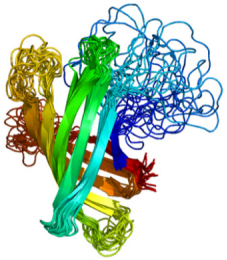
Modeling Ensembles of Transmembrane β-barrel Proteins

Jérôme Waldispühl[1,2,*], Charles W. O'Donnell[2,*], Srini Devadas[2], Peter Clote[3], Bonnie Berger[1,2]

1. Department of Mathematics, MIT, Cambridge, USA
2. CSAIL, MIT, Cambridge, USA
3. Department of Biology, Boston College, Chestnut Hill, USA

Contact: jeromew@mit.edu
* Equal contribution



IMA Workshop: Protein Folding
January 14-18, 2008

Overview

Objective: Compute statistical properties of ensembles of structures rather than predicting a single structure.

Target: Transmembrane β-barrel proteins.

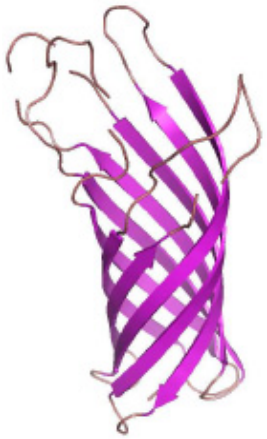
Method: Calculate the partition function over TMB structures and analyze the Boltzmann distribution.

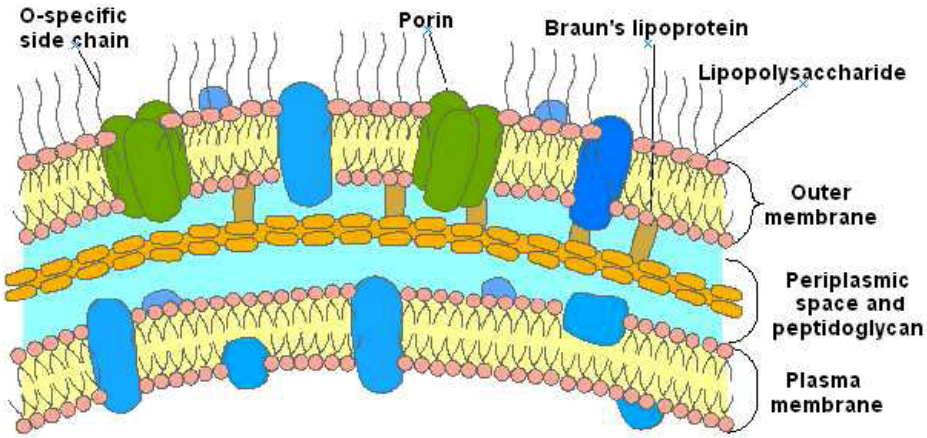
Principles:

- Describe the conformational space: Abstract template (grammar).
- Weight the structure: Energy function.
- Efficient algorithm: Dynamic programming.

Transmembrane β-barrel proteins

- TMBs: Found in outer-membranes (gram-negative bacteria, Mitochondria)
  Important for signaling, drugs, etc

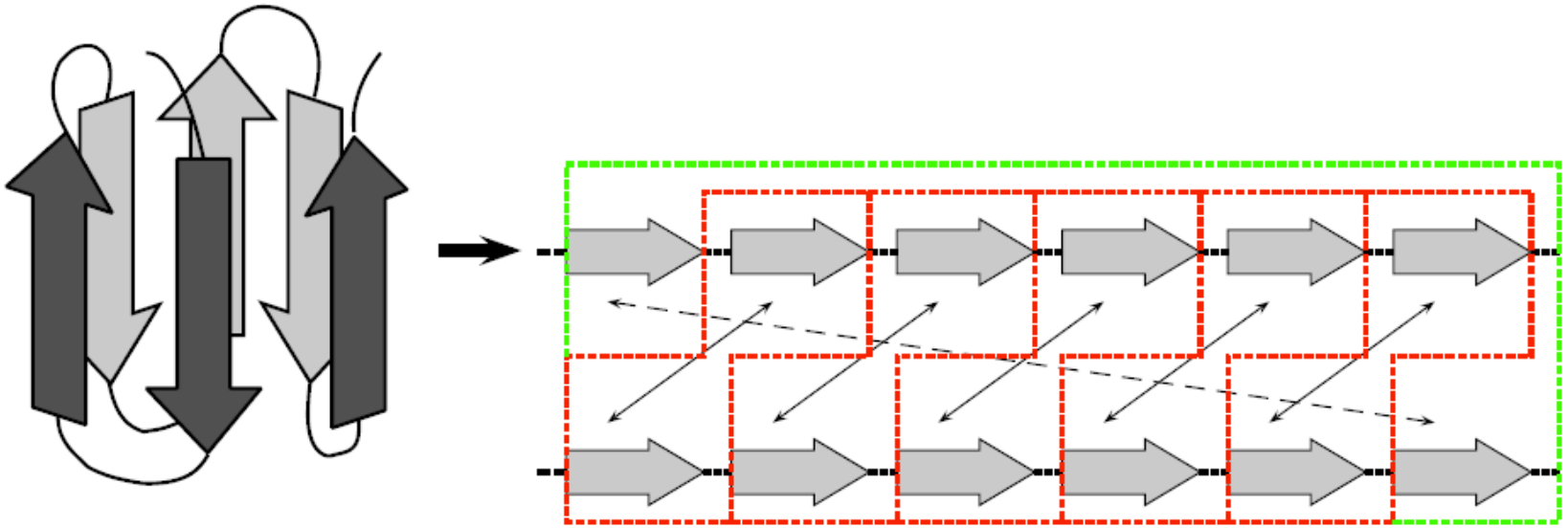




- Difficult to solve with X-Ray/NMR techniques,
  20 non-homologous structures in PDB.
- *TM barrel fold* highly conserved across species,
  but *high sequence variability* (Schultz'00).
- TMBs undergo conformational changes in vivo (Tamm'04).

Modeling with grammars

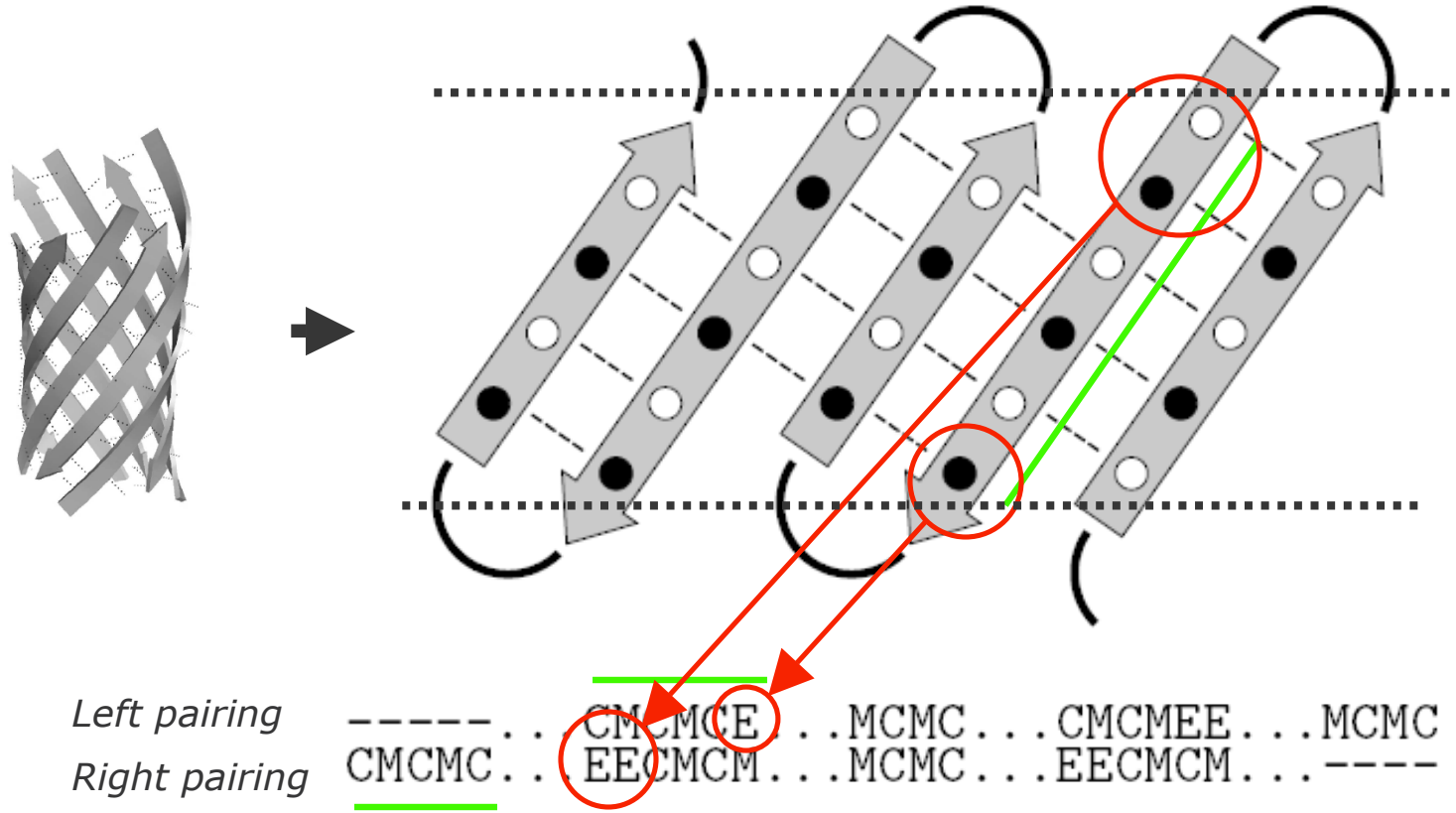
- 2-tape grammar model TMB protein containing only β–strands and loops/random-coils



Property: Anti-parallel strand pairs are isolated.

Modeling with grammars (2)

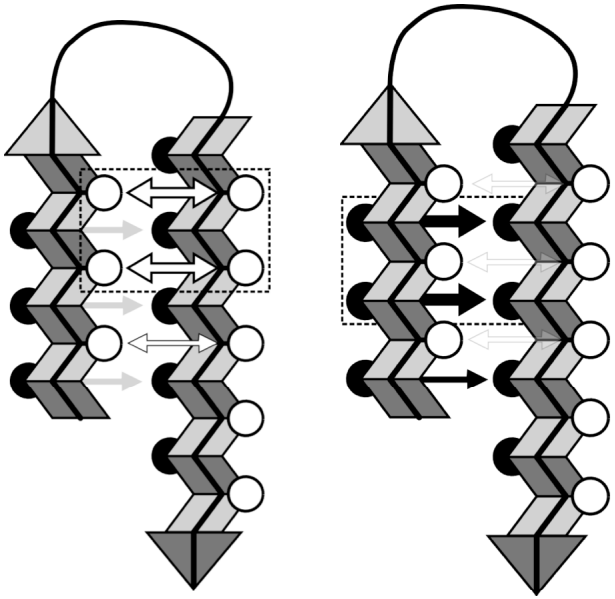
- Strand inclination (shear) and variable strand length handled by strand *extensions:*



- Distinguish side-chain orientation (`M`: Membrane, `C`: Channel)

Energy model

- Instead of pairwise interactions (BETAWRAP), consider stacking pairs:

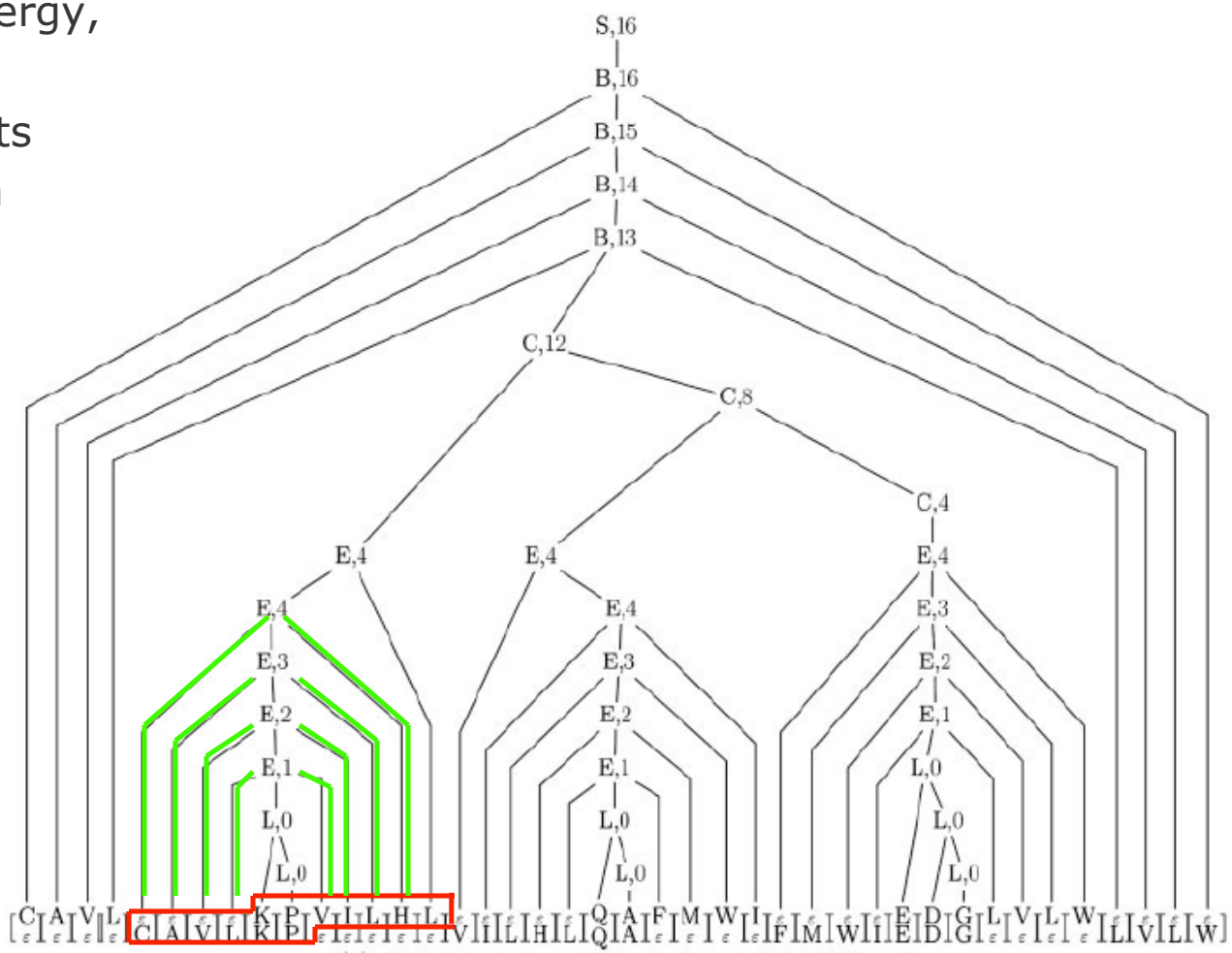


$$E(i,j,x \mid i+2, j+2) = -RT\log\left(p_{i,j,x \mid i+2,j+2}\right) - RT\log\left(Q_{tmb}\right)$$

- Requires 2×20^4 values of $p_{i,j,x}$ so must use *reduced alphabet.*
- Potentials computed from a dataset of *globular* proteins to overcome the small dataset problem.
- Distinguish interaction environment by similarity:
  Membrane=Buried, Channel=Exposed.

Multi-tape S-Attribute Grammars

- Parse of tree gives structure,
- Node labeled with energy,
- Additionnal constraints can be added to each node.



Boltzmann ensembles

- Boltzmann Partition Function $Q(s) = \sum_{s \in S(s)} e^{-E(s)/RT}$
  - Encodes statistical mechanical properties of the system:

$\langle E(s) \rangle = RT^2 \frac{\partial}{\partial T} \ln Q(s)$ $C = \frac{\partial \langle E(s) \rangle}{\partial T}$ $A = -RT \ln Q(s)$ $S = -\frac{\partial A}{\partial T}$

- Efficient algorithms using dynamic programming principles.
  grammar allows to use parsing algorithm (CKY, Earley, GCP…)
- **Output:**
  - Partition function value,
  - Stochastic backtracking: Structure sampling,
  - Residue interaction probability.
- **Allows:**
  - Whole structure prediction through clustering of samples,
  - Residue contact prediction,
  - Prediction of B-value (reproduce experimental observation).

Results: Residue contact probability

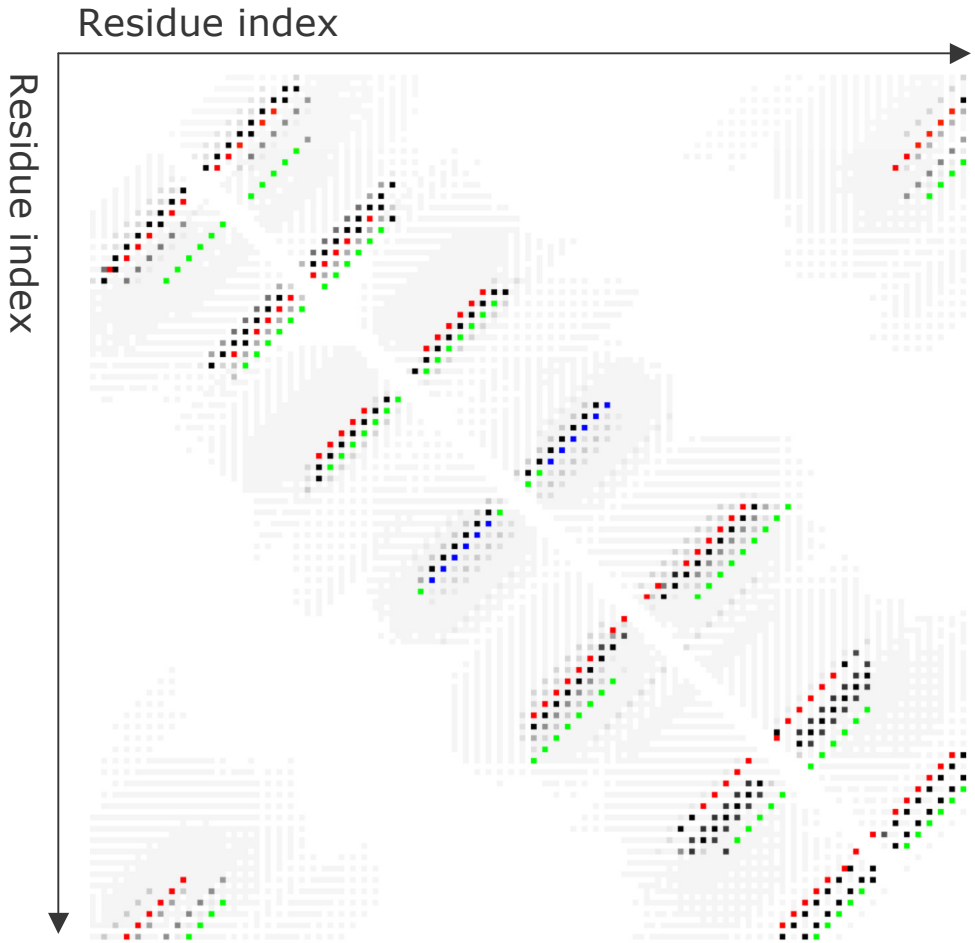
Partition function of all TMBs with contact (i,j):

$$Q(i,j) = \sum_{s_{(i,j)} \in S} e^{-E(s_{(i,j)})/RT}$$

Contact probability:

$$p(i,j) = \frac{Q(i,j)}{Q_{tmb}}$$

p(i,j) assembled in a *stochastic contact map*:



Red: Crystal structure
Green: M.F.E. structure
Upper triangle: Membrane
Lower triangle: Channel

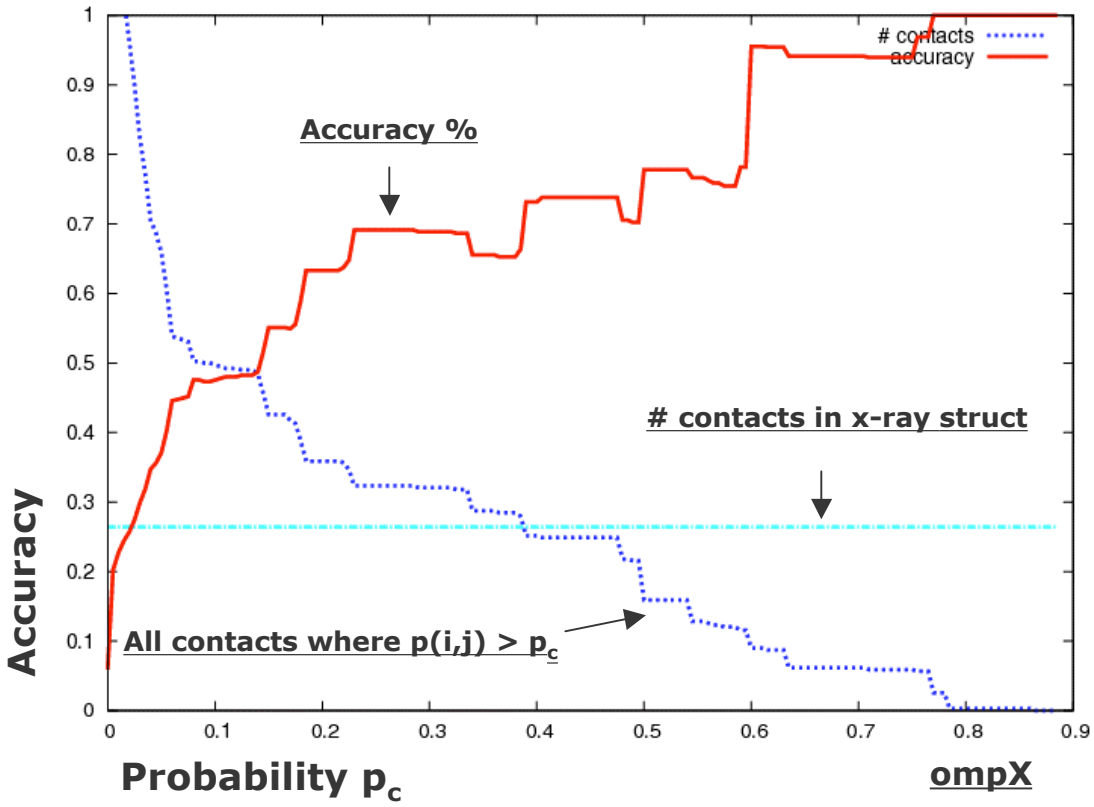
Can be used to help reconstruct 3D models (Grana'05, Punta'05)

Results: Residue contact probability (2)

Prediction of contacts by filtering $p(i,j) \geq p_c$

Coverage: TP/(TP+FN)
Accuracy: TP/(TP+FP)
F-measure:
(2×cov×acc)/(cov+acc)

Comparison with BETApro (general β-strand predictor; Cheng&Baldi, 2005)



F-measure peak	1QJ8	1P4T	1QJP	1THQ	1TLY	1K24	1I78	1QD6
partiFold	**0.66**	**0.38**	**0.27**	**0.18**	**0.43**	0.40	**0.27**	0.16
BETApro	0.49	0.14	0.22	0.05	0.08	**0.56**	0.15	**0.66**

Results: B-value prediction

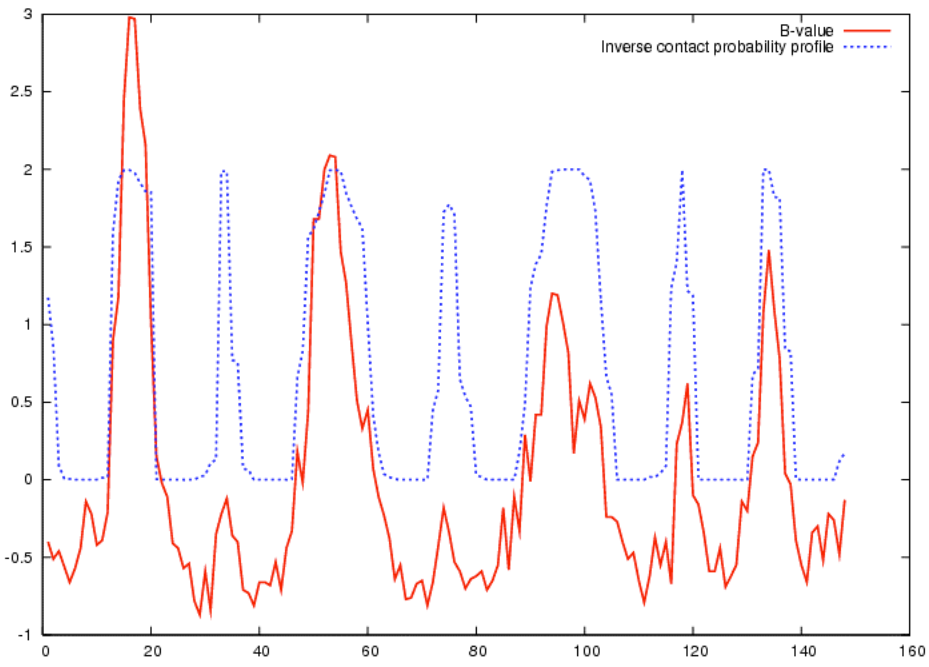
- Contact probability profile $P_{cp}(i)$
  - Frequency of β–strand pairing per residue

$$P_{cp}(i) = \sum_{j=1}^{n} p_{(i,j)}$$

- *Debye-Waller factor* (*B-value*) in x-ray crystal structures
  - Indicates uncertainty or disorder in crystal

- *Higher* values of $P_{cp}(i)$ and *B-value* indicate more flexible regions (eg. loops)

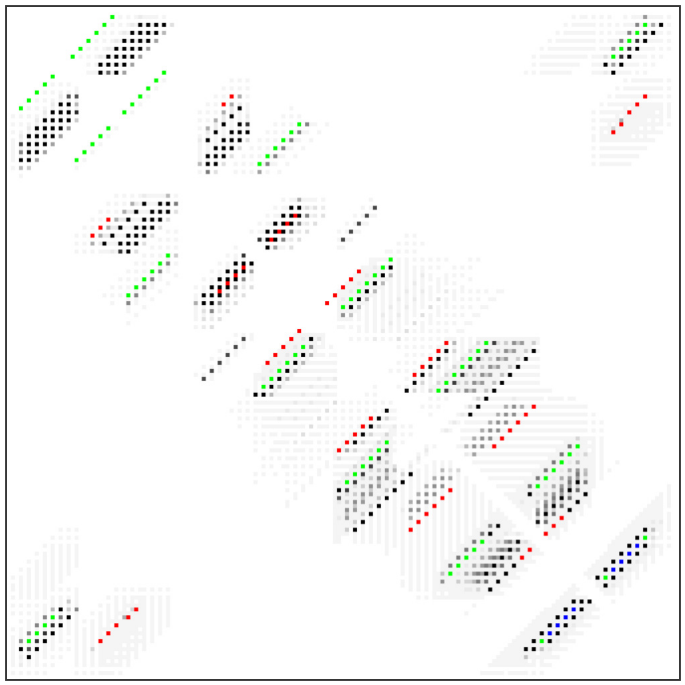
- Match the performance of PROFbval (Schlessinger&Rost,2005)

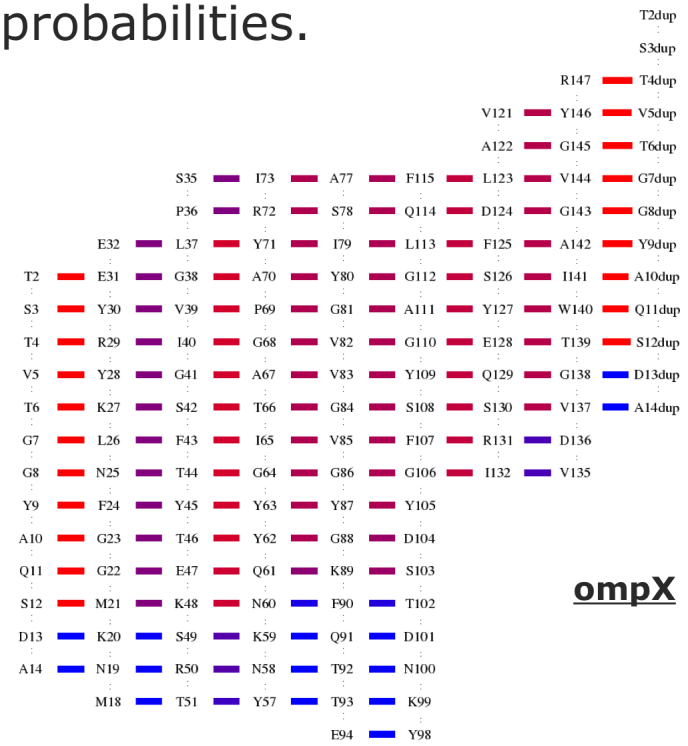


ompX

Results: Whole structure prediction

- Sample structures, identify substructure probabilities.
- Clustering gives multiple compact clusters of conformations



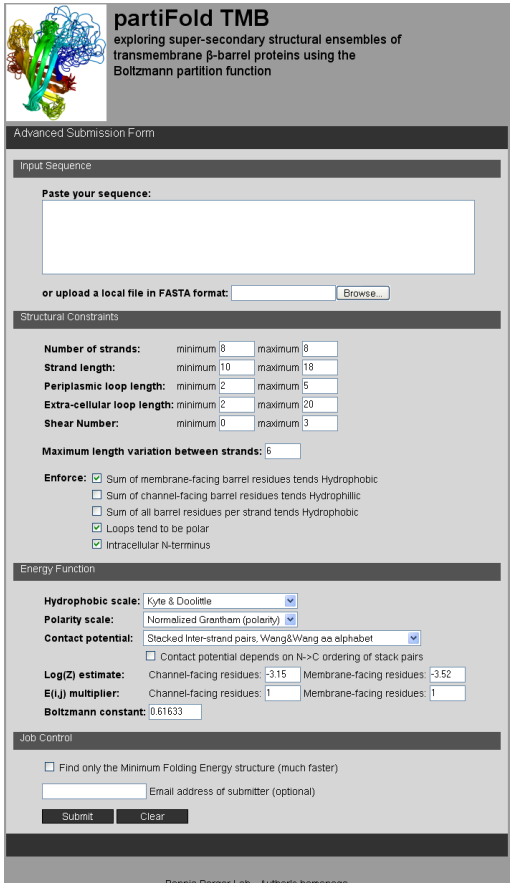


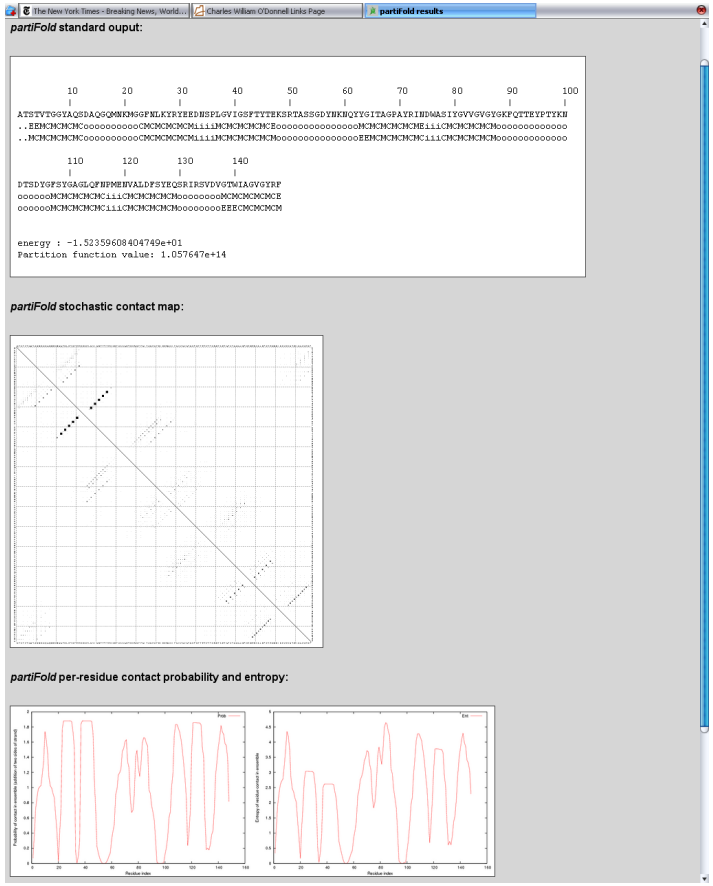
Red: Crystal structure
Green: M.F.E. structure

- Representants of clusters provide better candidates and outperform the *minimum folding energy* structure.

Webserver

http://partifold.csail.mit.edu





Tunable structural constraints and energy model, fast, permanently updated.
Binary distribution is also provided.

Acknowledgments

MIT
- Charles W. O'Donnell
- Mieszko Lis
- Nathan Palmer
- Srinivas Devadas
- Bonnie Berger

Whitehead
- Susan Lindquist
- Rajaraman Krishnan

Ecole Polytechnique
- Jean-Marc Steyaert

Boston College
- Peter Clote

References

- J. Waldispühl*, C.W. O'Donnell*, S. Devadas, P. Clote and B. Berger.
  Modeling Ensembles of Transmembrane β-barrel Proteins
  PROTEINS: Structure, Function and Bioinformatics, published online 14 Nov. 2007.
  doi:10.1002/prot.21788
  (* authors equally contributed)
- J. Waldispühl, B. Berger, P. Clote and J.-M. Steyaert,
  Predicting Transmembrane β-barrels and Inter-strand Residue Interactions from Sequence.
  PROTEINS: Structure, Function and Bioinformatics, vol. 65, issue 1, p.61-74, 2006.
  doi:10.1002/prot.21046
- J. Waldispühl and J.-M. Steyaert,
  Modeling and Predicting All-α Transmembrane Proteins Including Helix-helix Pairing,
  Theoretical Computer Science, special issue on Pattern Discovery in the Post Genome,
  p.67-92, 2005. doi:10.1016/j.tcs.2004.12.018

Current and future work presented in the poster:

Modeling structure ensemble of conserved β-sheet folds

Presenter: Charles O'Donnell