# Sequence structure relationship

Are there ways to study the sequence structure relationship besides

- protein folding experiment
- structure determination
- structure prediction

Introduce amino acid substitutions and study structural perturbations

More manageable

Narrow the scope of study by reducing the number of variables

Perhaps simpler to interpret the results than folding studies

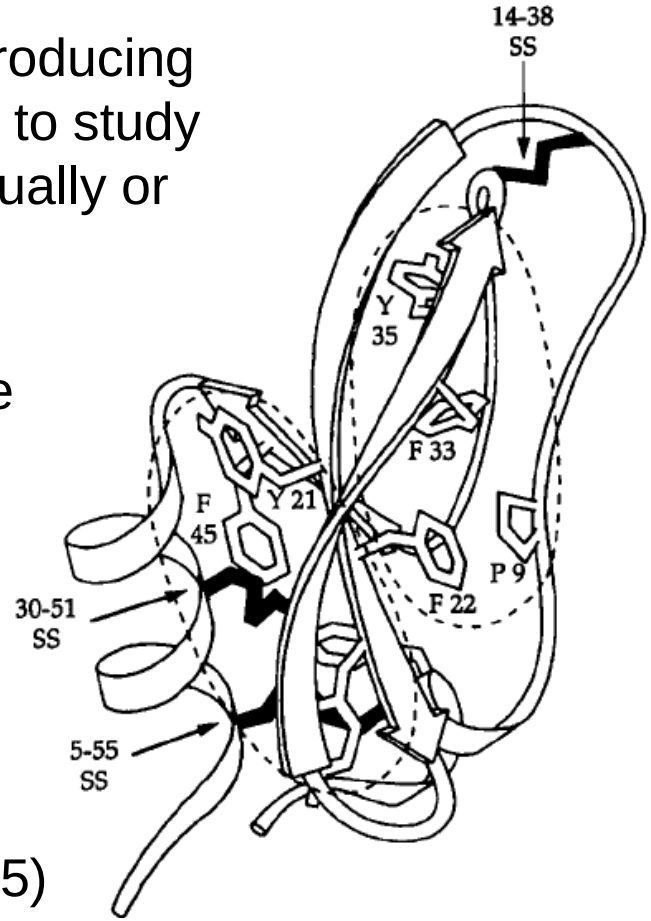
# Alanine scanning of BPTI

Ala is the smallest amino acid containing Cbeta

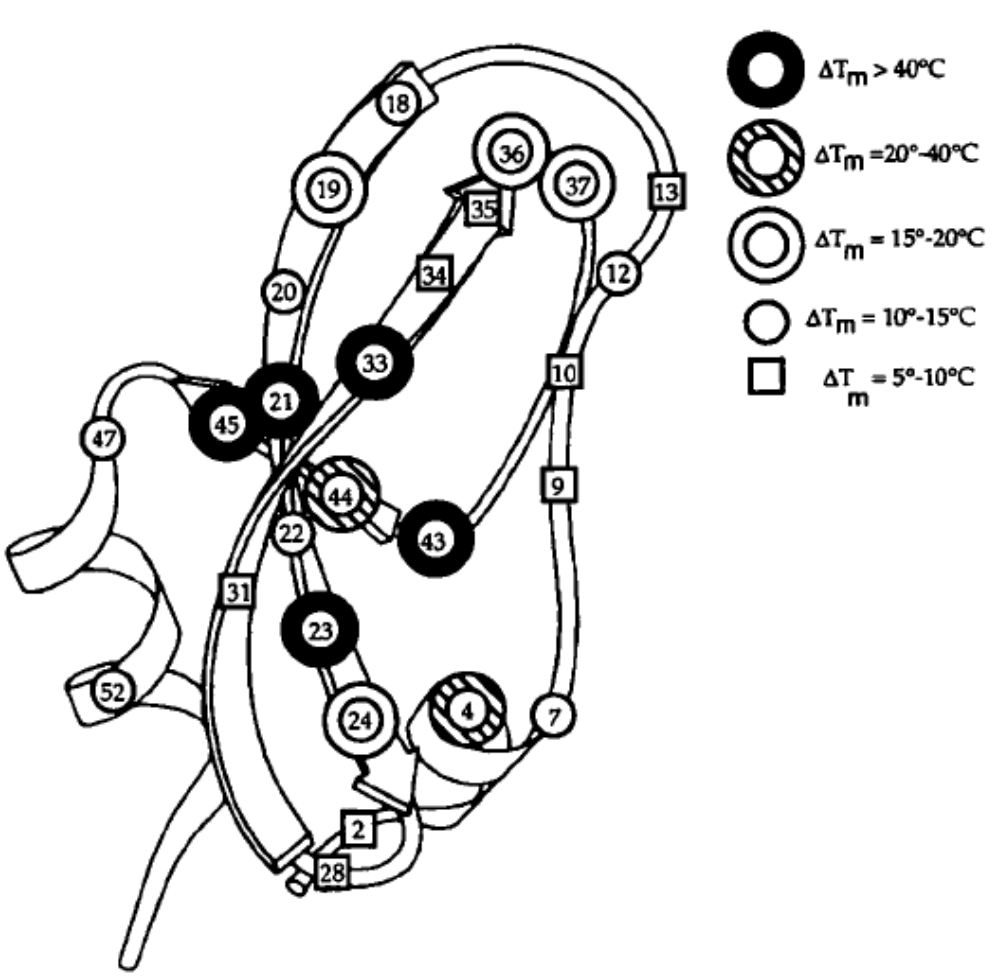
Ala substitution can probe the role of the side chain

**"Alanine scanning"** refers to systematically introducing an Ala mutation at a series of residues positions to study their structural and functional roles either individually or as a group

Systematically mutate every residue in BPTI one at a time to evaluate its contribution to stability




Yu et al, JMB 249, 388 (1995)



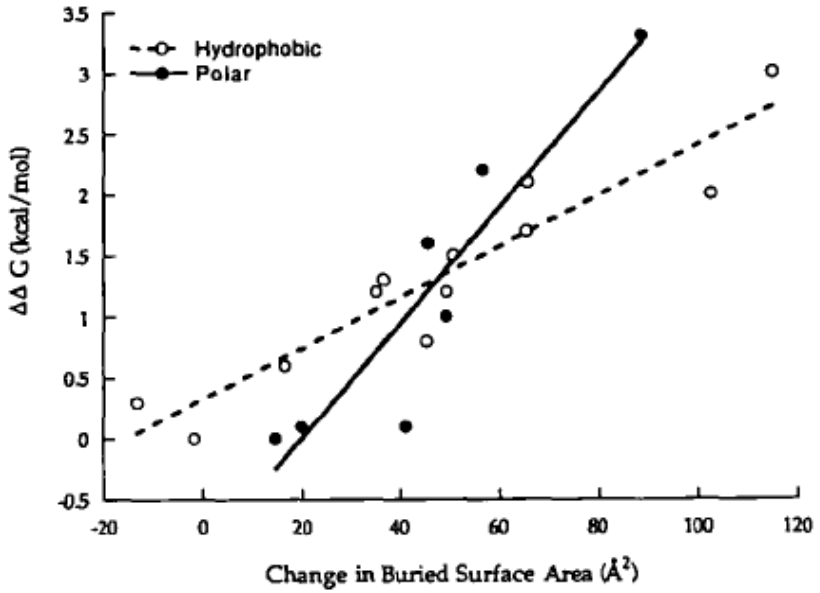




**Figure 5.** Relationship of $\Delta\Delta G$ of unfolding as a function of buried surface area. Substitutions of the hydrophobic residues (○, Leu, Val, Phe, Met, Ile and Pro) are plotted separately from substitutions of the polar residues (●, Ser, Thr, Asn and Gln). The straight lines are the least squares fit to the data. The slope of the lines are 20 and 50 cal/mol Å$^2$ ($r$ = 0.92 and 0.93) for the hydrophobic and polar residues, respectively. Solvent-accessible static surface area of each atom was calculated from the structure of BPTI (30A/51A) (Eigenbrot *et al.*, 1990).

# Paracelsus challenge

Paracelsus, 16th century Swiss alchemist

While the sequence of a protein should uniquely define the tertiary structure, it's not clear if the structural information is uniformly distributed throughout the sequence

Homology modeling is rooted in the assumption that a similarity in sequence implies a similarity in structure

Sequence similarity > 30% is routinely viewed as a sufficient condition for deducing evolutionary homology and is used to model unknown structures

Challenge: Is it possible to design two proteins with sequence identity of at least 50% that fold to two completely different topologies? Equivalently, is it possible to start from a protein of known structure and mutate no more than 50% of the sequence and arrive at a structure with a different structure?

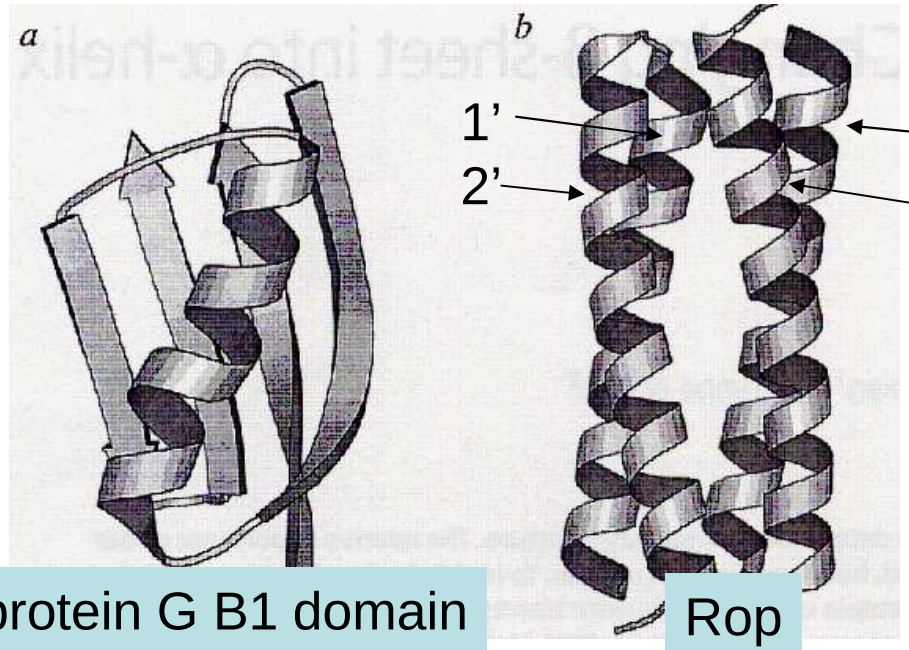
George Rose and Trevor Creamer (1994)

# Protein alchemy

The B1 domain of Protein G is a mixed alpha/beta protein of 56 amino acids

Rop is a homodimeric four helix bundle protein (helices 1 & 2 in the first subunit, and 1’ & 2’ in the second subunit) that facilitates sense-antisense RNA pairing by binding to the transiently formed hairpin pairs

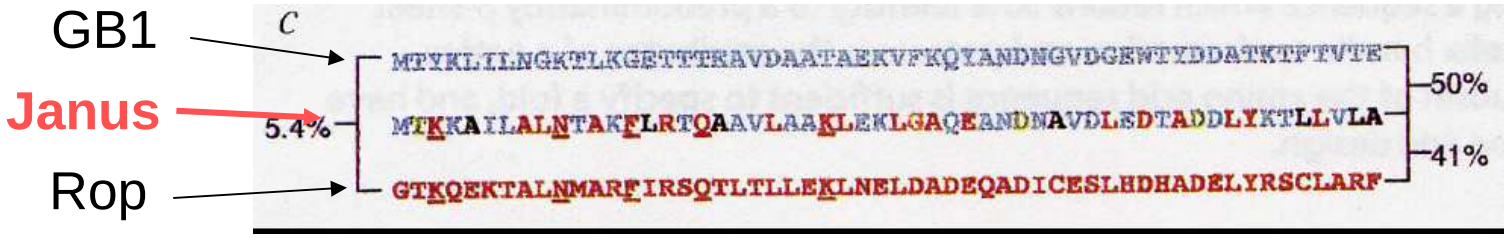
The first 56 amino acids of Rop form alpha helices while the last 7 amino acid are unstructured




protein G B1 domain

Rop

Dalal et al, NSB 4, 548 (1997)




Helices are easier to construct than beta strands

Align the two protein sequences

While retaining amino acids of high helix propensity in GB1, replace residues with high beta sheet propensity with amino acids that favor helix formation

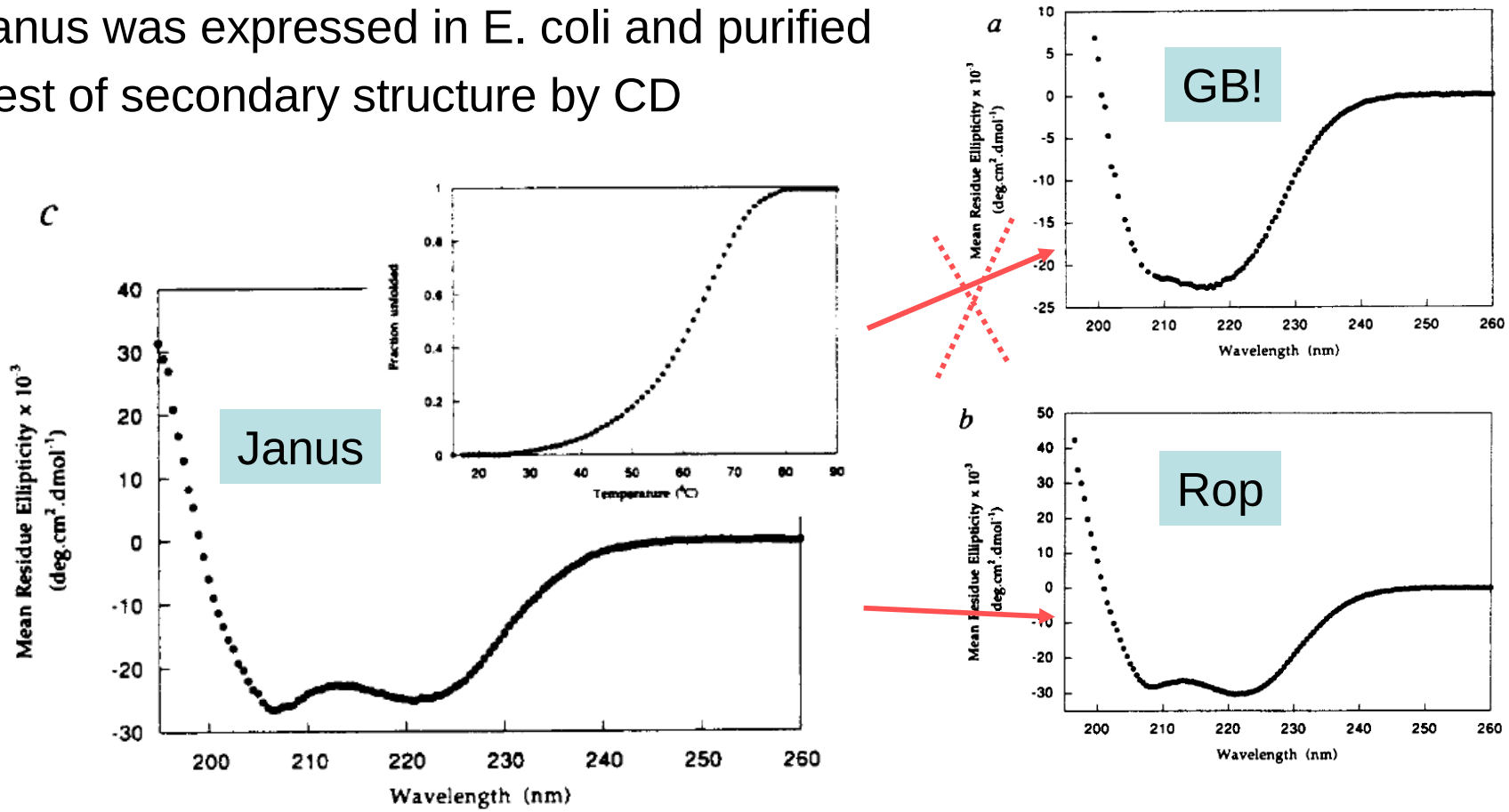
Also incorporate hydrophobic residues at "**a**" and "**d**" positions of the heptad repeat

Surface of subunit 1/1' mainly neutral or positively charged, surface of subunit 2/2' mainly negatively charged

Tyr at position 49 as diagnostic spectroscopic probe of folding

Janus was expressed in E. coli and purified

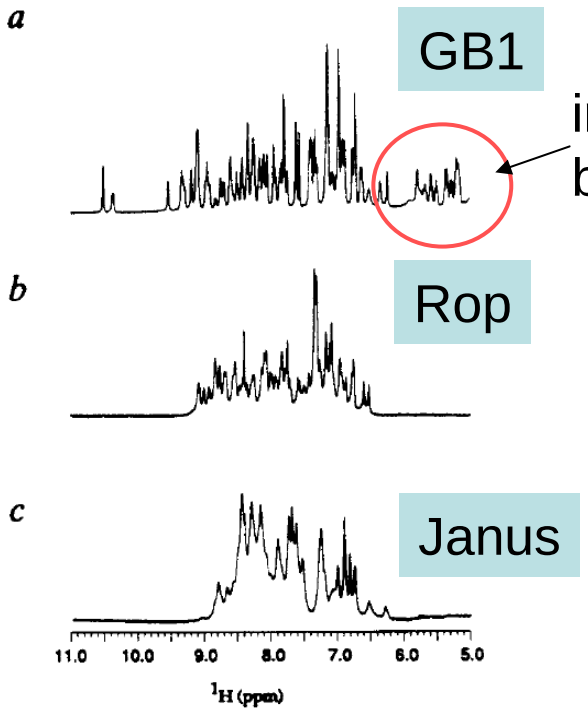
Test of secondary structure by CD




Thermal denaturation shows cooperative reversible unfolding with Tm=62℃

Amide protons are protected from exchange for > 15 days

1D and 2D NMR spectra Janus resembles Rop more than GB1 and has a well-packed core



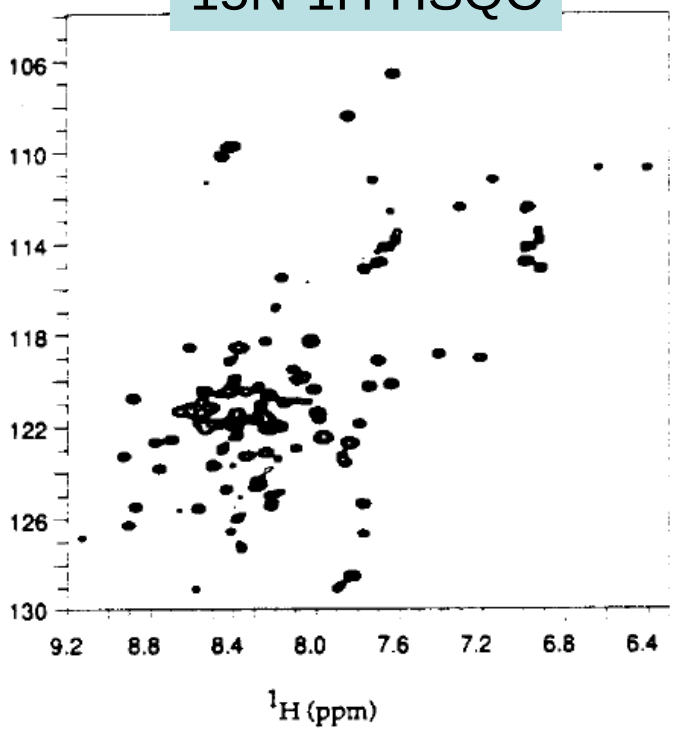

GB1

indicative of beta sheet

Rop

Janus

15N-1H HSQC




good separation of individual peaks suggests a well-folded 3D structure

# Formation of secondary structure

Propensities of amino acids to form particular secondary structures are influenced by local conformational preferences (e.g. helix propensity, beta strand propensity) as well as non-local factors

Can the same peptide sequence be induced to form one secondary structure in one context and another secondary structure in another context?

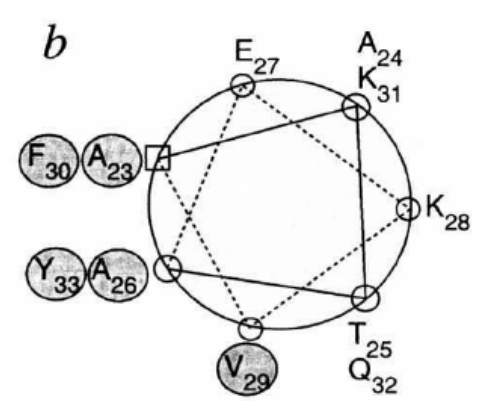
In particular, design a sequence that can form either an alpha helix or a beta strand depending on the context

When designing such sequences, it is important to preserve the hydrophobic nature of the residues

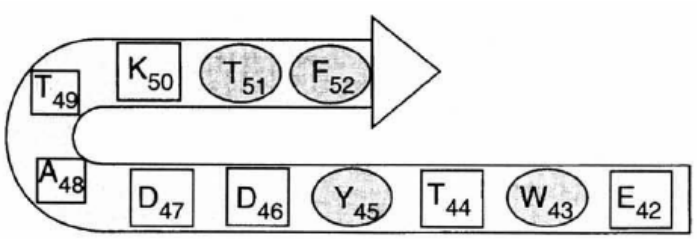
periodicity in helix: 3.6

periodicity in sheet: 2

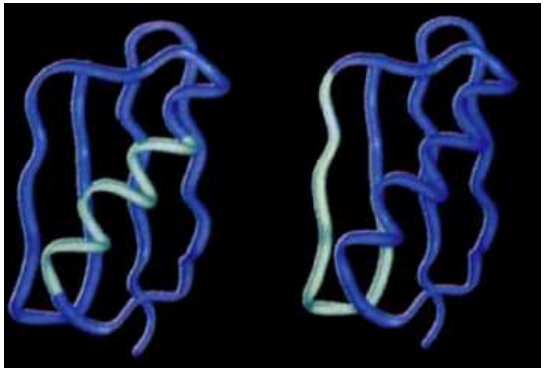
Within the protein G, B1 domain, find stretches of alpha helix and beta sheet residues with minimum amount of "conflict of interest"




GB1 23-33




GB1 42-52



Minor and Kim, Nature 380, 730 (1996)

Classify the residues according to the difference of roles placed they play in each sequence

i. residue is buried in one secondary structure but exposed in the other

Solution: adopt the identity of the buried residue

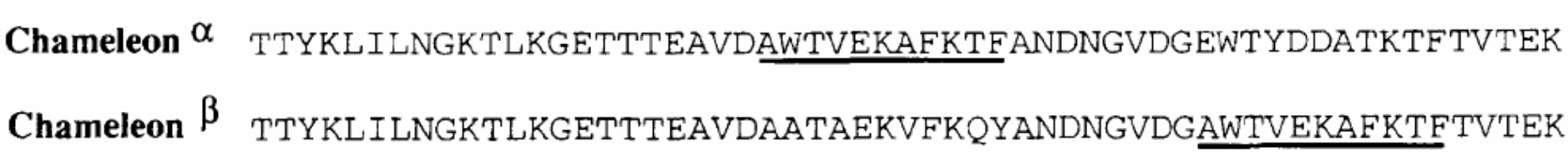
ii. residue is buried in both but is different in size and polarity

Solution: try several hydrophobic residue pairs in the wild type background

iii. no conflicts in size, hydrophobicity, or solvent exposure

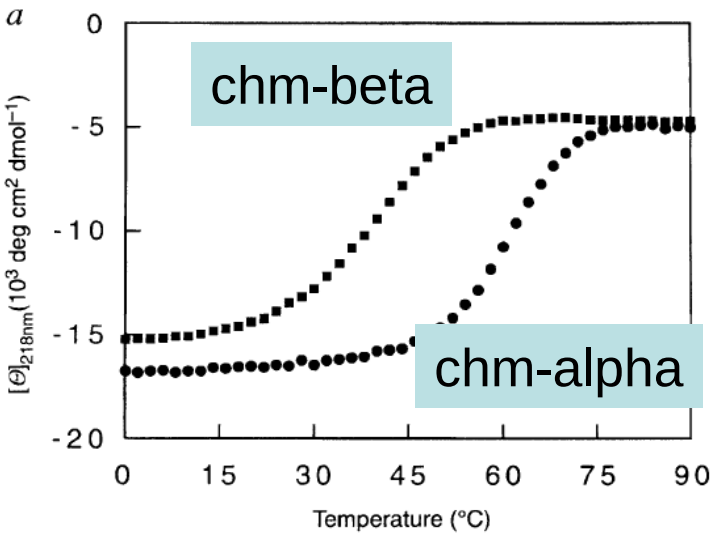
Solution: choose the residue from alpha helix

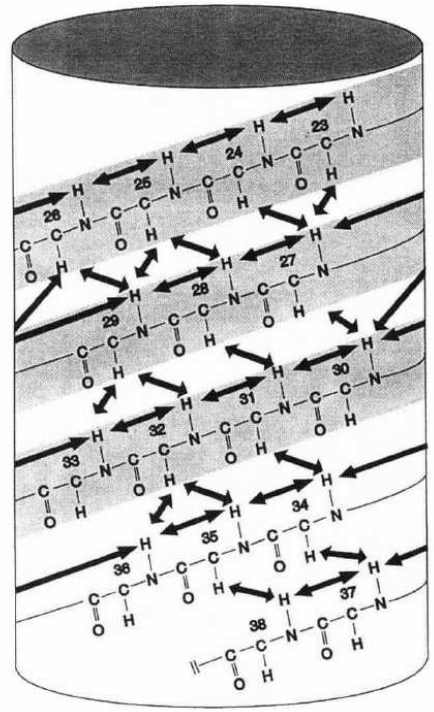
```
ALPHA HELIX 23-AATAEKVFKQY-33
 BETA SHEET 42-EWTYDDATKTF-52
  chameleon ..-AWTVEKAFKTF-..
```

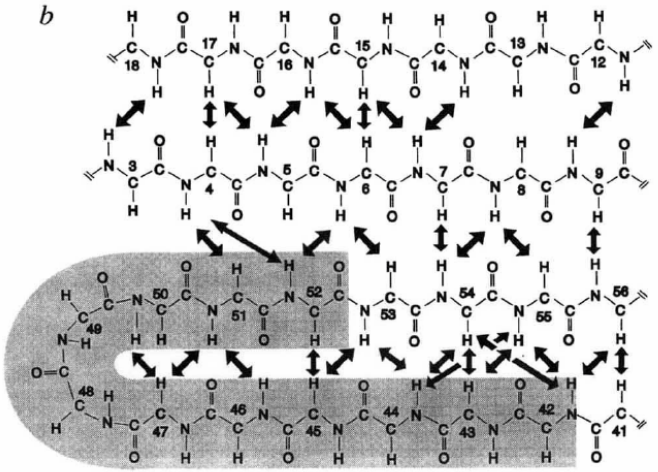


# NMR analysis of chameleon proteins

CD measurements show the proteins are well folded, whereas 0Nuclear Overhauser Effects (NOEs) shows that each folds to the predicted structure





# Evolution of protein fold

What is the minimum amount of sequence perturbation to change the protein fold?

Highly disruptive mutations in T4 lysozyme have not significantly altered the overall structure of the protein

- hydrophobic to charged substitutions
- insertion of secondary structural elements

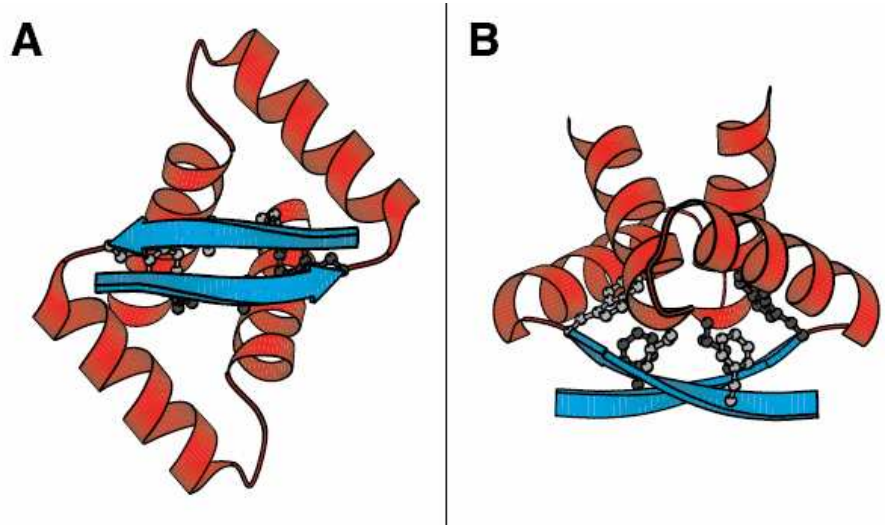
--Vetter et al, Protein Sci 5, 2399 (1996)

This seems to imply that only drastic or large-scale mutations are required to perturb the fold of a protein

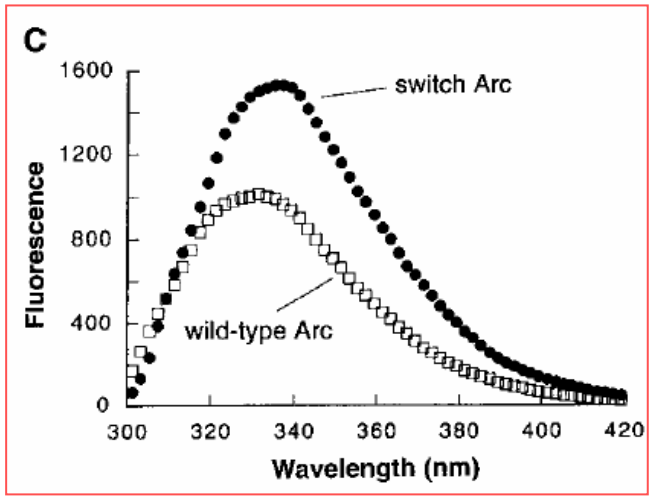
# “Switch” Arc repressor

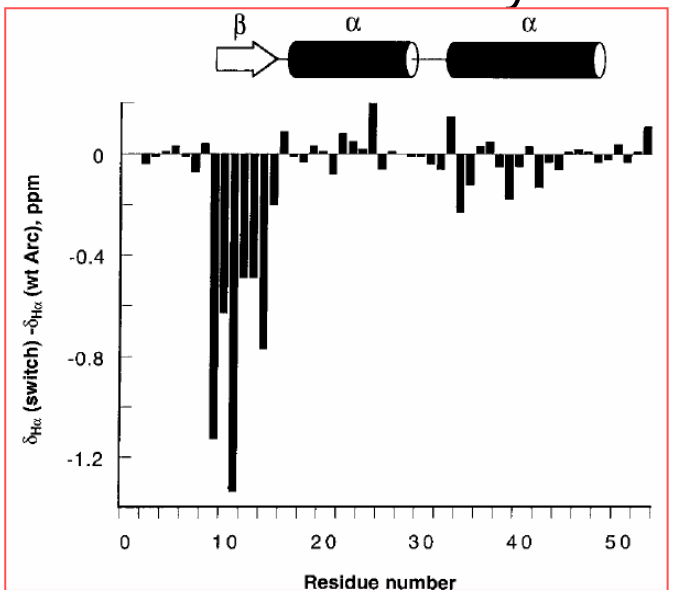
Arc repressor of bacteriophage 22 a homodimer formed by strand exchange and binds DNA through solvent exposed strand residues

Strand sequence is : 9-QF**NL**RW-14



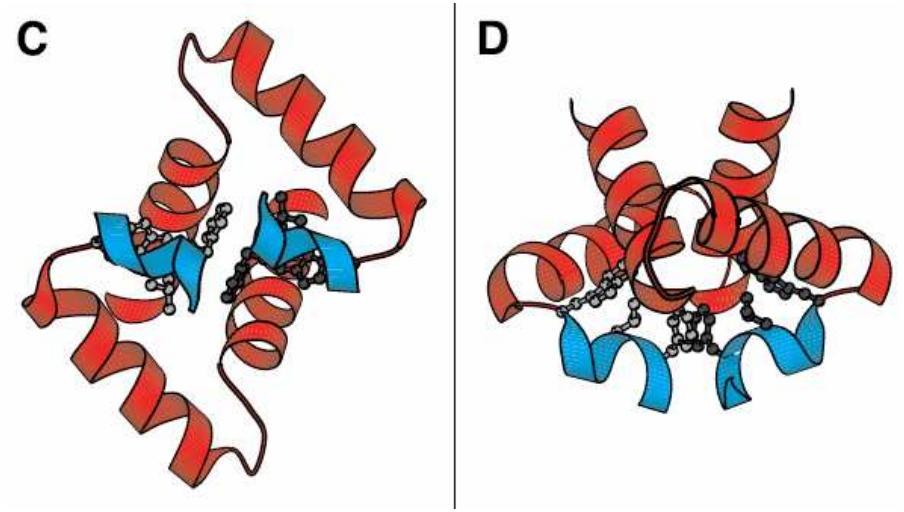

“Switching” N11 and L12 disrupts the hydrophobic pattern of the strand and creates structural changes that are observable by CD and NMR




Cordes et al, Science 284, 325 (1999)

NMR structure of “switch” Arc shows the strand has been replaced by a short helix



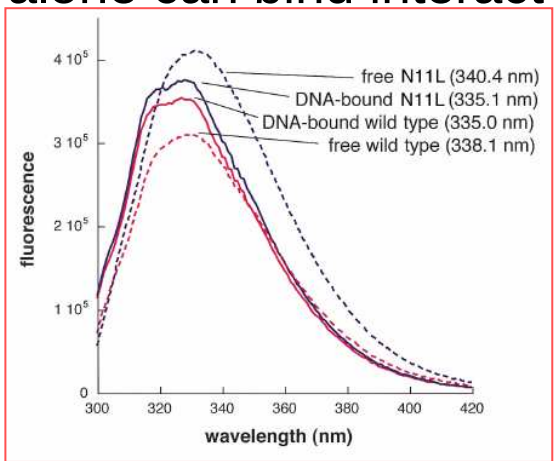

A simple swapping of a hydrophobic residue and a hydrophilic residue resulted in a change of the secondary structure

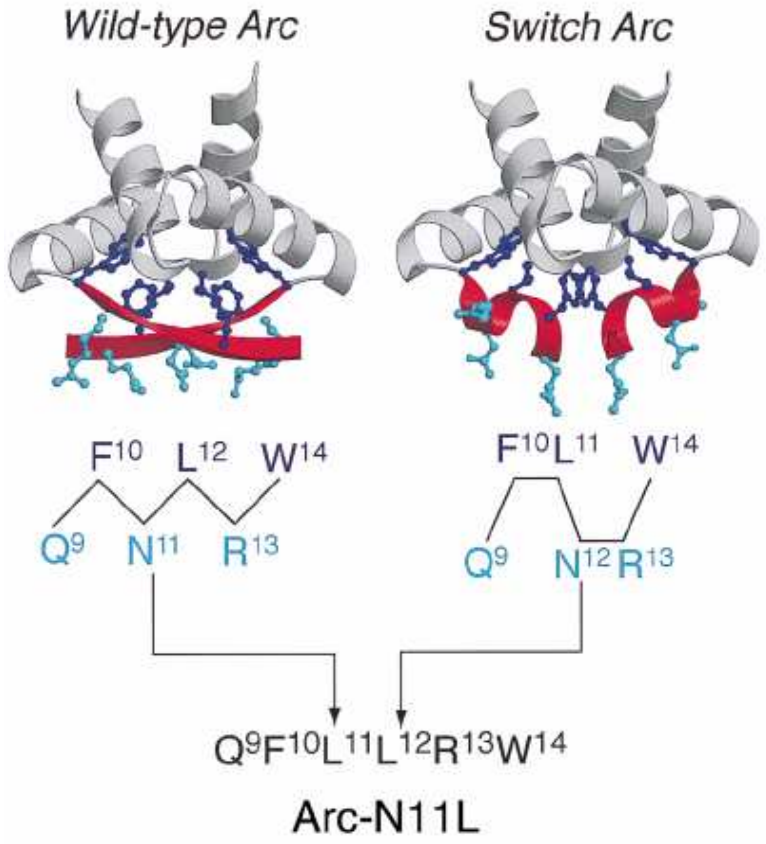
# Protein fold in motion

Modifying the hydrophobicity pattern has changed the lone beta strand of wild type Arc repressor to a helix of the double mutant “switch” Arc

A single point mutation at N11 to L11 produces a structure that alternates between the wild type conformation and switch Arc conformation on millisec time scale (observed by NMR)

Addition of DNA stabilizes the wild type conformation, which alone can bind interact with DNA




Cordes et al, NSB 7, 1129 (2000)