Protein core

The amino acids in the interior are on average more hydrophobic than the residues on the surface

Protein core is the part of a folded protein with **zero solvent accessibility**

Accessible surface area (ASA) corresponds to the area traced out by the center of a probe sphere of radius 1.4 Å as it is rolled over the surface of the molecule of interest

"**NACCESS**" from J. Thornton computes the ASA

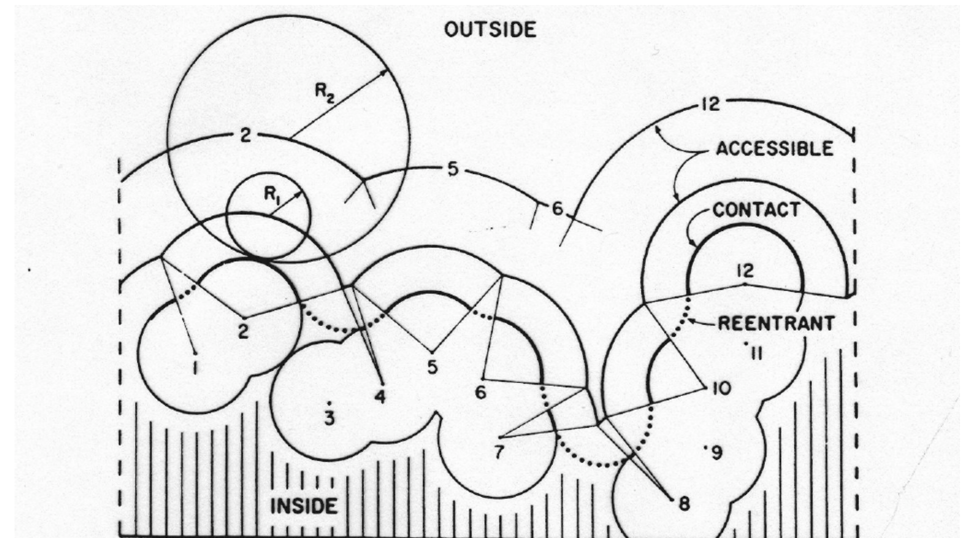


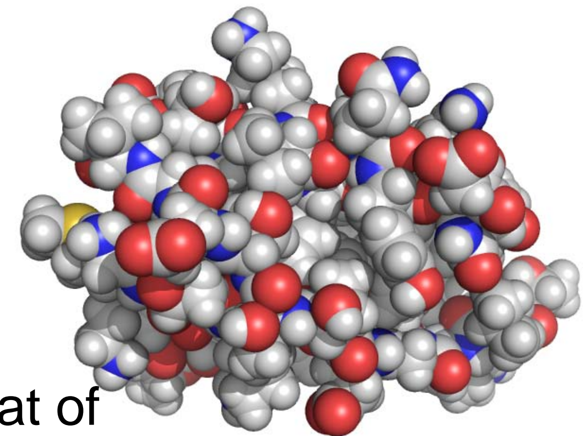
FIG. 7. Schematic representation of possible molecular surface definitions. A section through part of the van der Waals envelope of a hypothetical protein is shown with the atom centers numbered. The accessible surfaces generated by two probes of difference size, R_1 and R_2, and the geometrical definition of contact and reentrant surfaces are shown. (Reproduced with permission from Richards.[22])

Example output of naccess

```
REM  Relative accessibilites read from external file "standard.data"
REM  File of summed (Sum) and % (per.) accessibilities for
REM RES _ NUM      All-atoms   Total-Side   Main-Chain    Non-polar    All polar
REM                ABS   REL    ABS   REL    ABS   REL    ABS   REL    ABS   REL
RES LEU A 999   234.57 131.3 158.83 112.5  75.74 201.9 159.59 112.1  74.98 206.4
RES MET A1000   153.35  79.0 137.88  88.0  15.47  41.2 139.63  88.5  13.72  37.8
RES MET A1001   193.67  99.8 166.14 106.1  27.53  73.4 166.14 105.3  27.53  75.8
RES HIS A1002   133.92  73.2 118.50  80.6  15.42  43.1  80.36  82.7  53.56  62.5
RES SER A1003    56.05  48.1  37.72  48.3  18.33  47.7  32.20  66.3  23.86  35.1
RES GLN A1004   164.98  92.4 147.81 104.8  17.16  45.8  56.66 108.5 108.31  85.8
RES LYS A1005    72.98  36.3  72.98  44.7   0.00   0.0  47.24  40.5  25.74  30.6
RES ARG A1006   133.72  56.0 128.23  63.7   5.49  14.6  40.64  52.2  93.09  57.8
RES VAL A1007     1.54   1.0   0.00   0.0   1.54   4.1   0.22   0.2   1.33   3.7
RES VAL A1008     2.87   1.9   2.87   2.5   0.00   0.0   2.87   2.5   0.00   0.0
RES VAL A1009     0.00   0.0   0.00   0.0   0.00   0.0   0.00   0.0   0.00   0.0
RES LEU A1010     0.00   0.0   0.00   0.0   0.00   0.0   0.00   0.0   0.00   0.0
```

```
ATOM      1  N   LEU A 999       0.041 148.800  54.967 49.906  1.65
ATOM      2  CA  LEU A 999       0.374 147.376  55.166  6.484  1.87
ATOM      3  C   LEU A 999       0.199 146.548  53.888  0.757  1.76
ATOM      4  O   LEU A 999      -0.859 146.033  53.570 25.076  1.40
ATOM      5  CB  LEU A 999      -0.477 146.729  56.231 17.074  1.87
ATOM      6  CG  LEU A 999      -0.139 146.132  57.568 14.939  1.87
ATOM      7  CD1 LEU A 999       0.286 144.647  57.433 58.243  1.87
ATOM      8  CD2 LEU A 999       0.947 146.895  58.330 62.090  1.87
ATOM      9  N   MET A1000       1.294 146.400  53.124  4.553  1.65
ATOM     10  CA  MET A1000       1.193 145.597  51.902  3.260  1.87
ATOM     11  C   MET A1000       1.681 144.162  52.120  1.747  1.76
ATOM     12  O   MET A1000       2.482 143.861  53.013  9.171  1.40
ATOM     13  CB  MET A1000       2.016 146.267  50.824  1.099  1.87
ATOM     14  CG  MET A1000       1.364 147.492  50.072 33.384  1.87
ATOM     15  SD  MET A1000       2.921 148.211  49.477 36.200  1.85
ATOM     16  CE  MET A1000       3.093 147.338  47.917 63.938  1.87
```

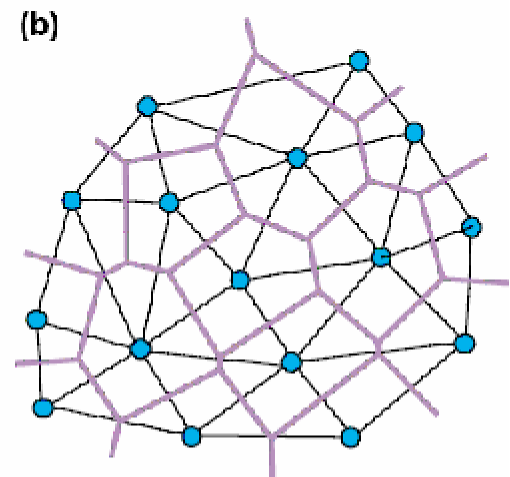
Crystal structures show that protein main chain and side chains occupy almost every available space



The interior has a packing density comparable to that of organic solid and is denser than organic liquid

- packing density = van der Waals volume/Voronoi volume
  - van der Waals volume is the volume actually taken up by atoms
  - Voronoi volume is the sum of the atomic volume, inner voids and the surface of empty spaces (as defined by the molecular surface)
- organic solid 0.68 – 0.8
- protein 0.72 – 0.77

If the core is so well packed, then is it difficult to find other combinations of amino acids to repack the protein core just right?



Poupon, COSB 14, 233 (2004)

Core residues are often conserved among related proteins

Structure and **stability** of a protein are strongly dependent on the arrangements and compositions of the hydrophobic core

The tendency for hydrophobic groups to aggregate is seen as "**hydrophobic bond**"

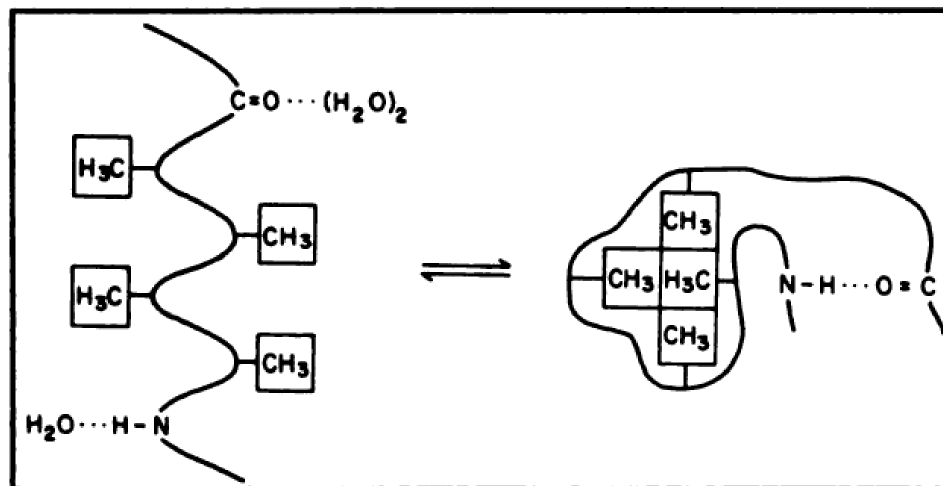
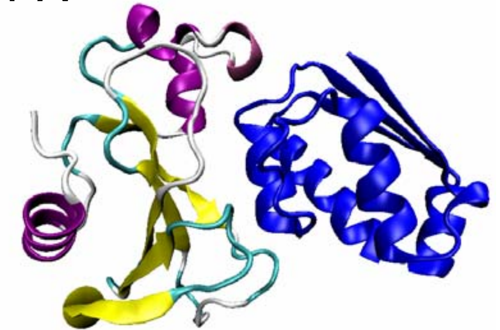


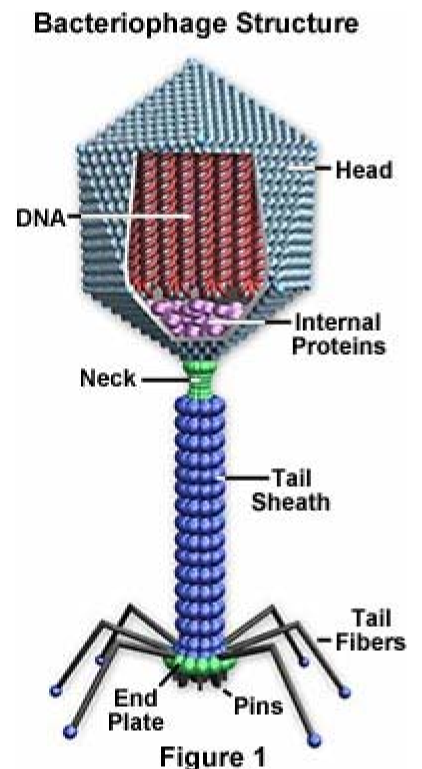
Figure 1. Schematic illustrating the formation of a hydrogen bond and a hydrophobic bond in the folding of a protein.

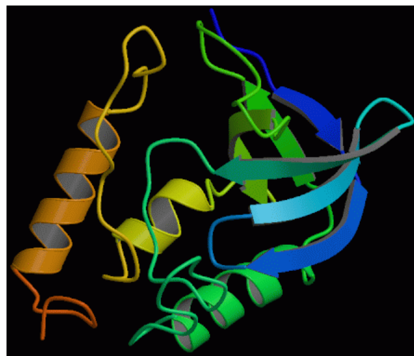
Pace et al, Faseb J 10, 75 (1996)

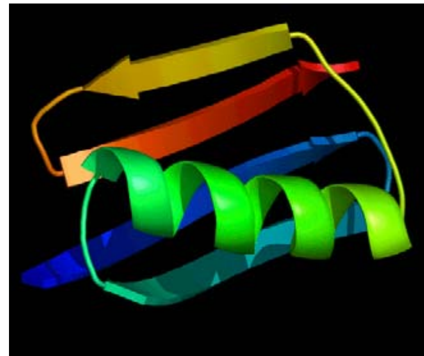
Model proteins in protein engineering studies

- **T4 lysozyme** : hydrolyzes the glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine of peptidoglycan in (bacterial) cell wall
- **Hen egg white lysozyme**
- **Barnase/barstar** : ribonuclease from bacteria Bacillus
  – hydrolyzes RNA
- **Ribonuclease A**
- **Staphylococcus nuclease**
- **Protein G, L** : binds immunoglobulin
- **Lambda repressor** : binds DNA from bacteriophage
- **Bovine pancreatic trypsin inhibitor (BPTI)**









Core packing

What factors are important when designing the protein core?

The core may either be described as:

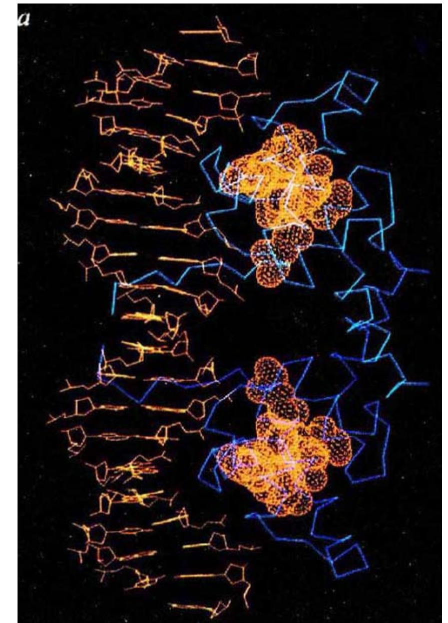
–**3D jigsaw puzzle**, where every amino acid side chain has a unique place to go

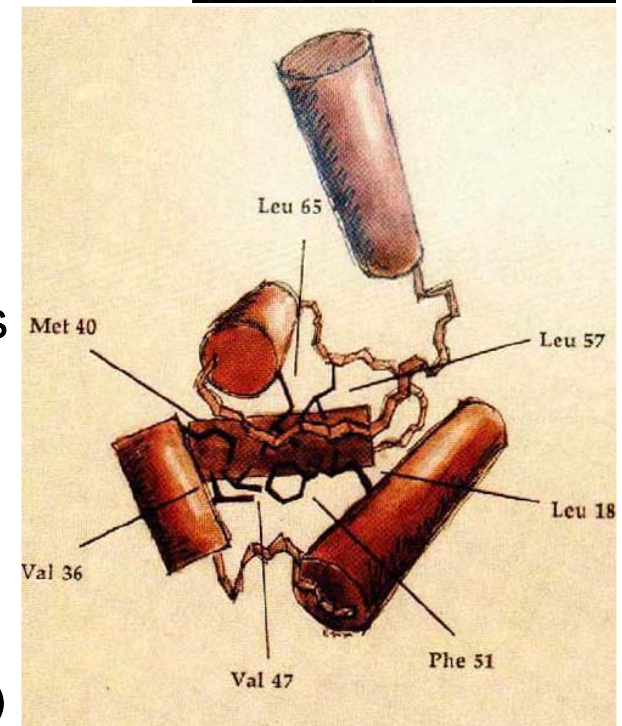
–"**Oil drop**" that can alternate among many equivalent packing arrangements

Introduce random substitutions at seven positions corresponding to the protein core of the N-terminal domain of lambda repressor, and select for mutants that retain activity

Assay: functional lambda repressor will make bacteria resistant to lambda phage

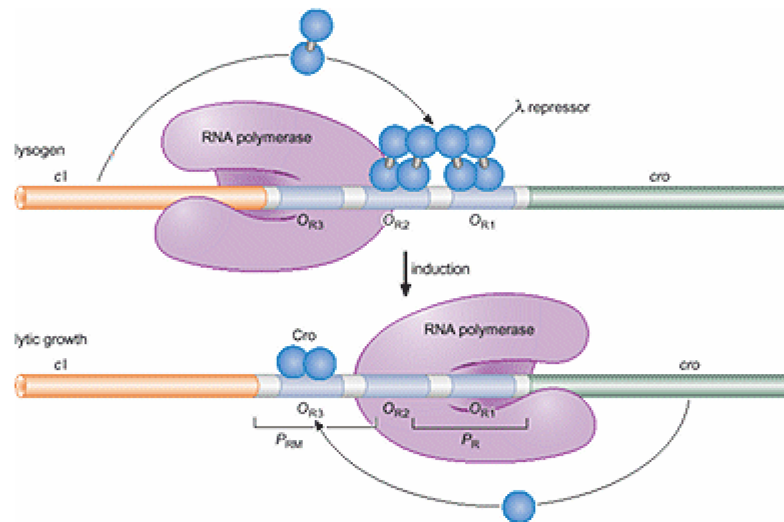
Lim & Sauer, Nature 339, 31 (1989)

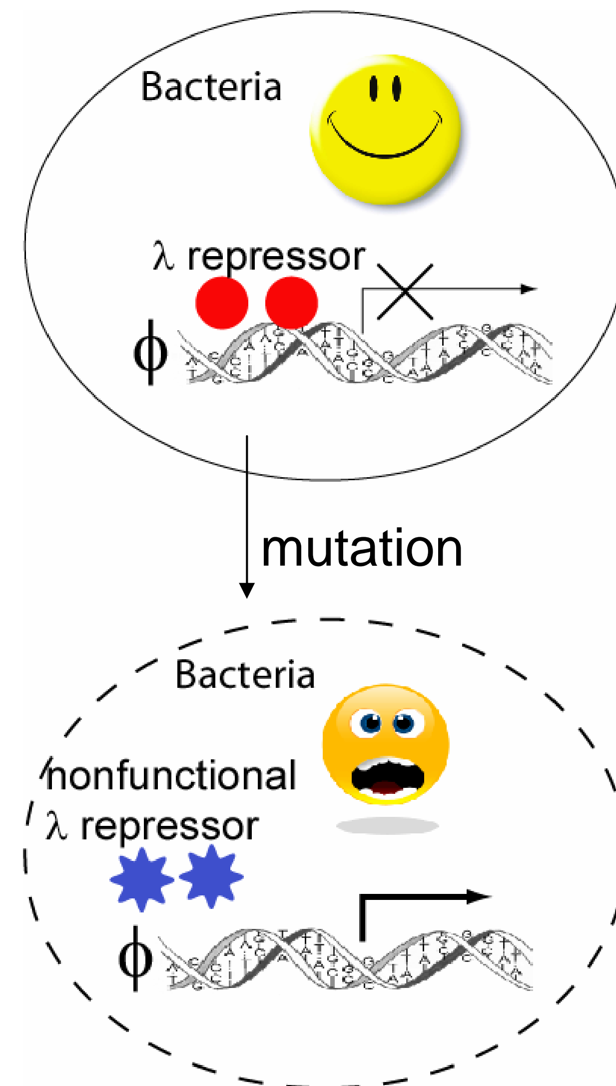




High throughput assay

development is often critical to the success of many protein engineering projects





The substitutions that retain activity consists of substitutions with :

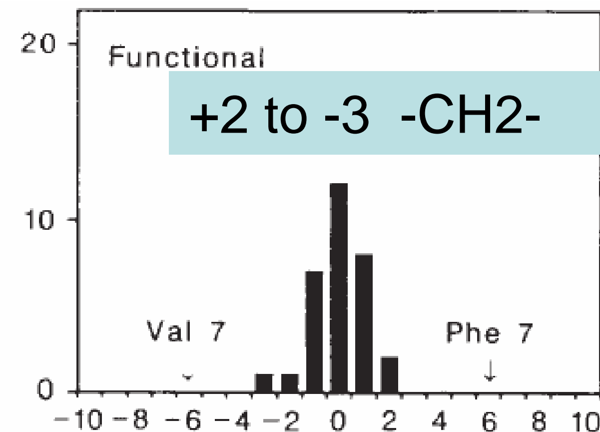
- Ala, Cys, Thr, Ile, Val, Leu, Met, Phe
- Favorable free energies of transfer from water to organic solvent

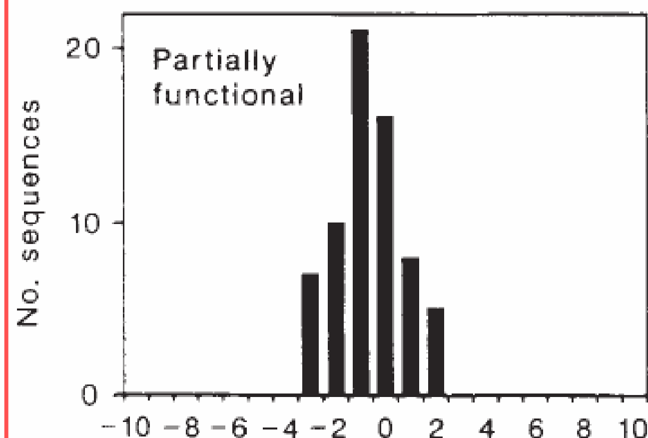
Combination of volume, hydrophobicity, and steric is a good indicator of functional sequence

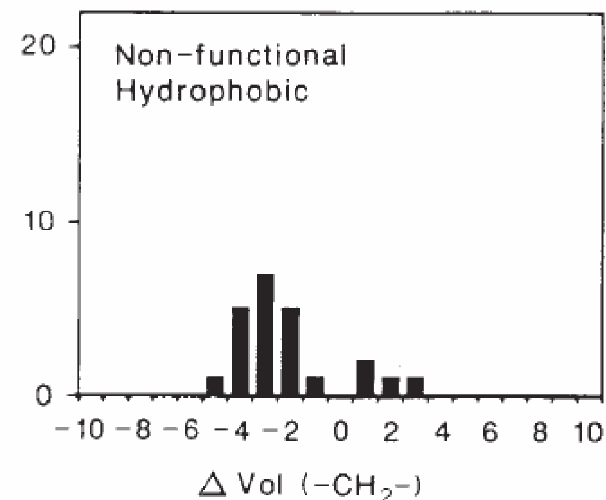
Fully functional

18	36	40	47	51	57	65	Vol	$\Sigma\Delta G_t$
L	V	M	V	F	L	L	549	14.4
L	**I**	M	V	F	L	L	567	15.2
L	V	**L**	V	F	L	L	549	15.0
L	V	M	**I**	F	L	L	567	15.2
L	V	M	V	**L**	L	L	538	14.3
L	V	M	V	**M**	L	L	538	13.6
L	V	M	V	F	**I**	L	549	14.5
L	V	M	V	F	**M**	L	549	13.8
L	V	M	V	F	L	**F**	560	14.5
L	V	M	V	F	L	**M**	549	13.8
L	V	M	V	F	L	**V**	531	13.7
L	**I**	M	V	**L**	L	L	556	15.1
L	**M**	M	V	**I**	L	L	556	14.4
L	V	**L**	**I**	F	L	L	567	15.8
L	**L**	M	**I**	F	L	L	585	15.9
L	V	M	**L**	**L**	L	L	556	14.9
L	V	M	**T**	**L**	L	L	526	13.0
A	**L**	M	V	**L**	L	L	500	13.0
I	**L**	M	V	**L**	L	L	556	15.1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	518	13.7
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	548	15.7
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	585	16.6
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	574	15.7
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	574	15.6
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	574	14.5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	544	13.6
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	556	14.4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	538	14.3
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	574	15.7
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	526	13.1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	L	538	14.4

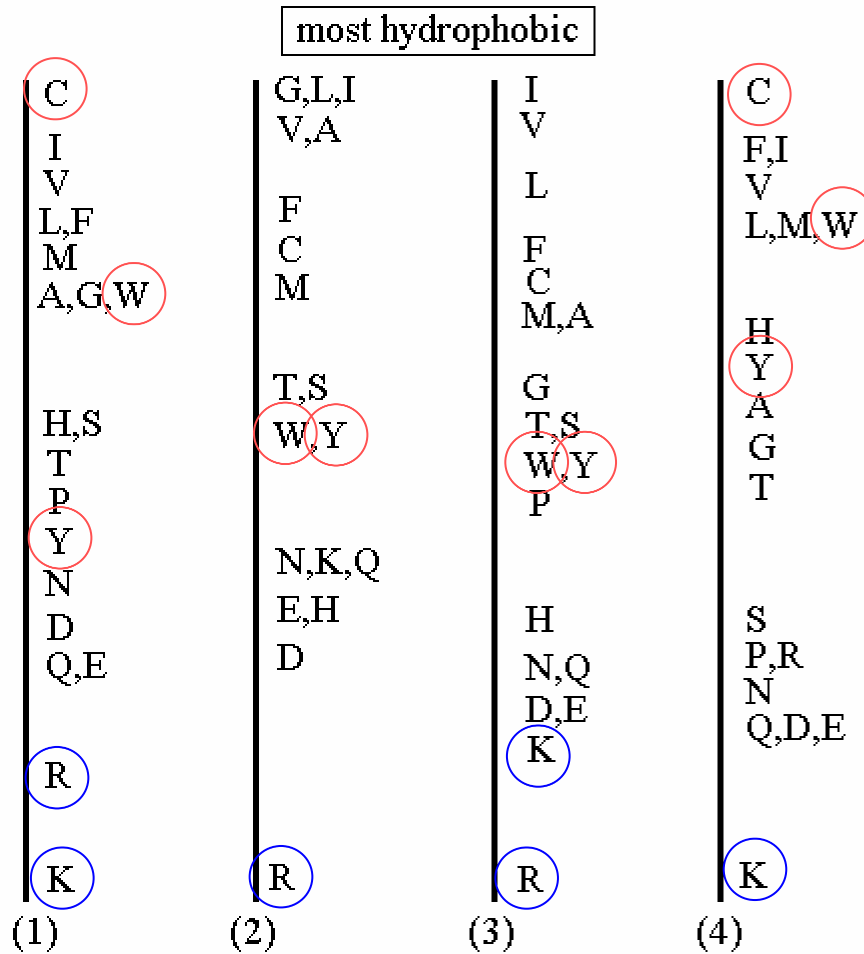
Applied constraints	Experiment	I (V36 M40 V47)
None		8,000 (100%)
Volume		5,848 (73%)
Composition		512 (6%)
Volume and composition		425 (5%)
Volume, composition, and steric:		
active sequences		~110 (1.4%)
fully functional sequences		~20 (0.3%)







Hydrophobicity scale



1. Janin, Nature, 277(1979)491
2. Wolfenden et al, Biochemistry 20(1981)849
3. Kyte and Doolite, JMB 157(1982)105
4. Rose et al, Science 229(1985)834

Designing the hydrophobic core

Barnase is a ribonuclease from bacteria *Bacillus*

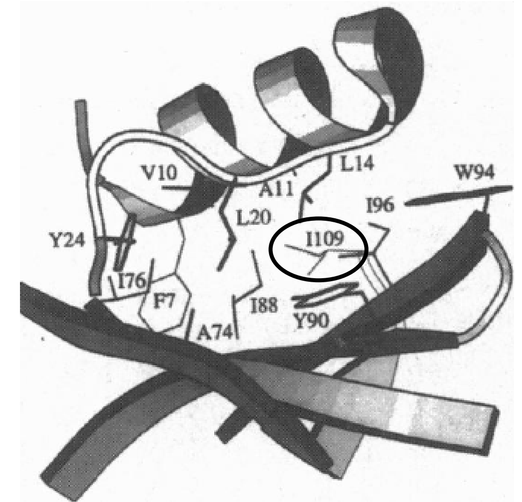
Assay: If barnase is expressed in the absence of its inhibitor barstar, the protein will degrade RNA in the cell and thus kill the cell

The assay is sensitivity enough to detect a mutant protein with > 0.2% of the activity of wild type

Randomly mutate 12 of the 13 core residues to other hydrophobic residues

23% of all mutants retained enzymatic activity

Hydrophobicity is a sufficient criterion for constructing a core that is capable of supporting enzymatic activity



Axe et al. PNAS 93, 5590, (1996)

barnase

How does protein respond to core mutations

Substituting a hydrophobic residue with another hydrophobic residue may induce a volume change

Less disruptive substitutions (i.e. large residue to small residue) also introduces a volume change which further destabilizes the mutant

Construct six “cavity creating” mutants of T4 lysozyme

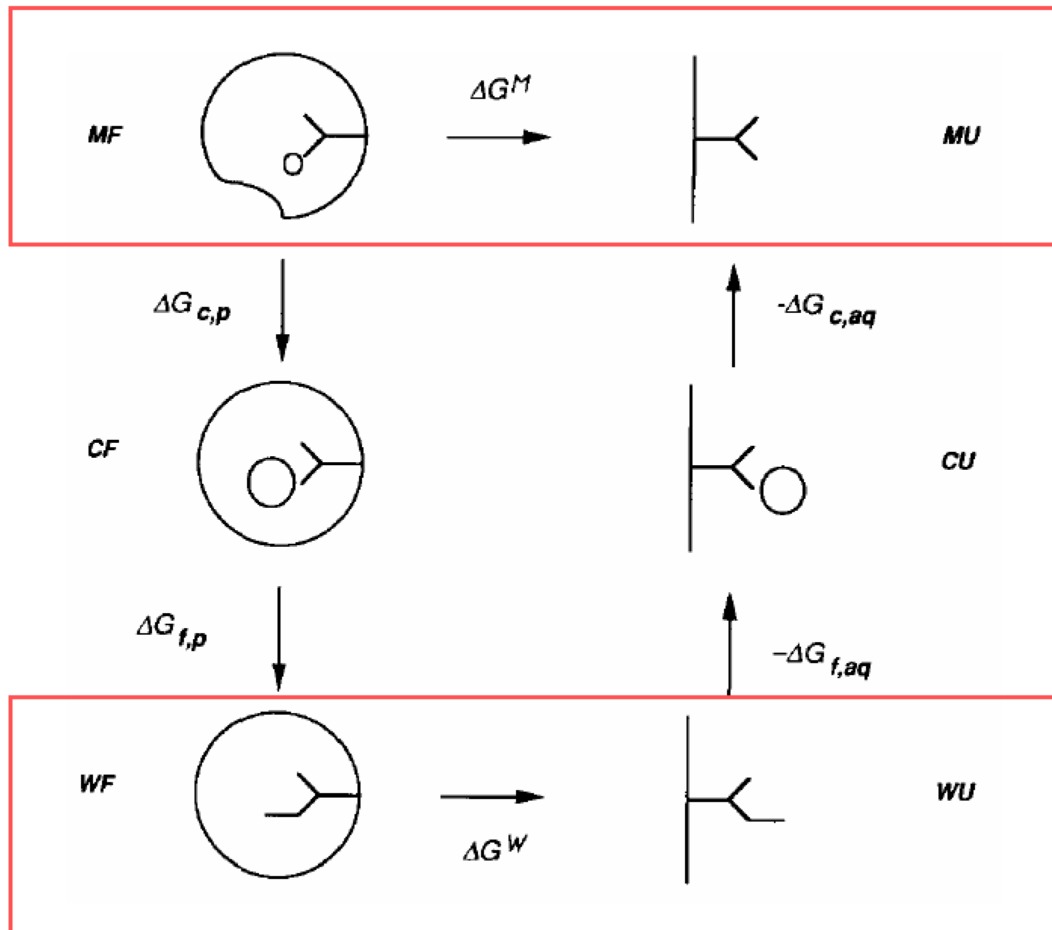
L46A, L99A, L118A, L121A, L133A, F153A

Lysozyme	T_m (°C)	ΔH (kcal mol^{-1})	$\Delta\Delta G$ (kcal mol^{-1})
WT	53.5	130.0	0.7
WT*	51.8	118.8	0.0
L46A	43.2	89.8	−2.7
L118A	39.6	75.5	−3.5
L121A	42.5	81.0	−2.7
L99A	36.1	79.6	−5.0
L133A	42.9	91.5	−3.6
F153A	39.5	74.9	−3.5
L99A/F153A	10.0	6.0	−8.3

Eriksson et al, Science 255, 178 (1992)

Thermodynamic cycle

How to rationalize the stability difference between mutant with wild type



Lee, Protein Sci 2, 733 (1993)

For Leu ➔ Ala mutations, a cavity of ~ 24 Å^3 remains

For Phe ➔ Ala mutation, a cavity of ~ 150 Å^3 remains

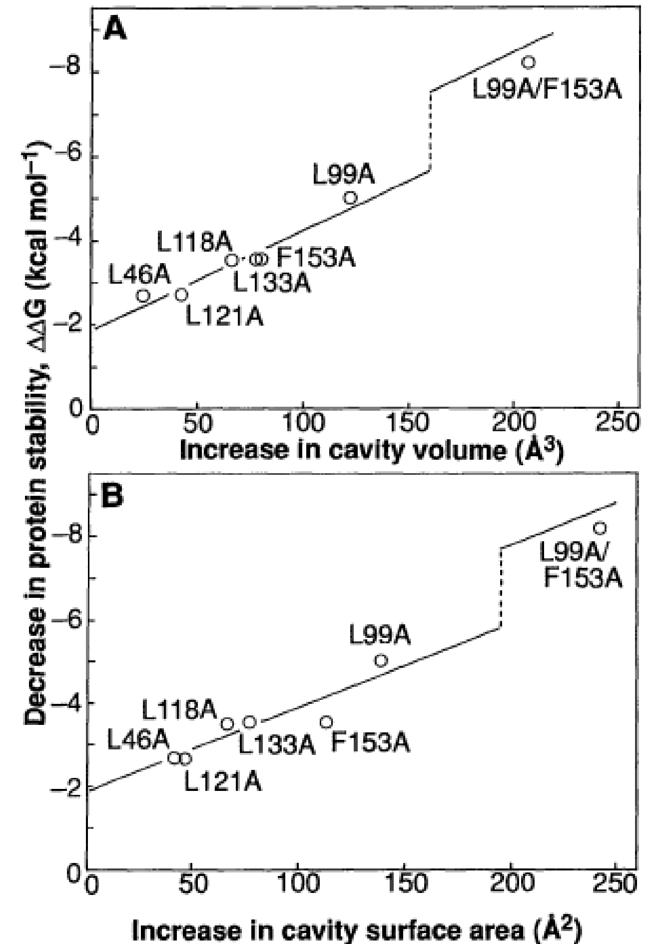
Parameterize the relationship between the decrease in stability ($\Delta\Delta G$) and the cavity volume (ΔV) or cavity surface area (ΔS) by a straight line

$$\Delta\Delta G = a + b\Delta V$$

$$\Delta\Delta G = c + d\Delta S$$

a = c = - 1.9 kcal/mol
(roughly the transfer free energy of leucine from water to organic solvent with respect to alanine)

b = - 0.024 kcal/mol/Å^3
d = - 0.020 kcal/mol/Å^2



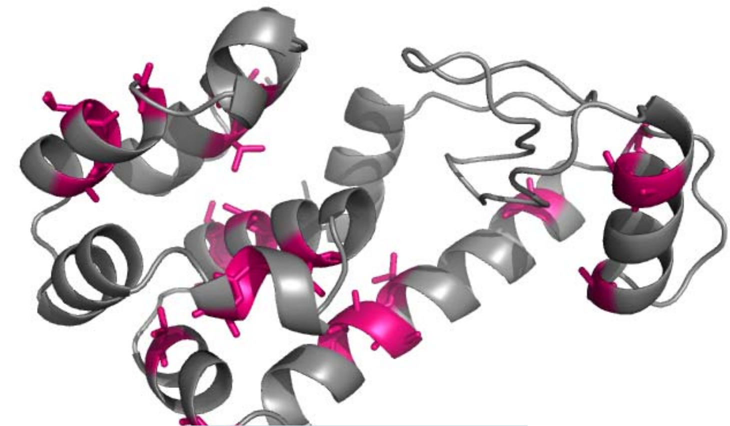
Energetics of burying a hydroxyl group

Introduce a hydroxyl group by replacing 9 Ala with Ser and 3 Val with Thr in T4 lysozyme

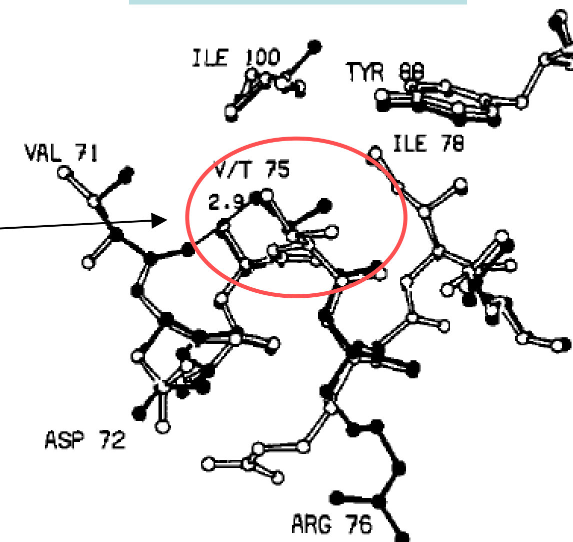
Measure the stability of the mutants in vitro

Most destabilizing mutation : V149T ($\Delta\Delta G$ = 2.8 kcal/mol)

Crystal structures of the mutants are similar to that of wild type



T4 lysozyme

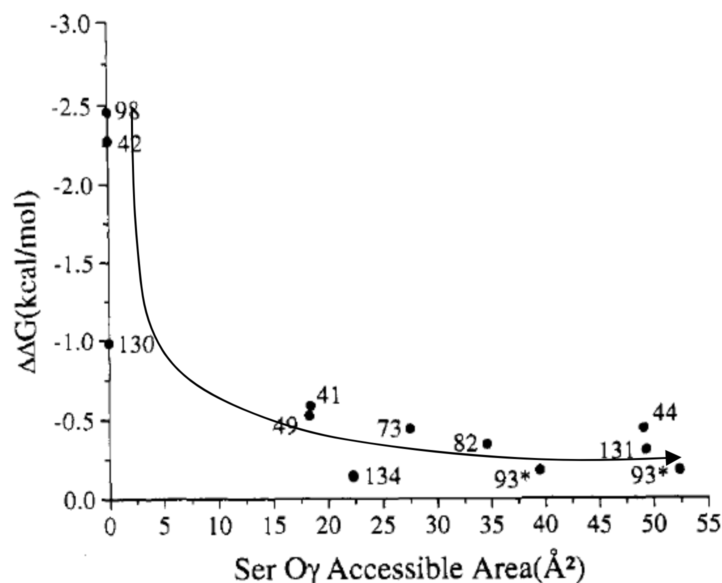


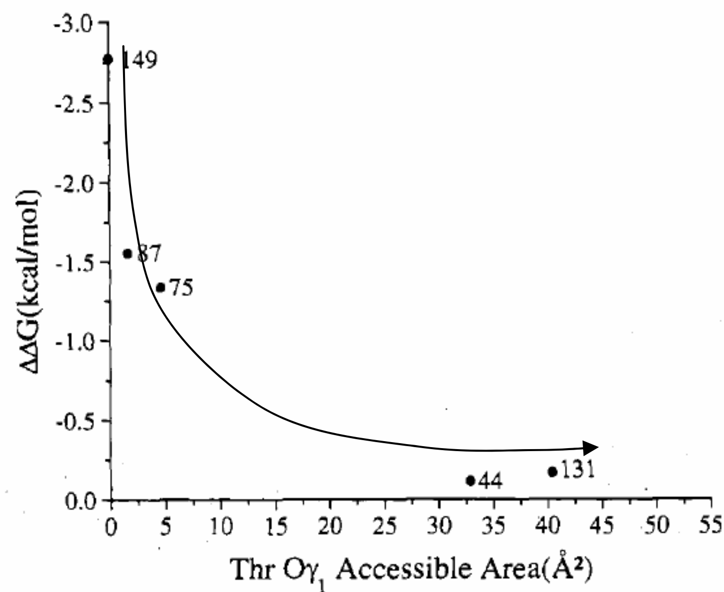
Blaber et al, Biochemistry 32, 11363, (1993)

The cost of burying a hydroxyl group depends on solvent accessibility

If the residue is exposed, the mutant is destabilized by < 0.5 kcal/mol

If the residue is fully buried, the mutant is destabilized by ~ 1- 3 kcal/mol



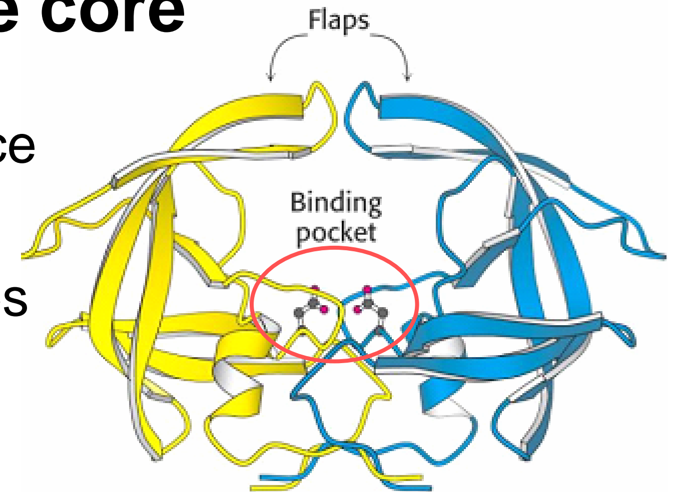


Blaber et al, Biochemistry 32, 11363, (1993)

Charged residues in the core

Charged residues are usually found on the surface

Exceptions include catalytically important residues



HIV protease

How bad is it to bury a charged residue in the core?

- Depends on the polarizability of the core
- Typically epsilon (dielectric constant) is assumed to be 2 – 8
  - Gilson et al, JMB 184, 503 (1985)
- But can be as high as 12
  - Dwyer et al, Biophys J 79, 1610 (2000)

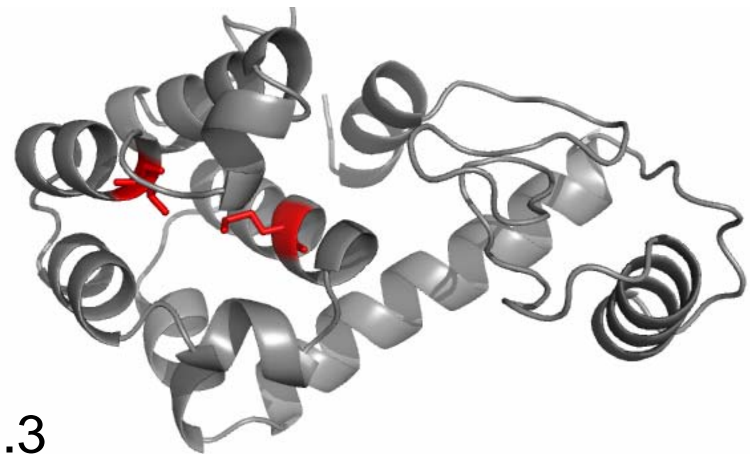
$$E_{elec} \propto \frac{q_1 q_2}{\varepsilon r_{12}} = \frac{\left(q_1 / \sqrt{\varepsilon} \right)\left(q_2 / \sqrt{\varepsilon} \right)}{r_{12}} = \frac{\widetilde{q_1}\,\widetilde{q_2}}{r_{12}}$$

Energetics of burying a charged group

What are the energetic and **structural** consequences of burying a charged group in the core?

Introduce a charged residue in the core of T4 lysozyme

- Met102➔K (M102K) : 35% activity
- Leu133➔D (L133D) : 4% activity



T4 lysozyme

M102K is less stable by 6.9 kcal/mol at pH 5.3

pKa of K102 ~ 6.5

L133D is less stable by 5.7 kcal/mol at pH 6.5

pKa of D133 ~ 6.2

Dao-pin et al, Biochemistry 30, 11521 (1991)

Buried salt bridges

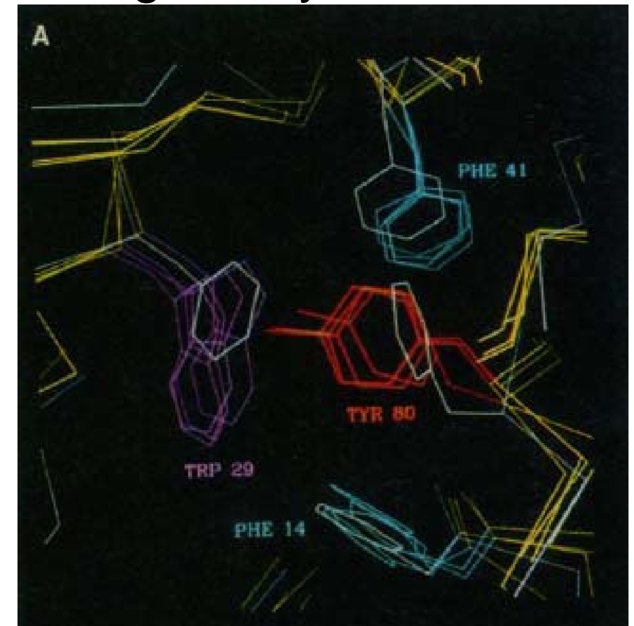
Hydrophilic and charged residues are tolerated in the core when they pair up and all hydrogen bonding needs are met

– Ion pairs or salt bridge network

Are buried ionic pairs equivalent to hydrophobic pairs?

Is Arg *** Glu (salt bridge) equivalent to a van der Waals pair, e.g. Leu – Val ?

Computation suggests burying salt bridges may be energetically unfavorable since it involves desolvation of ionic groups



Moore et al, Science 240, 314 (1988)

Arc repressor of bacteriophage P22 is a homodimeric DNA binding protein

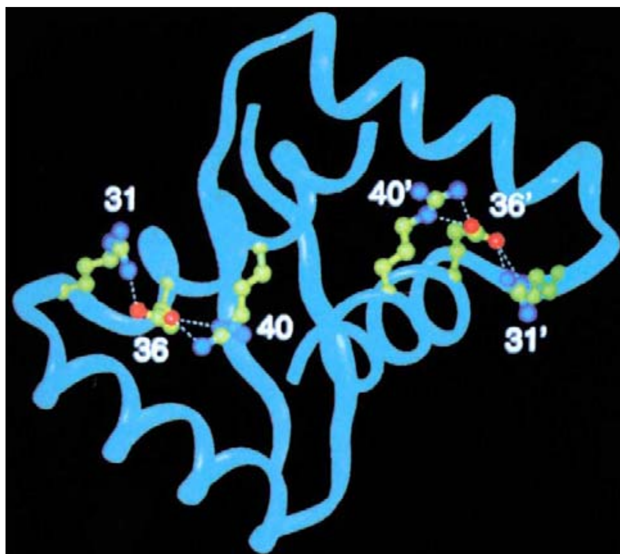
Contains a buried salt bridge network contiaining

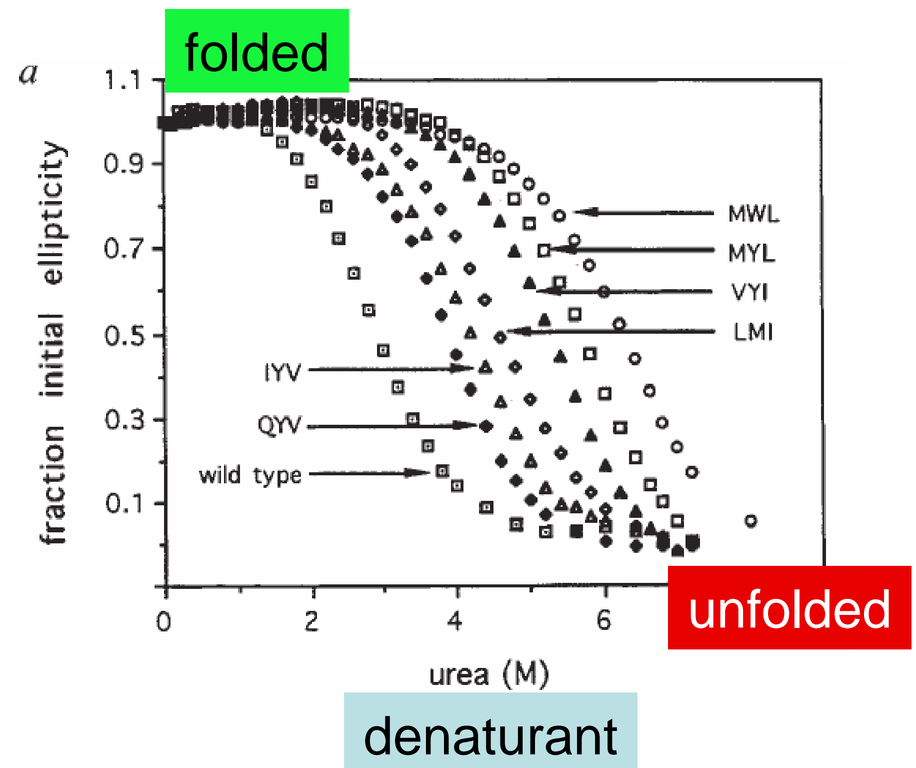
- R31 (37% solvent accessible)
- D36 (0%), and R40 (27%)

Introduce random mutations at these three positions and look for activity

20 x 20 x 20 = 8,000 possible mutants

Assay : functional Arc dimer will make E.coli resistant to P22



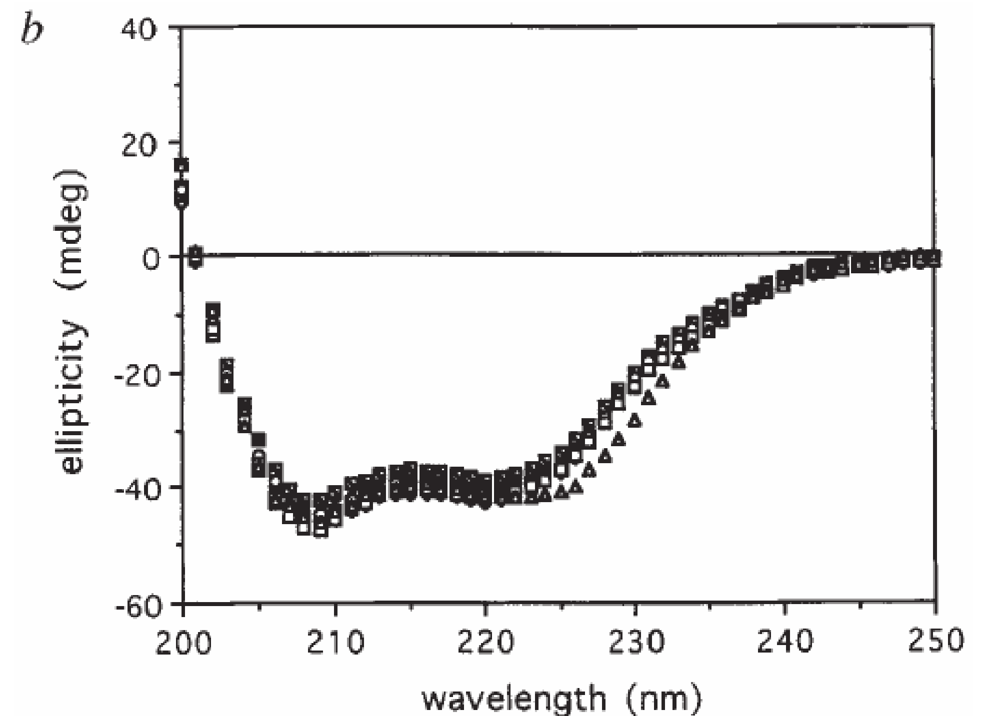


Waldburger, et al NSB 2, 122 (1995)

M31-Y36-L40 is 3.9 kcal/mol more stable than wild type and has full activity

Simple hydrophobic interaction can contribute more to stability than buried salt bridges, while offering conformational specificity required for function

However, internal salt bridges may confer specificity by discriminating against alternate conformations



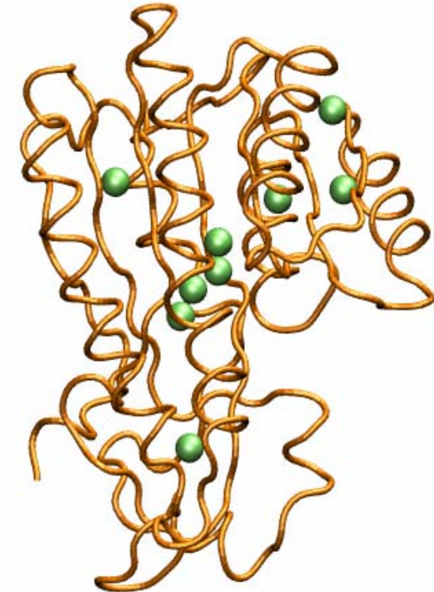
Buried water

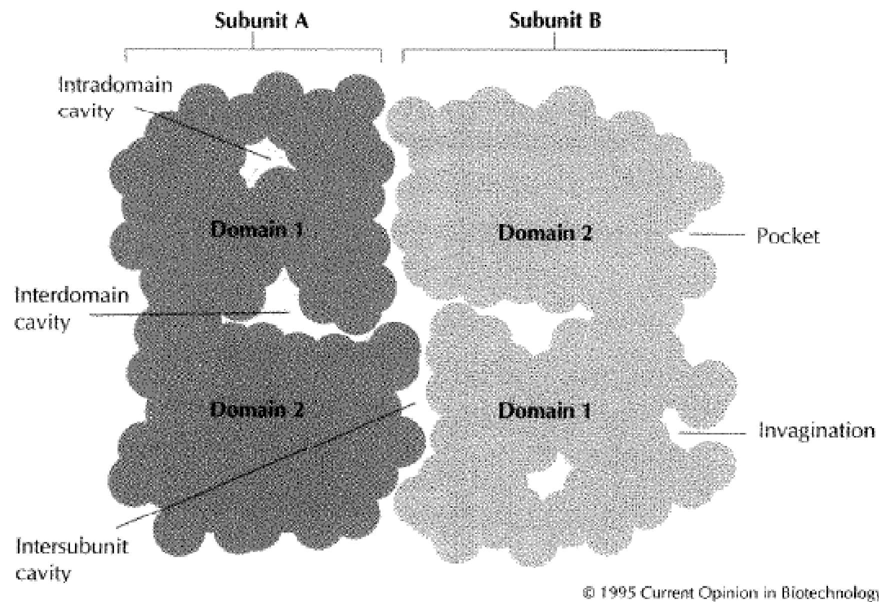
There are many “packing defects” and cavities in the protein core

Some cavities contain water molecules

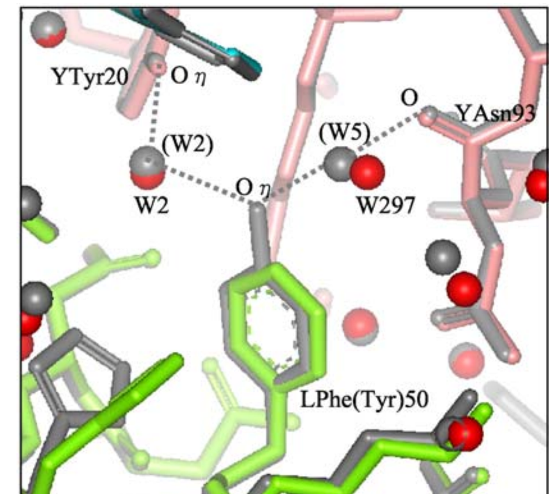
Hubbard et al., Protein Eng 7, 613 (1994).

Williams et al., Protein Sci 3, 1224 (1994)





Hubbard & Argos, Curr Op in Biotech 6, 375 (1995)



Structural stability from bound water

Bound water is close to energy neutral

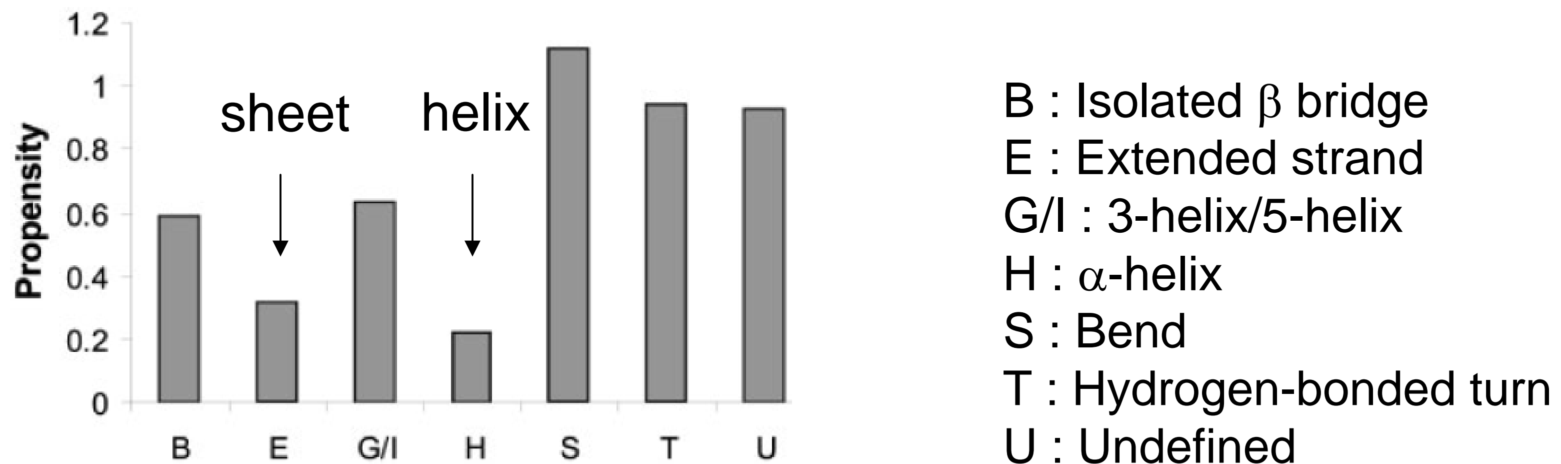
entropic cost of immobilizing a water molecule ~ 2 kcal/mol at 300 K

All amino acids (even hydrophobic residues) contain polar atoms

Buried polar atoms must be H-bonded

Water can satisfy the H-bonding needs of turn/loop/coil residues

Distribution of bound water varies with secondary structure



Park & Saven, Proteins 60, 450 (2004)

Destabilizing barnase with a single mutation

Identifying a single point mutation that abolishes function can provide information on the sequence-structure-function relation

All inactivating substitutions are:

- replacement of a catalytically important side chain
- replacement of a substantially buried side chain
- introduction of Pro residue
- replacement of a Gly residue

Use a functional assay based on the RNase activity of barnase to identify structure-function coupling that isn't apparent from the structure

G52V mutant has 1000 fold reduction in activity in vivo, and is destabilized by 8.4 kcal/mol compared to wt

Destabilization is likely to due to steric clash with neighboring side chains

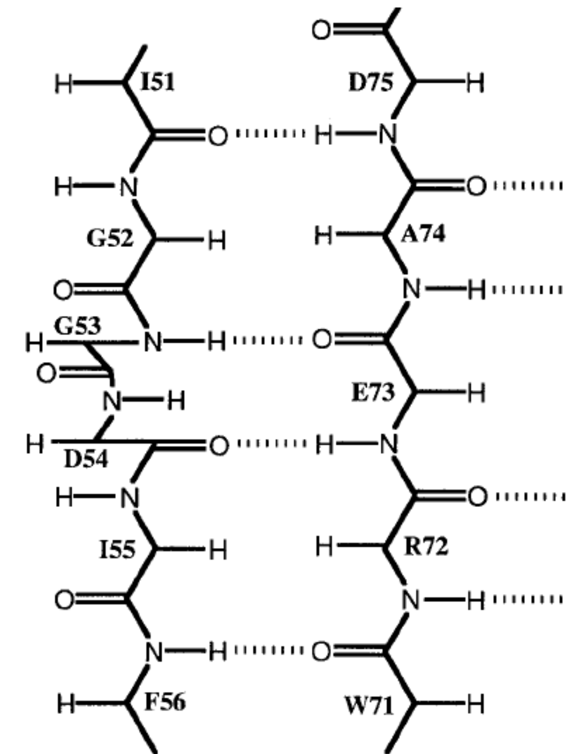


Figure 1. Irregular hydrogen bonding between the first two strands of the barnase β-sheet (Lubienski, 1994). The left strand (strand 1) contains the β-bulge involving residues 53 and 54.

Axe et al, JMB, 286, 1471 (1999)

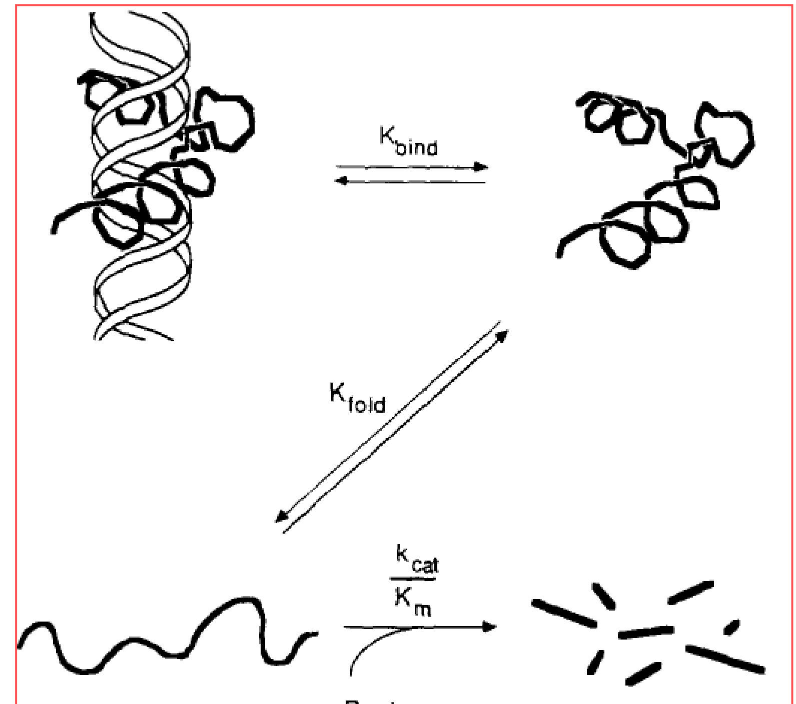
Stability and proteolytic susceptibility

We would expect that a less stable protein would be more readily degraded in vivo

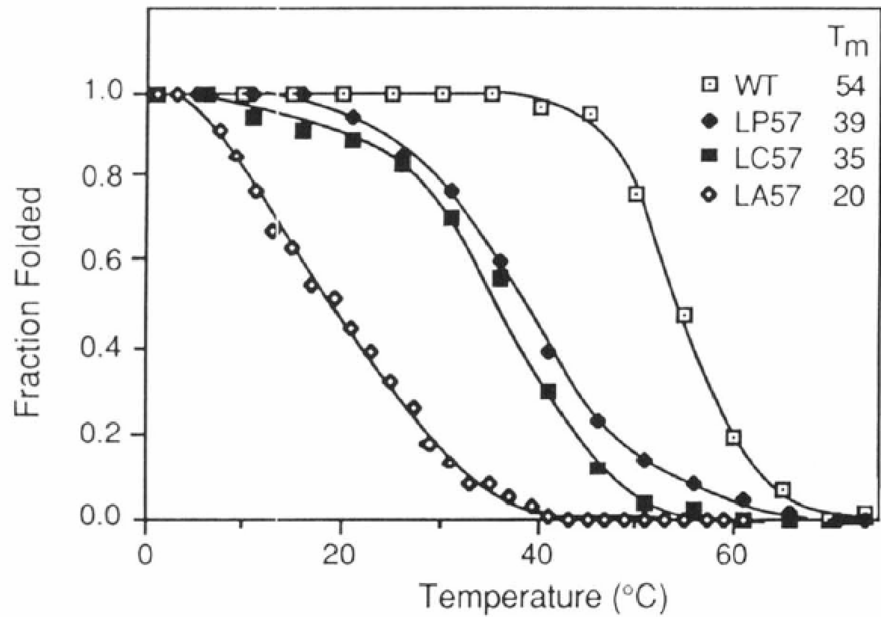
Do pulse-chase experiment of lambda repressor

e.g. label proteins synthesized between T=30 – 32 min with radioactive **35S**

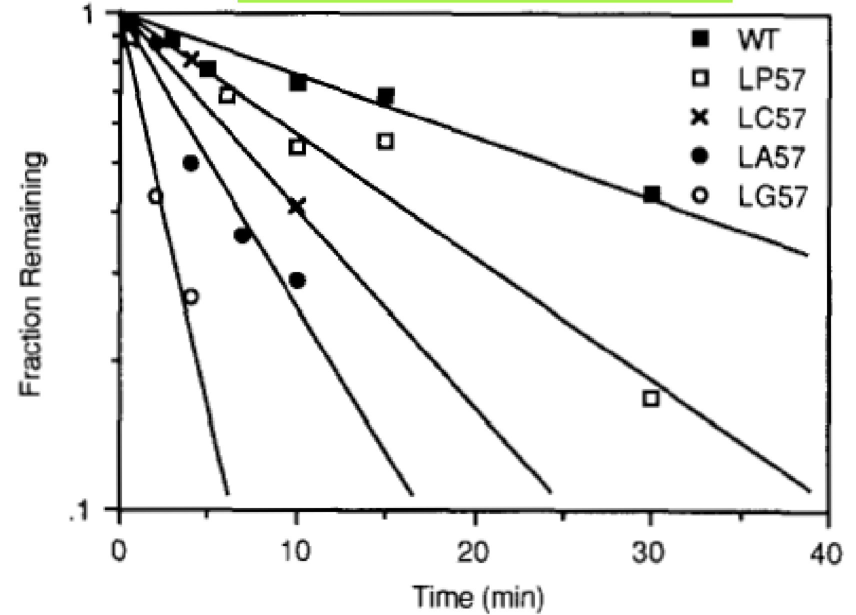
Monitor the amount of the protein over time and correlate with thermal stability



Parsell et al, JBC 264, 7590 (1989)



detected using 35S



A complementary second mutation (ND52) increases the stability of a mutant as well as the half life of the protein

Thermal stability influences the average lifespan of a protein molecules since proteolysis likely occurs while the protein is in an unfolded state

Disulfide bond

Disulfide bond is a **covalent bond** formed between two cysteine side chains with the bond energy of ~ 70 kcal/mol

There are strict structural requirements for ideal disulfide geometry

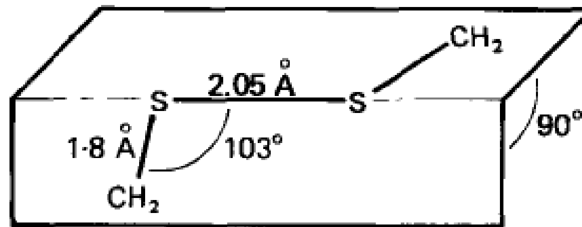


Fig. 1. *Optimal stereochemistry of a disulphide bond. The two possible torsion angles about the disulphide bond of +90° and −90° are equally favourable.*

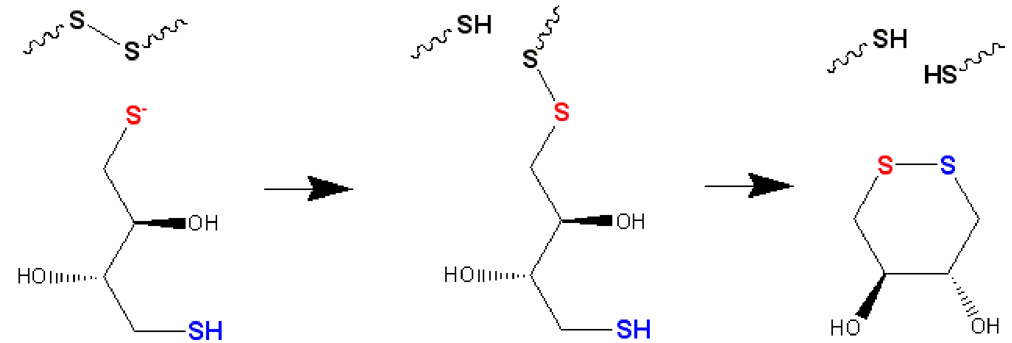
Creighton, BioEssays 8, 57 (1988)

Both isomers (right-handed and left-handed) are observed in natural proteins

Particularly important in small proteins that lack genuine hydrophobic cores

Protein disulfide isomerase (PDI) catalyzes internal disulfide exchange and helps correct wrong disulfide bonds that may form during folding

Disulfides may be “reduced” using reducing agents containing thiols (e.g. βME, DTT)



In the presence of a thiol, the disulfide undergoes an exchange reaction :

R1-S-S-R1 + R2-S$^-$ ↔ R1-S-S-R2 + R1-S$^-$

A disulfide bond can stabilize a protein by 2 – 5 kcal/mol by reducing the conformational flexibility of the unfolded peptide chain, and thus destabilizing the denatured state of a protein relative to the folded state

$$\Delta S = -2.1 - \frac{3}{2}\ln(n)$$

Betz, Protein Sci 2, 1551 (1993)

However, a crosslink can also affect the folded states—the effect of a crosslink on conformational stability depends on the change in the effective concentration of the thiols between the unfolded and folded states

Removing native disulfides

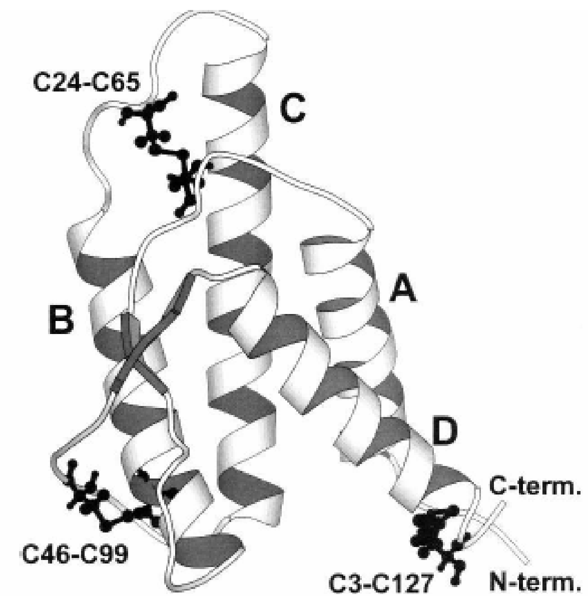
Removing disulfide bonds usually destabilizes the protein

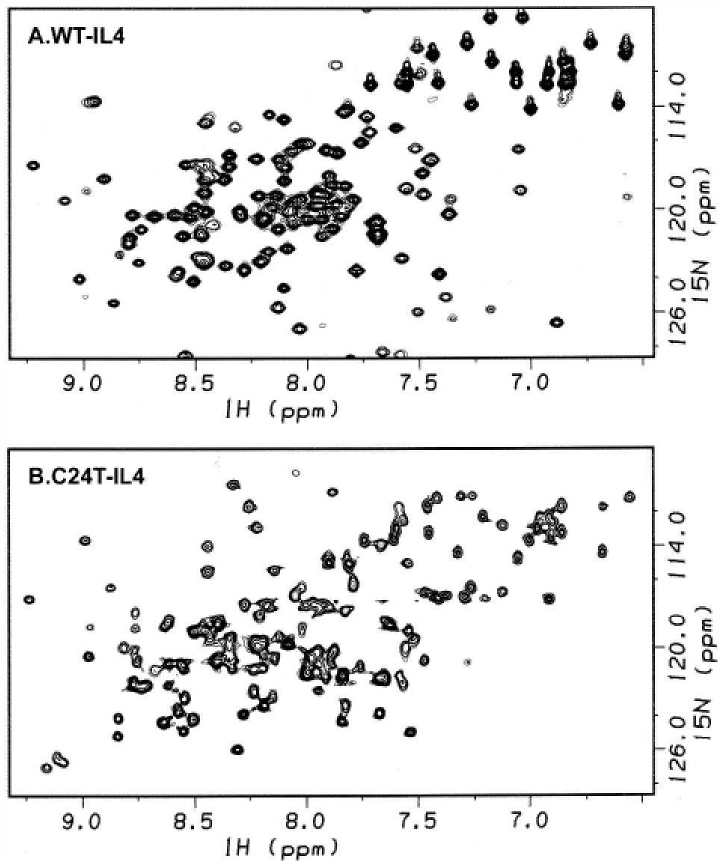
HEW lysozyme has three disulfides. Removing these disulfides destabilizes the protein and reduces the melting temperature by 25 °C

- Cooper et al, JMB 225, 939 (1992)

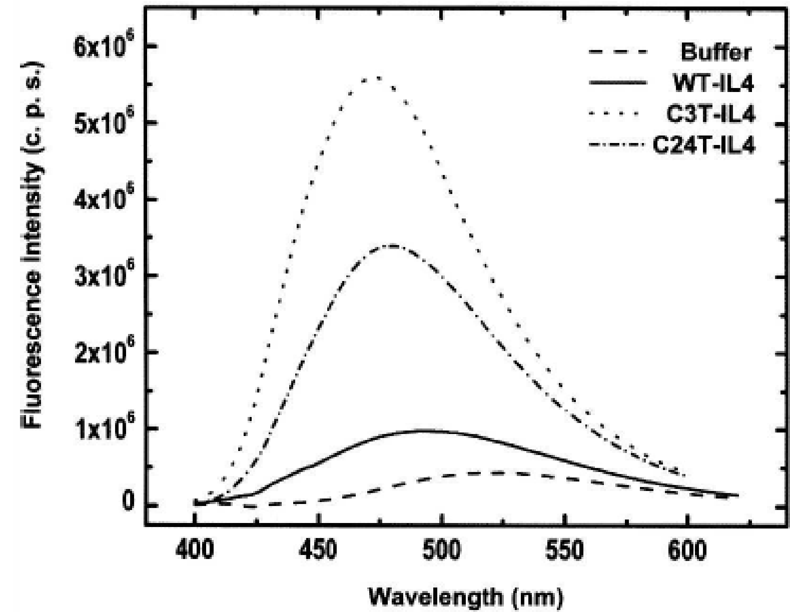
Removing disulfides in interleukin-4 significantly disrupts the integrity of its hydrophobic core

- Vaz et al, Protein Sci 15, 33 (2006)





Changes in NMR crosspeaks suggest altered dynamics



Binding by **ANS** (8-anilino-1-naphtalenesulphonic acid), which fluoresces when bound to the hydrophobic patches of a protein

Enthalpy v. entropy

Do disulfides stabilize proteins only by reducing the entropy of the denatured state?

Doig and Williams (JMB, 217, 349 (1991)) argues that disulfide bonds destabilize folded structure entropically but stabilize them enthalpically

C77A and C77/95A mutants of human lysozyme are destabilized by ~ 4.6 kcal/mol, most of caused mostly by an enthalpy change

- Kuroki et al, Biochemistry 31, 8323 (1992)

Differential scanning calorimetry can measure thermodynamic parameters, including

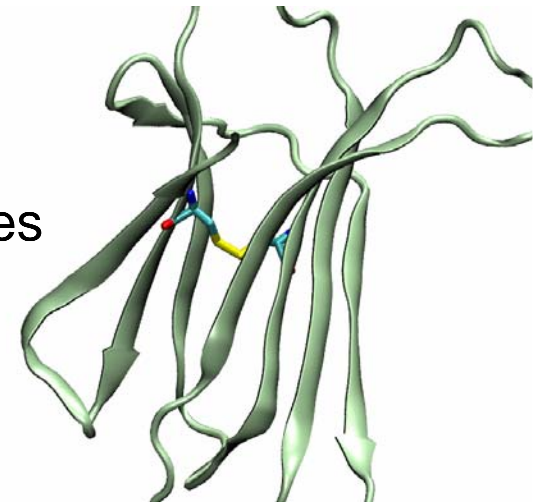
- melting temperature (T_m)
- enthalpy change (ΔH)
- heat capacity change (ΔCp)

$$\Delta_d H(T) = \Delta_d H(T_d) + \Delta_d C_p (T - T_d)$$

$$\Delta_d S(T) = \left[\frac{\Delta_d H(T_d)}{T_d}\right] + \Delta_d C_p \ln\left(\frac{T}{T_d}\right)$$

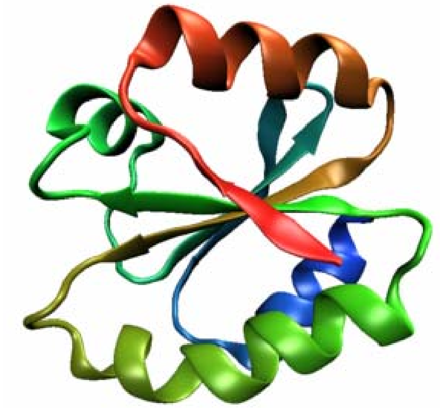
$$\Delta_d G(T) = \Delta_d H(T) - T\Delta_d S(T)$$

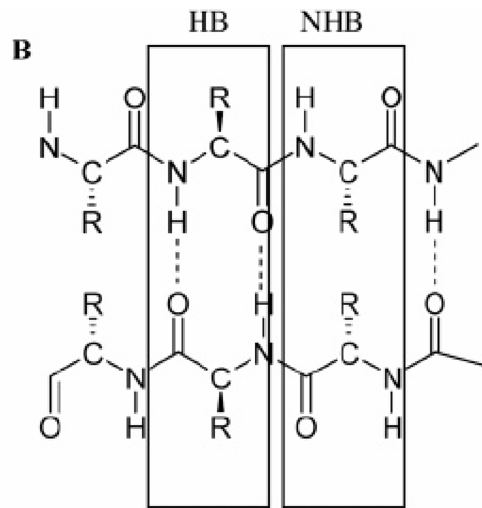
Stabilization by disulfide

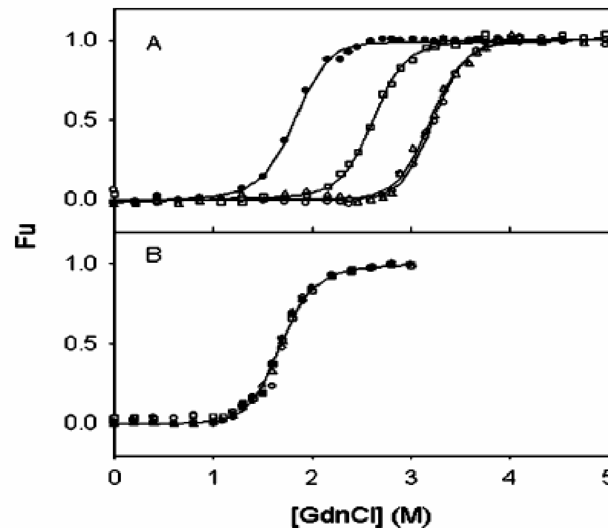


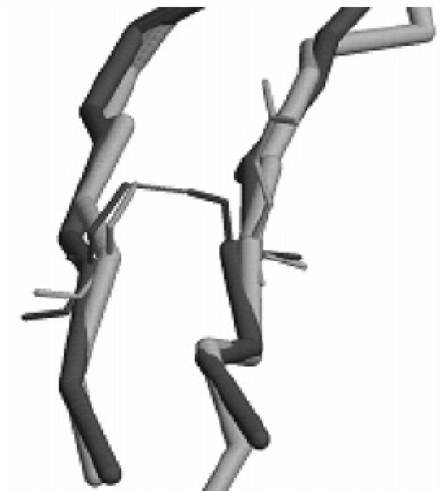
If removing a native disulfide destabilizes a protein, does introducing a new disulfide bond stabilize the protein?

Interstrand disulfides are rare (3% of all disulfides) and usually occur between non hydrogen bonded pairs of antiparallel strands







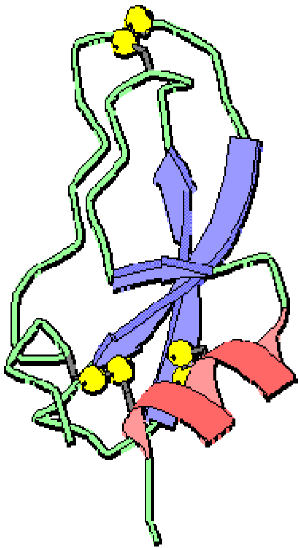


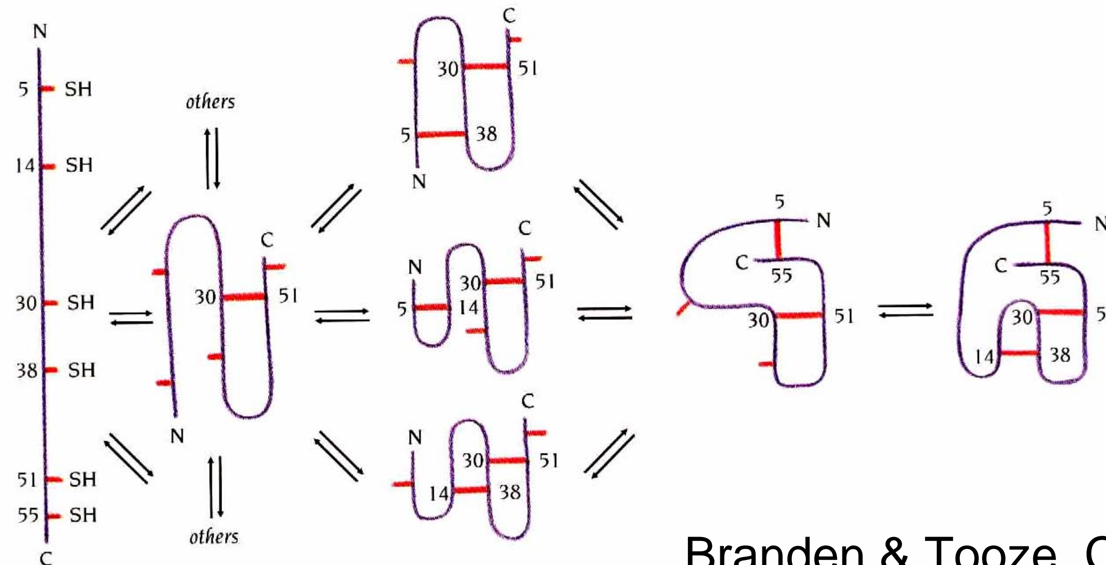
Introducing cysteins in thioredoxin can change stability

- Chakraborty et al Biochem 44, 14638 (2005)

Folding of BPTI

- Model system for studying disulfide bond formation—58 residues
- Contains three disulfides for stability
  - Reduction of all three disulfides results in complete unfolding
- Formation of disulfides poses a challenge since there are many potential combiantions





Branden & Tooze, Ch. 6