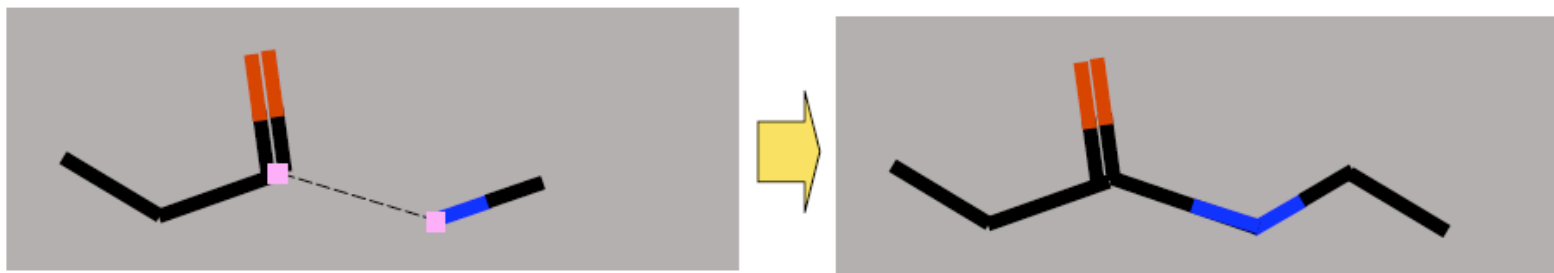
Run the tutorial *verbatim*.

Feel free to run it a second time using your own settings, exploring changes in the choice of template, alignment, energy minimization, etc.

But do it exactly as stated the first time.

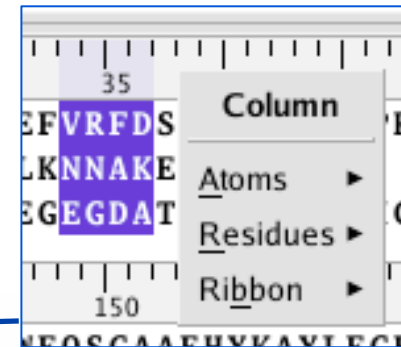
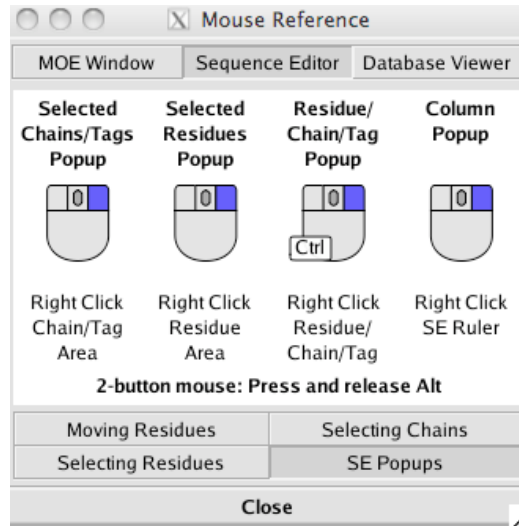
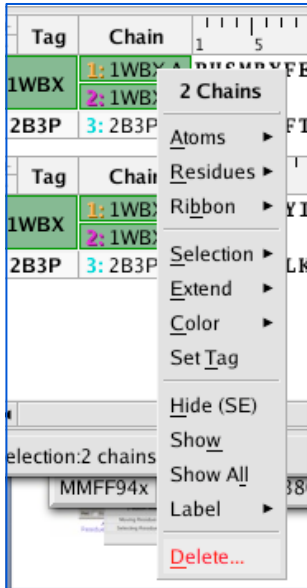
Supplementary: How to make a new peptide bond

Zoom in on splice points.



Delete extra oxygen, if present.
Select C, N. **Build | single bond**.
Select N. **Window | Atom manager** (or cntrl-a)
Select the atom. Set geometry to sp². Apply.

Supplementary slides: Pop-up menus in SEQ window. Use meta key or right-mouse.



Selected Chains
Popup

Residue Column Popup

Chain
Popup

Residue Popup

Selected Residues Popup

