

Molecular Modeling -- 2020
**Lecture 22 -- Protein folding and
design and protein-protein
docking**

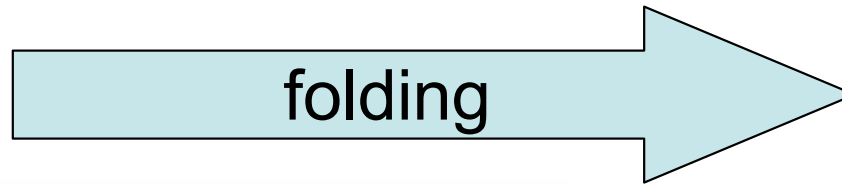
Tuesday, April 14

**First order of business:
term project**

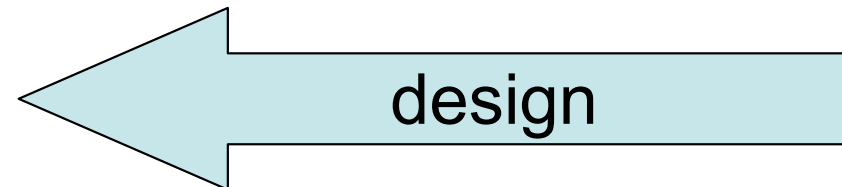
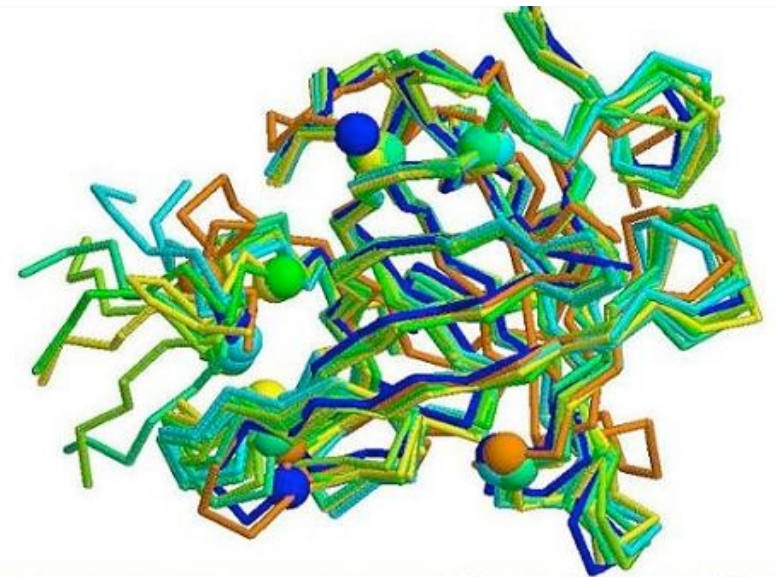
Why design proteins?

- Proteins can do anything!
- Protein-based drugs lack many of the side-effect issues of small molecule drugs.
- Proteins are immunogenic.
- Proteins are environmentally friendly.

Protein folding/ protein design



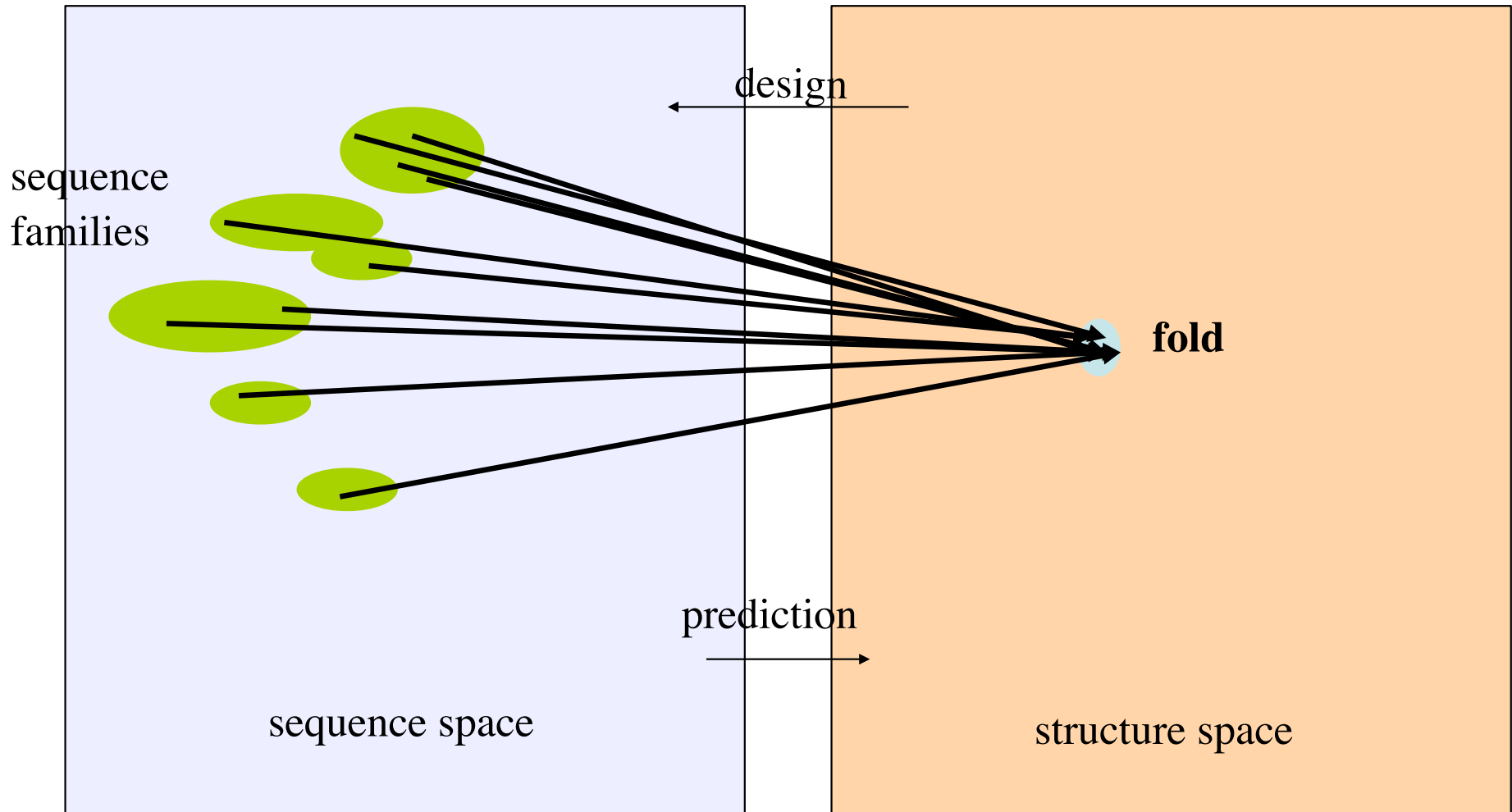
```
Q5E940 BOVIN -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 HUMAN -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 MOUSE -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 RAT -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 CHICK -----M*PREDRATWKSNYFMKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 RANSY -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMQOIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--SALE 76
Q7ZUG3 BRARE -----M*PREDRATWKSNYFLKIIQLDDY*PKCFIVGADNVGSKOMOTIRLSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 ICTPU -----M*PREDRATWKSNYFLKIIQLNDY*PKCFIVGADNVGSKOMOTIRLSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 DROME -----M*PRENKAAWKAQYFIKVVLEDFE*PKCFIVGADNVGSKOMONIRTSRGL-AVVLGKNTMMRKAIRGHLENN--PALE 76
RLA0 DICDI -----MSGAG-SKRKKLFIEKATKLF*TT*YDKMIVAEADFVGN-SLOQIRKRSIRGI-GAVLMGKNTMIRKVIIRDADSK--PELD 75
Q54LP0 DICDI -----MSGAG-SKRKNVFIEKATKLF*TT*YDKMIVAEADFVGN-SLOQIRKRSIRGI-GAVLMGKNTMIRKVIIRDADSK--PELD 75
RLA0 PLAF8 -----MAKLSQOQKQMYIEKLSL*IQOYSKILLVHV*DNVGN*QMASVYRKSIRGK-AVVLGKNTMIRKVIIRDADSK--PELD 76
RLA0 SULAC -----MIGLAVITTKKIAKWKYDEVAELTEKIKTKHTIIIANIEG*PADKLIHEIRKKLRGK-ADIKVTKNLIFGIAAKNAG--LDYK 79
RLA0 SULTO -----MRIMAVITQERKIAKWKLEVEKLEKIKRKYHTIIIANIEG*PADKLIHEIRKKLRGK-ADIKVTKNLIFGIAAKNAG--LDYK 80
RLA0 SULSO -----MKRLALALQKQKVASWKLLEVEKLETELKIKNSNTLLIGNLEGG*PADKLIHEIRKKLRGK-ADIKVTKNLIFGIAAKNAG--LDYK 81
RLA0 AERPE MSVVSIVLQGMKREK*IP*PKKTLMLRELELEFSKRYVFLADIT*CP*P*YVYRVKRLKPKK-YDMMVAKKRIILRAMKAAGLE--LDDN 86
RLA0 PYRAE -HMLAIGKRRYVTRQ*PARKKIVSEAT*ELLQK*YVYVFLDLHGLS*RIIHEVYRLERY-GVIKIIPK*LLFKIAFTK*VYGG--IPAE 85
RLA0 METAC -----MAEERHHT*EH*IPQWKDEEENIKELIQSHK*V*GMV*IEGLLATKWKIRRDLDKDY-AVLKVSRLNLLERAINLQIG--ETIP 78
RLA0 METMA -----MAEERHHT*EH*IPQWKDEEENIKELIQSHK*V*GMV*IEGLLATKWKIRRDLDKDY-AVLKVSRLNLLERAINLQIG--ESIP 78
RLA0 ARCFU -----MAAVR*GS--DPEKYRVAEEIKRMIS*SKPVVAIVSFRNV*PAG*Q*MKIRREFR*GK-AEIKVKNLLEKALDAG--GDYL 75
RLA0 METKA MAVKAK*GPPS*EY*PKVAEWKREVEKLEKELMDEYENGLVDLE*IPAP*LOEIRAKLRERD*TIIRMSRNTLMR*IALEEK*LDER--PELE 88
RLA0 METH -----MAHVAEWKKEVEQLHDLIK*EYEVGIANLADIPAROLOKMRQTLRDS-ALIRMSKTLISLAL*EKAGREL--ENVD 74
RLA0 METTL -----MITAESEHKIAPWKLEEVNKLKELKLNQOIV*ALVDMMEVPAROLOEIRDKIR-PTMLKMSRNTLIEKAI*KEVABEETGNPEFA 82
RLA0 METVA -----MIDAKSEHKIAPWKLEEVNKLKELKLNQOIV*ALVDMMEVPAROLOEIRDKIR-DQMLKMSRNTLIEKAI*KEVABEETGNPEFA 82
RLA0 METJA -----METKVAHVAEWKLEEVNKLKELIKSKPVVAIV*VDMMDVPAP*LOEIRDKIR-DKVKLRMSRNTLIEKAI*KEVABEELNPKLA 81
RLA0 PYRAB -----MAHVAEWKKEVEELANL*IKSYPIALVDVSSMPAYPLSQMRR*LI*RENG*LLRVSRLNLLIE*LAIKKAA*ELGKPELE 77
RLA0 PYRFU -----MAHVAEWKKEVEELANL*IKSYPIALVDVSSMPAYPLSQMRR*LI*RENG*LLRVSRLNLLIE*LAIKKAA*ELGKPELE 77
RLA0 PYRKO -----MAHVAEWKKEVEELANL*IKSYPIALVDV*AG*VPAYPLSKMRDKLR-GKALLRVSRLNLLIE*LAIKKAA*ELGKPELE 76
RLA0 HALMA MSASESEK*ET*IP*EWKQEEVD*AI*VEMIESYESVGVVNT*AG*IP*ER*LODMRRDLHCT-AELRVSRLNLLERALDDVD--DGLE 79
RLA0 HALV MSASEV*RT*EV*IP*QWKREVE*VDF*IESYESVGVV*G*V*AG*IP*ER*LODMRRDLHCT-AELRVSRLNLLERALDDVD--DGLE 79
RLA0 HALSA MSASEEQ*RT*EE*V*PEWKRQEV*AE*VLDLLET*YDSVGVV*NT*G*IP*ER*LODMRRDLHCT-AELRVSRLNLLERALEEAG--DGLD 79
RLA0 THEAC -----MKEVSQKKELVNETIRKASRSVAIVD*AG*IR*RO*ID*IRGKNR*GK-INLKVIKKLL*LLFKALENLGD--EKLS 72
RLA0 THEVO -----MRKNPKKKEIYSELA*DD*ITKSKAV*IVD*IK*Q*VR*ROMODIRAKNRDK-VKIKVKKLL*LLFKALDLSIND--EKLT 72
RLA0 PICTO -----MTEPAQKKID*FVKKLENEINSR*VAA*V*SK*GR*NN*E*FK*IR*NS*IRDK-ARIKVSAR*LL*LL*LA*EN*GK--NNIV 72
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90
```



sequence

structure

Sequence space maps to structure space like this



.. many-to-one.

Thus design is "easier" than prediction.

Short history of protein design

Site-directed mutagenesis, engineering (J. Wells, 1980's-90's)

Coiled coils, helix bundles (DeGrado, 1980's-90's)

Binary patterning (Hecht, 1990's)

Extreme protein stabilization (Mayo, 1990's)

Binding pocket design (Hellinga, 2000)

New fold design (Kuhlman & Baker, 2002-4)

Protein-protein interface design (Gray & Baker, 2004)

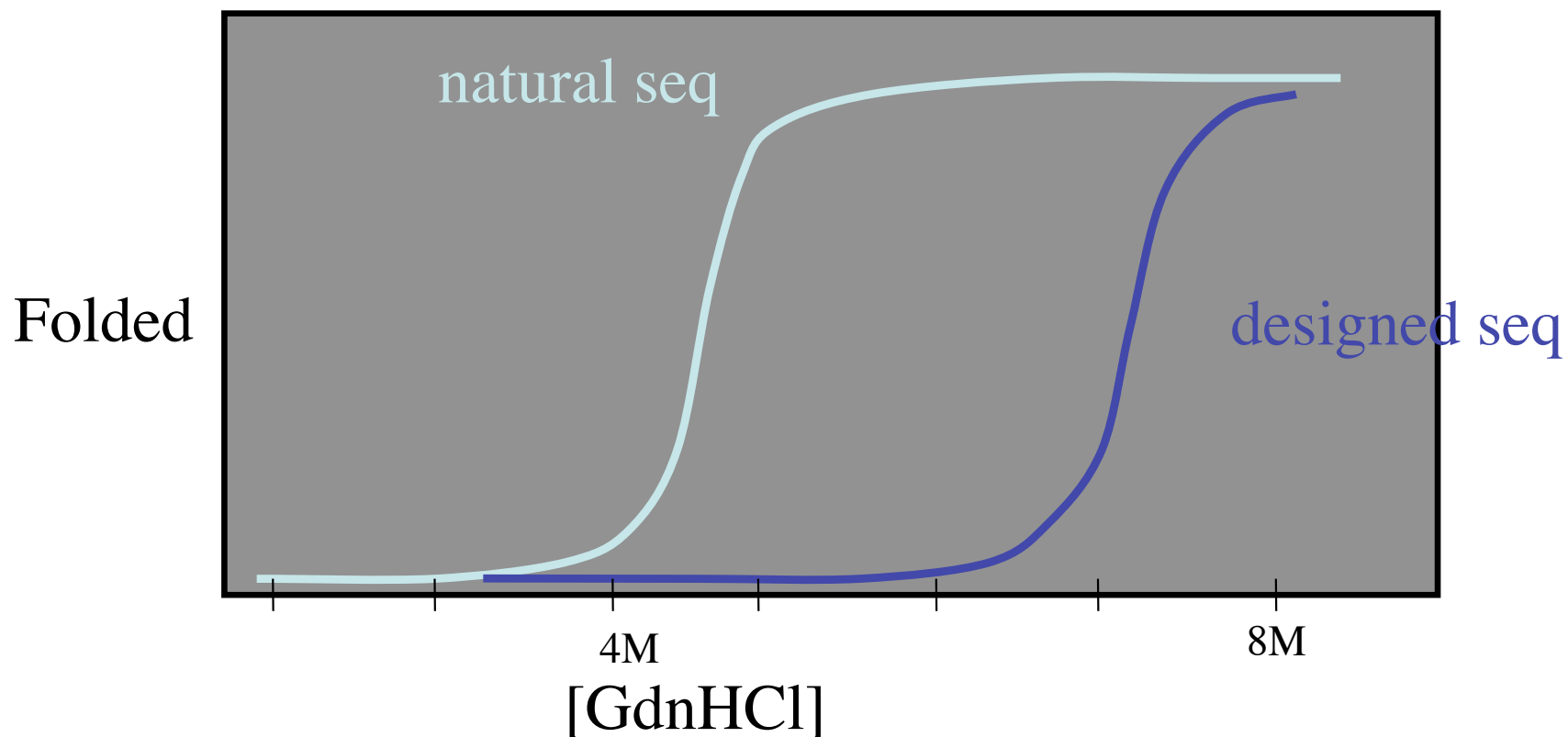
Open source protein design algorithm EGAD (Pokala, 2005)

Enzyme design (Baker, 2008)

Flexible backbone protein design (Kortemme, 2009)

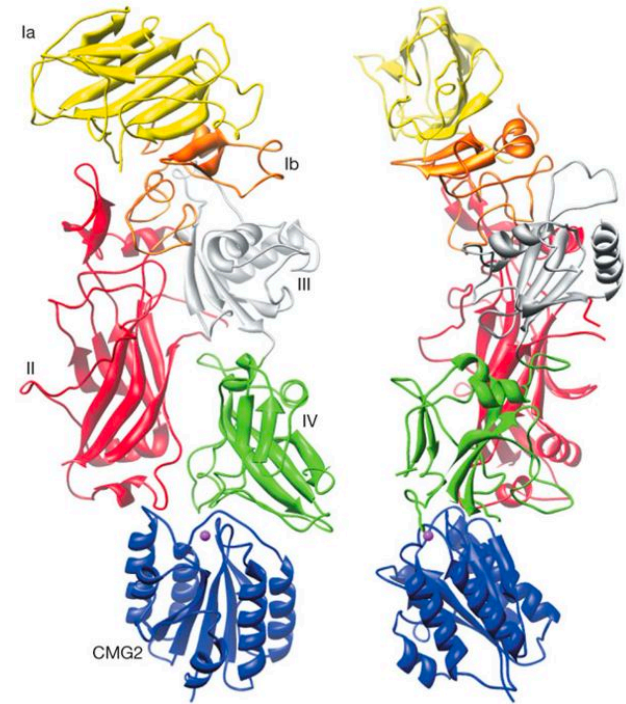
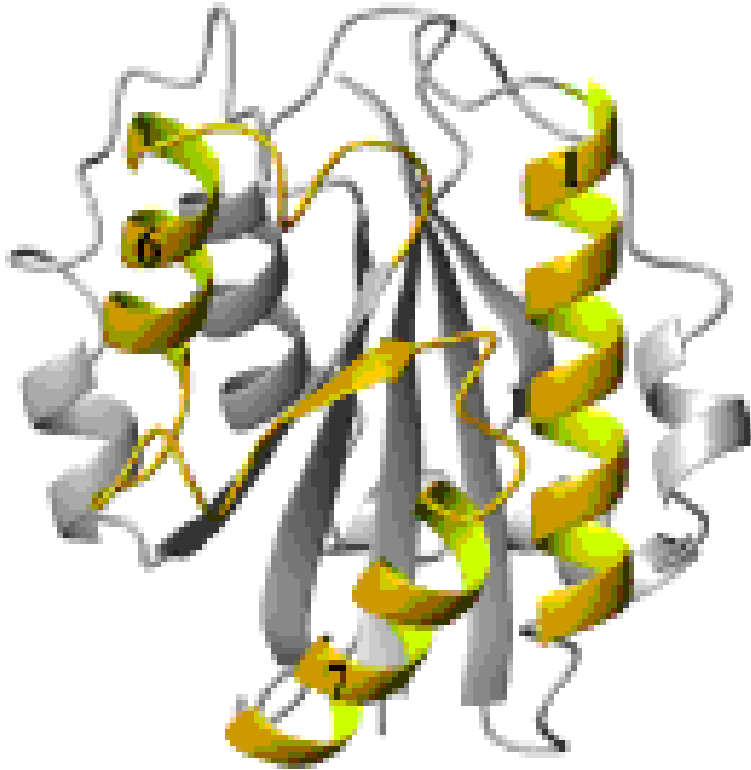
Continuous rotamer design (Donald, 2012)

Proteins can be made super-stable



Malakauskas SM and Mayo SL (1998) "Design, Structure, and Stability of a Hyperthermophilic Protein Variant." *Nature Struct. Biol.*, **5**, p.470.

Distinct conformational states can be stabilized.



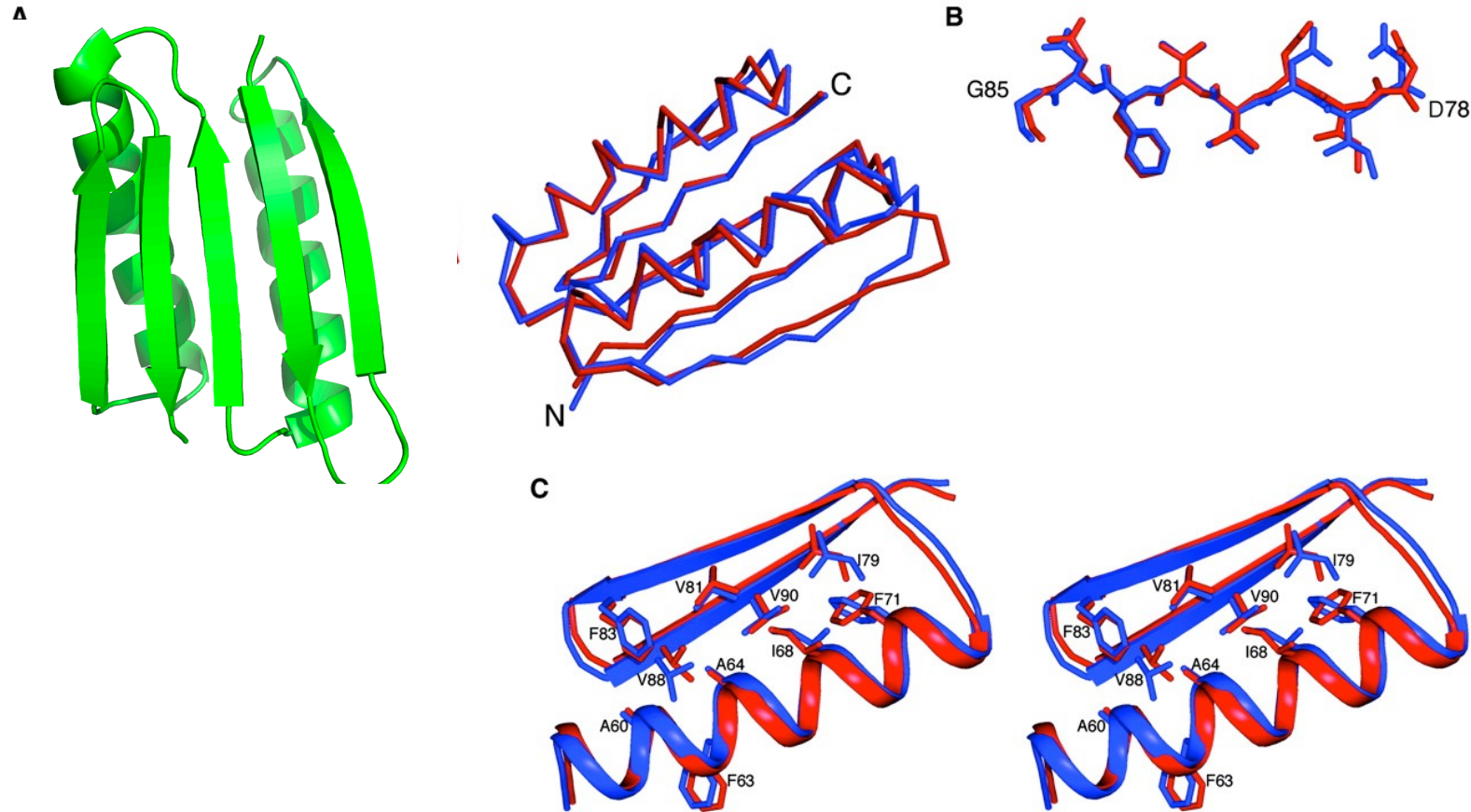
α M β 2 integrin I domain in 2 conformations

2 crystal structures are known. They differ in the highlighted region.

Shimaoka *et al* designed sequences for each form, **open** and **closed**. The two designs were shown to have different physiological properties.

Shimaoka, M., Shifman, J. M., Takagi, J., Mayo, S. L., Springer, T. A. (2000)
“Computational design of an integrin I domain stabilized in the high affinity conformation.” *Nature Struc. Biol.* 7(8), 674-678.

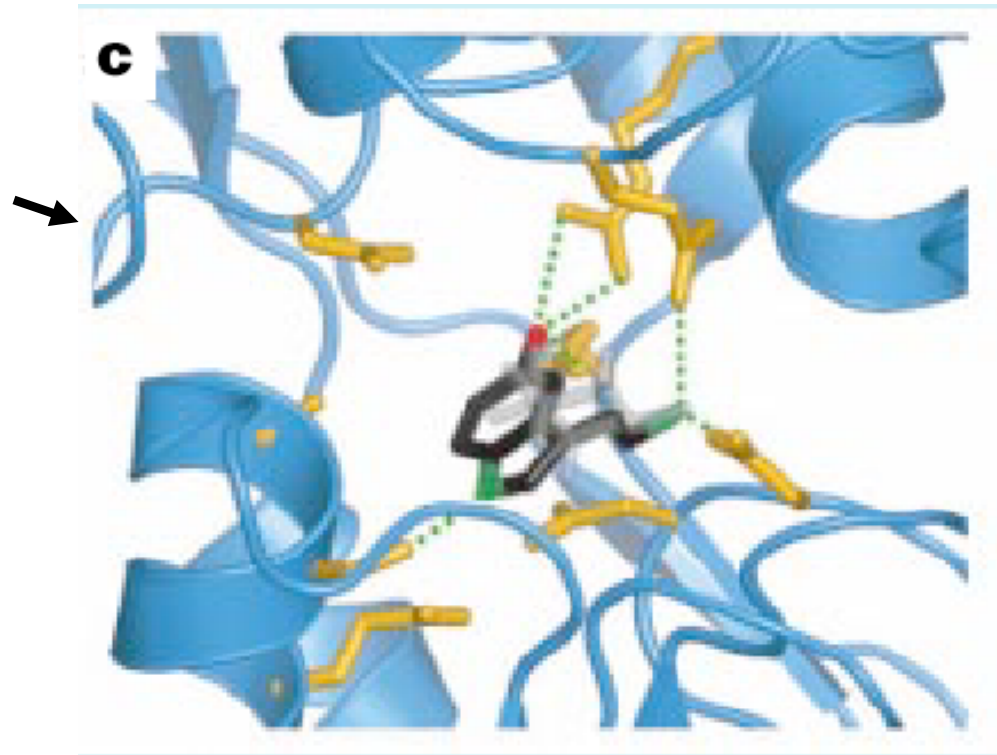
New folds have been designed



New proteins can be designed that have never been seen before. The designs are accurate (compare red and blue above) and they are highly stable.

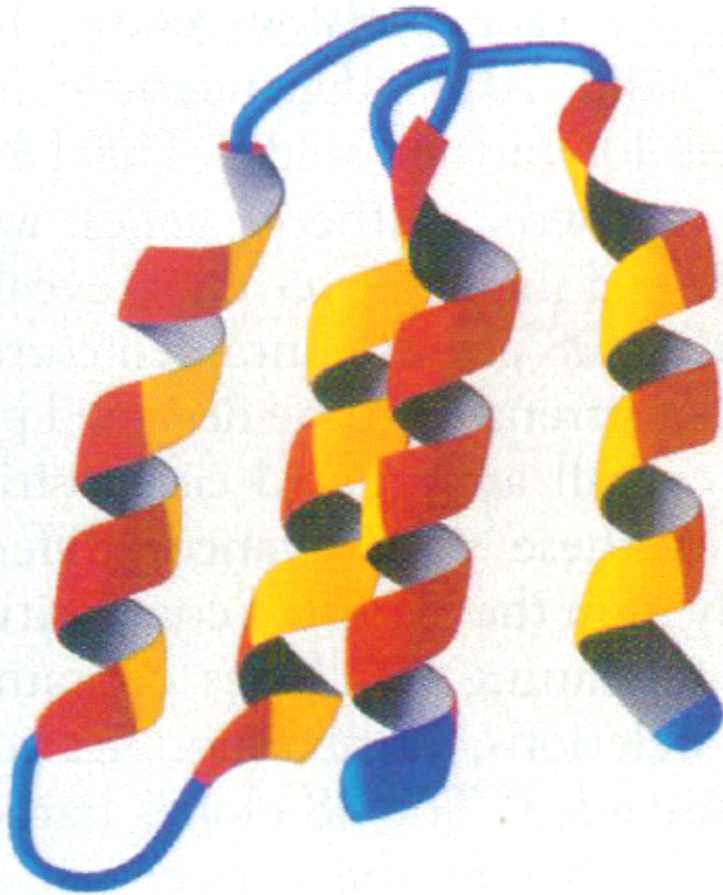
New binding sites can be designed

Used to bind arabinose, now it binds serotonin.

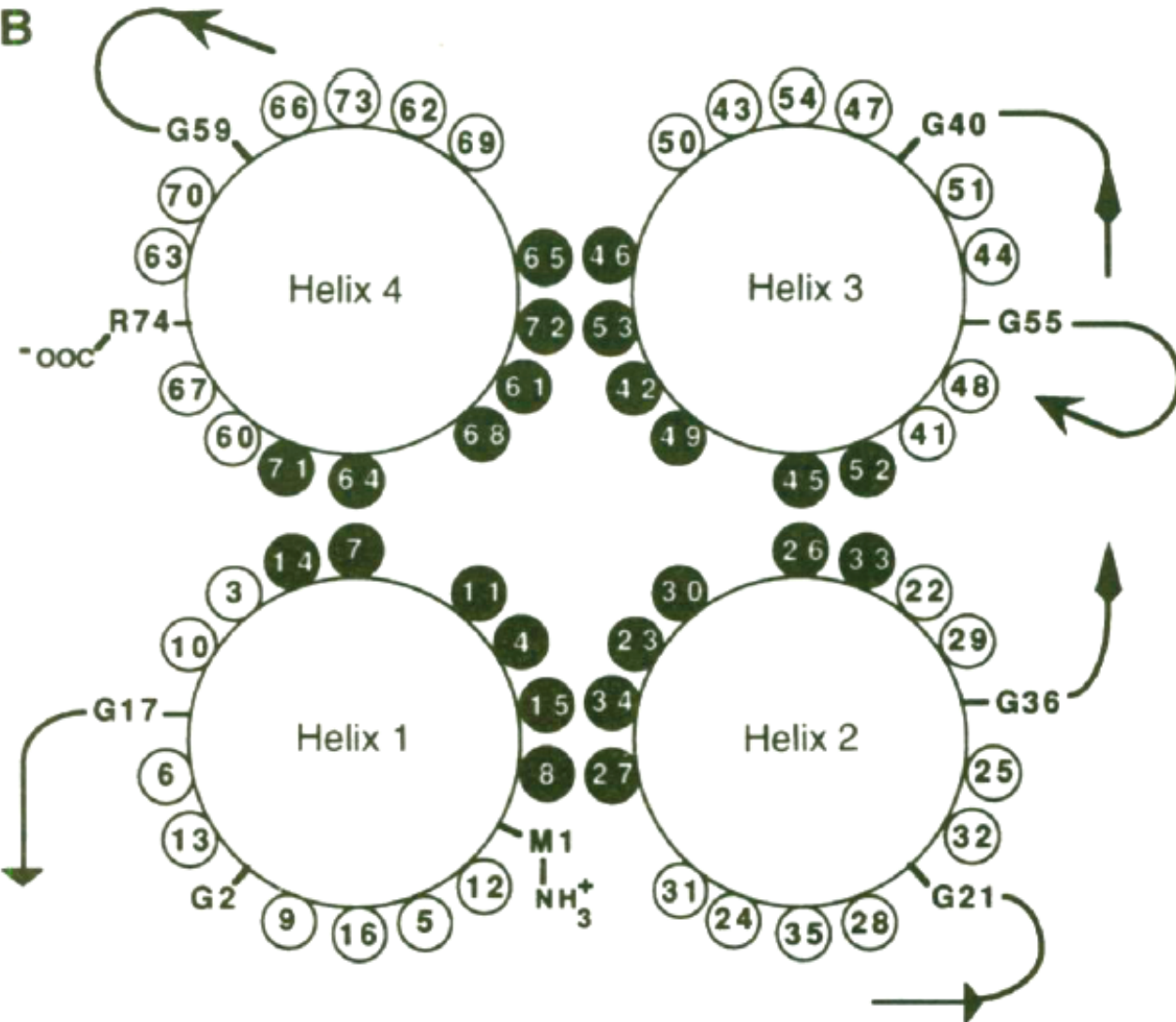


Sequence patterning is often enough to get the protein to fold.

A



B



(Kamtekar et al, Science, 1993)

All of these sequences fold to 4-helix bundle.

Helix 1	Helix 2	Helix 3	Helix 4	
MGDLENLLEKFEQLIKGPDSG	KLNHVQELQELVQGPSGG	KLKNLLNDFEDLINGPRSG	GNVQQLLKKLQQMIQR	B
..EIEDLLQKLQELME	..KIQKIIKVNELMQ	..DLHNLINKLDDVMQ	..KMHDLIDDLHLLLN	F
..DLKKLMDKVNMI	..KFNHILKELKQIMN	..KLDHFMEEMNKFLK	..ELHDVLHKLHVM	G
..EVQEVFKELEQLL	..ELNKMFEKVNNL	..KLEHVMEFNDMVE	..KLKEFIQEMQHLL	I
..DMKEVLKKLEQML	..NLQELMENVQDILD	..QLEEIMKNLENLL	..DIQNLIKEMQNFL	K
..ELEEVFQKFEDLL	..NIQNLVHEIHHFFN	..NFHEVVKELNKL	..KVKQFMNQFQOM	N
..KMENMIQELEQLL	..HFQQLLNE	..DLDKFLKELEELL	..QVQQLLQQLK	U
..ELKQLLQELKEM	..ELKNIMNQFQELLE	..QLKHLIEQLQQL	..EVQNLVEQLQ	Y
..QIQQMLENLKELL	..HLEHLFBEELQ	..KFQQFFQQLK	..DFKKFLKNIDQ	Z
..EFNEMLKEMHHF	..QFENVFNDMQK	..KLKQMMDEIHQ	..QIHQLMNHFNQ	8
..NMDKMLEQLQKIL	..EVHLLLEEFQEL	..HVENLLKEMK	..DVQNLLQQLIE	10
..KMKEVIQQLKHL	..QIKDVLQQFKQ	..KLEKMVEEFQQL	..EIKHVVNKFQQL	11
..KVEELFEEIEEIM	..EVQDLFEQLHH	..KMDHIMKQLQK	..HLNKLQKIEQL	12
..EFHEFVKNMQHLL	..NIQHFLHKIQQV	..ELDKVLHELK	..HMNQFLKQFEQ	13
..EMEKFMEKMEEMI	..DIHHVVKKMED	..KIDKLMKVVHE	..QFKEVFNQVHE	15
..DVEEVFQEKMQE	..QVQQVLKKNHMM	..KLEELLEELNN	..QLKQLLKFQDM	16
..ELDQLLQQVEDLL	..KFHQLLLEEMK	..ELEHLMQQFEH	..QFKDMLKQLQEL	17
..QLDEILEEIEQLM	..NLDKFIQKIKEI	..QLDKMMNELQEL	..HINQIFKELNQL	24
..QLNQMLQQVHQLL	..KLQNLMQQVQQM	..KLEELMEKLQK	..QFQNLQFHQLKQ	30
..DLQHIHKIHQLV	..HVQHIMHMHNL	..QMDEVLQEMQN	..HVKNVFEEMQNL	49
..EIDQVVQEMHKV	..HLKNSMDQIQNI	..NMENLLEQLEE	..DLQKIVHDFDK	51
..NLDELFEEKQMLE	..HIKDLMENLKQ	..ELEELFKEIEDL	..KLHQILQEIEDL	52
..EVENILKQLKELV	..NLKDLINQLKQ	..ELDHFLKQLKEL	..QVKQIVHHIQHL	60
..EMDNILDELQMM	..QINEFVHHLNEM	..EIKQIIDEMDQL	..QIEHLIQKFEHL	63
..EFQEMLKEMEDL	..ELEQLFEHIQE	..KLDQLLEEVND	..QLHELLQDMHHL	76
..DLEDMLEHMQLLQ	..HMEQLLDKQEV	..DMQKLLNDVKE	..DFQNLHQLIHN	83
..KLEKLMHQFQQLV	..DIKHLMNEMKHL	..ELHNFLHNLEH	..NVQKLVQDVQHL	85
..KLNDLLEDLQEV	..HLQNVIEDIHD	..KLQEMMKEFQQV	..NIKEIFHHLEEL	86
MGHLEEILNEMQEMLDG	QVKKILNELNQMLEG	PSGGHMQNIKLNHLKFL	QVHQIFEKLHKFFHR	90

Polar residues

Nonpolar residues

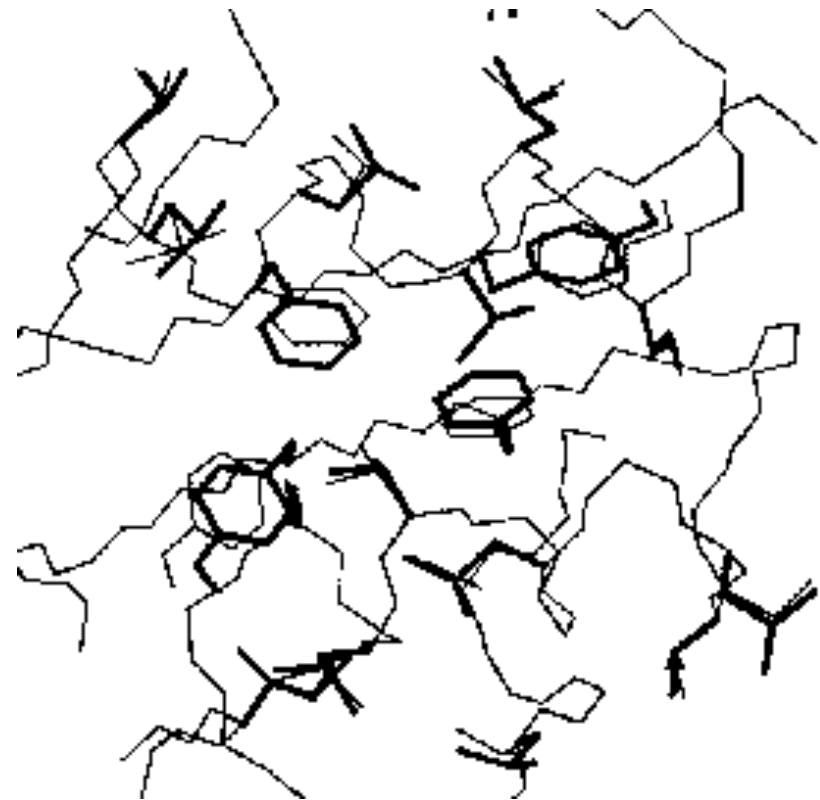
Computational design of sidechains using Dead-End Elimination

1. Select positions for mutating.
2. Choose a *palette* of amino acids at those positions.
3. For all selected positions, all palette amino acids, try all **rotamers**.
4. Chose the sequence whose rotamers give the lowest energy.

rotamer libraries facilitate protein design

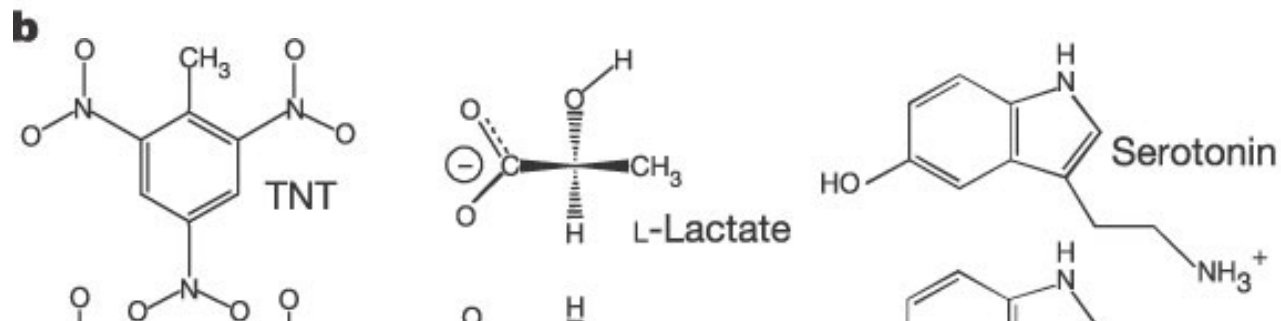
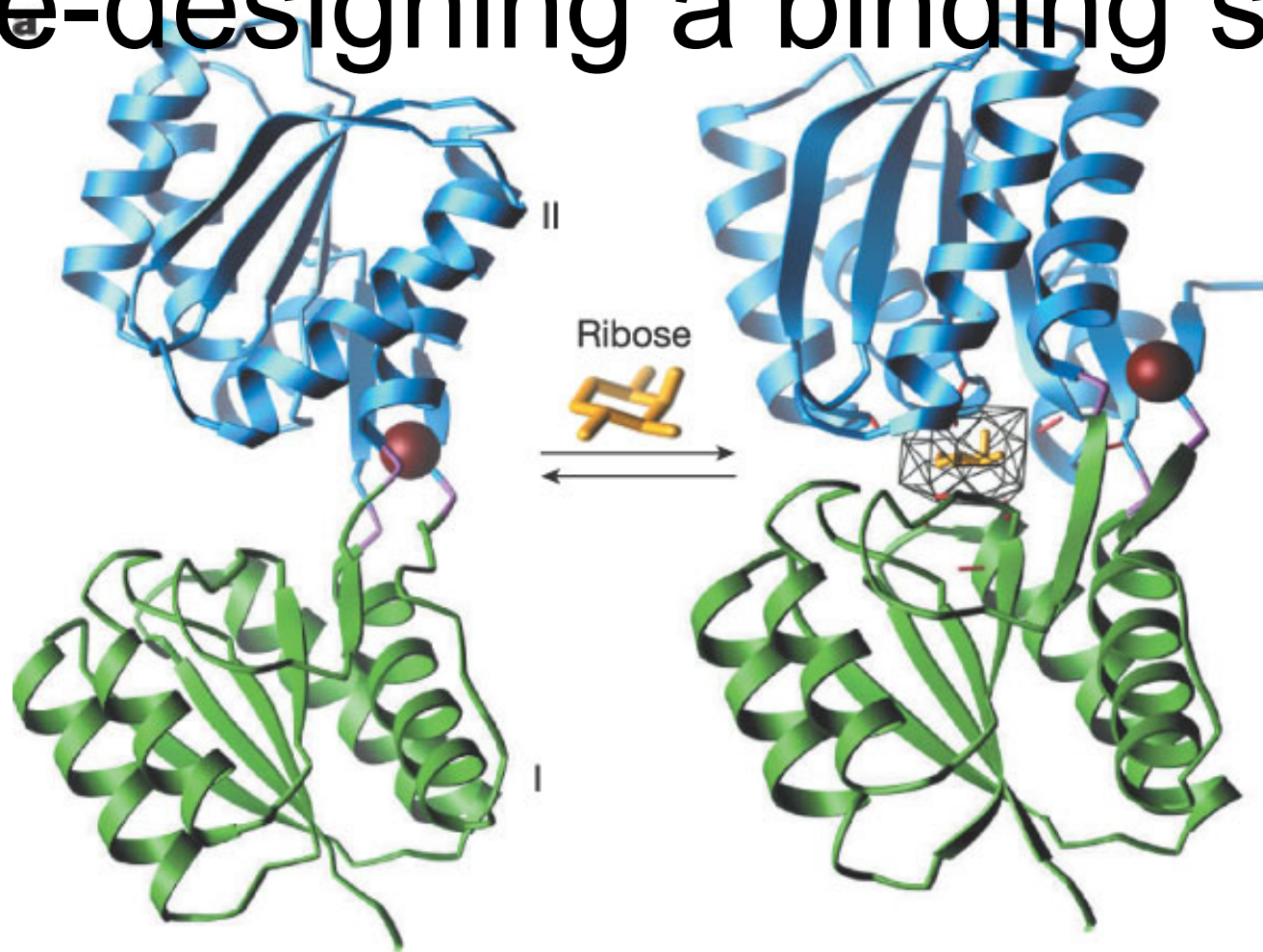
Given the sequence and only the backbone atom coordinates, accurately model the *positions of the sidechains*.

fine lines = true structure
thick lines = side chain predictions using the method of Desmet et al.

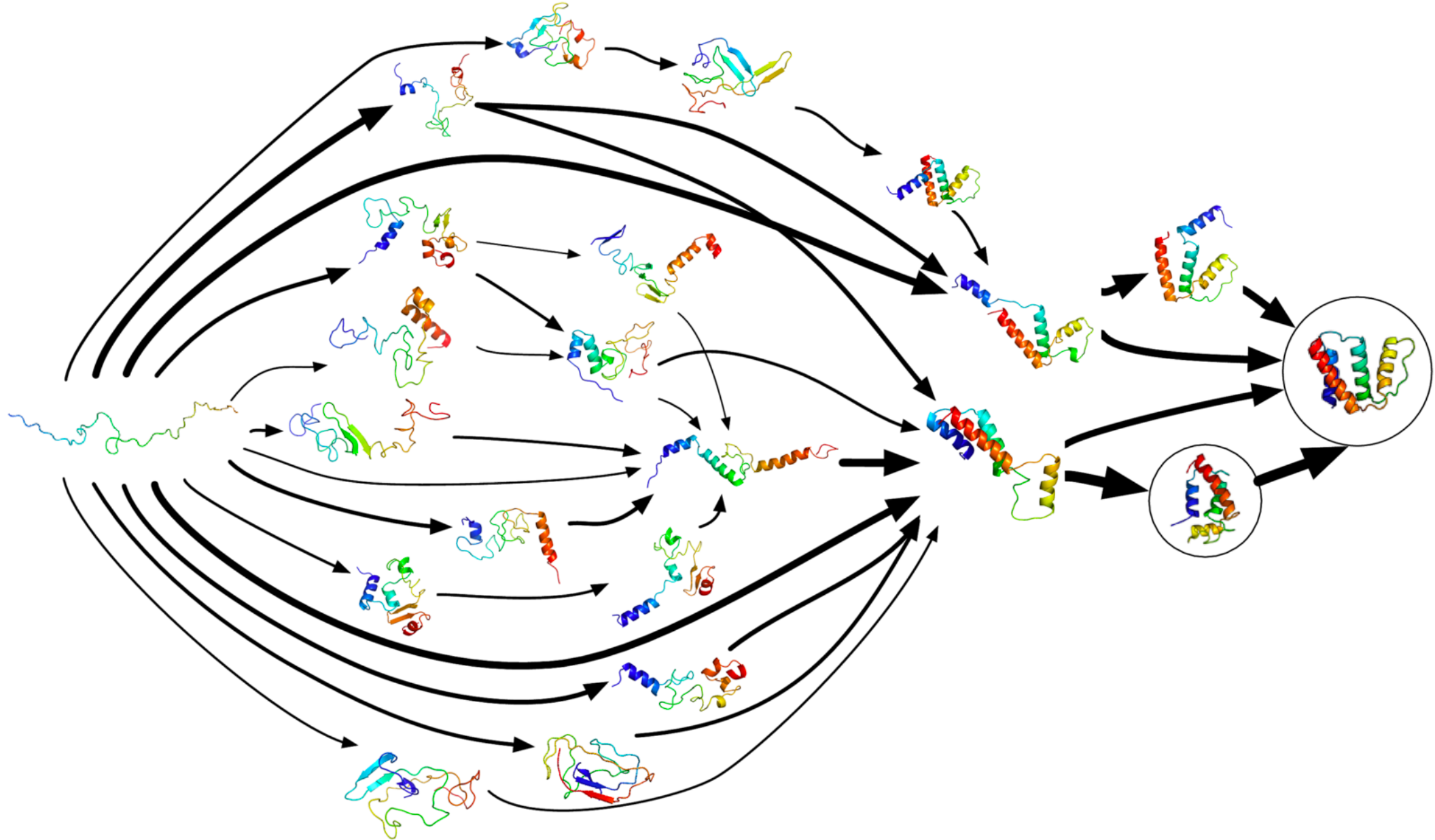


Side chain prediction using the **Dead-End Elimination** algorithm

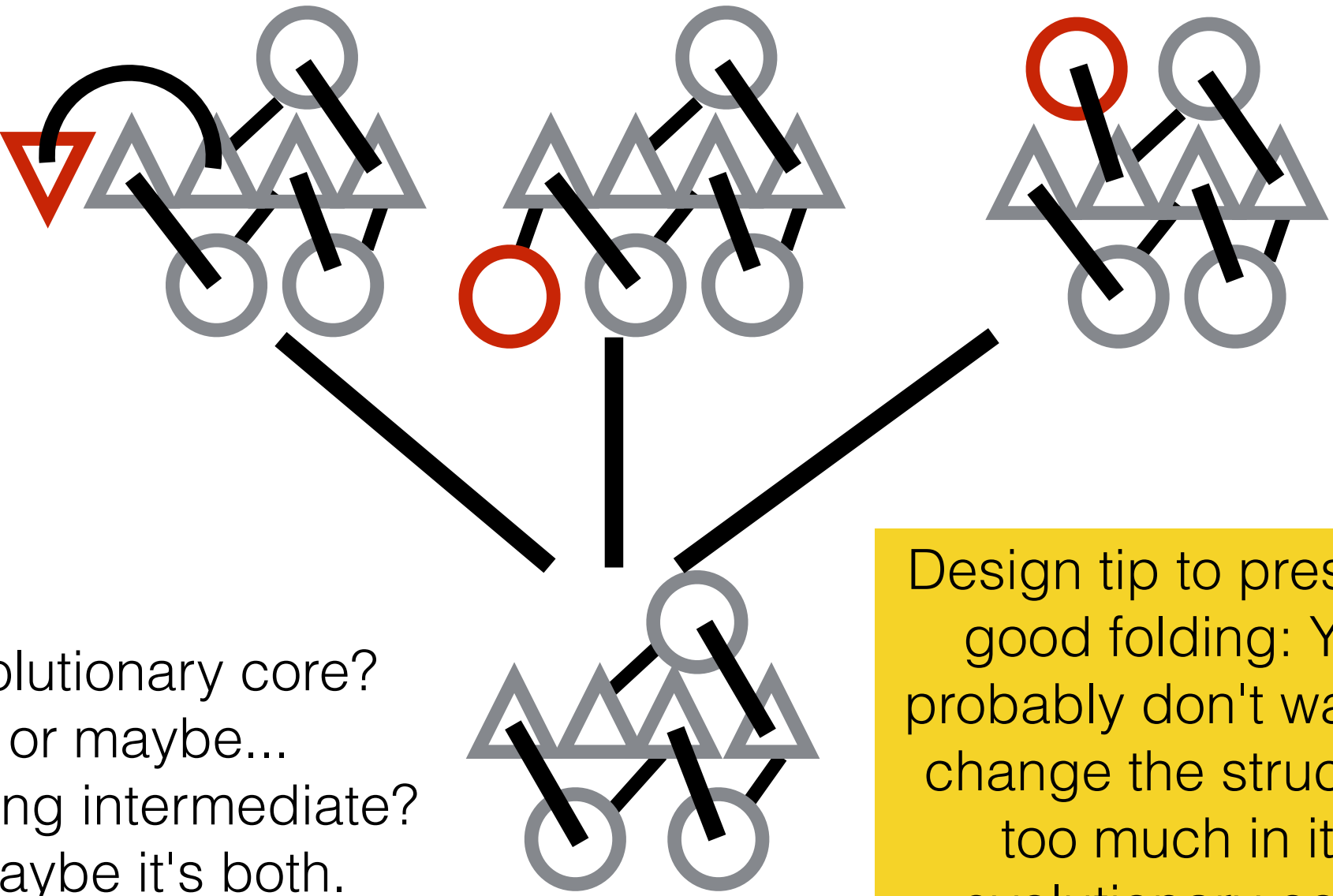
Re-designing a binding site



Folding pathway matters



Many proteins share common core structures (evolutionary cores)



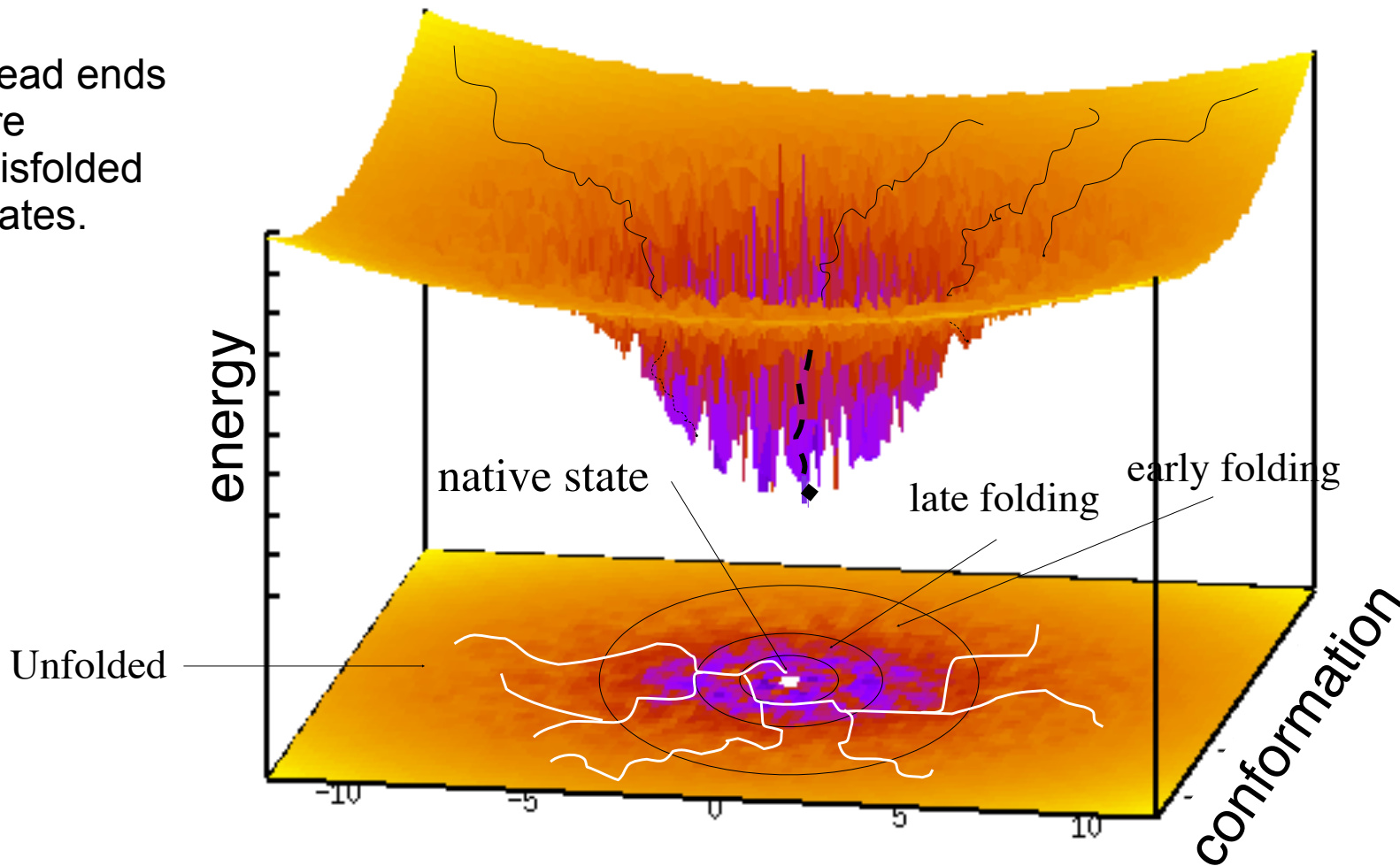
Evolutionary core?
or maybe...
Folding intermediate?
Maybe it's both.

Design tip to preserve good folding: You probably don't want to change the structure too much in its evolutionary core.

The protein folding energy landscape: *the folding funnel*

Dead ends are misfolded states.

Many paths to the folded state.

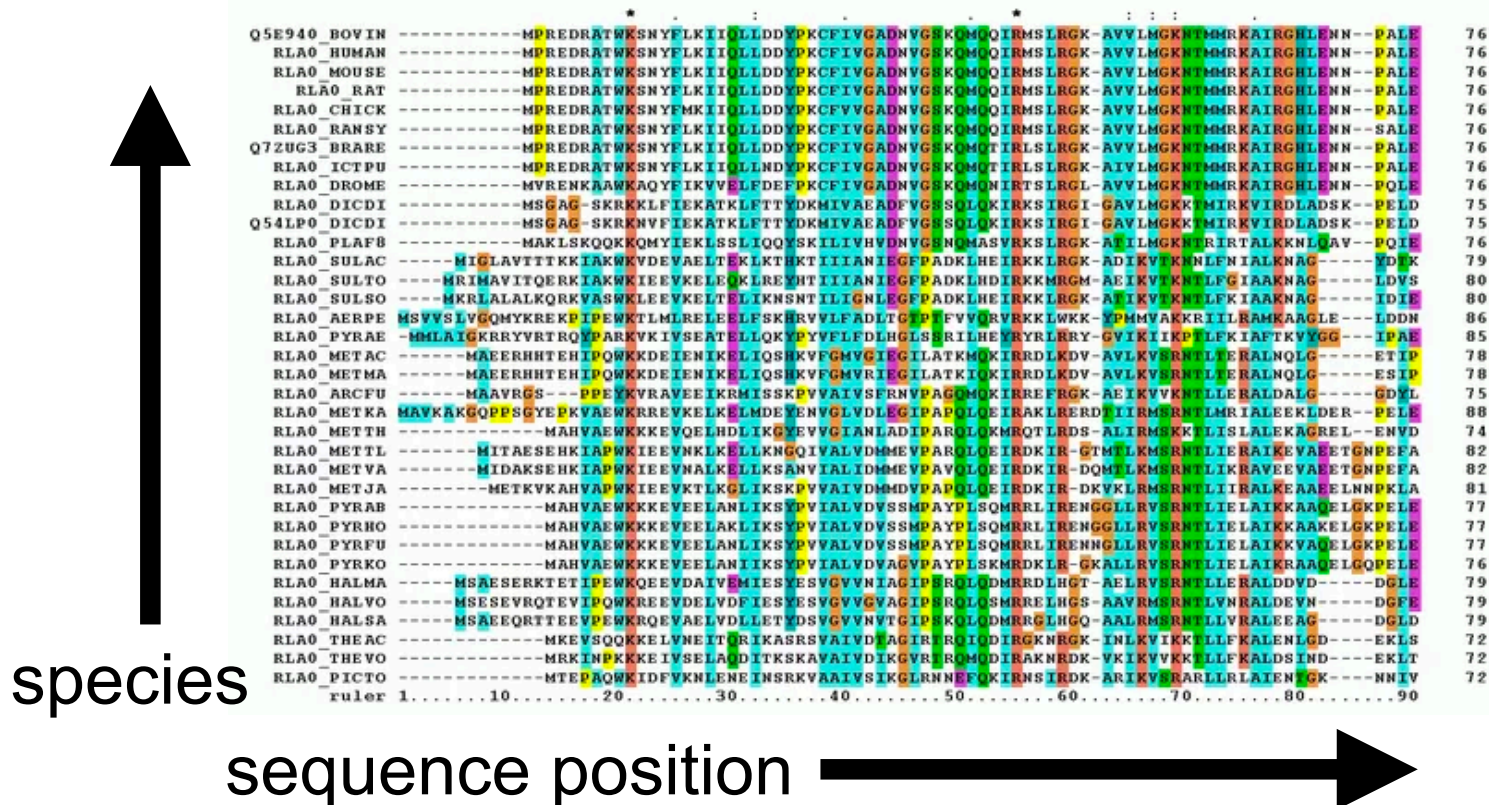


Design tip to preserve good folding: You probably don't want create **non-native folding intermediates**. Don't put a helix pattern where it is supposed to be sheet, and vice versa.

Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science*, 8(6), 1166-1180.

Garcia-Manyes, S., Dougan, L., Badilla, C. L., Brujić, J., & Fernández, J. M. (2009). Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proceedings of the National Academy of Sciences*

Requirements for folding are encoded in the sequence history



In protein design, we want to work within the amino acid palette defined by millions of years of evolution.

The protein folding pathway must be maintained

Design tip to preserve good folding: You probably don't want to change the structure too much in its **evolutionary core**.

Design tip to preserve good folding: You probably don't want create **non-native folding intermediates**. Don't put a helix pattern where it is supposed to be sheet, and vice versa.

In protein design, we want to work within the **amino acid palette** defined by millions of years of evolution.

Exercise 22.1 -- manual docking

- Download **receptor** 1TIM
- Download ligand **template** 1F2X
- Remove waters and other extraneous molecules.
- Fix the **receptor**. Unfix the **ligand**.
- Display as ribbons. Hide all atoms.
- Select the **ligand**.
- **Dock** using **alt-middlemouse**, **shift-alt-middlemouse**, **shift-middlemouse**, and **middlemouse** dragging.
- Done when you have *maximized number of ligand/receptor distances between 5Å and 8Å*

The philosophy of expert protein design

- Well-trained intuition is much faster than a random search.
- There are many, many, many right answers. We don't need the very best one.
- No force field is perfect anyway. We can't avoid the need for experimental confirmation. (See Lecture 18)

Work on Homework 4