A realistic statistical coil model for the unfolded state

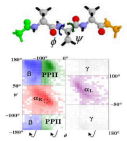
Abhishek Jha[1,2,3], Andres Colubri[1,2], Muhammad H. Zaman[1,2], Tobin R. Sosnick[2,4], Karl F. Freed[1,3]

[1] Department of Chemistry, The University of Chicago, Chicago, IL 60637
[2] Institute of Biophysical Dynamics, The University of Chicago, Chicago, IL 60637
[3] The James Franck Institute, The University of Chicago, Chicago, IL 60637
[4] Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637

Abstract

Unfolded proteins are the starting state for mechanistic and thermodynamic studies of folding and therefore its description is a central issue in protein folding. Here we present a statistical model for the unfolded state that reproduces known properties.

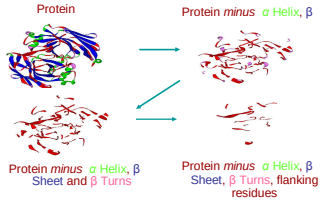
The coil library has been used in earlier studies to detangle the determinants of protein structures and investigate intrinsic conformational preferences of amino acids. Here we present a stringent definition of the coil library, thanks to the huge PDB at our disposal. More importantly, using the new coil library obtained from a non- redundant database, we show that the conformational preference of residues located in coil regions can accurately reproduce *both* helix and β-sheet propensities.



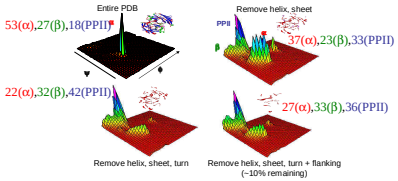
Ramachandran map

Hence, these propensities can be rationalized solely on local effects. Equally surprising, the influence of nearest neighbors is just as important in many cases. Finally, we find that the largely ignored Polyproline II is, in fact, the default conformation of the polypeptide.

Our Coil Library

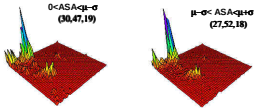


PPII dominates in the coil library



Probability distribution in the Ramachandran plane for all residues (except Gly and Pro) for different coil libraries.

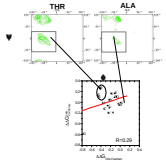
Preference for PPII is an intrinsic polypeptide property Not solely determined by solvation as is often proposed.



The reduction in steric overlap and the accessibility of the backbone for hydrogen bonding likely are the major operational principles underlying conformational preferences, whether the chain resides in water or within a protein.

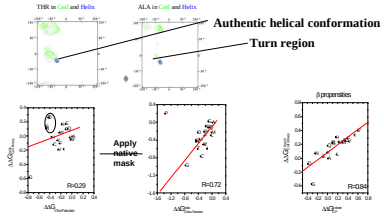
α, β propensities are largely encoded by local interactions

Origin of helical and sheet propensities



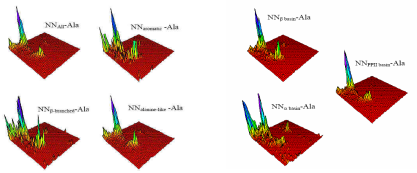
ALA less helical than THR?

This anomaly can be resolved if we apply a "mask" i.e. look only in the region that corresponds to the authentic helical conformation. Using this method, we can reproduce the secondary structure propensities from coil library.



Breakdown of Flory Isolated Pair Hypothesis

Φ,ψ distribution of all residues depends on the identity as well as the conformation of its neighbors



(A) Ramachandran map for alanine from the coil library, averaged over all neighbors (except Pro and Gly), for cases where the upstream residue is β-branched (Val, Ile, Thr), aromatic (His, Trp, Phe, Tyr), or the rest (alanine-like). **(B)** Same as (A) but where the upstream residue is in the β, α, or PPII basins.

References

- Rose *et al.* PNAS 2004
2. Data from Mohana-Borges *et al*, JMB 2004
3. Experimental Data: Shortle *et al* Science 2001; Biochemistry 2004

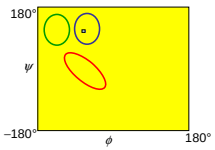
Acknowledgements

Burroughs Wellcome Fund Interfaces Fellowships, NIH

Structure of unfolded polypeptide chains:

An unfolded state ensemble is generated using this coil library and is subjected to excluded volume constraints.

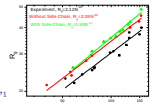
Calculated results are in good agreement with experiments for chemically denatured proteins for both the radius of gyration and NMR residual dipolar couplings (RDC). The agreement with the NMR data improves when the backbone conformational preferences for each amino acid include correlations with the chemical and conformational identity of neighboring residues. The most stretched members of the ensemble of unfolded state structures are found to contribute most to the RDC signal, while the sign of the RDCs for unfolded state proteins follows from the preponderance of polyproline II and b conformers over helical basin preferences for all residues in the unfolded state. Although the ensembles match experimental observables, they do not display evidence of native-like topology.



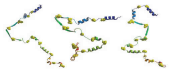
- Choosing basin (α,β or PPII): using NN effect
- Choosing ϕ and ψ in basin: Simulated Annealing
- Excluded volume effect

Global dimensions of polypeptides

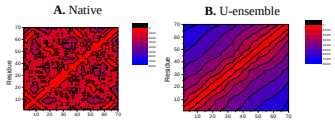
This test is necessary but not sufficient. Reproduction of scaling behavior of global dimensions as a function of size only confirms the "self-avoiding random walk" nature of any polymer.



A "contrived counterexample"[1]



No evidence of native-like topology in unfolded ensemble

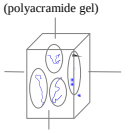


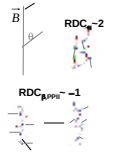
Contour plot of average distance between the C_β–C_β of residues for (A) X-Ray structure of Ubiqutin and (B) unfolded ensemble generated for the same protein using our model. Red indicated that the residues are in contact and blue signifies residues far away from each other.

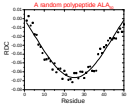
NMR Residual Dipolar Couplings: local backbone conformation

$RDC_{NH} \propto (3\cos^2\theta - 1)$

Measures the orientation of NH vectors in a weakly oriented medium (polyacramide gel)

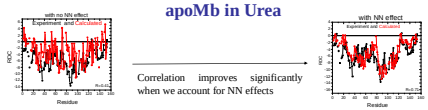






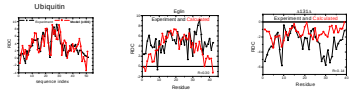
For an ideal random coil, the expected RDC pattern would be a flattened bell shaped curve. A deviation from this pattern as observed in experiments is an indication that the unfolded state of proteins have some structural content.

apoMb in Urea

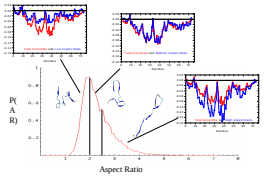


Correlation improves significantly when we account for NN effects

Sometimes yes, sometimes no



Most stretched conformations contribute most to RDC signals



Conclusions

Coil library – old idea, but still useful

PPII is preferred conformation

α, β propensities are locally determined

Unfolded state: *Statistical* rather than a *random* coil

Preliminary version of a web-server has been set up to generate initial conformations for dynamics simulations