Amino acid alphabet

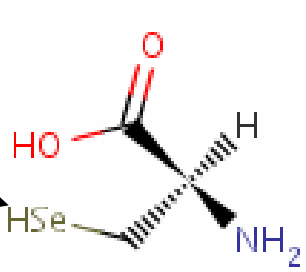
Twenty amino acids are used in all proteins with the exception of:

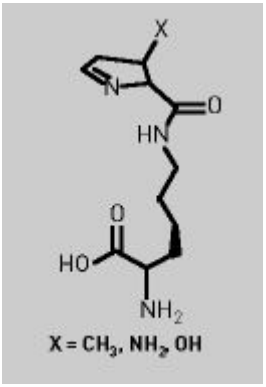
selenocysteine: occurs in enzymes from all three kingdoms (archae, eubacteria, eukaryotes) and is coded by the “amber” stop codon (UGA)

Synthesized in situ from Ser-tRNA (no aminoacyl transferase for selenocysteine)

pyrrolysine: occurs in some methanogenic archaea in enzymes as part of their methane-producing metabolism

selenium instead of sulfur





The relative abundances of individual amino acids on primordial earth were probably different than they are today and thus the amino acid alphabet may have consisted of fewer (or more) than 20 amino acids

- Osawa et al Microbiol Review 56, 229 (1992)

Also protein structures are often tolerant to amino acid substitutions

How did the current amino acid alphabet evolve?

The standard alphabet may consist of the smallest and cheapest amino acids that form a functional protein library

Helps identify homologous proteins and is crucial for homology modeling

The sequence structure relationship may become more apparent if proteins can be engineered using fewer than 20 amino acids

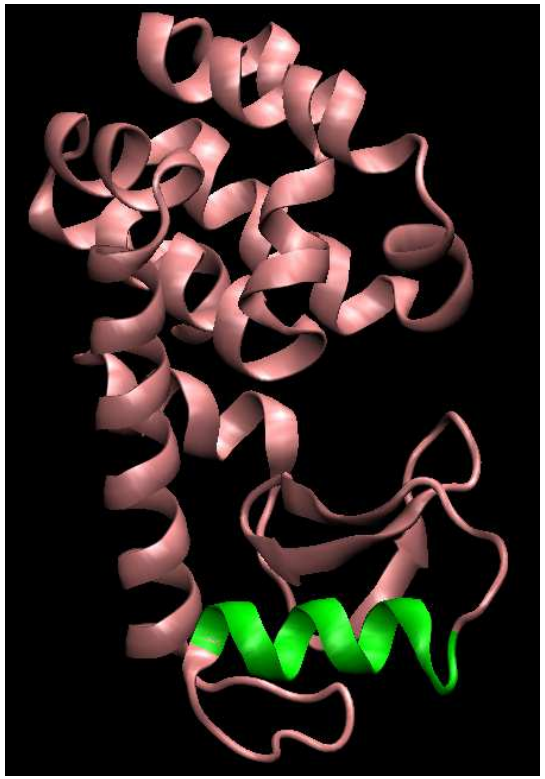
T4 lysozyme with A40-49

Can we engineer functional protein with well defined structure using fewer than 20 amino acids and what is the minimum number of amino acids required for the task?

Mutate multiple residues simultaneously to see if a large set of amino acids can be altered in combination to a much smaller set without perturbing the structure and function irreversibly

Residues 39 to 50 T4 lysozyme form an alpha helix

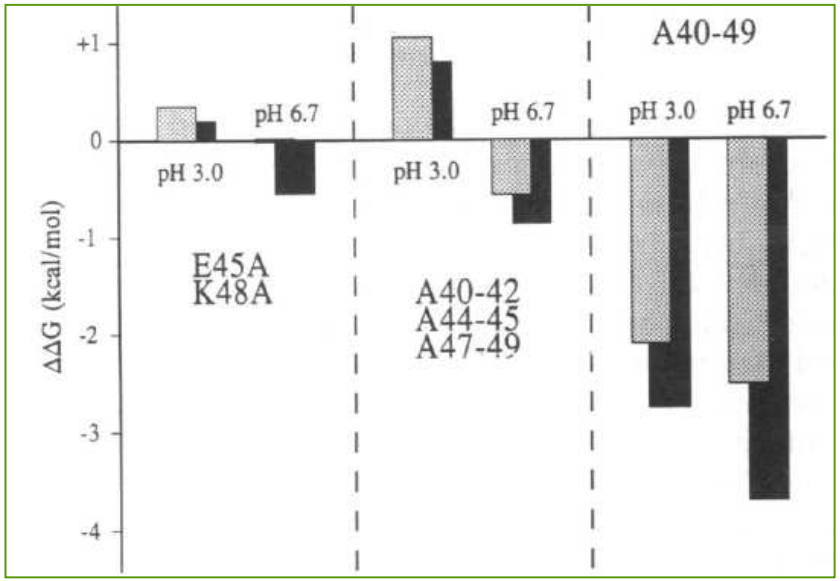
Mutate residues 40 to 49 with 10 consecutive Ala

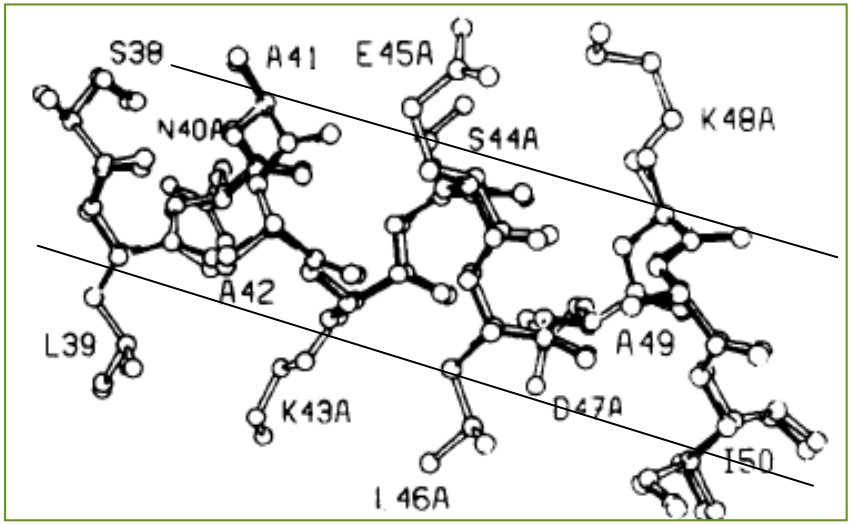


Heinz et al. PNAS 89, 3751 (1992)

Depending on the number of substitutions, the mutants are destabilized by ΔTm = - 8.5 – 10.7 ℃ and $\Delta\Delta G$ = 2.1 – 2.5 kcal/mol at pH 3.0 and 6.7

Yet they preserve the overall the structure

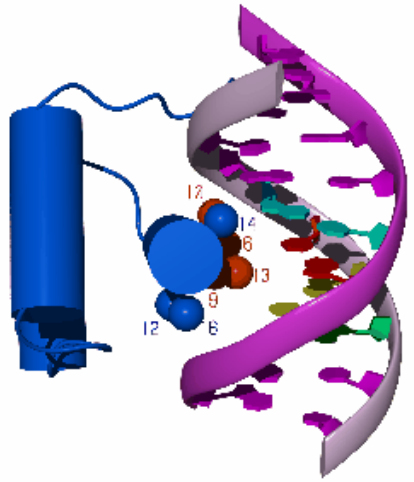




main chain atoms overlap

Minimal homeodomain

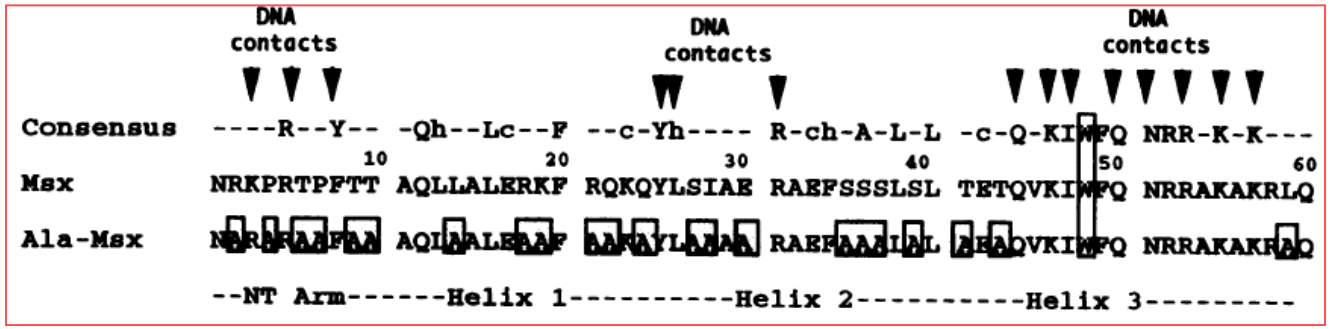
Homeodomain is a conserved DNA binding domain that is ubiquitous among developmental regulatory proteins from nematodes to humans



Consists of 3 alpha helices and a long N-terminal arm

A **consensus** sequence can be generated by sequence comparison and identify residues that are important for tertiary structure and DNA binding activity

Homeodomain proteins also exhibit sequence specificity contributed by nonconserved residues



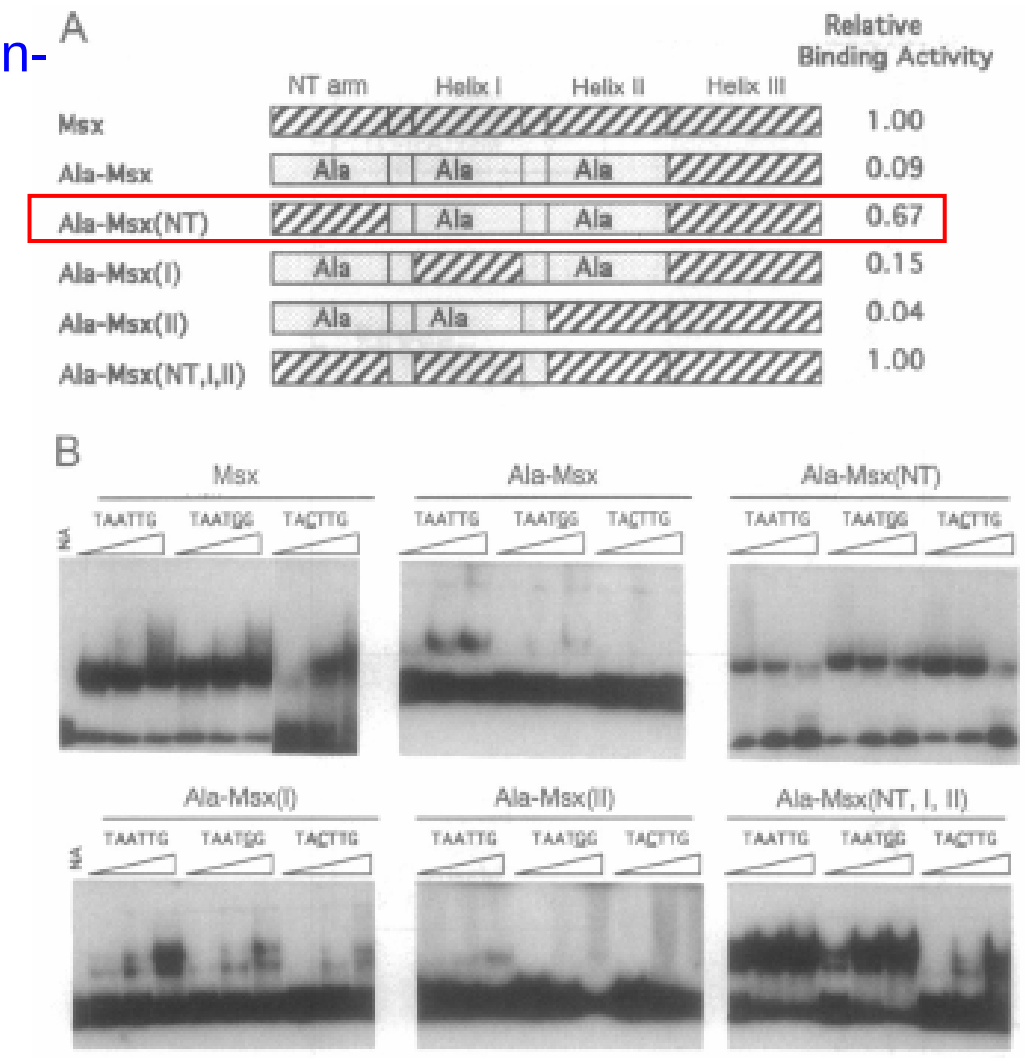
Shang et al, PNAS 91, 8373 (1994)

Goal: Design a “minimal” homeodomain containing all consensus residues, in which all non-conserved residues have been substituted with alanine

Consensus residues contain sufficient sequence information to construct the “scaffold” for specific DNA recognition

Non-conserved residues in the N-terminal arm are required for high affinity DNA binding as well as for tertiary structure

It is possible to design a vastly simplified high affinity mutant with a large number of Ala



gel shift assay

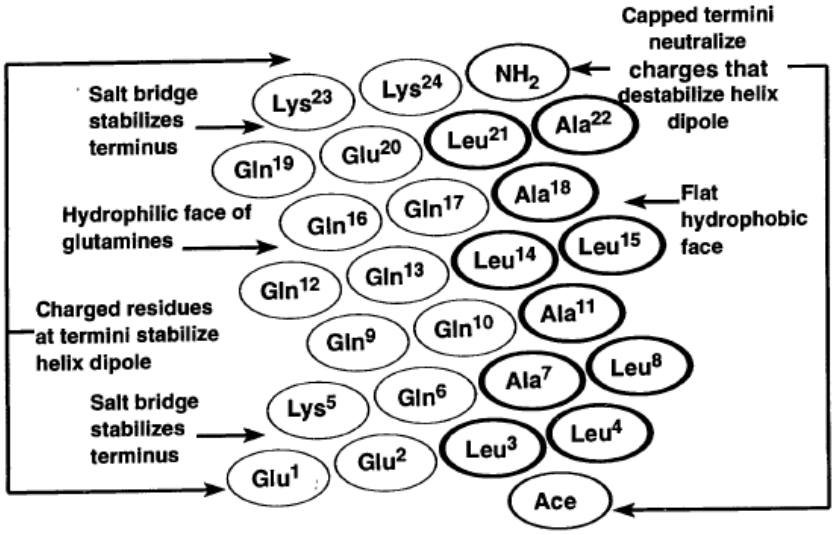
“Peptitergent” to solubilize membrane protein

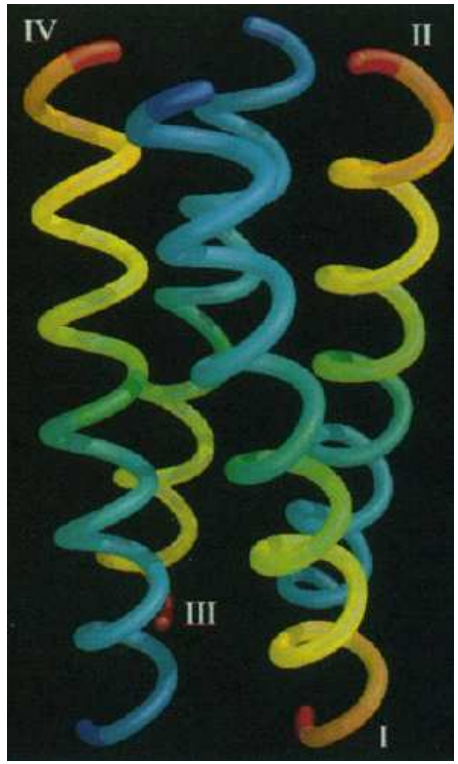
Structural determinants of alpha helical peptides are well understood

Alpha helical peptide of 24 residue peptide-detergent PD1 (“peptitergent”) was designed to self-associate and to solubilize transmembrane proteins

- Schafmeister et al, Science 262, 734 (1993)

PD1 is amphipathic with Ala and Leu on one side





Helix bundle with 7 amino acids

Four helix bundles are versatile. Design and characterize an antiparallel four helix bundle DHP1 using an alphabet of seven amino acids

- Interhelical loops consist of 3, 4 and 3 glycines
- Lys, Glu, Gln, Ala, Leu, Gly, Ace (capping residue)

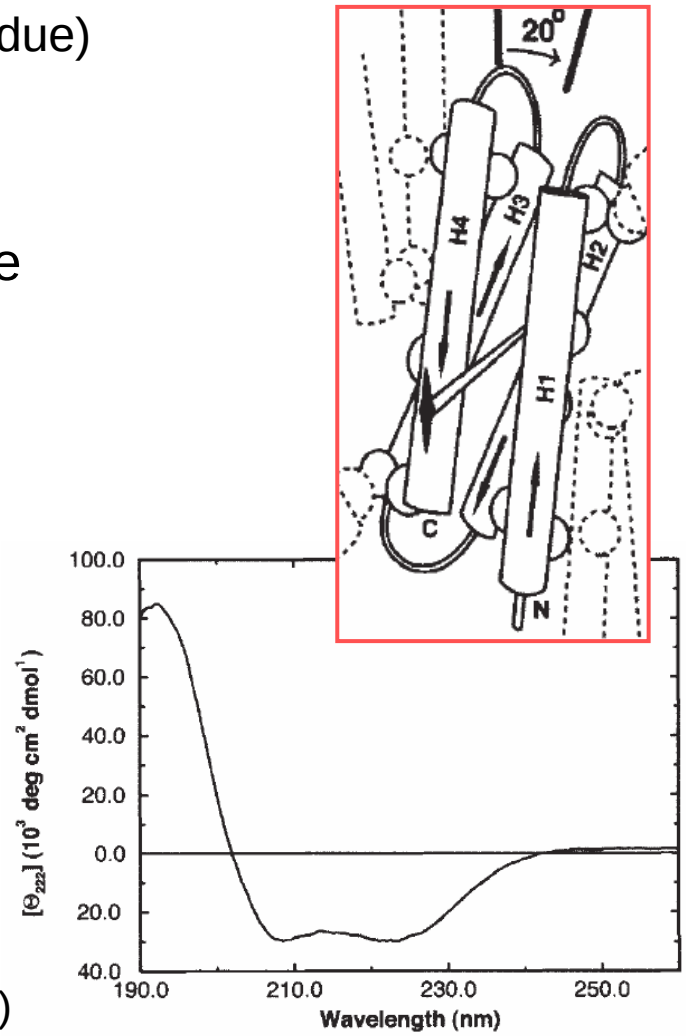
Helices cross each other at -20°

Hydrogen/deuterium exchange shows that amide protons are shielded from solvent exchange with a protection factor > 1.6 x 10^4

Cf. For molten globule proteins, the protection factors are < 100

In native proteins, between 1000 – 5 x 10^8

Sufficiently ordered structure can be designed using only seven amino acids



Schafmeister et al, NSB 4, 1039 (1997)

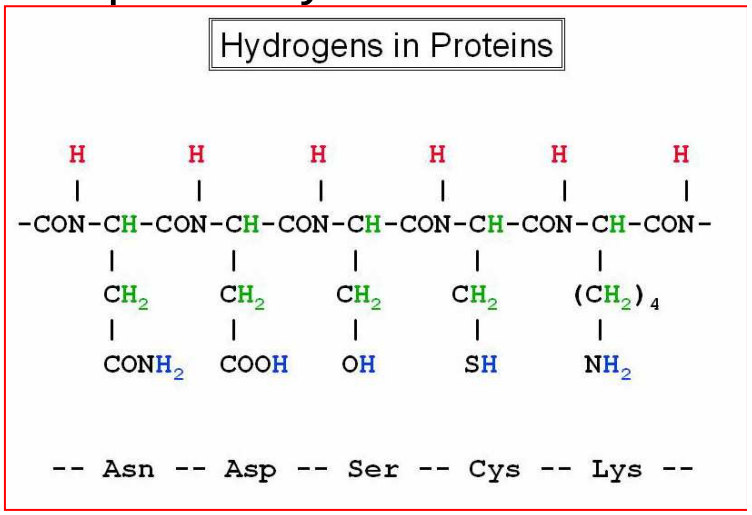
Hydrogen-deuterium (H/D) exchange

Amide protons constantly exchange with the solvent

If the protein is submerged in heavy water (D_2O instead of H_2O), then after losing its original hydrogen an amide nitrogen will pick up a deuterium

The rate of exchange can be monitored by mass spec or by NMR

green: no exchange
blue: too fast
red: just right



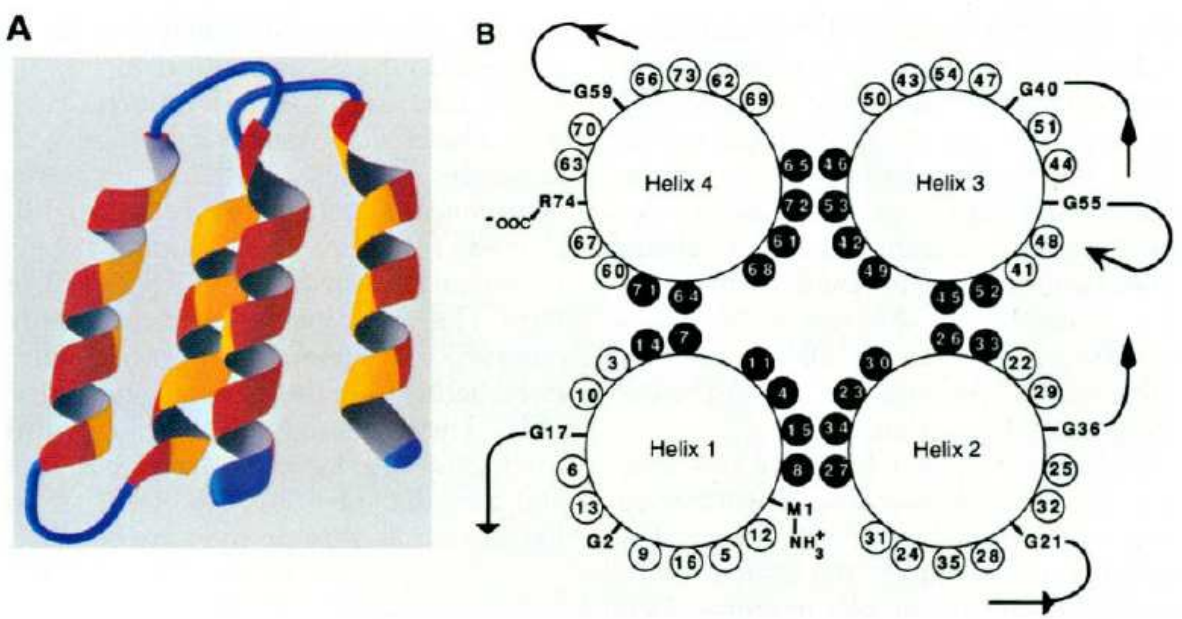
H/D exchange can be used to :

- Measure the rate of folding and unfolding
- Solvent accessibility at protein-protein interfaces
- Determine disordered regions of native, folded protein—hence stability as well

Binary patterning

Is it possible to design a protein by specifying just its hydrophobicity pattern without the amino acid identities?

The hydrophobicity pattern on a helix repeats itself roughly every 3.6 amino acids. Test if a randomized peptide with the hydrophobicity pattern of a helix actually folds to a helical protein



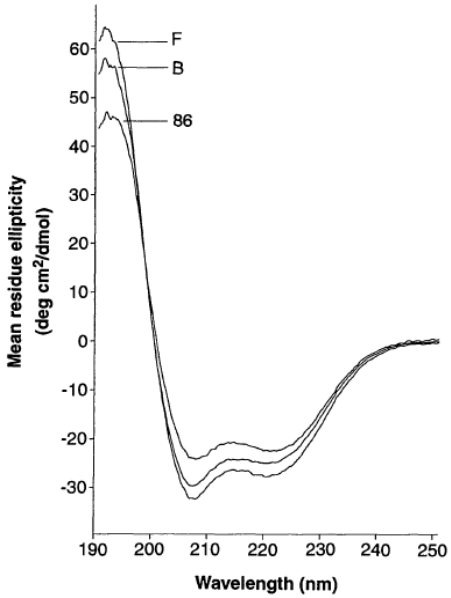
Kamtekar et al, Science 262, 1680 (1993)

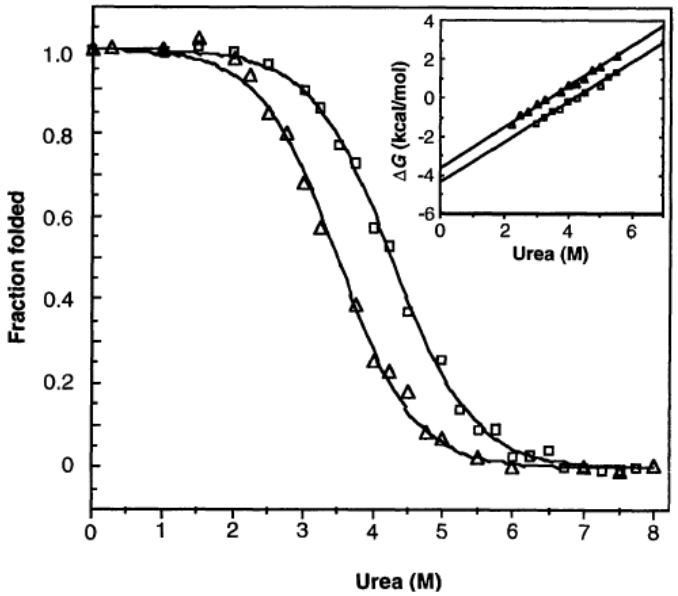
Fig. 1. (**A**) Ribbon diagram (*43*) showing the periodicity of polar (red) and nonpolar (yellow) residues in a four-helix bundle. (**B**) Head-on representation of a four-helix bundle. Individual helices are shown in the helix wheel representation with a repeat of 3.6 residues per turn (*44*). The binary pattern of the design is illustrated by showing generic polar residues as white circles and generic nonpolar residues as black circles (*39*). The identities of the N-cap and C-cap residues at the ends of each helix are shown explicitly. The interhelical turns are represented by arrows.

Binary patterned library

- Identify the residue positions required to be hydrophobic and hydrophilic
- Synthesize genes to introduce F, L, I, M or V at 24 hydrophobic positions, and D, E, K, N, Q, and H at 32 hydrophilic positions

total diversity: 5 ^ 24 + 6 ^ 32 = 4.7 x 10^41



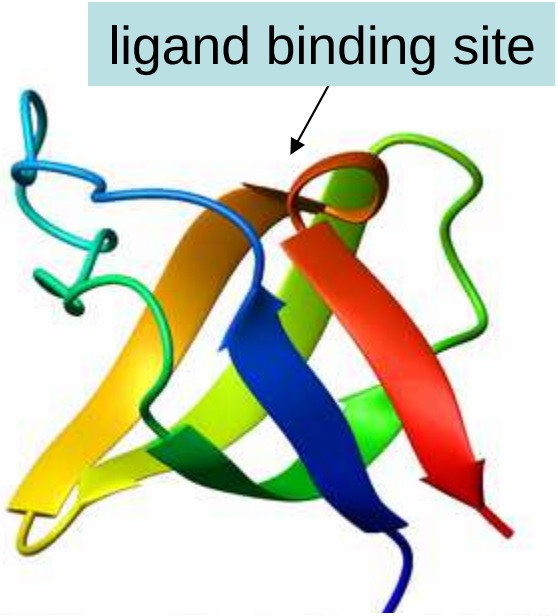


Beta sheet protein with reduced alphabet

Helices and helical domains are in general easier than beta sheets

Can a beta sheet protein be similarly engineered using a reduced alphabet?

SH3 domain is a small beta barrel like structure that binds a proline-rich peptide



Replace all residues not involved in ligand binding with a small set of amino acids including I, K, E, A, G, but maintaining

WT	TFVALYDYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGRTGYIPSNYVAPSD
FP1	EFIAIYDYKAETEEDLTIKKGEKLEIIEK E GDWWKAKAIGSGEIGYIPANYIAAAE
FP2	EFIAIYDYEAKEKEDLTIKKGEKIEIIEK E GDWWKAKAIATGKIGYIPSNYIAAAE

Riddle et al, NSB 4, 805 (1997)

Enzyme from 9 amino acids

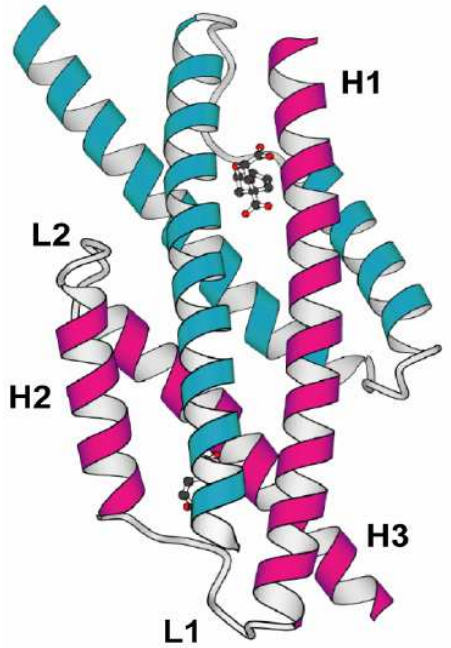
Can a reduced amino acid alphabet support enzymatic catalysis?

Designing/engineering an enzyme is difficult because of the precise positioning of residues required for catalysis

Chorismate mutase (CM) catalyzes the rearrangement of chorismate to prephenate and is essential for the biosynthesis of aromatic residues

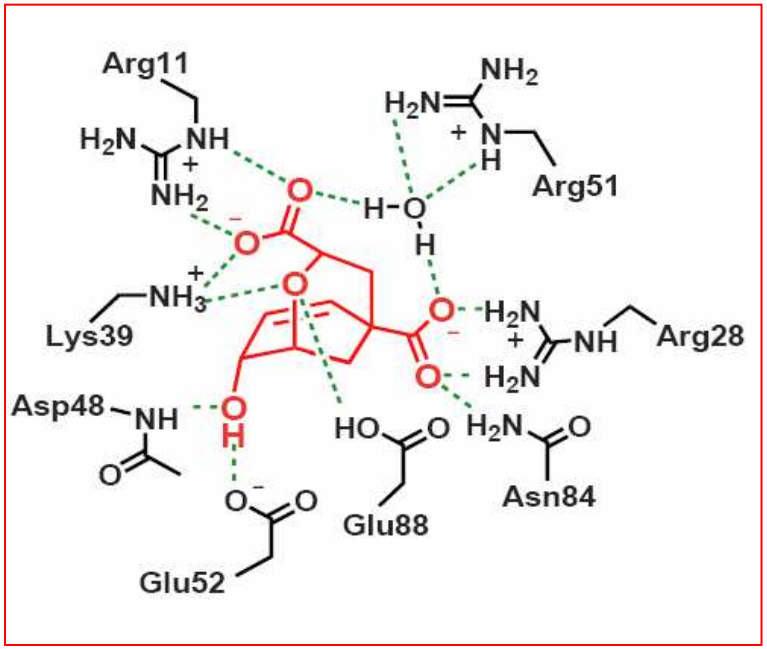
The enzyme is a domain-swapped homodimer of three helices and two loops

Assay: express a mutant CM in a strain deficient in the enzyme ("auxotrophic") and select for viable clones



1. Active site: **Q88**, R11, R28, R51
   - E > N > D
   - kcat is 3-fold lower than wild type
   - kM is 40 fold higher
2. Loop #2: **H66, V68**
3. Loop #1: **G43-I44-P45-I46**

   D, E, N, K, R, F, I, L, M



i. + ii. mutant
iii. wild type
iv. pos control
v. neg control