Interplay between native state topology and sequence in two-state protein folding

Daniel Trotter, Stefan Wallin

Department of Physics and Physical Oceanography,
St John's, Newfoundland

MEMORIAL
UNIVERSITY

Two-state protein folding

- Many globular proteins of ~50-120 amino acids

- Two structurally distinct states, U and N, separated by a single free energy barrier (TS)

- *Minimal models* for folding.
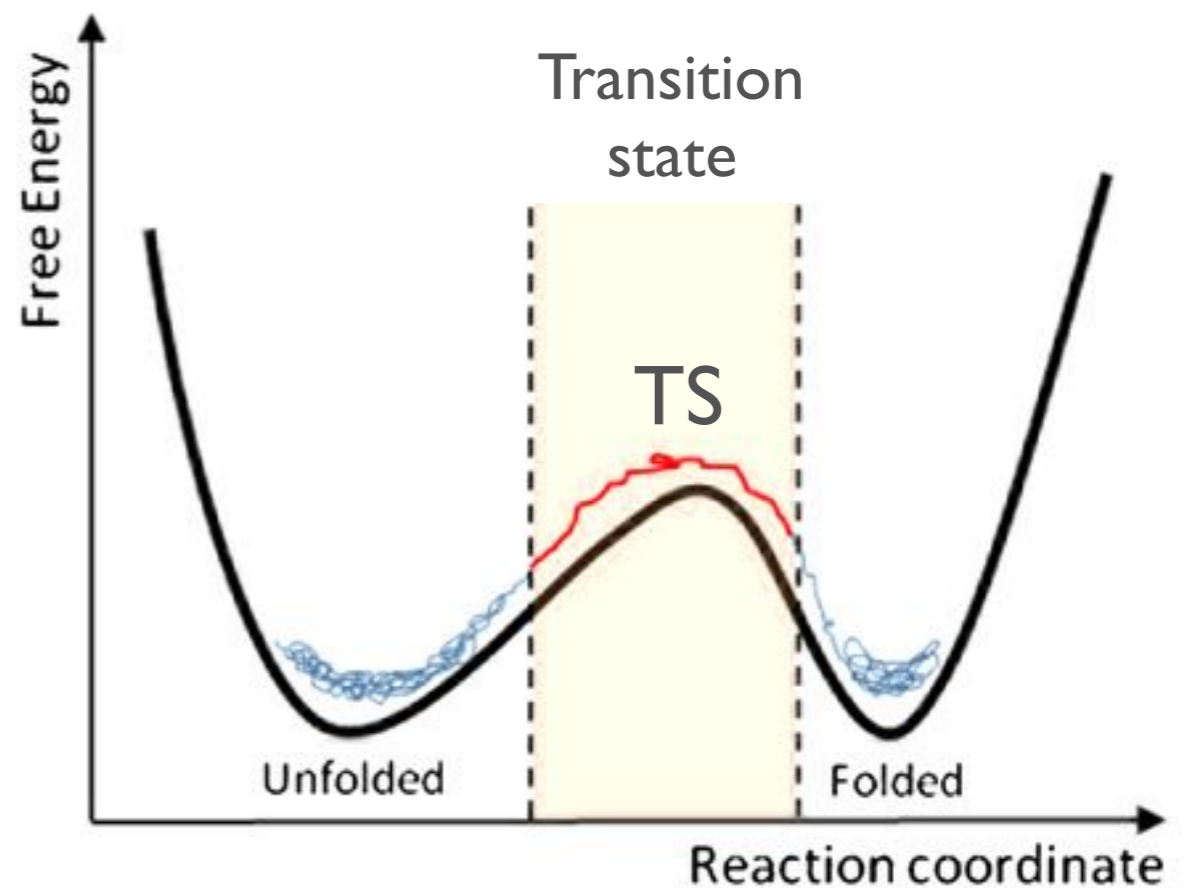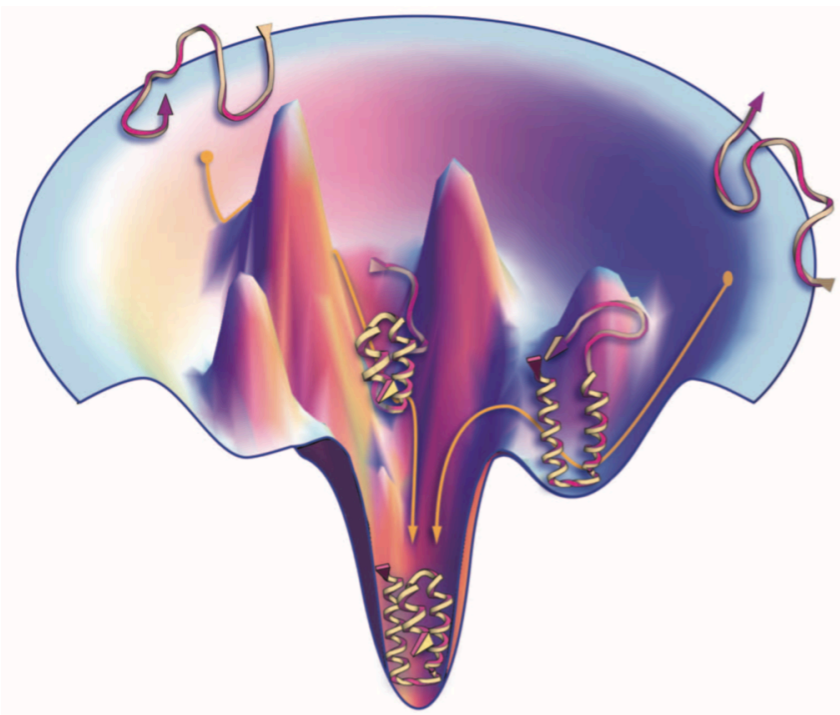
Folding rate $k_f \propto \exp\left(-\Delta G / k_B T\right)$

Effects of topology and sequence on protein folding

Topology. Folding rates k_f span ~6 orders of magnitude

Higher number of "local" vs "nonlocal" contacts means higher folding rate.

RCO = average sequence separation |i - j| between contacts ij in native structure.

Plaxco, et al. 1998

Sequence. Effects pronounced for some topologies but not others.

Two SH3 domains (Fyn and Spc) with only ~30% sequence identity but conserved TS.

Zarrine-Afsar, et al. 2005
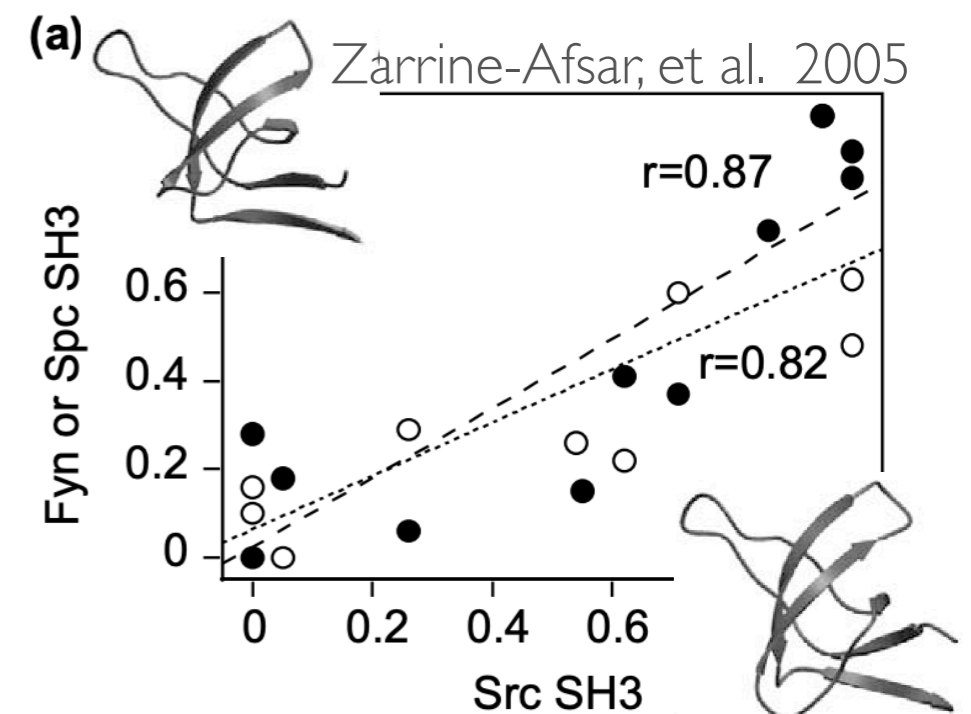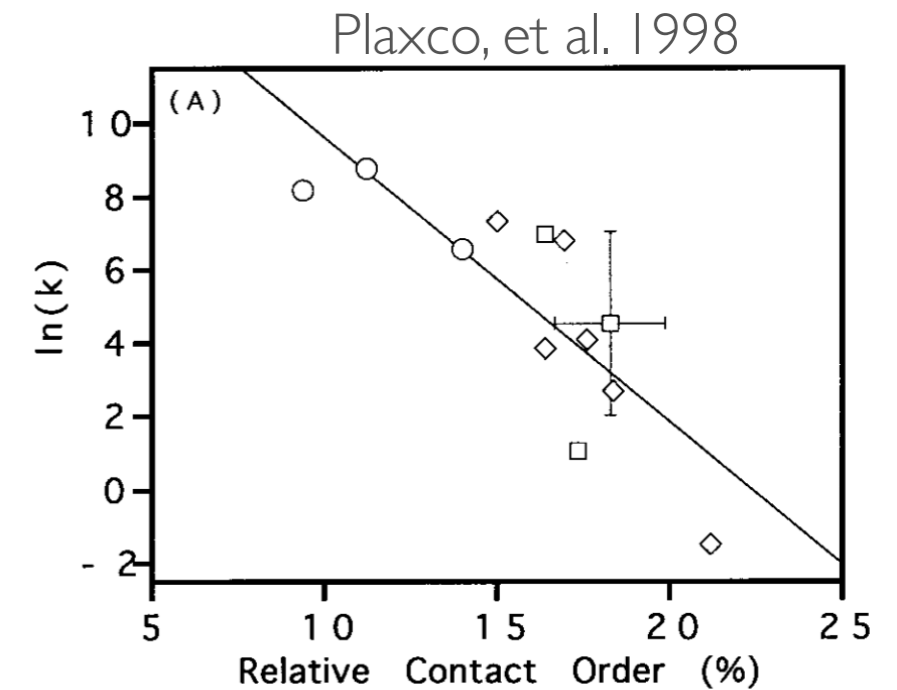
Why is folding into "non-local" folds, e.g. **β**-barrels, more robust to sequence changes than folding into "local" folds, such as **α**-helical bundles?

Coarse-grained "C$_\beta$ model" for protein folding

All-atom backbone/single-site sidechain representation.
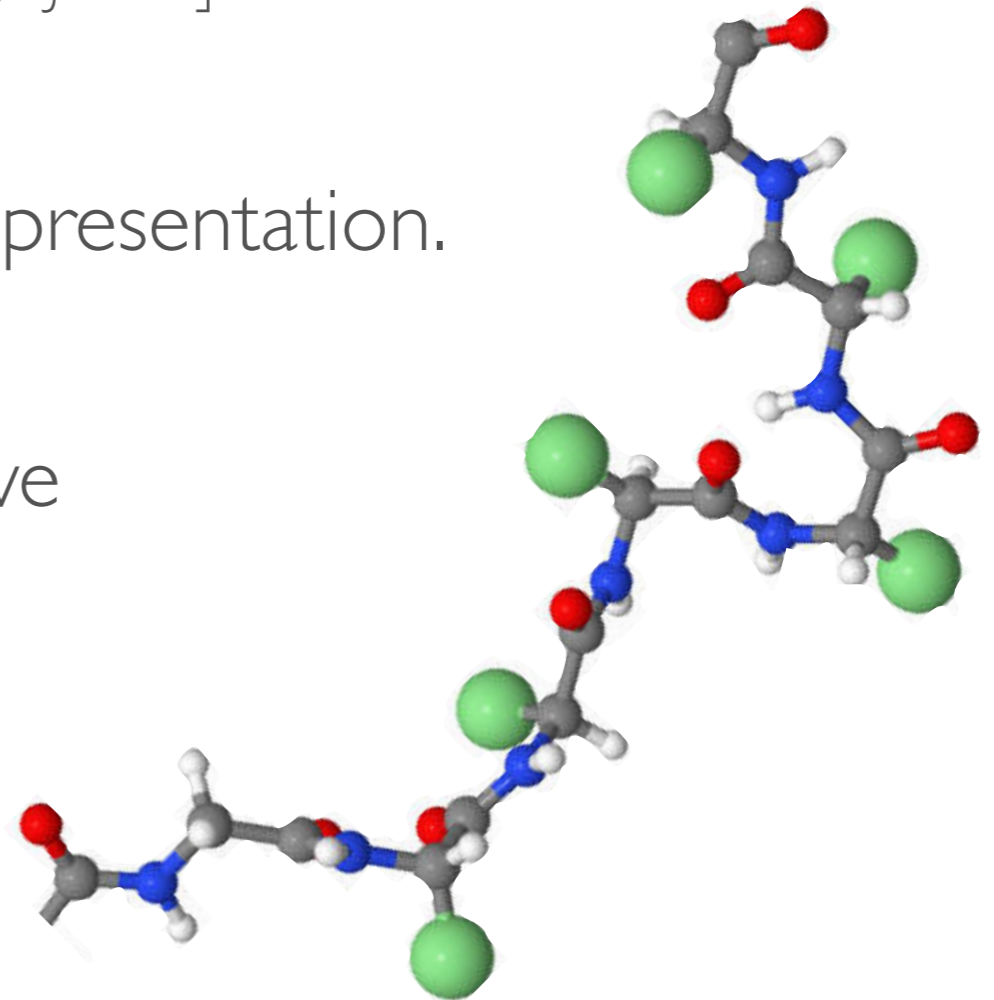
Potential energy function based on effective *hydrophobic forces* and *hydrogen bonding*.
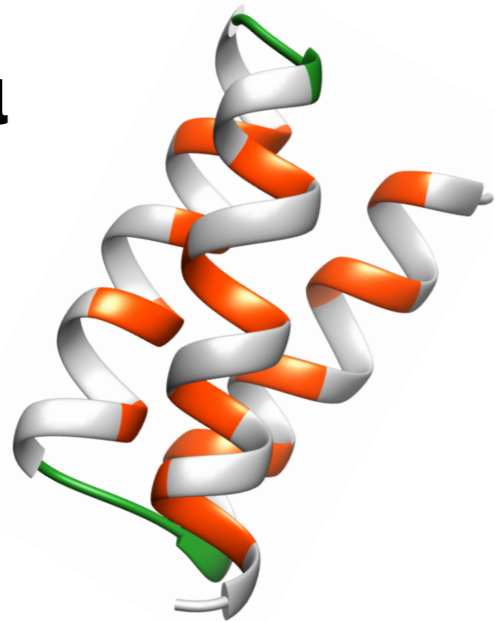
I. Sequence-based.

   — 3 amino acid types: hydrophic/polar/turn.
   — not "Go-type" or structure-based.

II. Model sequences fold into realistic protein folds
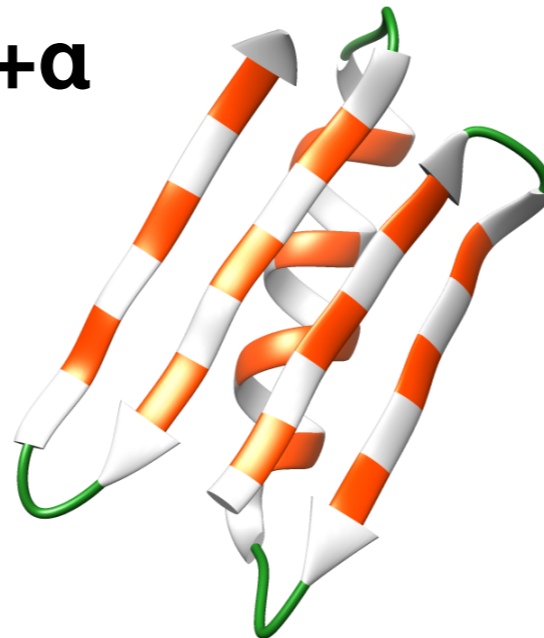
   — both **α**-helix and **β**-sheet structure

3α

4β+α

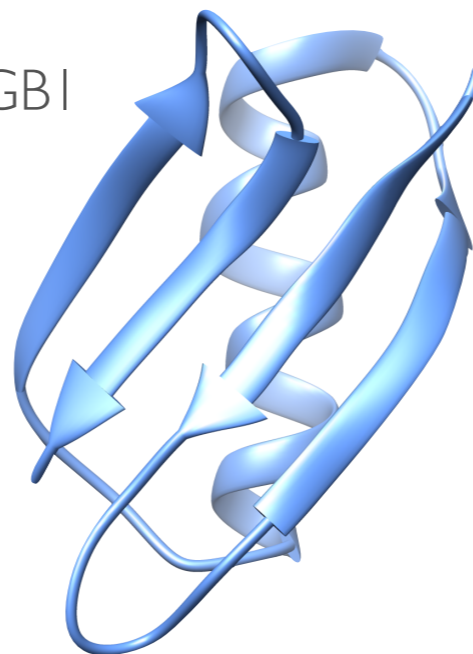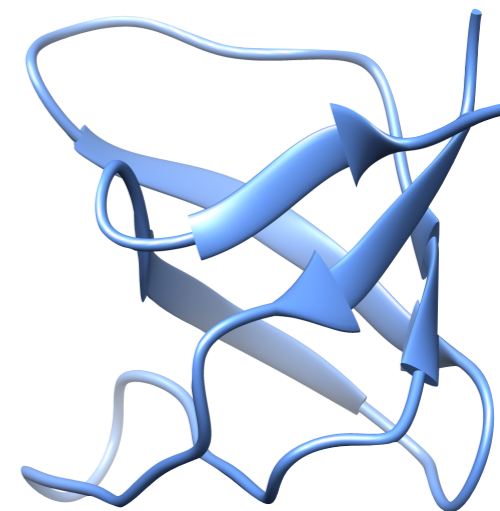**β-barrel**

$L = 54$

$L = 54$

$L = 35$

PDB id: 1BDD
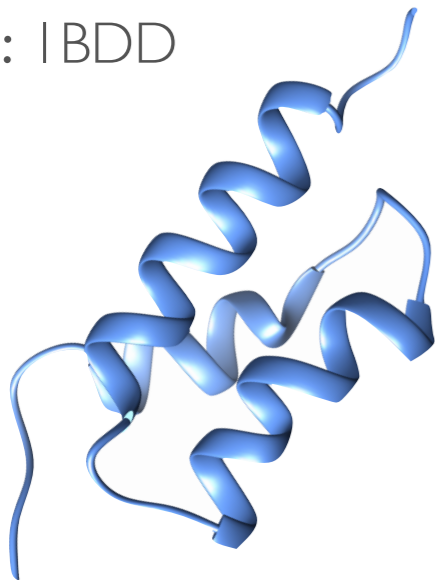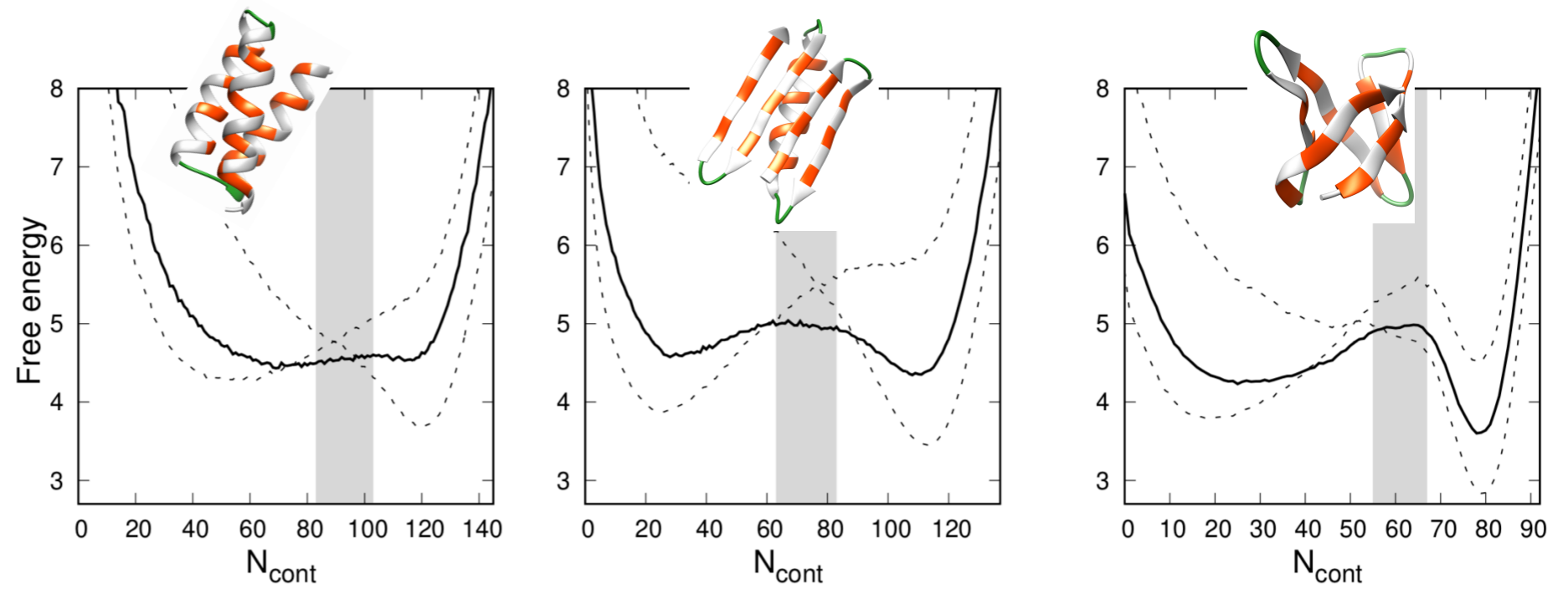
2GB1

1SHF

A domain of
protein G

B domain of
protein G

Fyn SH3 domain

Model proteins exhibit topology-dependent folding

Rank-order of cooperativity follows RCO trend at folding temperature Tf.

Extract transition state ensemble (TS) from peak barrier location.

Exploring the sequence effects on protein folding

1) Generate all possible hydrophobic/polar *single*/*double*-point mutants

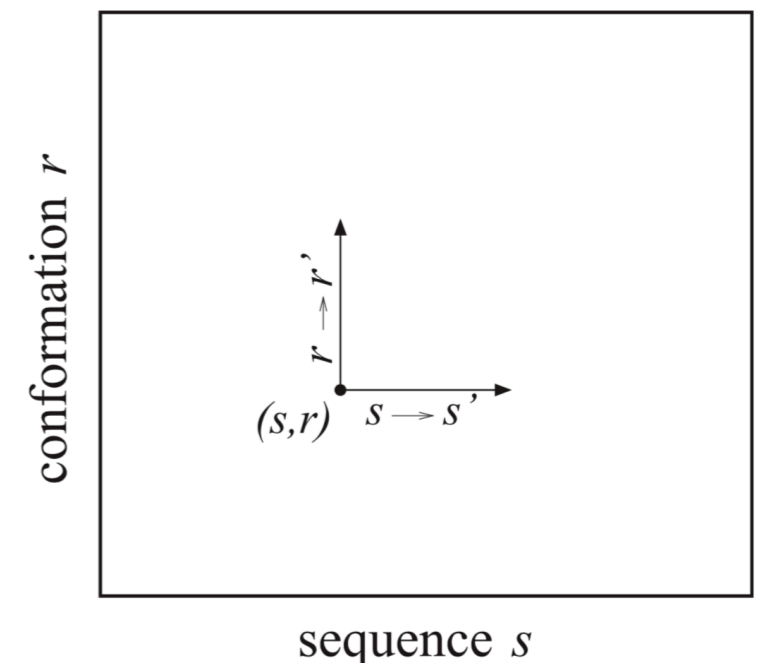$\approx$400-1200 possible such mutant sequences per protein

2) Determine equilibrium behaviour of all mutants at $T = T_f$.

Simulate the joint probability distribution $\quad P(s,r) \propto e^{-E(s,r)/k_{\mathrm{B}}T+h(s)}$
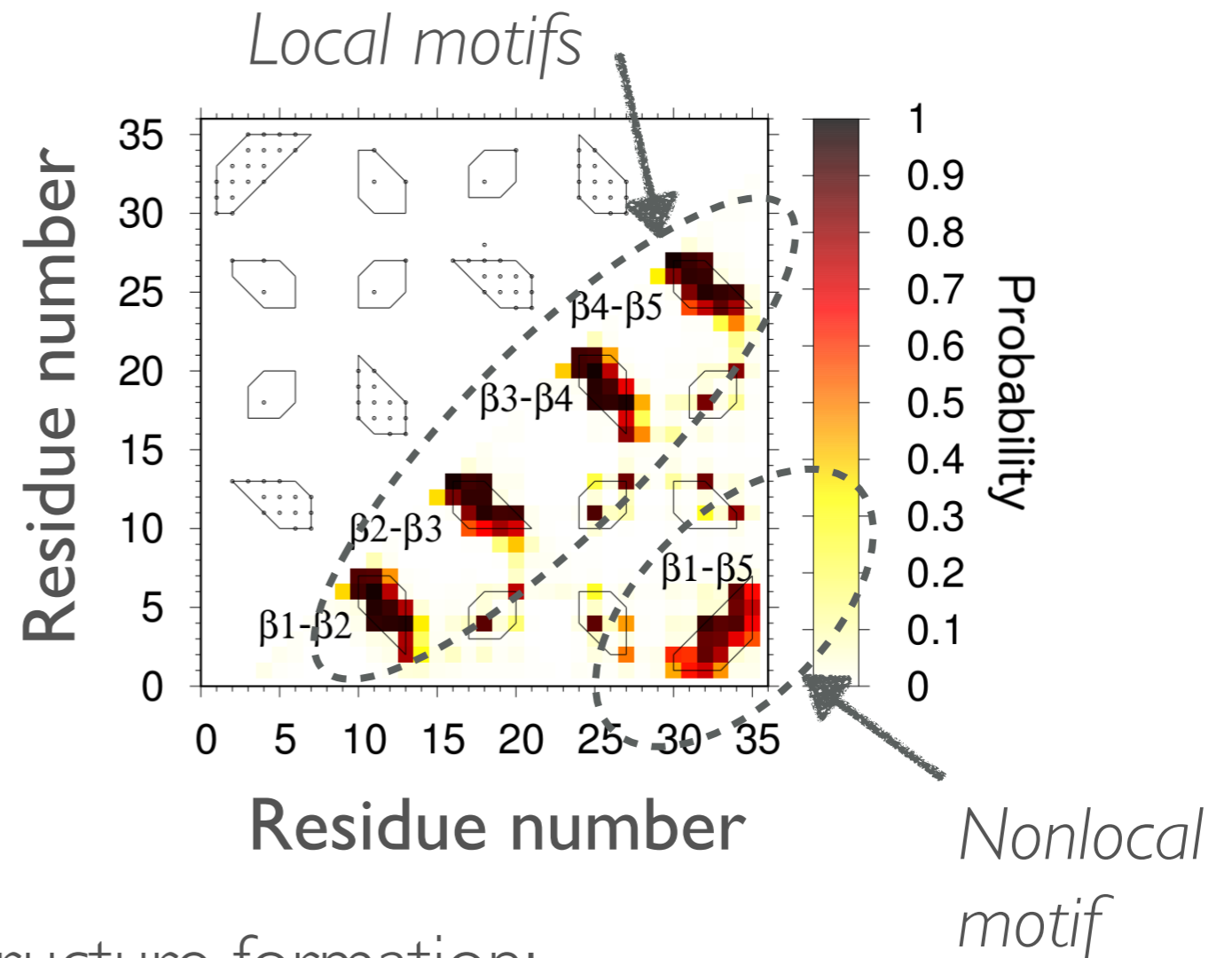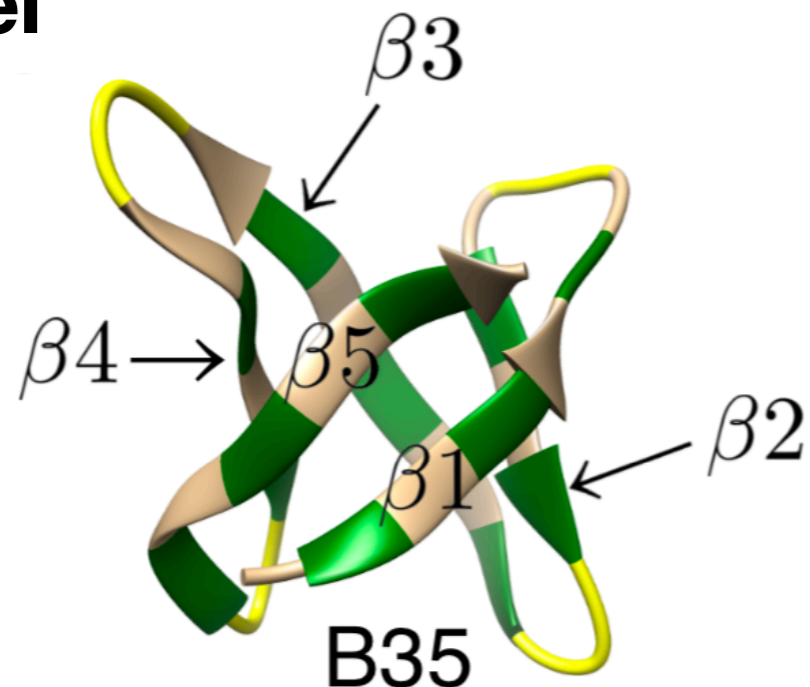
1. Conformational update

pivot

2. Sequence update

seq s → seq s'

conformation r

(s,r) $r \to r'$ $s \to s'$

sequence s

[Aina and Wallin, JCP 147 095102 (2017)]

Monitoring structure formation during folding

β-barrel
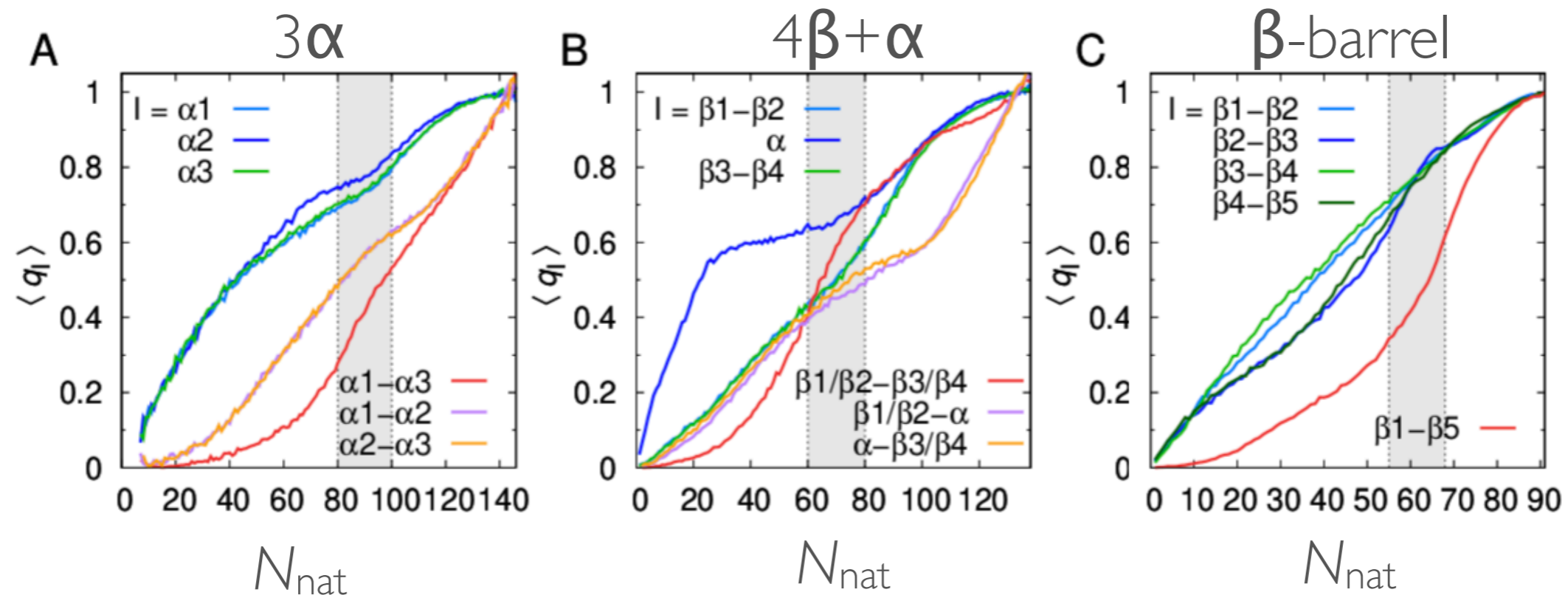

Two different variables describe structure formation:

q_I = fraction of native contacts formed in motif I

ϕ_i = fraction of native contacts formed for residue position i

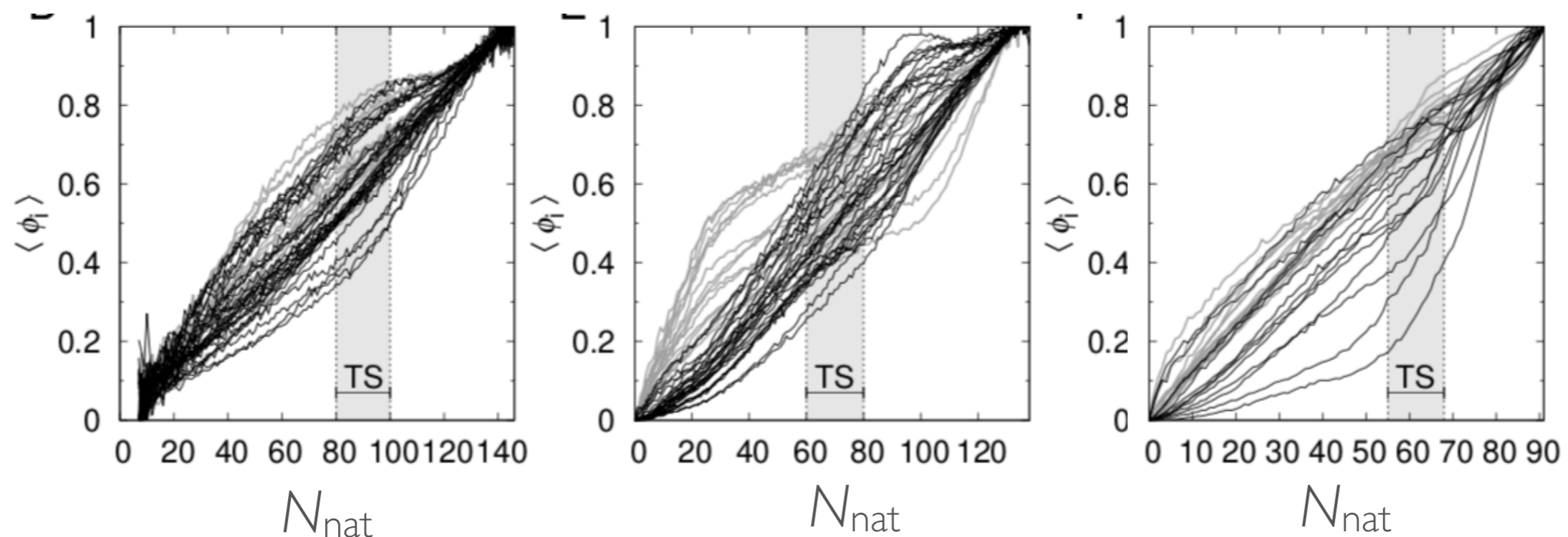Overall folding progress:

N_{nat} = total number of native contacts formed

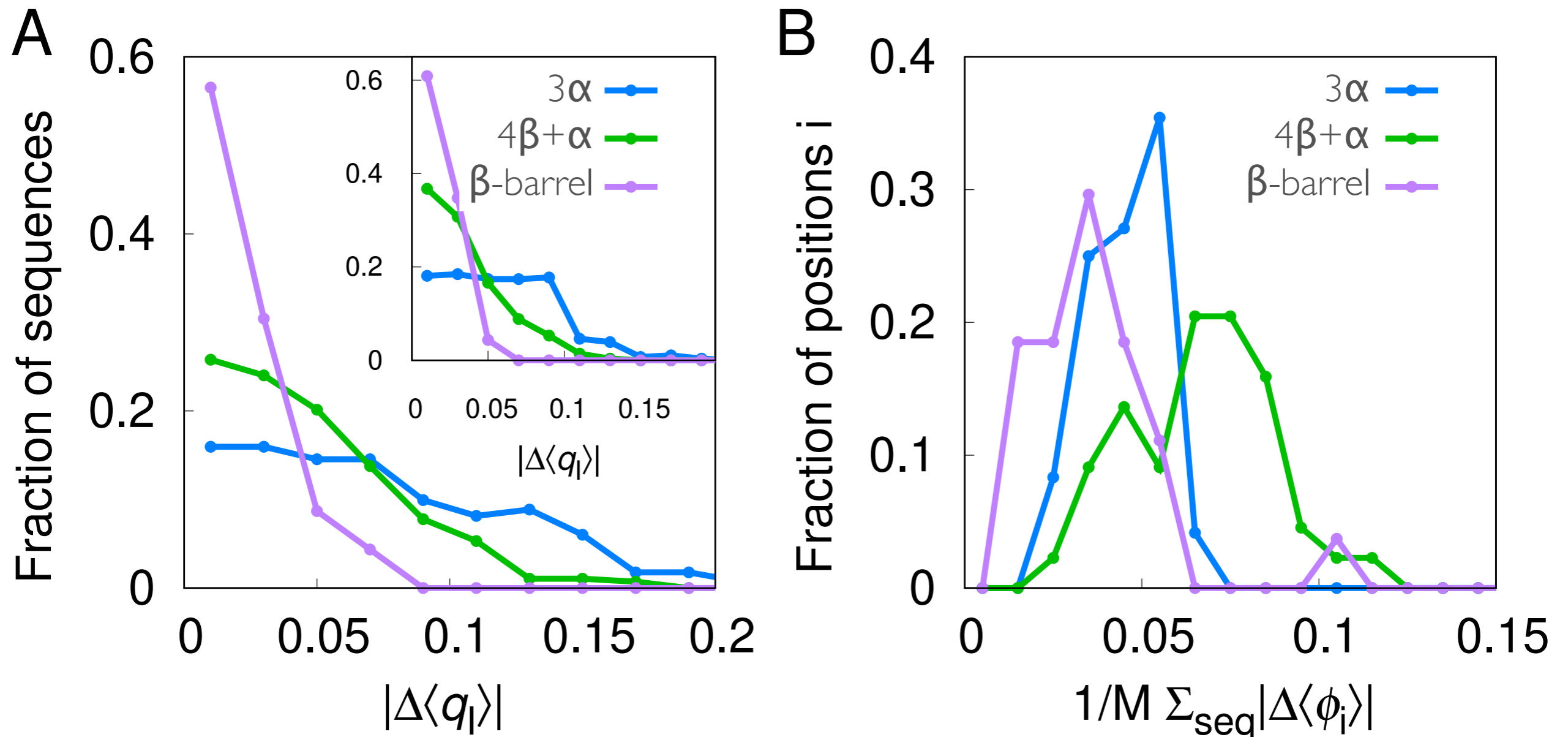# Formation of *nonlocal* contacts drive cooperativity in folding

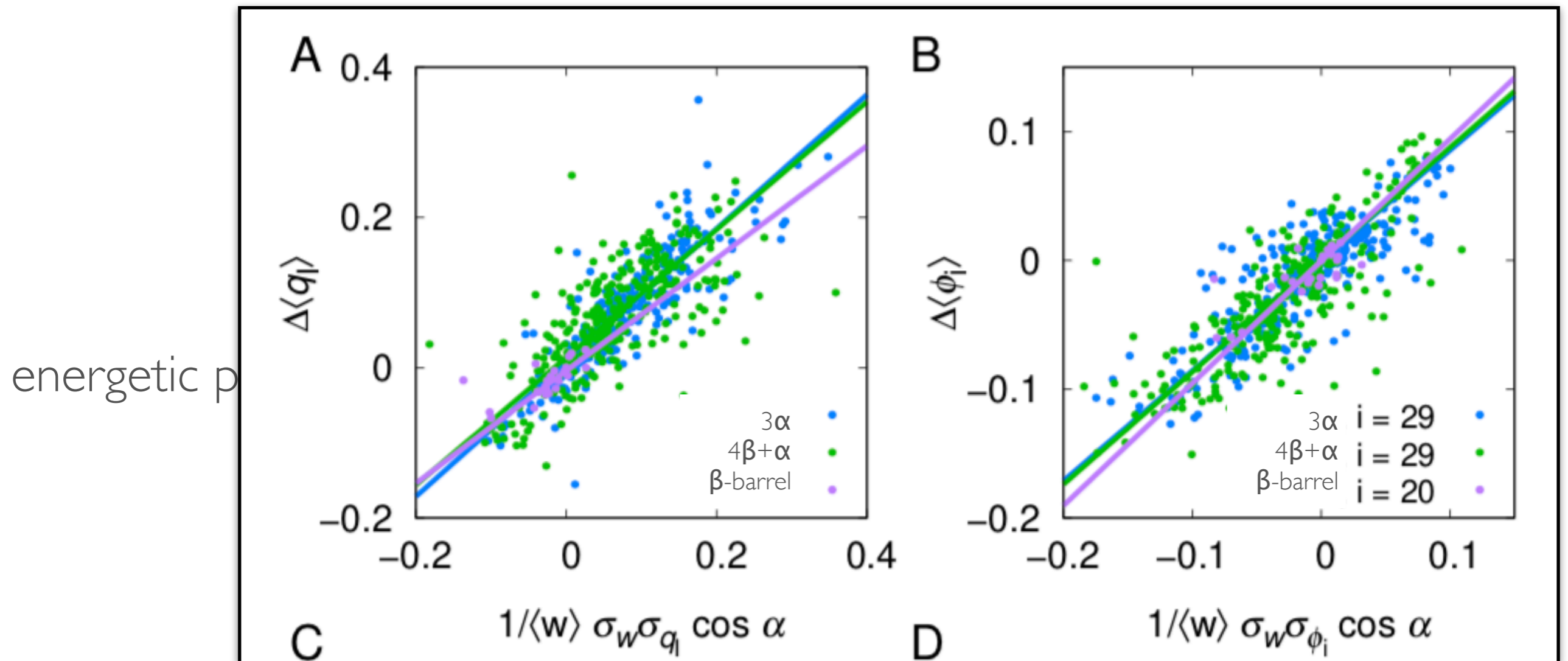# … and underlie a greater ϕ-value diversity in all-**β** protein

Sequence effects on folding TS

Mutations induce shifts in equilibrium quantities, e.g., $\Delta\langle q_I\rangle = \langle q_I\rangle - \langle q_I\rangle_0$

- q_I and ϕ_i-values of β-barrel protein least perturbed by mutations

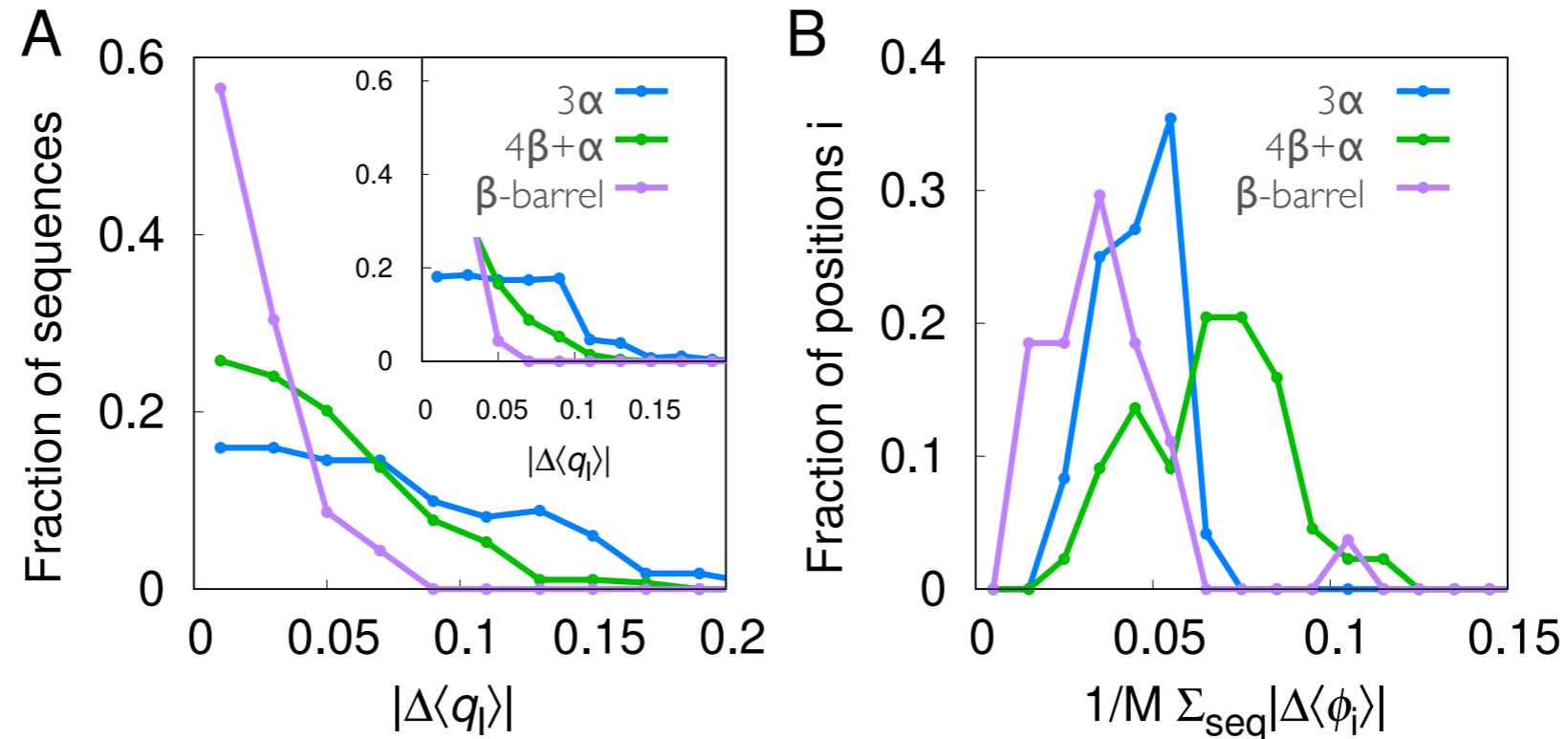# Can features of the "parent" protein explain the observed mutational response?

energetic p

A

$\Delta\langle q_l\rangle$

3α •
4β+α •
β-barrel •

$1/\langle w\rangle\ \sigma_w\sigma_{q_l}\cos\alpha$

B

$\Delta\langle\phi_i\rangle$

3α i = 29 •
4β+α i = 29 •
β-barrel i = 20 •

$1/\langle w\rangle\ \sigma_w\sigma_{\phi_i}\cos\alpha$

C

$$\Delta\langle X\rangle = \frac{1}{\langle w\rangle_s}\sigma_w\sigma_X\cos\alpha$$
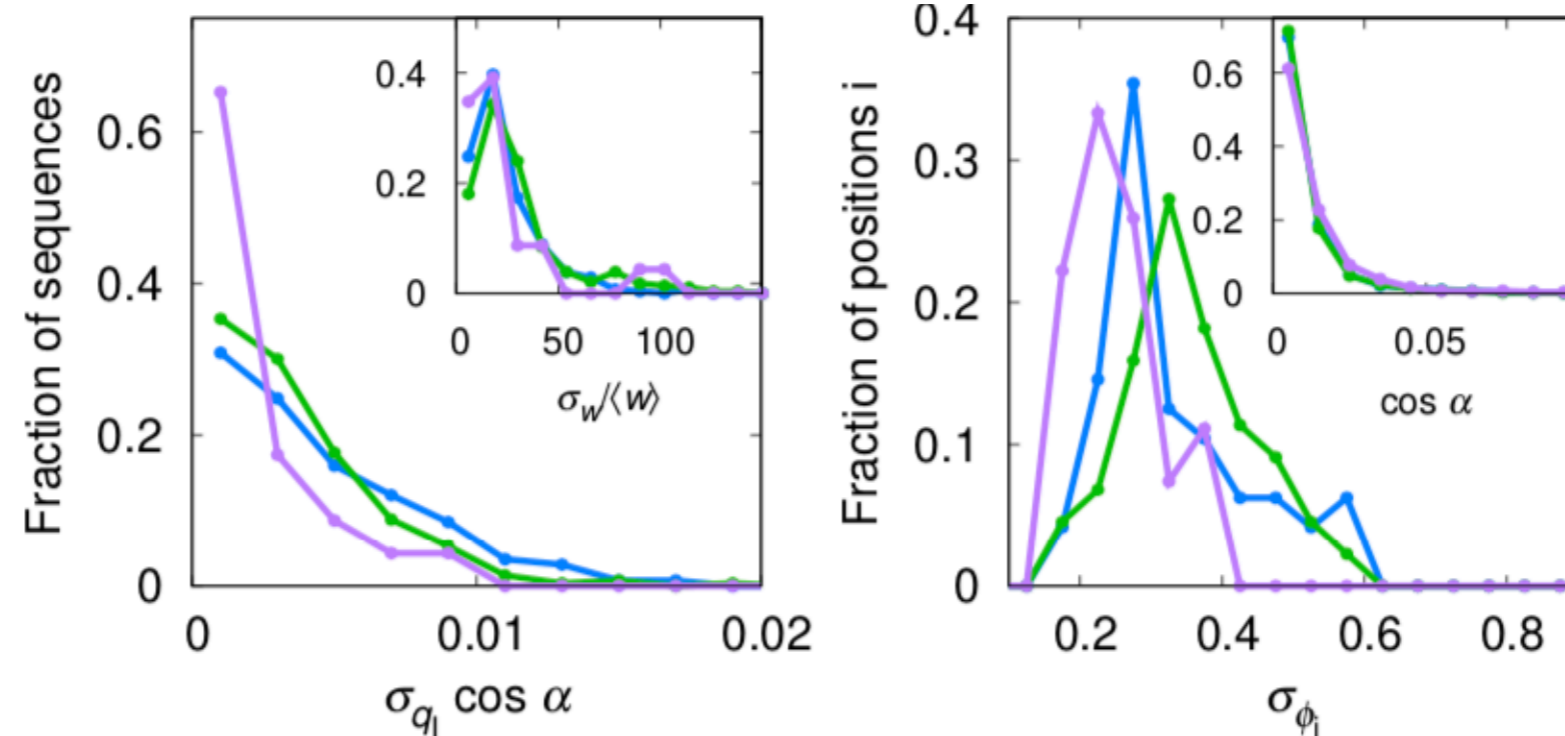
D

relevant observable X (q_l or $\boldsymbol{\phi}_i$)

Can features of the "parent" protein explain the observed mutational response?
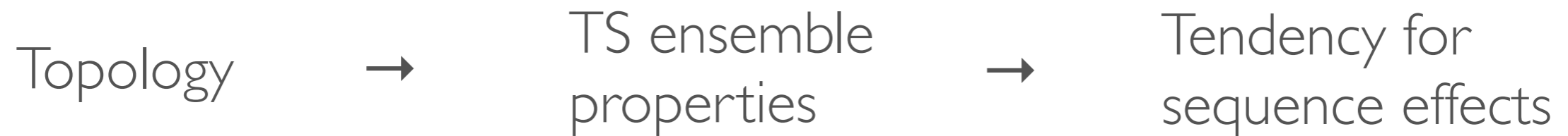
Observed mutational response on TS…

…can be explained by factors $\sigma_X \cos \alpha$ alone

Conclusion

Topology and sequence effects in protein folding are coupled:

Topology $\rightarrow$ TS ensemble properties $\rightarrow$ Tendency for sequence effects

I. Conformational diversity. TS of all-β proteins more structurally restricted than for all-α proteins, leading to weaker mutational response.

In particular, ϕ-values at positions with a broad distribution $P(\phi)$ should tend to diverge with sequence.

II. Mutational-energetic correlations. Conformational variations in TS of all-β proteins less "aligned" with energetic changes than in all-α, again weakening mutational response.