Module 2 overview

lecture	*lab*
1. Introduction to the module	1. Start-up protein eng.
2. Rational protein design	2. Site-directed mutagenesis
3. Fluorescence and sensors	3. DNA amplification
4. Protein expression	4. Prepare expression system
SPRING BREAK	
5. Review & gene analysis	5. Gene analysis & induction
6. Purification and protein analysis	6. Characterize expression
7. Binding & affinity measurements	7. Assay protein behavior
8. High throughput engineering	8. Data analysis

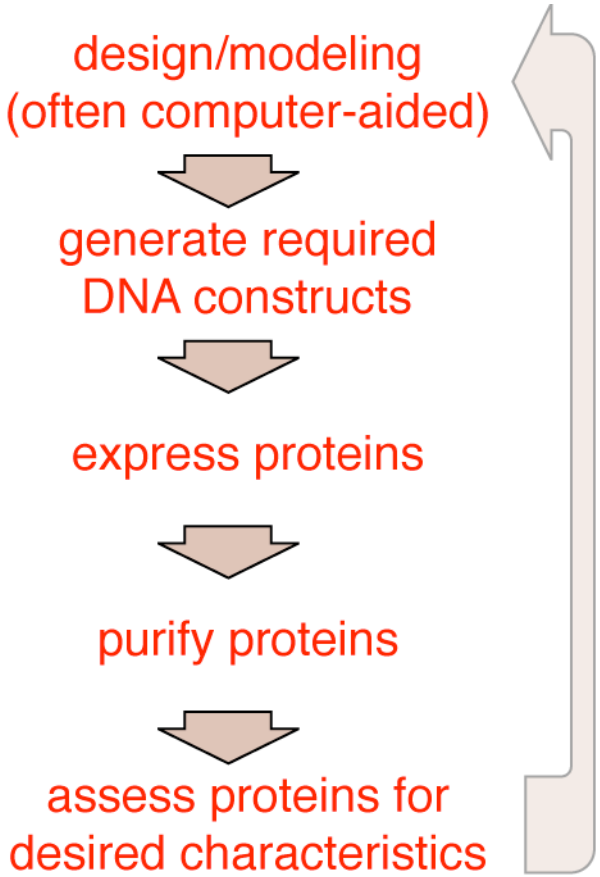
Lecture 2: Rational protein design

I. "Blob-level" protein design
   A. Engineered fusion proteins
   B. Knowledge required for blob-level engineering

II. Protein engineering at high resolution
   A. Modifying existing proteins
   B. *De novo* protein engineering
   C. Knowledge needed for high-resolution design
   D. Computational modeling

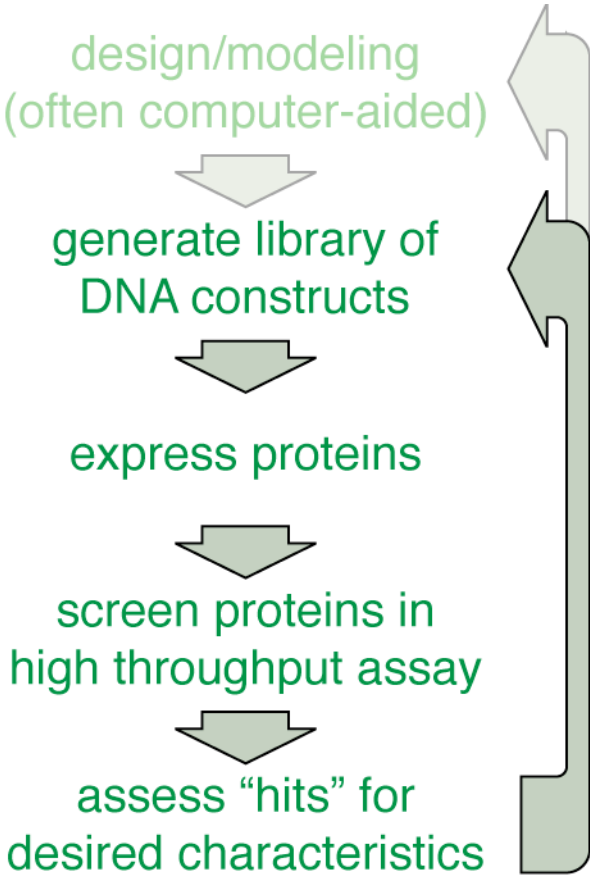
Rational protein design:

Knowldege-based, deterministic engineering of proteins with novel characteristics



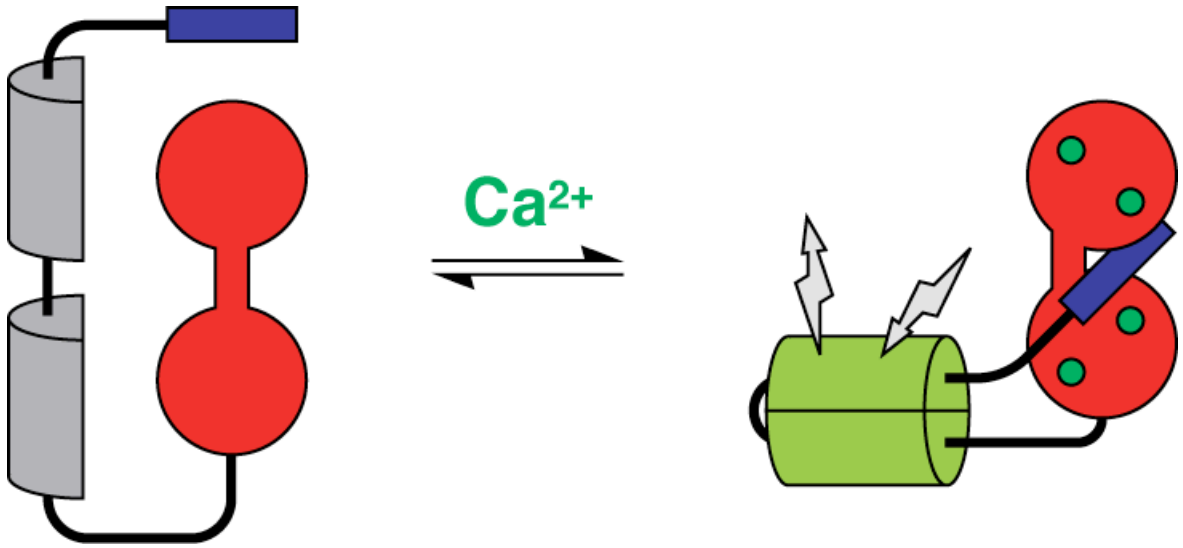
"Irrational" high throughput protein engineering:

Selection for desired properties from libraries of random variants

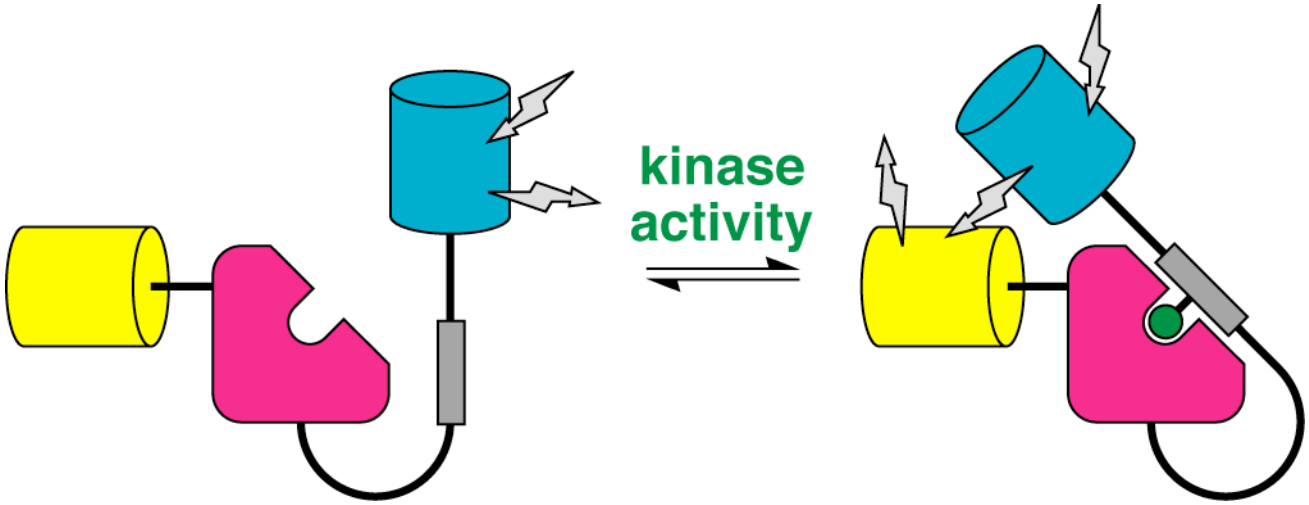


"Blob-level" protein design

- Basic idea is to combine protein units of defined function (domains) to engineer a fusion protein with novel functionality
- Examples include sensors, signal transduction components, transcription factors, therapeutics, *etc.*



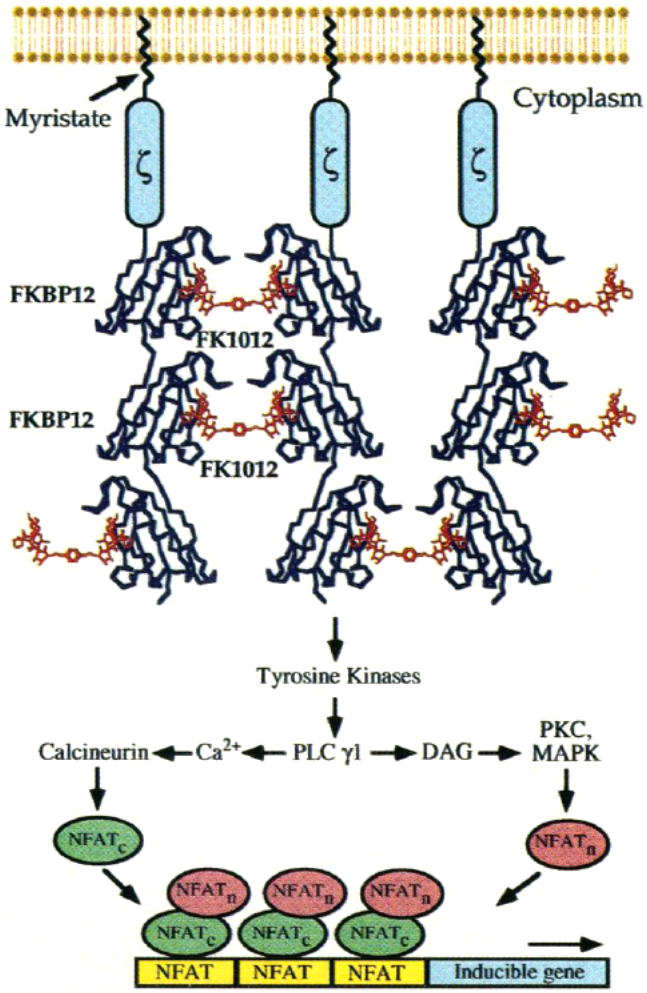
GFP-based approaches extend to other sensors:



Ting *et al.* (2001) *Proc. Natl. Acad. Sci. USA 98:* 15003-8

Can you think of other sensors one could construct based on this design strategy?

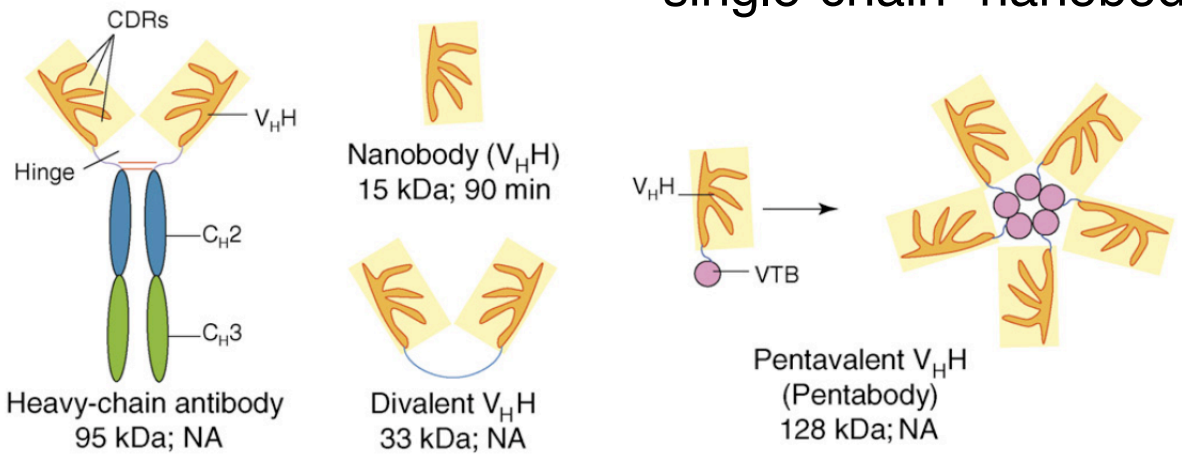
An early “synthetic biology” project–signal transduction triggered by a small molecule dimerizing agent:



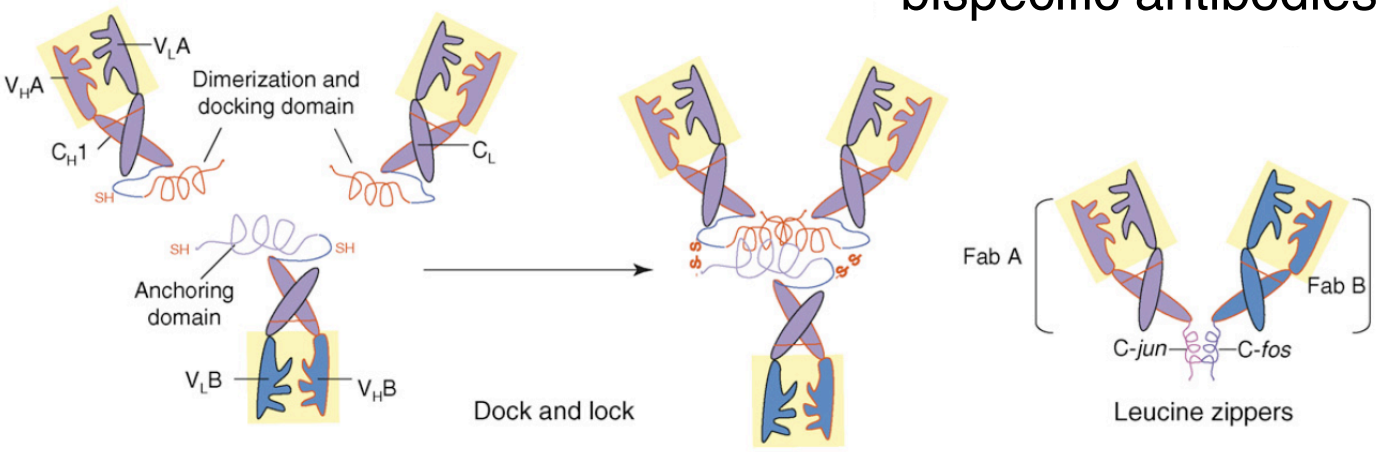
Spencer *et al.* (1993) *Science 262:* 1019-24

Engineered antibodies as therapeutic agents:

single-chain “nanobodies”



bispecific antibodies

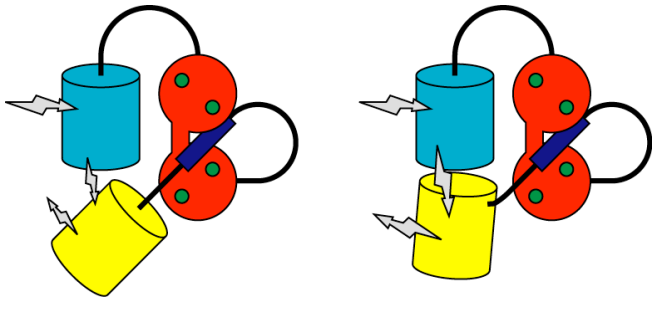


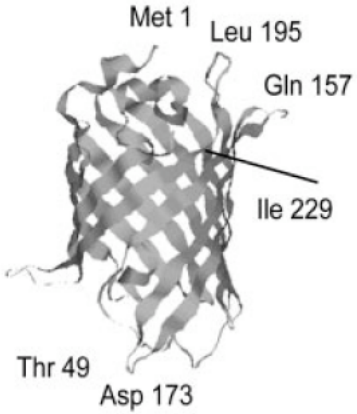
Jain *et al.* (2007) *Trends Biotechnol. 25:* 307-16

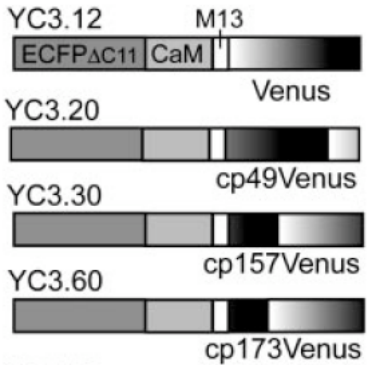
What knowledge is required for “blob-level” protein engineering?:

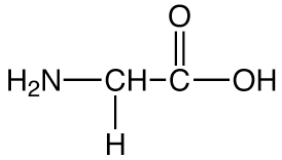
- rough geometry of protein domains (low resolution structure)
- secondary structure, if insertions or disruptions are planned
- desired linker properties (length, flexibility, hydrophilicity)

Example: CaM-based calcium sensors

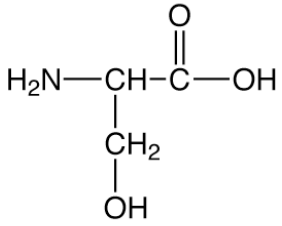




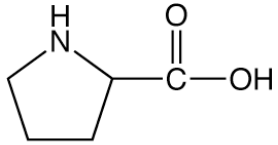




glycine (flexible)



serine (hydrophilic)



proline (rigid)

Nagai *et al.* (2004) *Proc. Natl. Acad. Sci. USA 101:* 10554-9

What we've called "blob-level" design is useful for combining functionalities associated with individual protein domains–but what if we want to create new functionalities or make subtle manipulations?

"Which brings us to my next point."

Protein engineering at high resolution

- Alter/tune properties of proteins by making structurally or computationally informed changes at the amino acid level
- In some cases, produce entirely new proteins based on predictions of structure and function from amino acid sequence
- Can be “rational” when combined with structural information and/or computational modeling approaches
- Can be “irrational” when combined with high throughput screening and random mutagenesis *(to be discussed later in the module)*

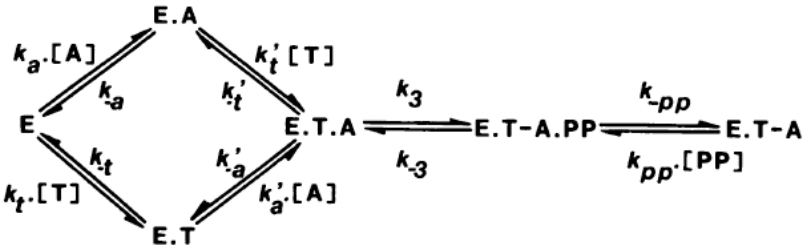
This is what we are doing in the lab for this module!

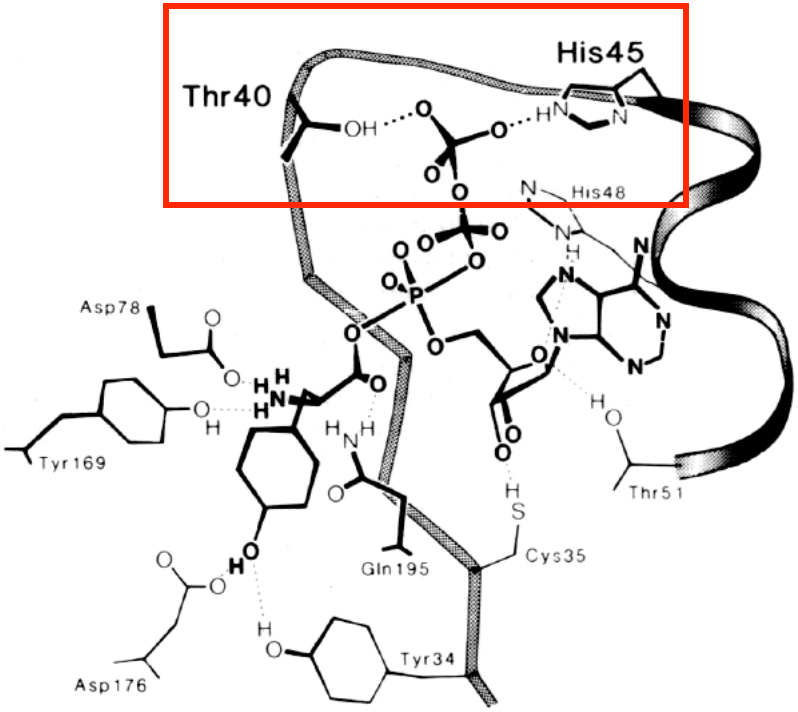
1. We looked at the CaM & GFP structures and made predictions about which point mutations would shift the calcium affinity of pericam.
2. We are now going to produce the mutant genes and proteins, and assay purified molecules for desired properties.
3. If we had more time, we might then go on and make a new round of predictions/mutant proteins, to continue the process of tuning the calcium affinity.

Classic example: tyrosyl-tRNA synthetase, engineered to study mechanism of catalysis

$$E + \text{tyrosine} + ATP \rightarrow E{\cdot}Tyr{-}AMP + PP_i$$

$$E{\cdot}Tyr{-}AMP + tRNA^{Tyr} \rightarrow E + Tyr\text{-}tRNA^{Tyr} + AMP$$





Enzyme	k_3,* s^{-1}	K_S for tyrosine, μM	K_S for ATP, mM
Tyrosyl-tRNA synthetase†	38	12	4.7
Tyrosyl-tRNA synthetase(His-45 → Gly-45)	0.16	10	1.2
Tyrosyl-tRNA synthetase(Thr-40 → Ala-40)	0.0055	8.0	3.8
Tyrosyl-tRNA synthetase(Thr-40 → Ala-40; His-45 → Gly-45)	0.00012	4.5	1.1

Leatherbarrow *et al.* (1985) *Proc. Natl. Acad. Sci. USA 82:* 7840-4

Rational design can also be used to stabilize proteins–general route to improvement of function/utility

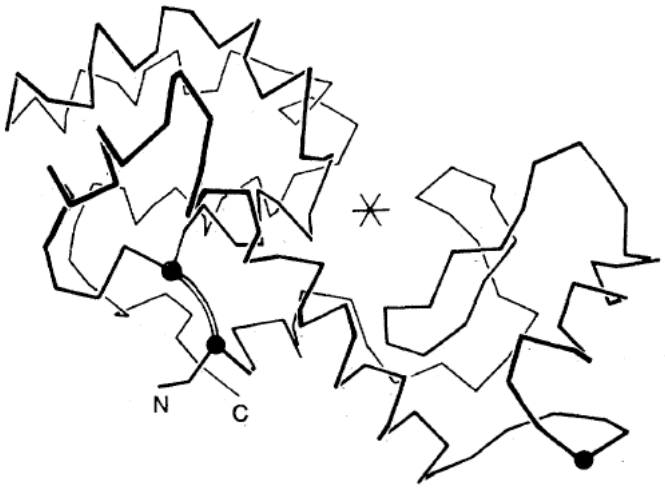
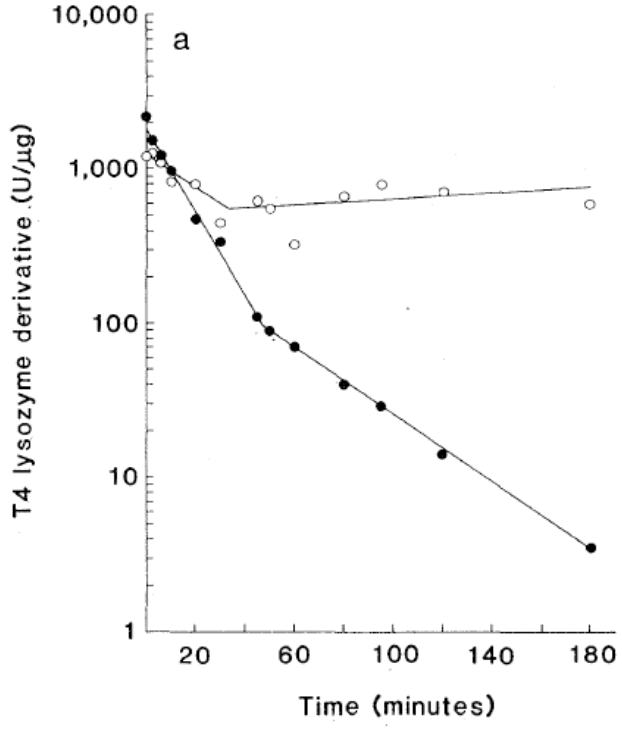


Fig. 1. Computer graphics simulation of T4 lysozyme ($Ile^3 \rightarrow Cys$) α-carbon chain, showing the amino- and carboxyl-chain termini (N and C, respectively), the three cysteines (●), and the active site (star). Cys^3 and Cys^{97} are connected by a schematic disulfide.



Perry & Wetzel (1984) *Science 226:* 555-7

The “holy grail” of rational engineering is to design entire proteins *de novo* to fold into a defined shape (and ideally carry out a function)

Simplest task is to design peptides with defined 2° structure

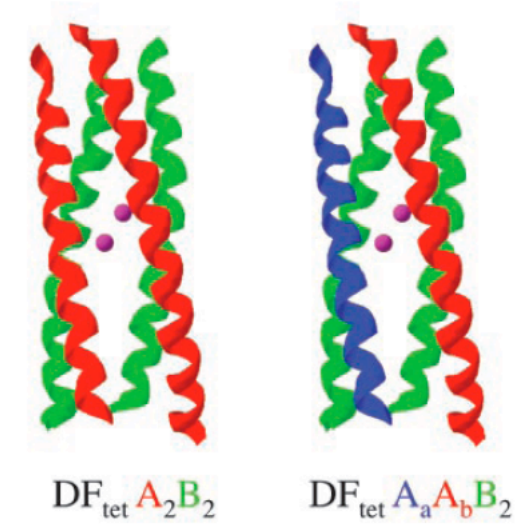
Amino Acid	$f_\alpha{}^b$	$P_\alpha{}^c$	$f_{\alpha i}{}^b$	$P_{\alpha i}{}^c$	$f_\beta{}^b$	$P_\beta{}^c$	$f_c{}^b$	$P_c{}^c$
Ala	0.522	1.45	0.272	1.59	0.167	0.97	0.311	0.66
Arg	0.282	0.79	0.115	0.67	0.154	0.90	0.564	1.20
Asn	0.263	0.73	0.090	0.53	0.113	0.65	0.624	1.33
Asp	0.351	0.98	0.090	0.53	0.137	0.80	0.514	1.09
Cys	0.278	0.77	0.056	0.33	0.222	1.30	0.500	1.07
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

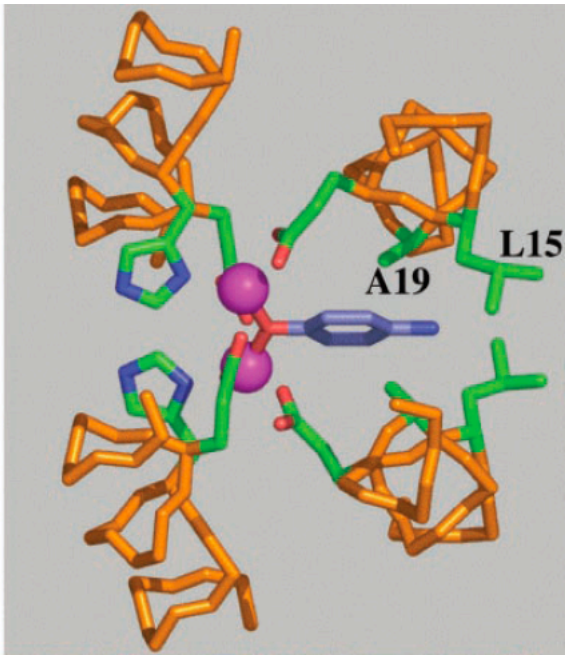
Related task is to predict 2° structure from sequence

```
          .         .         .         .         .         .
     MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADG
helix <--------------------->     <-------->         <----------->
sheet EEEEEEEE                EEEEEEEEEEE           EEEEE
turns   T   TT      T      T  TT    T          TT               T

          .         .         .         .         .         .
     NGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDE
helix    <------------------------------>     <-------------------
sheet     EEEEEEEE                           EEEEEEEEEEE
turns TT           T      T T   T      T  T               T   TT

          .         .
     EVDEMIREADIDGDGQVNYEEFVQMMTAK
helix ---------->    <----------->
sheet                 EEEEEEEEE
turns            T T
```

Chou & Fasman (1974) *Biochemistry 13:* 222-45

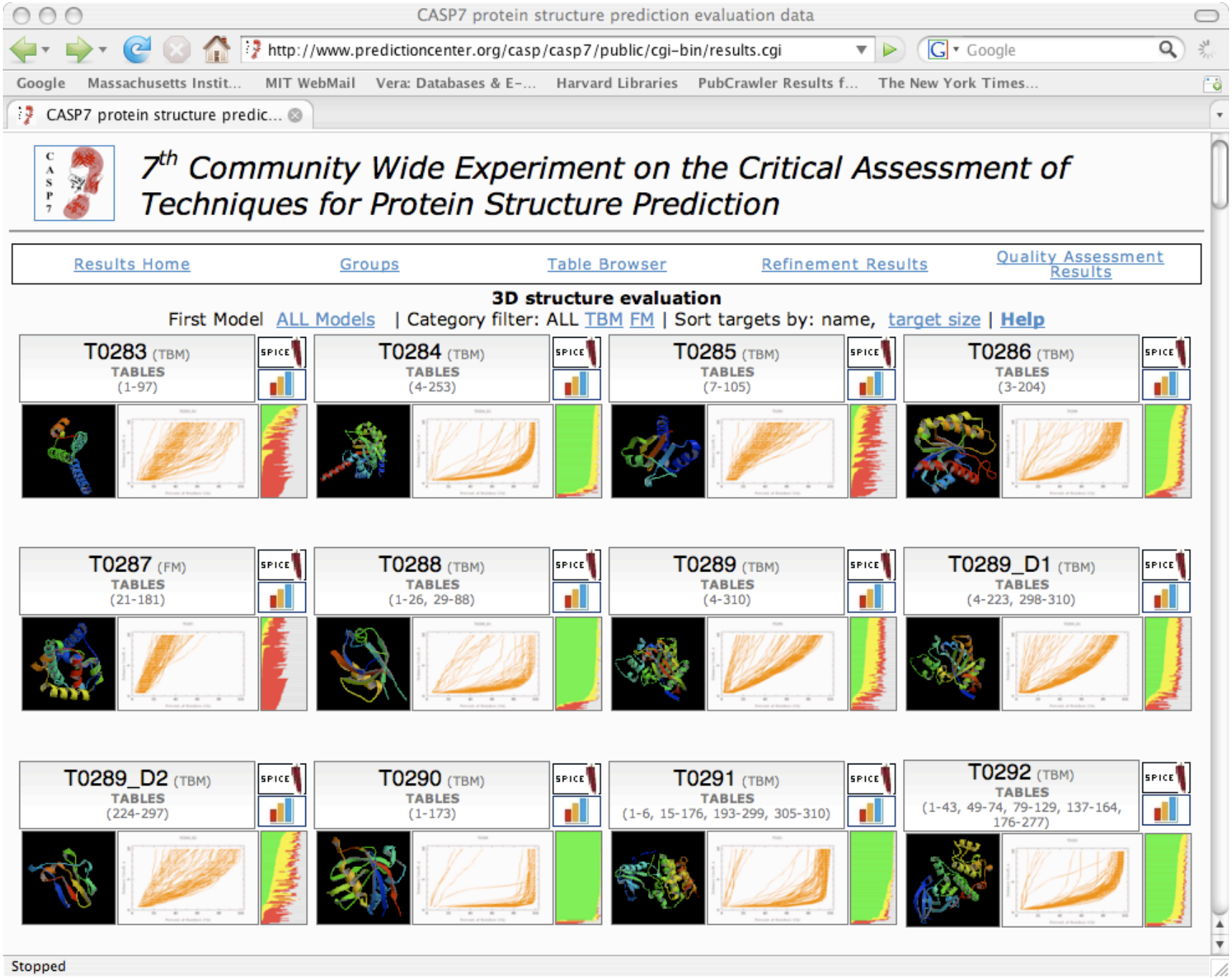
De novo design can be extended to 3° and 4° structure. Example is design of a functional enzyme from so-called coil-coil peptides:






```
                 g abcdefg abcdefg abcdefg abcdefg abcd
DFtetA:  Ac-K LKELKSK LKELLKL ELQAIKQ YKELKAE LKEL-CONH2
DFtetAa: Ac-E LKELKSE LKELLKL ELQAIKQ FKELKAE LKEL-CONH2
DFtetAb: Ac-K LKKLKSR LKKLLKL ELQAIHQ YKKLKAR LKKL-CONH2
DFtetB:  Ac-E LEELESE LEKILED EERHIEW LEKLEAK LEKL-CONH2
```

Kaplan & Degrado (2004) *Proc. Natl. Acad. Sci. USA 101:* 11566-70


CASP7 protein structure prediction evaluation data
http://www.predictioncenter.org/casp/casp7/public/cgi-bin/results.cgi
Google
Massachusetts Instit...
MIT WebMail
Vera: Databases & E-...
Harvard Libraries
PubCrawler Results f...
The New York Times...
CASP7 protein structure predic...
7th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction
Results Home
Groups
Table Browser
Refinement Results
Quality Assessment Results
3D structure evaluation
First Model ALL Models | Category filter: ALL TBM FM | Sort targets by: name, target size | Help
T0283 (TBM) TABLES (1-97)
T0284 (TBM) TABLES (4-253)
T0285 (TBM) TABLES (7-105)
T0286 (TBM) TABLES (3-204)
T0287 (FM) TABLES (21-181)
T0288 (TBM) TABLES (1-26, 29-88)
T0289 (TBM) TABLES (4-310)
T0289_D1 (TBM) TABLES (4-223, 298-310)
T0289_D2 (TBM) TABLES (224-297)
T0290 (TBM) TABLES (1-173)
T0291 (TBM) TABLES (1-6, 15-176, 193-299, 305-310)
T0292 (TBM) TABLES (1-43, 49-74, 79-129, 137-164, 176-277)
SPICE
Stopped

What knowledge is required for “high-resolution” protein engineering?:

- determination of 3D structure, for mutagenesis-based engineering
- knowledge of protein folding rules for *de novo* engineering
- computational modeling techniques usually required

Computational methods important for protein engineering:

- modeling & visualization
- energy/thermodynamic calculations
- searching conformation and sequence spaces
- comparison with known protein structures/sequences

The basis of more automated analysis of structural perturbations than our own “inspect and try” approach involves use of an energy function to evaluate plausibility of candidate structures:

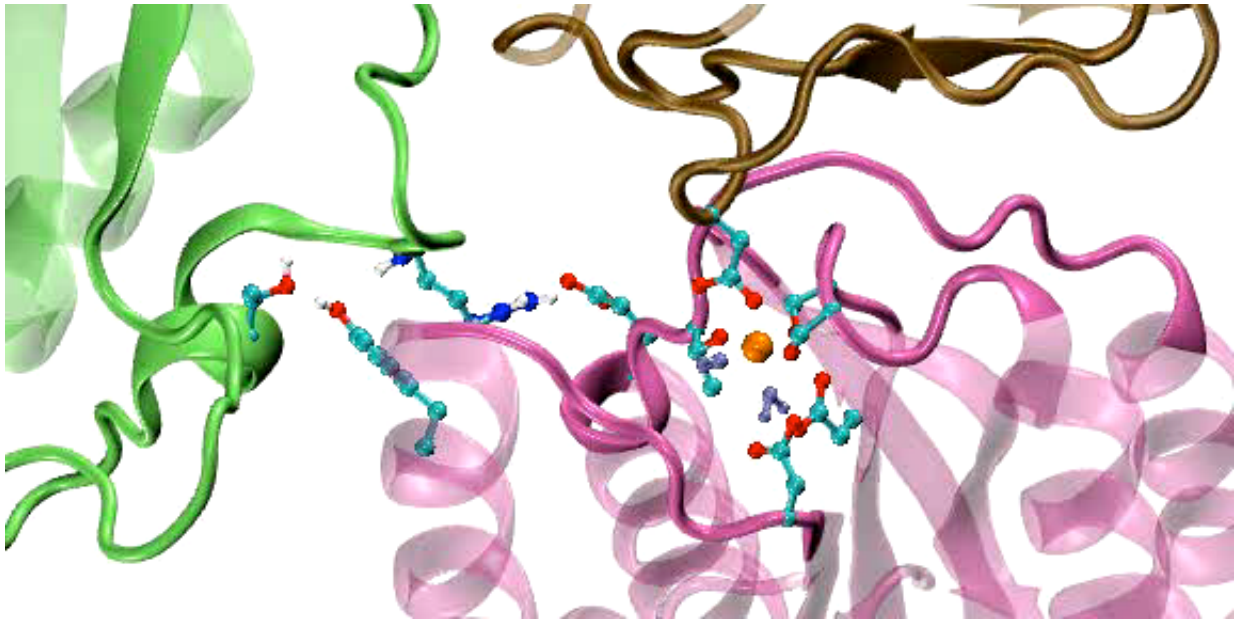
$$E_{tot} = E_{bond} + E_{angl} + E_{dihe} + E_{impr} + E_{VDW} + E_{elec} + E_{Hbond} + \ldots$$

This may be evaluated using a force field (*e.g.* CHARMM19) and atomic coordinates available from simulation or modified PDB file.

Computational techniques for investigation of specific structures:

- molecular dynamics: simulate physically plausible movements of a protein, with a "rule" that describes probability of motions in conjunction with the energy function at a given temperature
- energy minimization: gradually perturb a model protein structure to find a locally favorable structure (energy minimum) in the neighborhood of a starting structure
- both techniques can be applied after *in silico* mutagenesis, *e.g.* to anticipate the effect of mutation on stability or ligand binding

simulation of anthrax toxin dissociating from its receptor



www.ks.uiuc.edu/Gallery/Movies/