

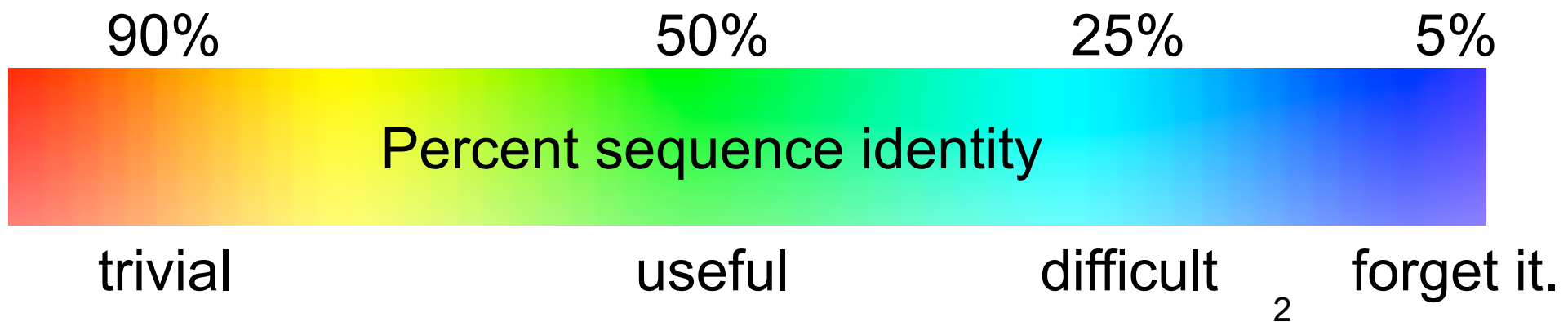
Molecular Modeling 2018-- Lecture 7

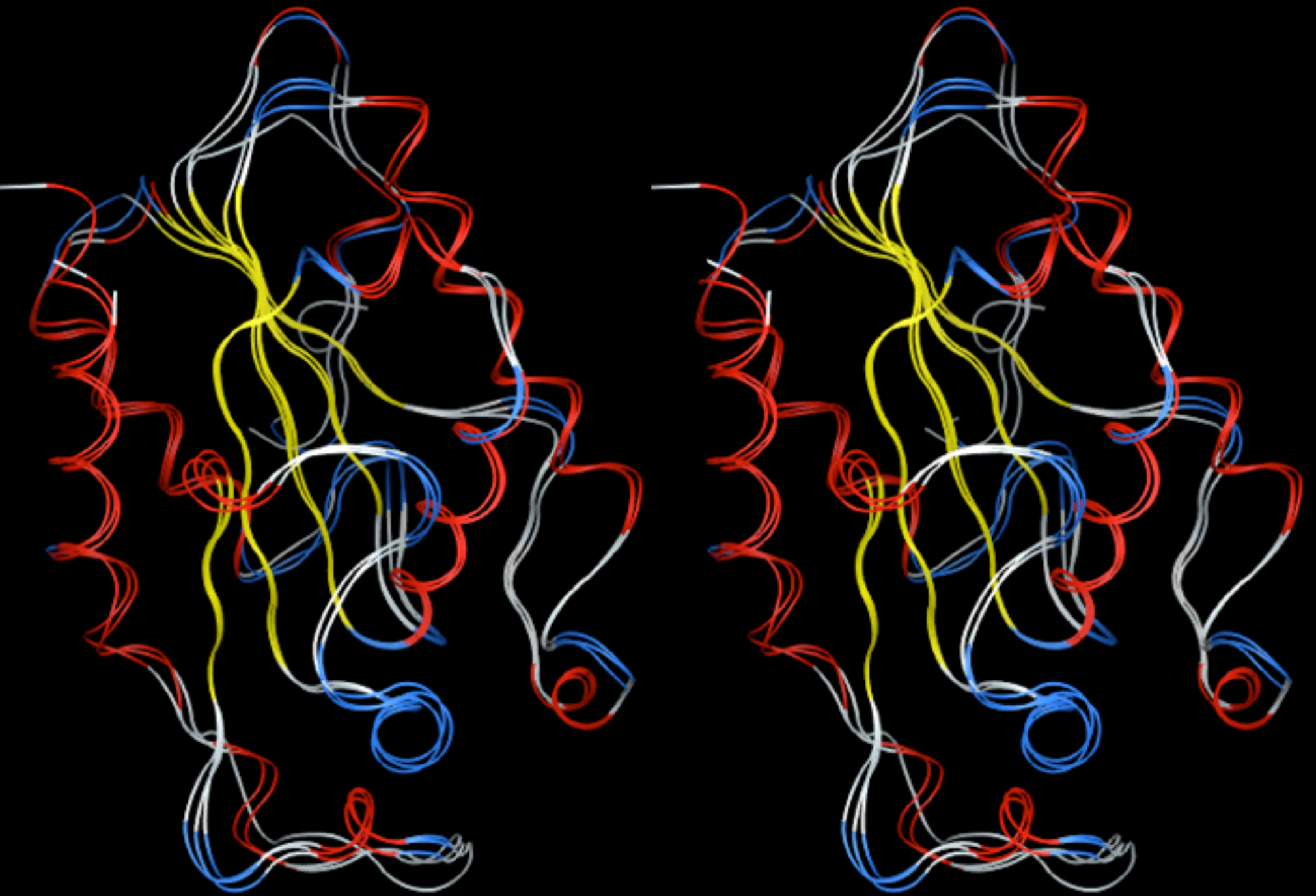
Homology modeling
insertions/deletions
manual realignment

Homology modeling

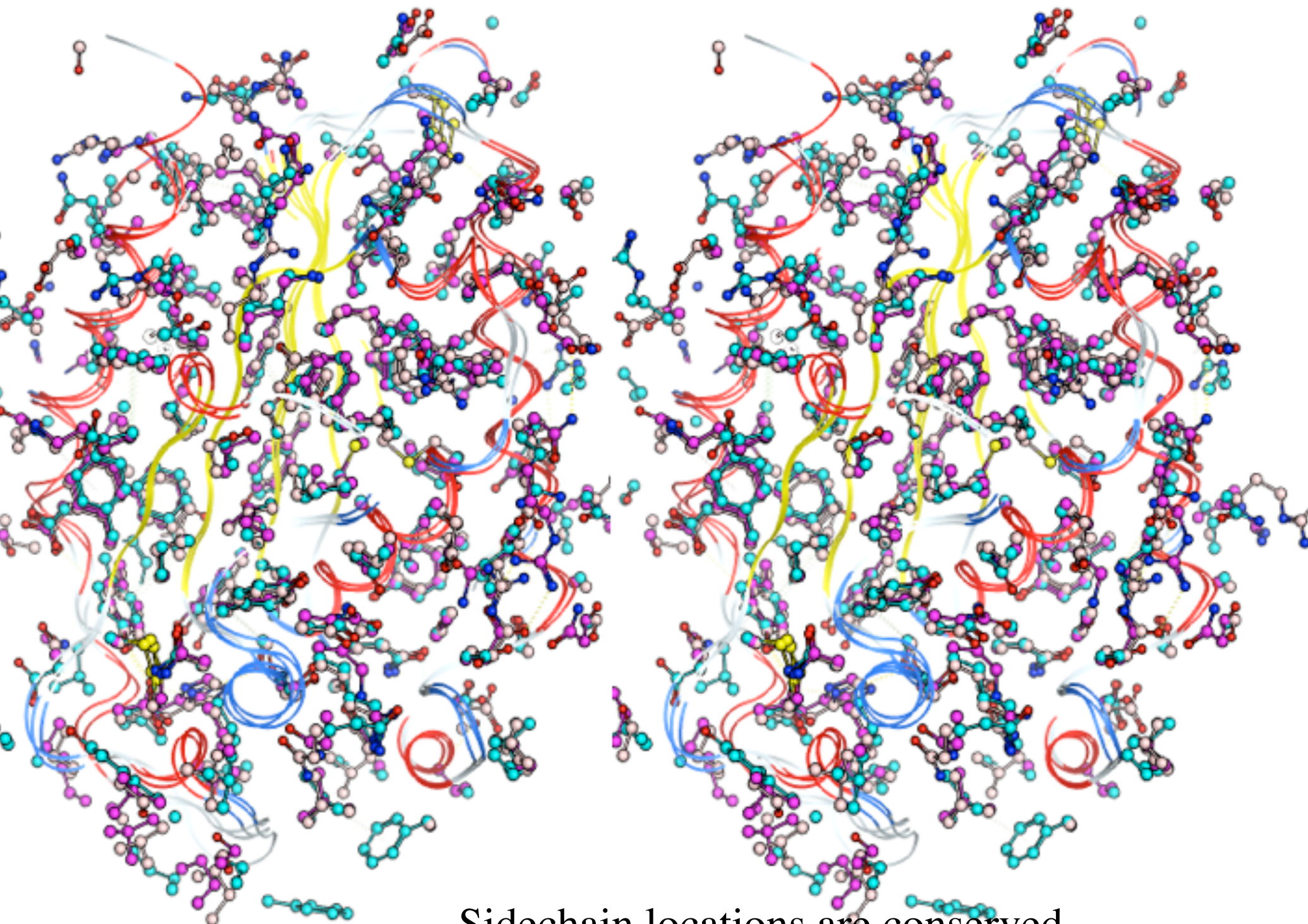
also called “comparative modeling”

- 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.

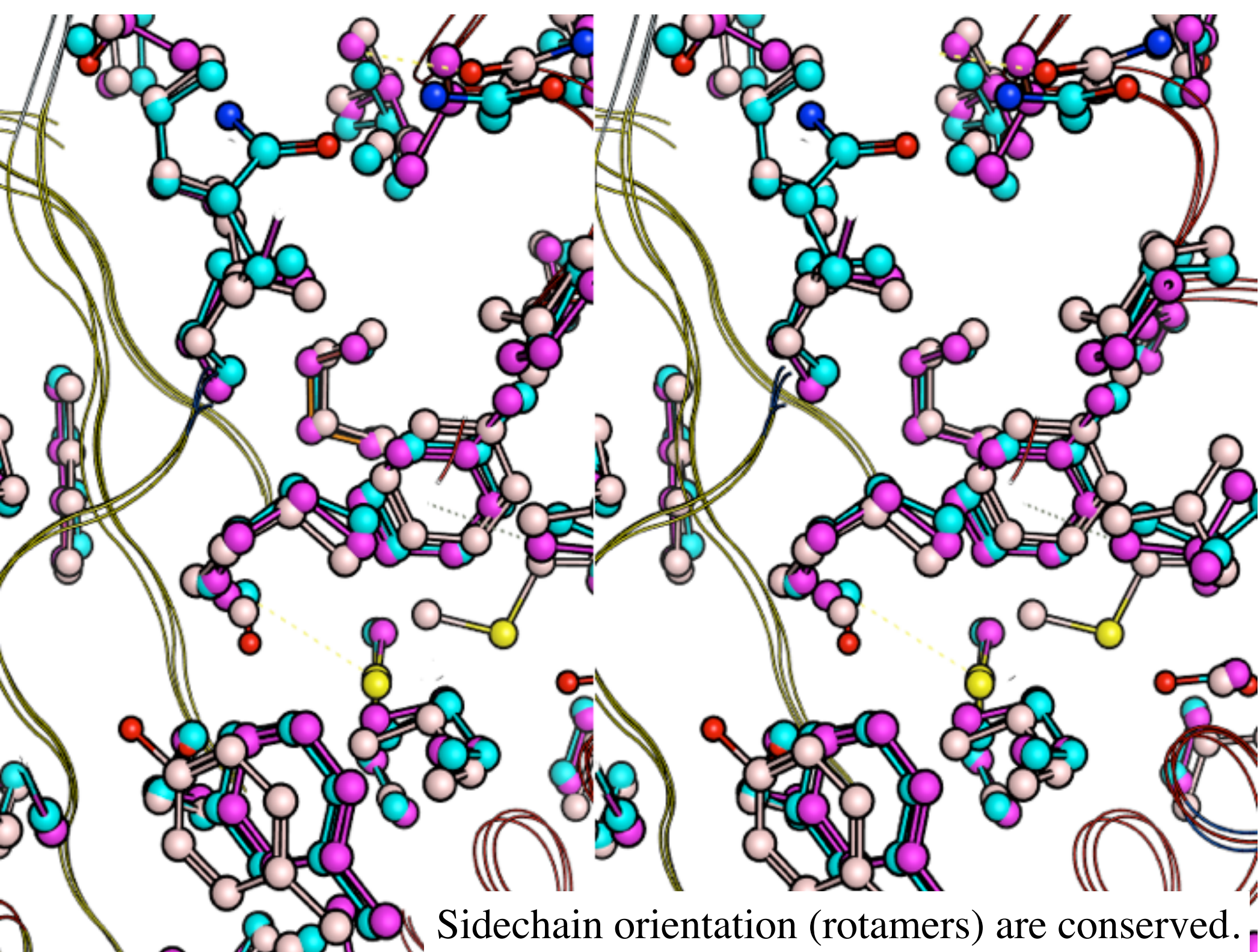




Global positioning of SSEs is conserved.



Sidechain locations are conserved.



Sidechain orientation (rotamers) are conserved.

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 ~90K crystal structures, but millions of sequences.

What an alignment means to a biologist

Given this alignment...

target

ACDEFG . . . HIKLMNPQRSTVWY

| | : | | | | : | | | | | :
| | : | | | | : | | | | | :

.CDDFGACDGHIIYIM..QQSTVWF

template

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.

What an alignment means to a modeler

Given this alignment...

target

ACDEFG . . . HIKLMNPQRSTVWY

| | : | | | | : | | | | | :

.CDDFGACDGHYIM . . QQSTVWF

template

• Modeler program should...

- Add Ala to the N-terminal Cys using energy minimization.
 - Keep the conserved Phe sidechain and backbone.
 - Cut out the four residue insertion and connect G to H.
 - Switch non-similar sidechains Y->K. Possibly move backbone.
- Possibly pick another alignment.
- Cut at M-Q, insert two residues, Asn-Pro.
 - Switch similar sidechains F->Y. Keep backbone fixed.

Aligned positions share a common spatial 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**.

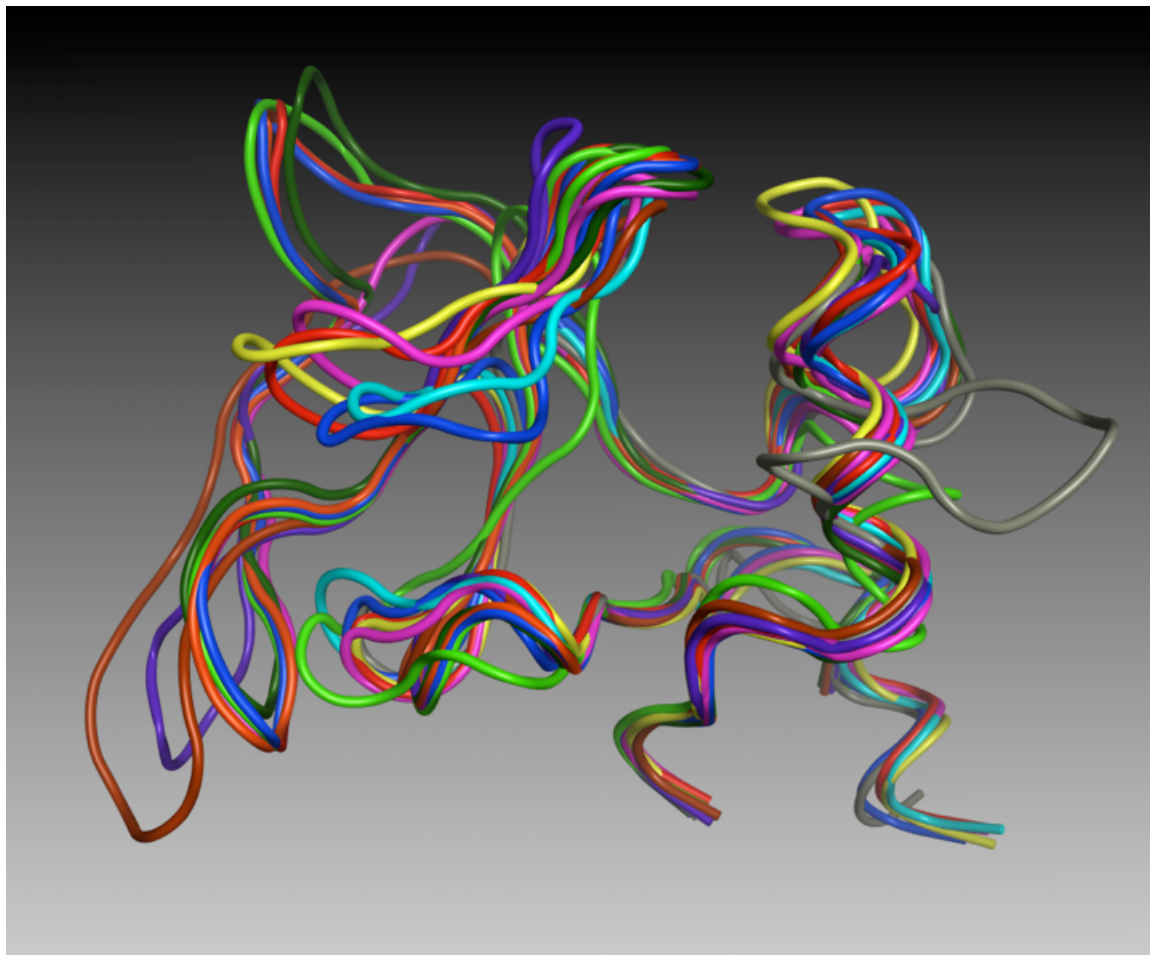
Specifically, How?

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.

Modeling jargon : Loop

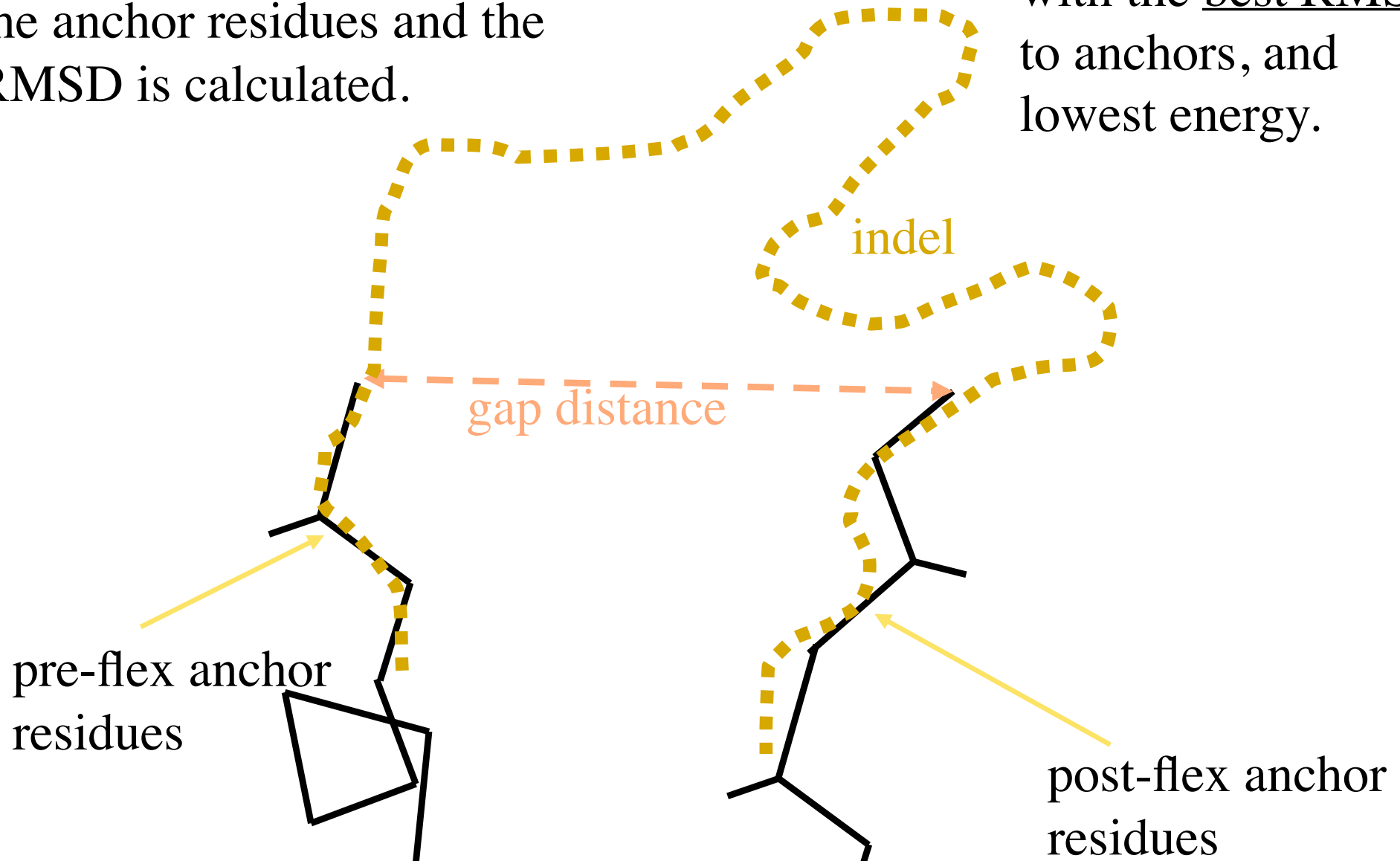
Three types of loops:

- **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.

Automated Loop Search

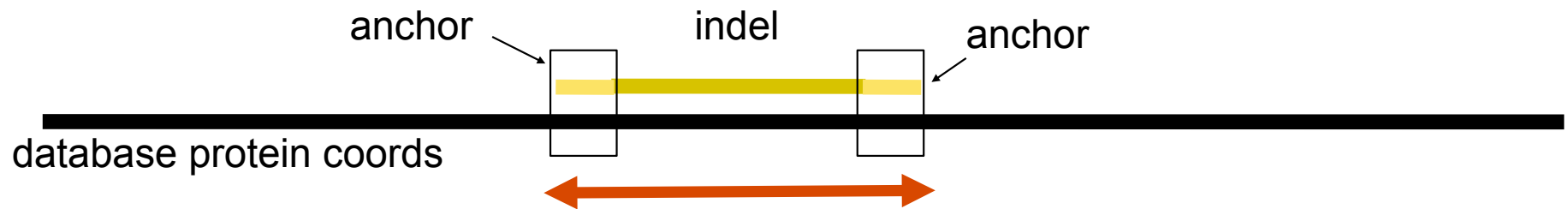
Loops of the right length in the database are superimposed on the anchor residues and the RMSD is calculated.

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

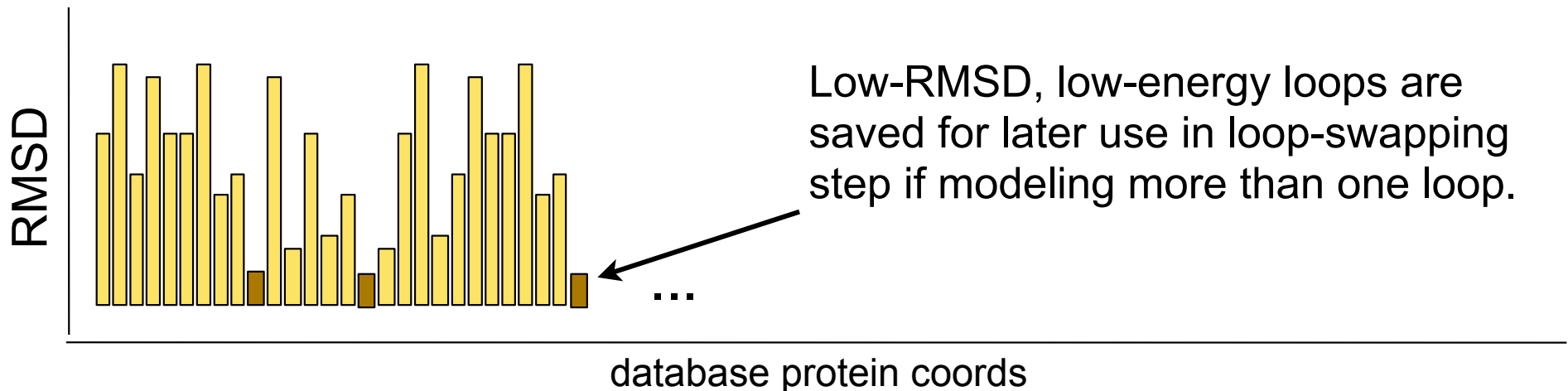


Automated loop search

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

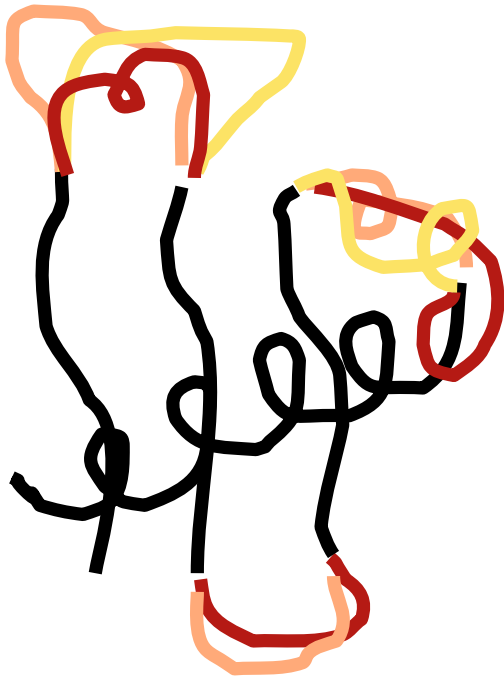


Anchor coords are superimposed on the database protein, and the RMSD is calculated. Low RMSD fragments are superimposed on the model to check the energy, after quick-n-dirty energy minimization.



Automated loop search

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.

The alignment tells MOE how to do the loop search

target

ACDEFG HIKLMNPQRSTVWY

| | : | | | | : | | | | | | :

template

.CDDFGACDGH IYIM . . Q_QSTVWF

MOE will align G to G, H to H, delete ACDG and attempt to connect G to H

MOE will align M to M and Q to Q and attempt to insert NP between M and Q

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 better loop search

target	ACDEF	G H	IKLMNP .	QRSTVWY
	:		:	:
template	. CDDE	. GACDGH .	IYIM . . Q .	QSTVWF

Align F to F, I to I, delete GACDGH and add 2-residue loop GH from a loop search.

2-for-6
instead of 0 for 4.

Align M to M, R to Q₍₂₎, delete Q and add a 3-residue loop NPQ from a loop search.

3-for-1
instead of 2 for 0.

Syntax

target

ACDEFG HIKLMNP . QRSTVWY

| | : | | : | | | | : | | | | :

template

. CDDF . GADGH . IYIM . . Q . QSTVWF


Gap: no target residues to model

Aligned: MOE will use template coordinates

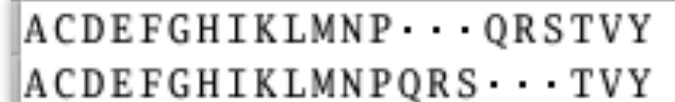
Unaligned: target residues will be modeled by a loop search

Exercise 5.1 Sequence editing practice

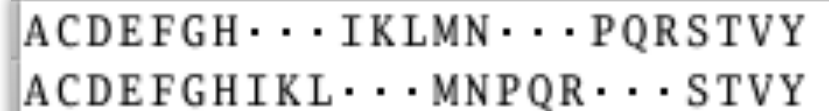
- Create a new sequence:
- SEQ: New / Text Editor
- Type ACDEFGHIKLMNPQRSTVWY
- Save: aa.seq (close Text Editor window)
- Open: aa.seq (type: raw_seq)
- Open: aa.seq (type: raw_seq)
- Try “abacus” mode: **middle-mouse**
- *Make this alignment*—->
- Use “side” move: **shift-middle-mouse**
- *Make this alignment*—->
- Use “slide” move **ctrl-middle-mouse**
- *Make this alignment*—->



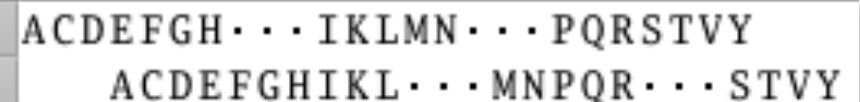
```
ACDEFGHIKLMNPQRSTVY
ACDEFGHIKLMNPQRSTVY
```



```
ACDEFGHIKLMNP...QRSTVY
ACDEFGHIKLMNPQRS...TVY
```



```
ACDEFGH...IKLMN...PQRSTVY
ACDEFGHIKL...MNPQR...STVY
```



```
ACDEFGH...IKLMN...PQRSTVY
ACDEFGHIKL...MNPQR...STVY
```

In class exercise 5.2: 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 mae file linked to “**realignment exercise**”

Alignment | Align

Hit OK

Go to **Window | Commands** (ctrl-space)

What is the percent identity?

- Use the appropriate scoring matrix
 - **SEQ: Alignment | Similarity..**
 - If %identity, as determined by the initial alignment, is X%, select BLOSUM<X>*, where <X> is the nearest value to X.
 - Re-align..

Alignment | Align (using Blosum30, gap start penalty 8).

*****Note location of deletion in helix.*****

What happens if you accept this alignment and run...

Protein | Homology model (set Models to 1)

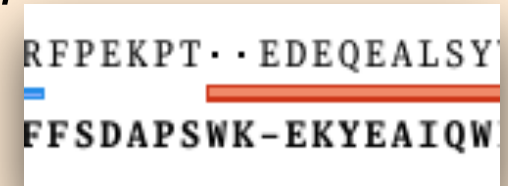
Wait for result. Inspect the resulting structure. Is the helix distorted? (yes!)

Re-open realignment exercise.

Alignment | Align (using Blosum30, gap start penalty 8).

Create this alignment--->

Protein | Homology model (set Models to 1)



RFPEKPT · · EDEQEALSY
FFSDAPSWK-EKYEAIQW

The image shows a sequence alignment. The top sequence is "RFPEKPT · · EDEQEALSY" and the bottom sequence is "FFSDAPSWK-EKYEAIQW". A blue vertical bar is under the first 'F' in the bottom sequence. A red horizontal bar is under the gap in the bottom sequence, which aligns with the two gaps in the top sequence. This indicates a deletion in the bottom sequence relative to the top sequence.

Wait for result. Inspect the resulting structure. Is the helix distorted? (no!)

Overrule automatic alignment when it...

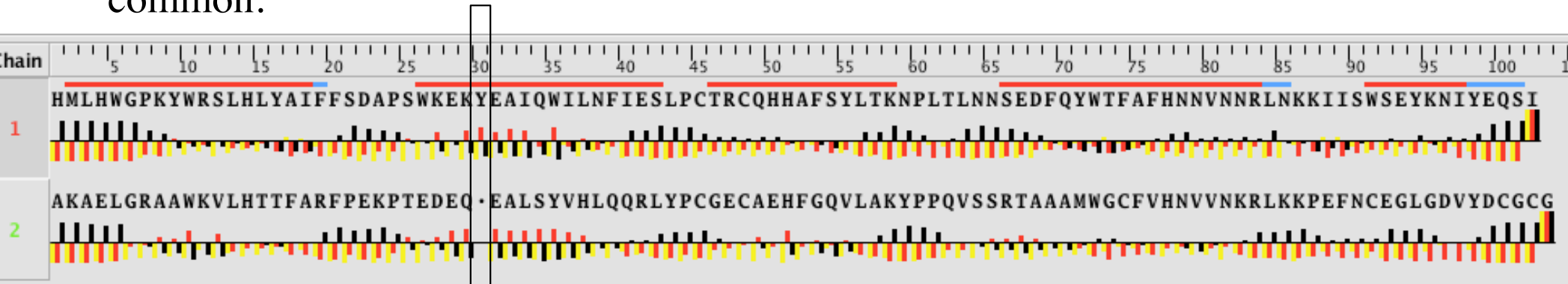
- puts an indel in the **middle of a helix**
- puts an indel in the **middle of a strand**
- puts the anchor residues of a deletion **too far apart**
- puts an insertion at a buried position
- puts an insertion in the middle of a helix
- forces a database search of a loop of length zero.
- forces a database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

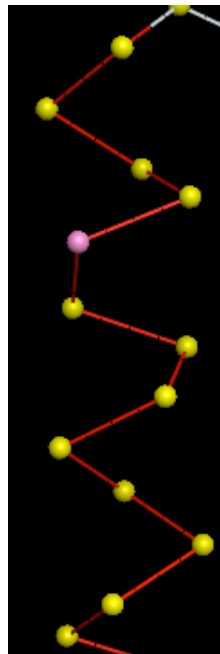
True indels within a contiguous alpha helix are rare. Alignment errors are pretty common.



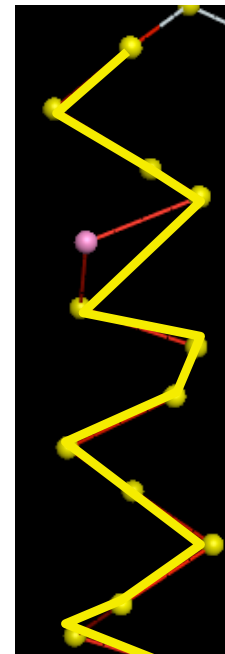
Deletion in the middle of a helix



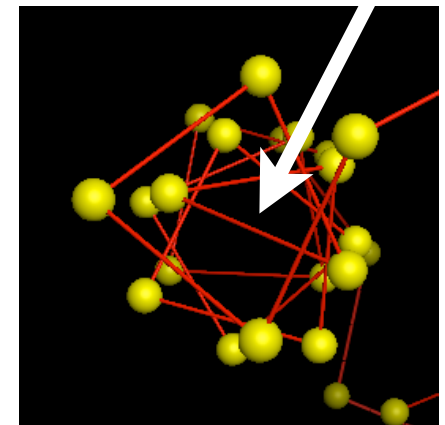
before deletion



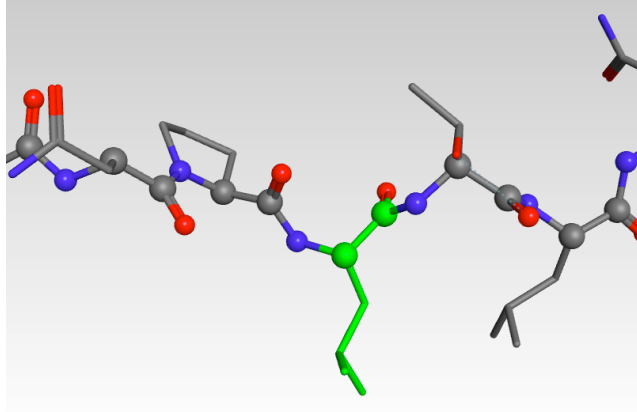
after



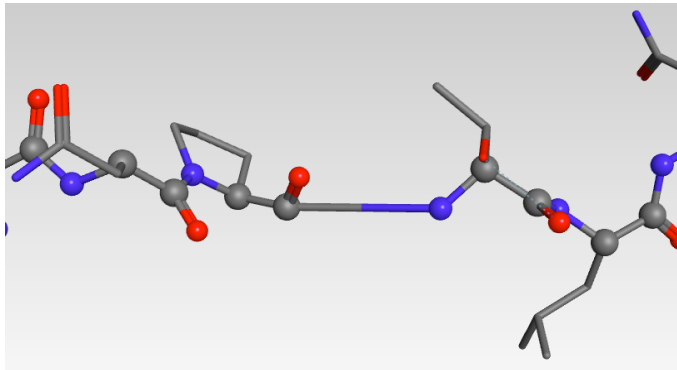
chain crosses through middle of helix!?



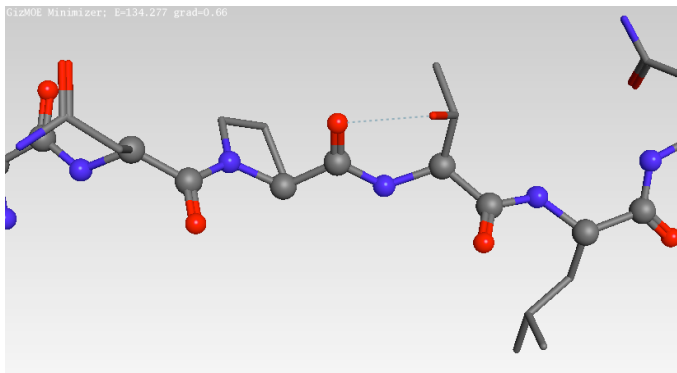
Deletion falls within a strand. What happens if you accept it?



Before deletion. Residue in green will be deleted.



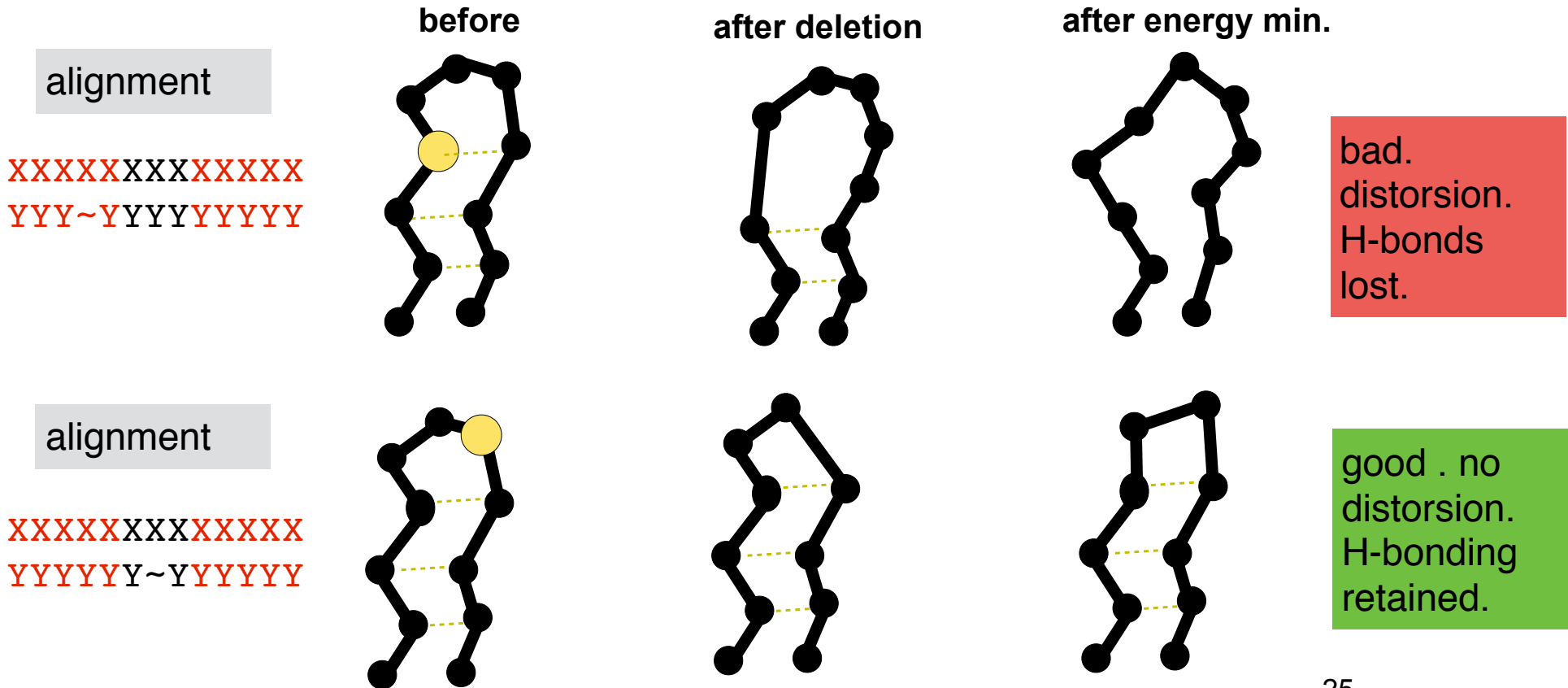
After deletion. Before energy minimization. Long peptide bond.



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

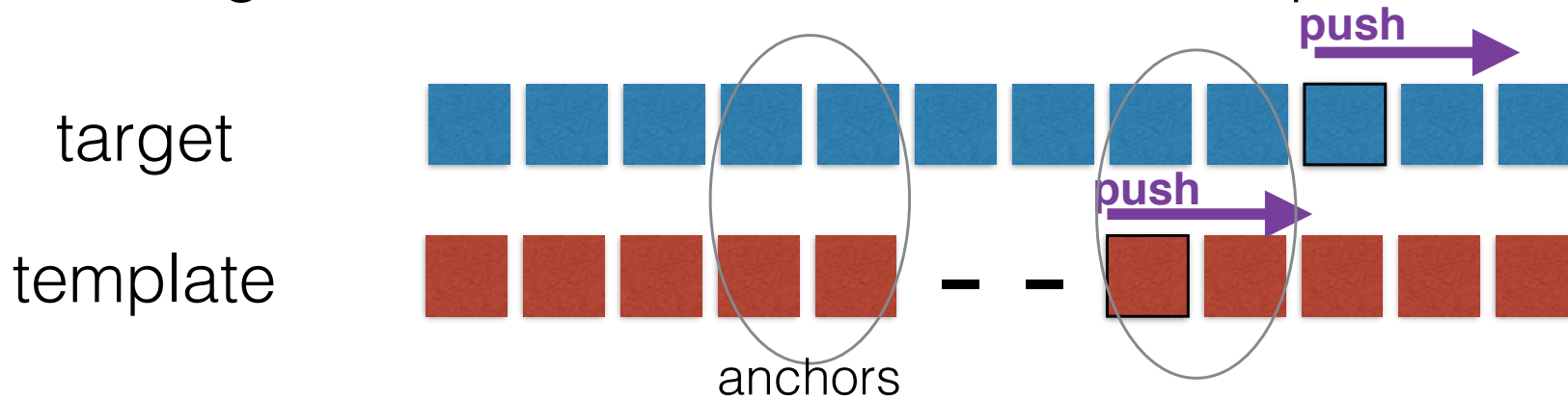
Move beta strand indel to turn

- 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.

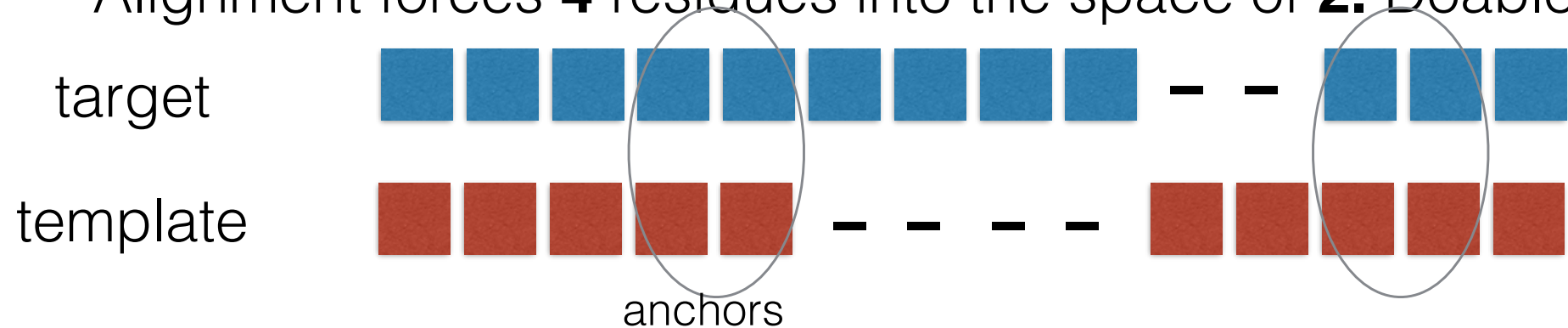


move the anchor points by "unaligning"

Alignment forces 2 residues into the space of ... zero!

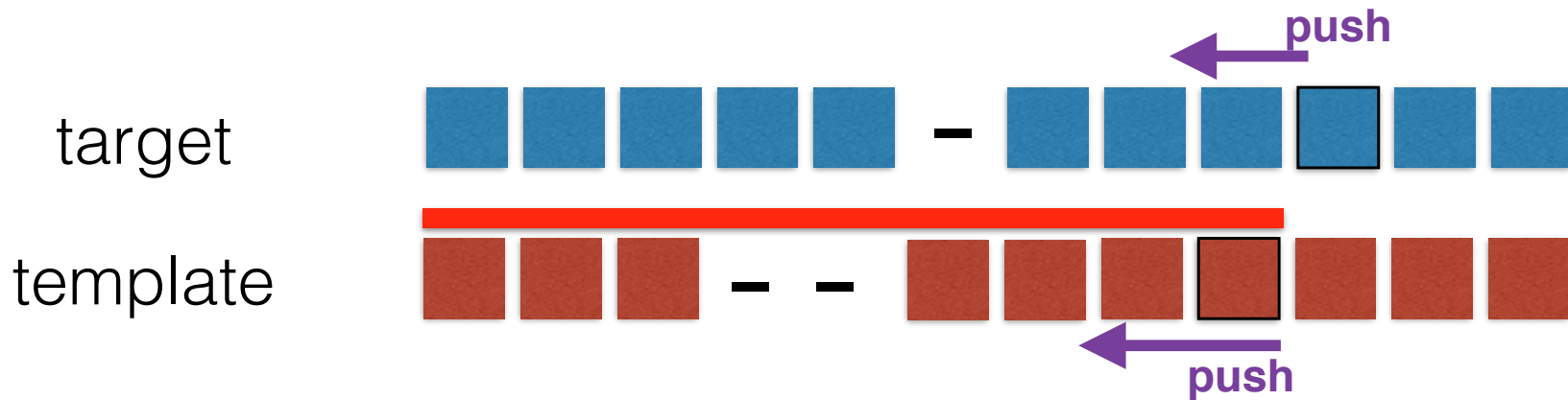


Alignment forces **4** residues into the space of **2**. Doable!

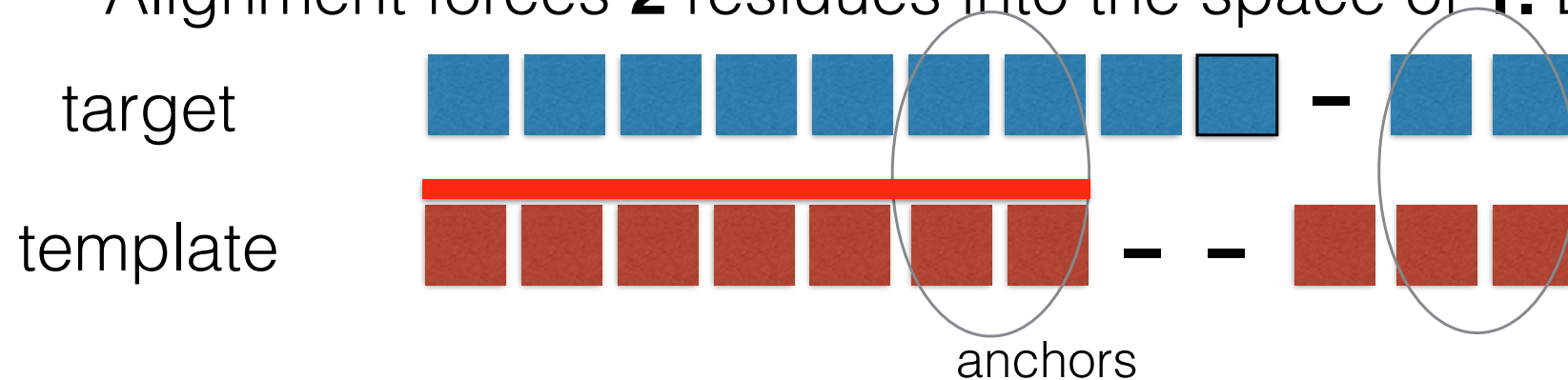


move insertion point to a turn

Alignment forces insertion into a helix!



Alignment forces **2** residues into the space of **1**. Doable!

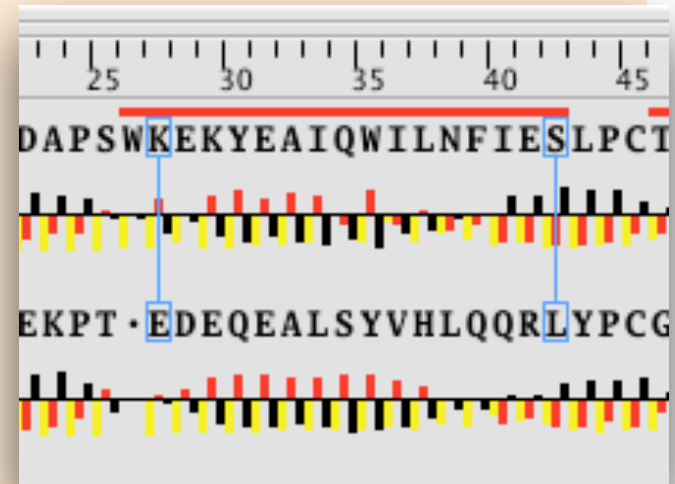


In class exercise 5.3: using alignment constraints.

SEQ: Alignment | constraints | constrain residues

Select pair of residues to constrain.

Hit “New constraint”



To remove:

SEQ: Alignment | constraints | manage constraint

select. delete. or..

SEQ: Alignment | constraints | unconstrained residues

select

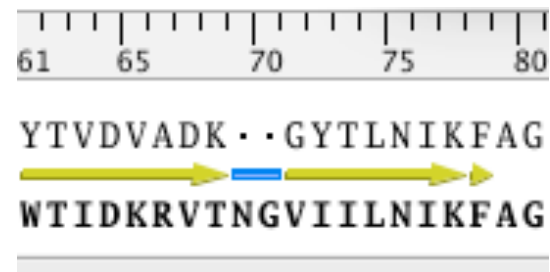
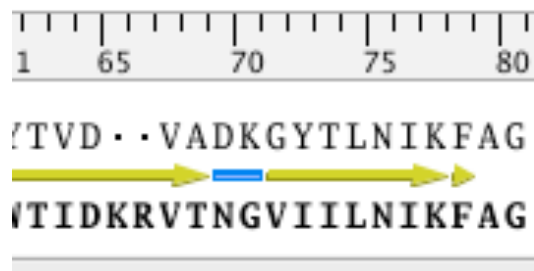
In class exercise 5.4: moving a deletion.

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.



Why move the deletion to the C-term, not N-term?

SEQ: Residues | Color: Similarity

C-term side less conserved, therefore less structurally conserved.

**Recommended:
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.

There is no grade for this exercise. I don't want to see it, but it is highly recommended that you do it. This avoids confusion later.

Review questions

- Which residues constrain a loop search?
- How are candidate loops selected from the database?
- What does alignment mean to a modeler?
- Why do we model the conserved residues first and the loops last?
- What is a SCR?
- What is a designated loop?
- What is a variable loop?
- What is a 4-for-2 loop search?
- What happens in homology modeling if a deletion occurs in a helix?
- Can energy minimization “fix” a bad alignment?