Computing structural and dynamic properties of biological systems at multiscale

2. Predicting structures, topology, and stabilities of bacterial outer membrane proteins and eukaryotic mitochondrial membrane proteins

Jie Liang

Bioinformatics Program

Dept. of Bioengineering

University of Illinois at Chicago

- Beta-barrel membrane proteins:
  - Outer membrane of gram-negative bacteria, mitochondria, and chloroplasts
- Many are ion channels
  - Membrane anchoring, enzyme activity, signaling and homeostasis, induction of apoptosis, bacterial virulence.
- Important for:
  - Understanding membrane protein folding and sorting
  - Delivery of molecules to cell
  - Ion and organic compound detection
  - Nanopore sensing of polymers and reaction
    - Reagent free DNA sequencing
  - Target cell death: cancer treatment
- Can be reliably identified from sequence analysis
  - VDAC: 19 strands, PapC: 24 strands

(Schultz, *BBA*, 2002; Wimley, *Prot Sci*, 2002; Majd et al, *Curr Op Biotech*, 2010)

Outline

1. Empirical Potential Function and Reduced State Model
   - Combinatorial analysis for reference state
   - Reduced configuration model and enumeration of states

   - Predicting protein-protein interactions and interfaces
   - Engineering oligomerization states

2. Computational Transfer Free Energy
   - Depth dependent profiles and cooperativity

   - Biological insight and hypothesis

3. Predicting Beta Barrel Membrane Proteins Structures from Sequences
   - Template free structural models
   - Loop structure prediction

1. Empirical potential function from combinatorial analysis of short strands

- Overall challenges:
  - Very limited structural data of nonhomologs!
  - Strong coupling effects in the short TM region
    - Confounding effects cannot be ignored
    - Simple model assuming replacement does not work (Thomas and Dill, 1996)

- No existing combinatorial models
  - Current studies: Bernoulli model, Markovian model

(Robin, Rodolphe, Schbath, 2005
DNA, Words and Models: Statistics of exceptional words)

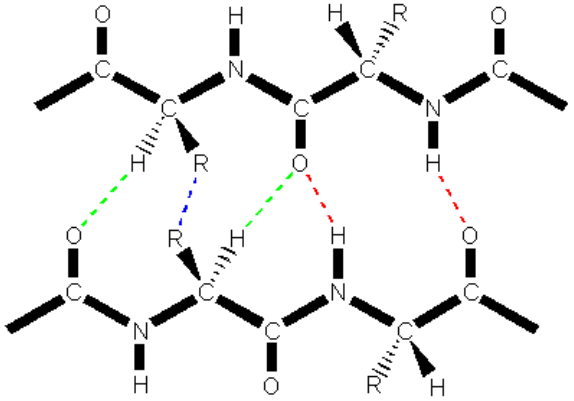
- More sensitive and specific combinatorial model and statistics are needed!
  - Shuffling model: Heavy combinatorial techniques required.

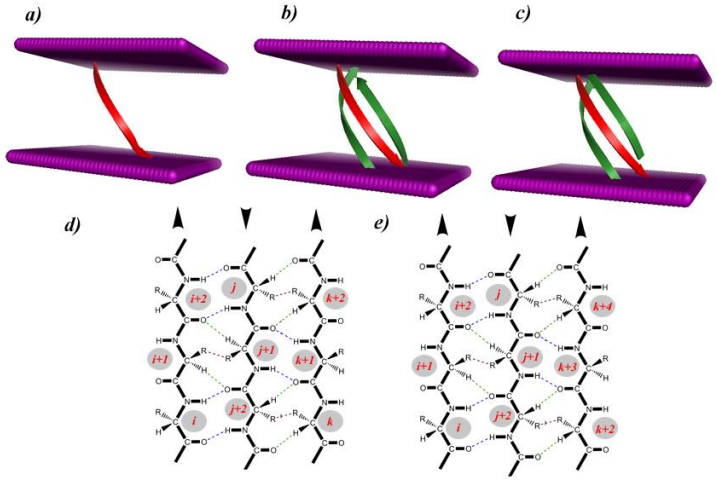
(Senes et al, 2000)

Physical Model:
Interaction Pattern of Antiparallel β-Sheets

Adjacent strands:

1. Strong H-bonds
   immediately across
2. Non-H-bond interactions
3. Weak C-O H-bonds
   across and one residue
   displaced on the strand





(Soluble and membrane proteins. Ho and Curmi, 1999, *J Mol Biol*)

MSIP: Membrane Strand Interface Pair propensity

- Interstrand contacts between *X* and *Y* residues:

$$P(X,Y) = \frac{f(X,Y)}{\mathbb{E}[f'(X,Y)]}$$

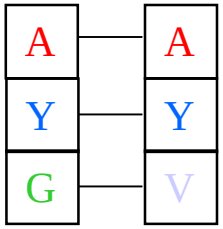
- **Null model:**
  - Two adjacent strands are permuted exhaustively and independently.
  - Each permutation is equally likely.
  - Generalization of shuffling model

- Calculating propensities for all 210 possible residue pairs
  - Separately for strong H-bond, weak H-bond, and non-bonded interactions.

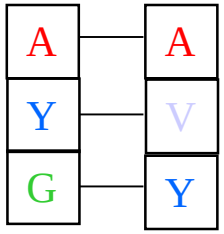
(Jackups, Jr and JL, *J Mol Biol*, 2005, 354:979–993)

(Jackups and JL, 2010, *IEEE/ACM Trans Comp Bio & Bioinf* 7:524-536)

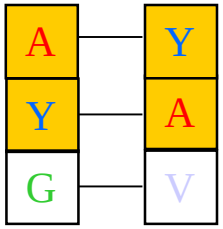
Example: AY contacts in 3-residue strand pair



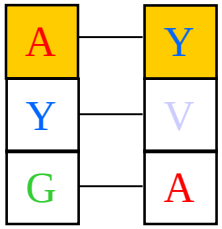
0 contacts



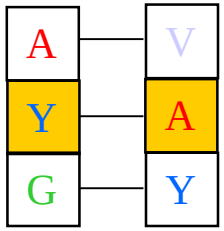
0 contacts



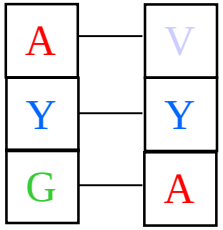
2 contacts



1 contact



1 contact



0 contacts

Expected number of AY contacts for this strand pair = 4/6 = 0.66.

(Jackups and Liang, 2010, *IEEE/ACM Trans Comp Bio & Bioinf* 7:524-536)

Contacts between residues of different type

- Expected frequency of X-Y contacts in all strand pairs:

$$\mathbb{E}[f'(X,Y)] = \sum_{sp \in \mathcal{SP}} \{\mathbb{E}[f'_{sp}(X,Y)] + \mathbb{E}[f'_{sp}(Y,X)]\}$$

$$= \sum_{sp \in \mathcal{SP}} \{\frac{x_1(sp) \cdot y_2(sp)}{l(sp)} + \frac{y_1(sp) \cdot x_2(sp)}{l(sp)}\},$$

- *p*-value is more difficulty, as $f'_{sp}(X,Y)$ and $f'_{sp}(Y,X)$ are dependent.

(Jackups and Liang, 2010, *IEEE/ACM Trans Comp Bio & Bioinf* 7:524-536)

Calculating *p*-value of X-Y contacts: Generalized Hypergeometric Distribution

- Trinomial function (a,b,c)! = (a+b+c)!/a!b!c!.
  - Define: $T(l, x_1, y_1) \acute{} (x_1, y_1, l\text{-}x_1\text{-}y_1)!$
- The random probability of *h* X-X contacts, *i* X-Ycontacts, *j* Y-X contacts, and *k* Y-Y contacts is:

$$\mathbb{P}(h,i,j,k) = \frac{T(x_1,h,i)\cdot T(y_1,j,k)\cdot T(l-x_1-y_1, x_2-h-j, y_2-i-k)}{T(l,x_2,y_2)}.$$

- Marginal probability of a total of $i + j = m$ X-Y contacts for computing *p*-value:

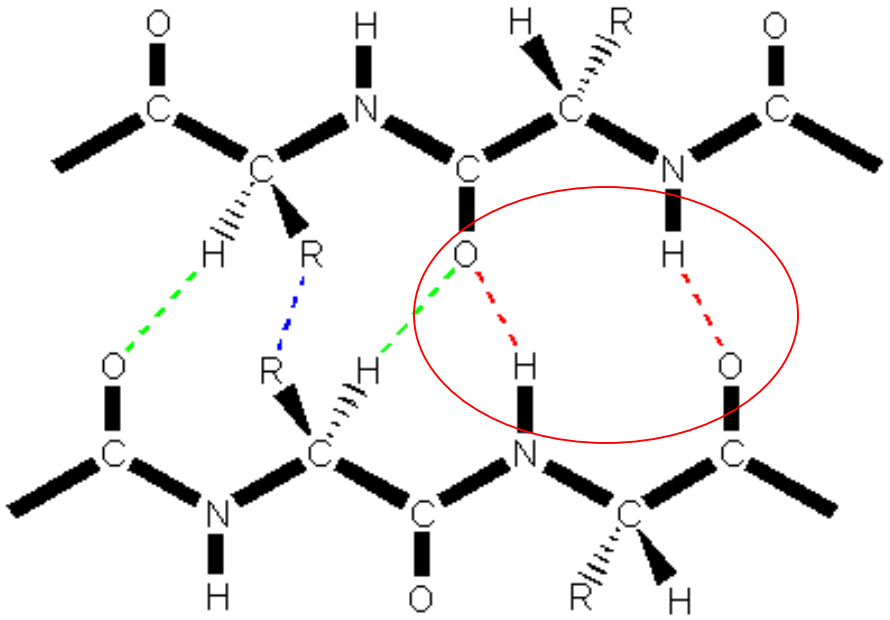
$$\mathbb{P}_{X,Y}(m) = \sum_{h=0}^{x_1} \sum_{i=0}^{x_1-h} \sum_{k=0}^{y_1-\binom{}{m-i}} \mathbb{P}(h,i,m-i,k),$$

(Jackups and Liang, 2010, *IEEE/ACM Trans Comp Bio & Bioinfo* 7:524-536)

Empirical Potential from Propensity for Interactions

H-bond and other 2 types of interactions

High		Low	
Pair	Odds	Pair	Odds
GY	1.48	YY	0.32
ND	2.49	NY	0.26
LY	1.42	KY	0.2
GF	1.76	VV	0.64
HK	3.48	HY	0
IY	1.58	RL	0.34
AV	1.37	PV	0
RP	4.00		
NN	2.05		
ET	1.58		
KS	1.64		
IM	2.06		
G-FWY			1.42
ILV-FWY			1.3

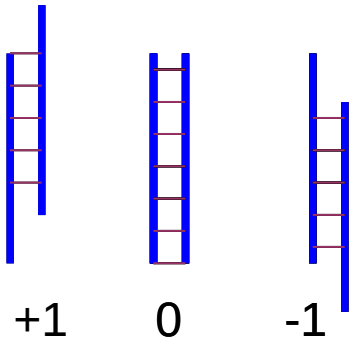


p-value < 0 .05

p-value < 0.10

Reduced Conformational Space and Energy Evaluation

- Through enumeration of conformation:
  - Each strand can have 7 conformations
- Energy evaluation:



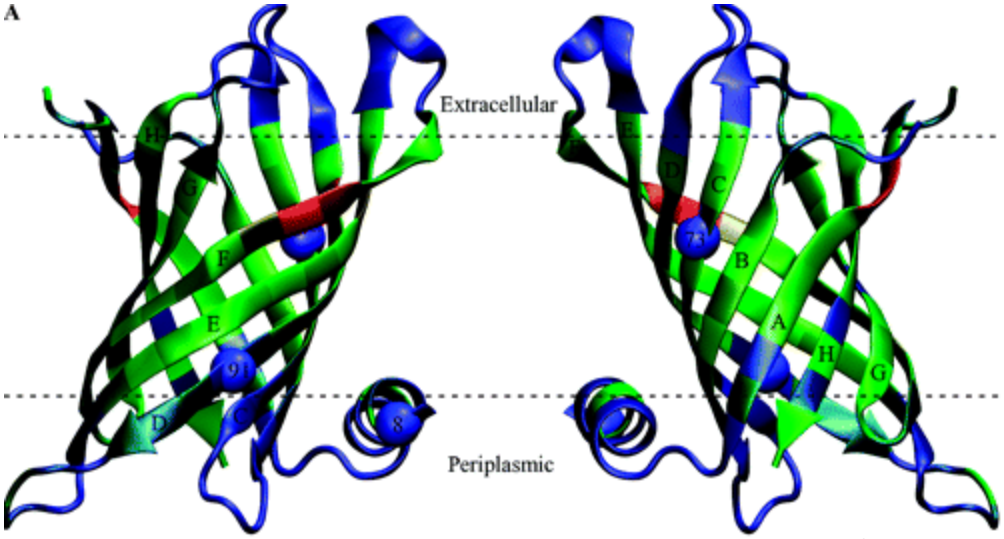
$$E(i,\boldsymbol{d}_i) = E_i(d_{i-1},d_i,d_{i+1}) = \sum_{k_i} E_1(k_i;d_i) +$$
$$\frac{\alpha}{2}[\sum_{k_i}\sum_{k_{i-1}} E_H(k_i,k_{i-1};d_i,d_{i-1}) + \sum_{k_i}\sum_{k_{i+1}} E_H(k_i,k_{i+1};d_i,d_{i+1})] +$$
$$\frac{\beta}{2}[\sum_{k_i}\sum_{k_{i-1}} E_{SC}(k_i,k_{i-1};d_i,d_{i-1}) + \sum_{k_i}\sum_{k_{i+1}} E_{SC}(k_i,k_{i+1};d_i,d_{i+1})] +$$
$$\frac{\gamma}{2}[\sum_{k_i}\sum_{k_{i-1}} E_{WH}(k_i,k_{i-1};d_i,d_{i-1}) + \sum_{k_i}\sum_{k_{i+1}} E_{WH}(k_i,k_{i+1};d_i,d_{i+1})]$$

$$E(\boldsymbol{d}) = \sum_{i=1}^{N} E(i,\boldsymbol{d}_i)$$

Naveed *et al*. *Proc Natl Aca Sci USA*. 2009 106(31):12735-12740.

Empirical Energy Function

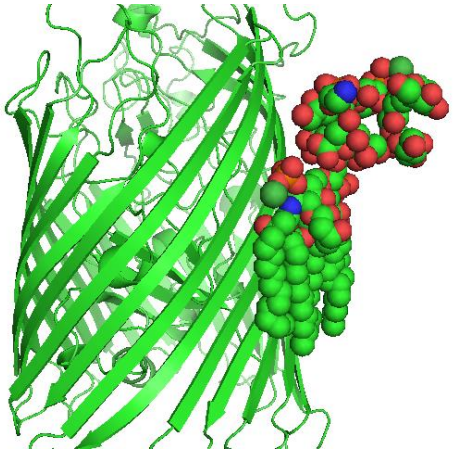
- Detection of structural anomaly:
  - PagP (pdb id : 1THQ)
    - Strand B & C interact with the alpha helix



Topology of an Outer-Membrane Enzyme: Measuring Oxygen and Water Contacts in Solution NMR Studies of PagP
*Ferenc Evanics, Peter M. Hwang, Yao Cheng, Lewis E. Kay, and R. Scott Prosser JACS **2006**, **128** (25), pp 8256–8264*

Identification of Weakly Stable Regions

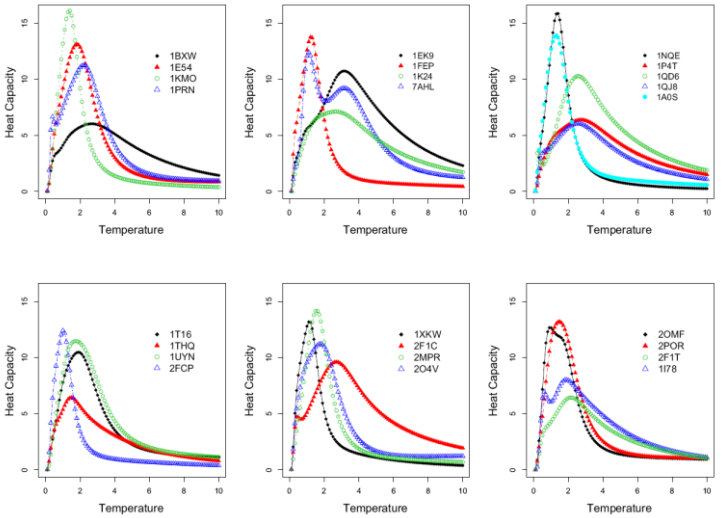
- Mechanism for stabilizing weakly stable regions
  - In-plugs
  - Out-clamps
  - Oligomerization
  - Protein-lipid interactions



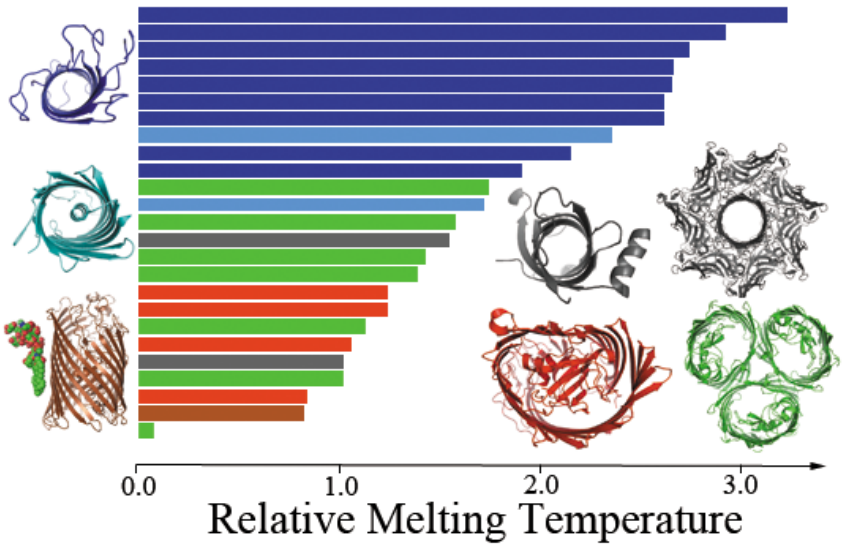
Naveed, Jackups Jr, and Liang. *Proc Natl Aca Sci USA*. 2009 106(31):12735-12740.

Heat Capacity and Melting Temperature

- Compute partition function through enumeration



- Relative melting temperature of TM region

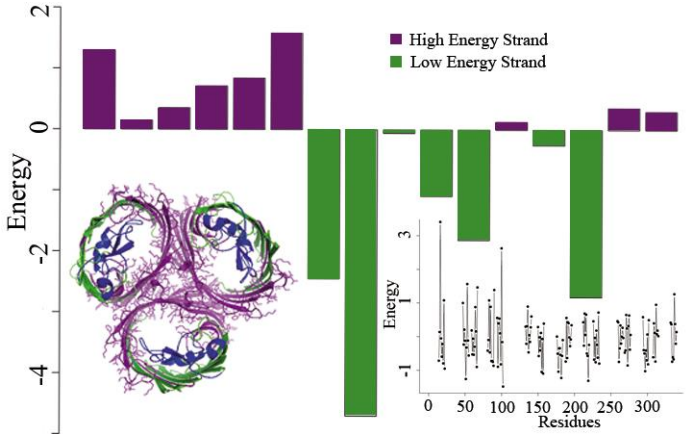


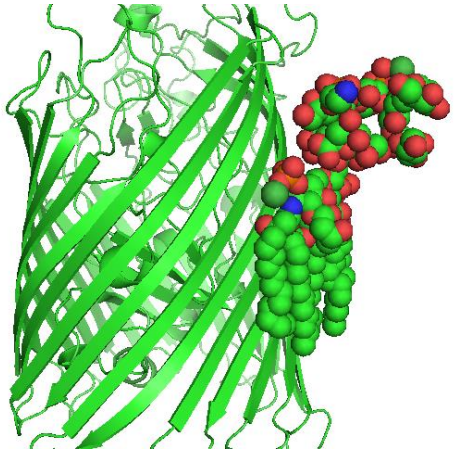
Naveed *et al*. *Proc Natl Aca Sci USA*. 2009 106(31):12735-12740.

Predicting Oligomerization State

- Can distinguish oligomers from monomers

- FhuA
  - Predicted to be an oligomer
  - Crystal structure is a monomer
  - Transient: reported to have several oligomeric states
- Accuracy
  - 100%





Naveed *et al*. *Proc Natl Aca Sci USA*. 2009 106(31):12735-12740.
Locher, K.P. and Rosenbusch, J.P. *Eur J Biochem.* 1997 247(3):770-775.

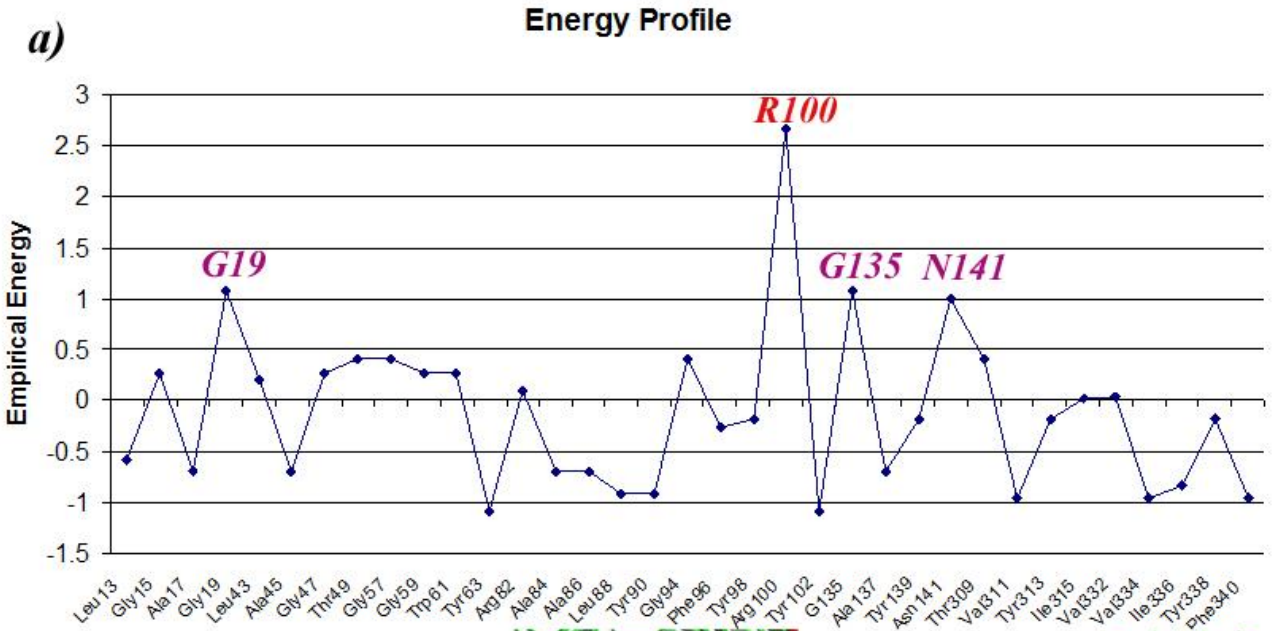
Predicting Protein-Protein Interaction Interfaces

- Continuous unstable regions form interface for protein-protein interactions.
- Identify interface through *neighbor correction*
  - Stringent criterion
- Accuracy
  - 78% when secondary structure is known.
  - 66% when using sequence only.
    - Secondary structure is predicted using ProfTMB.
- All 25 known β-barrels < 32% pairwise sequence identity
- Resolution range between 1.8 and 3.0 A^{o}
- Leave one out test

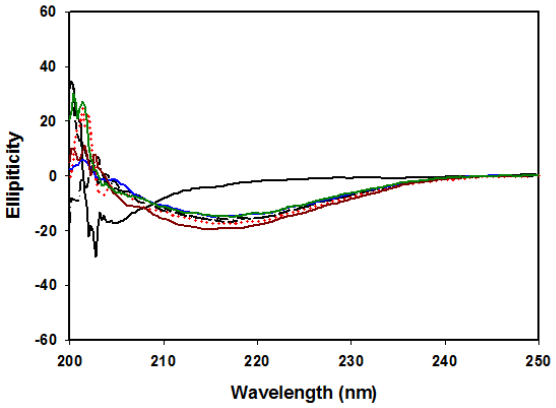
Naveed *et al*. *Proc Natl Aca Sci USA*. 2009 106(31):12735-12740.

Altering Protein-Protein Interactions in Bacterial Outer-Membrane: OmpF

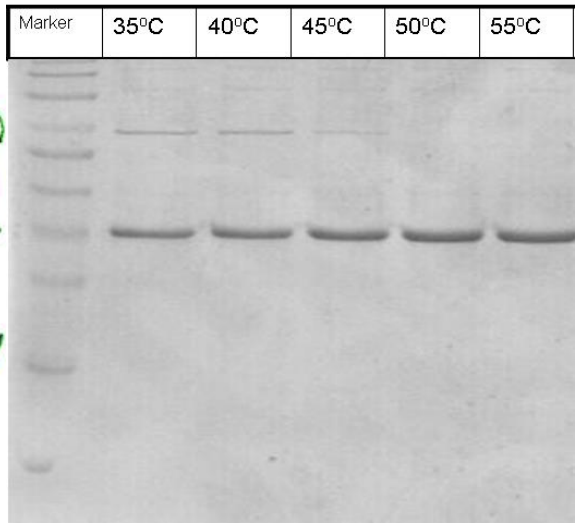
(with Linda Kenney) (Naveed et al, J Mol Biol, 2012)



CD

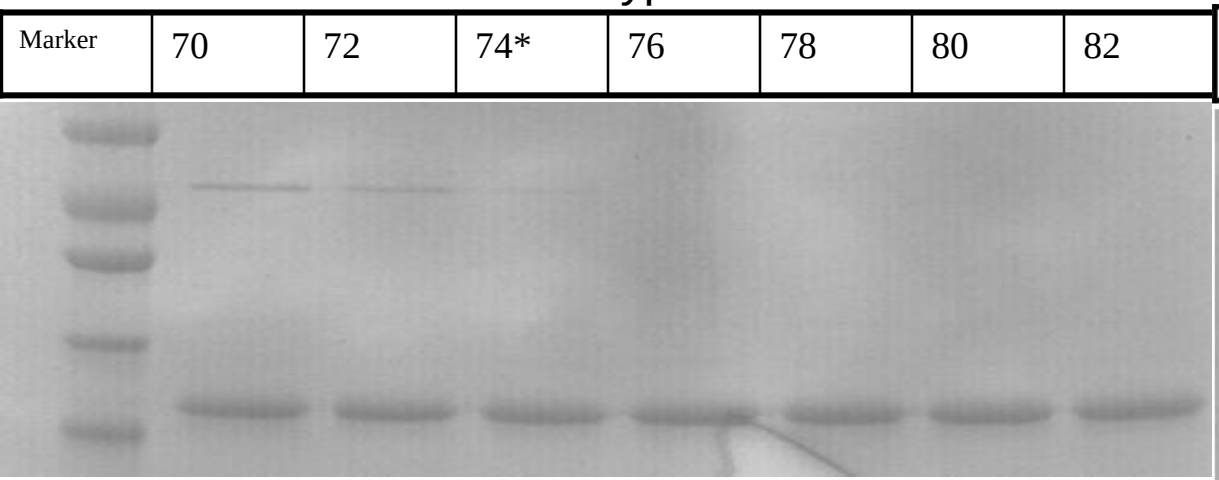


G19W

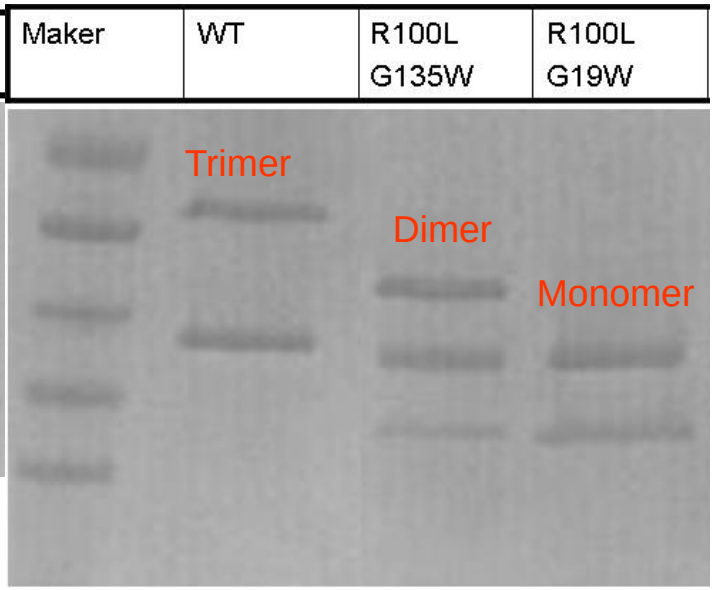


Tm of Oligomerization

Wild Type



RT Oligomeric State



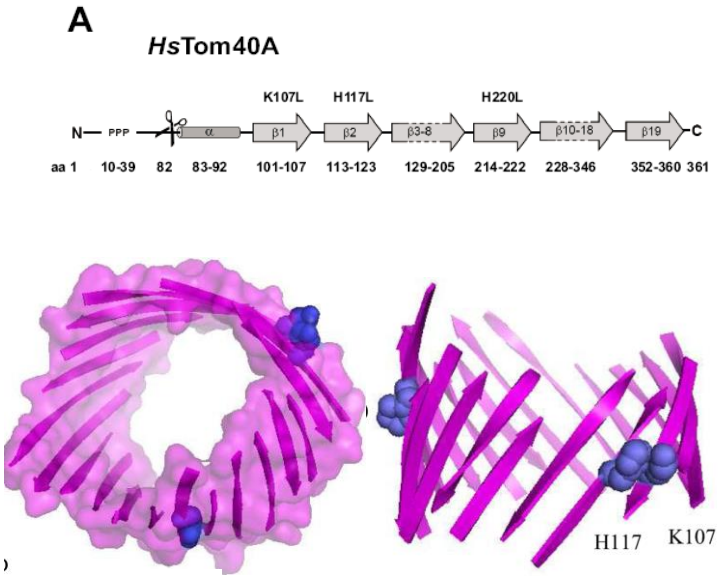
Mutant	Dissociation Temperature ($^{\circ}C$)	Δ T	Oligomerization State
Wild Type	$73(\pm)1$	–	Trimer
G19W	$47(\pm)3$	26	Trimer
R100L	$52(\pm)3$	21	Trimer
G135W	$52(\pm)3$	21	Trimer
N141W	$67(\pm)3$	6	Trimer
R100L/G135W	$44(\pm)1$	29	Dimer
R100/LG19W	–	–	Monomer

(with Linda Kenney)

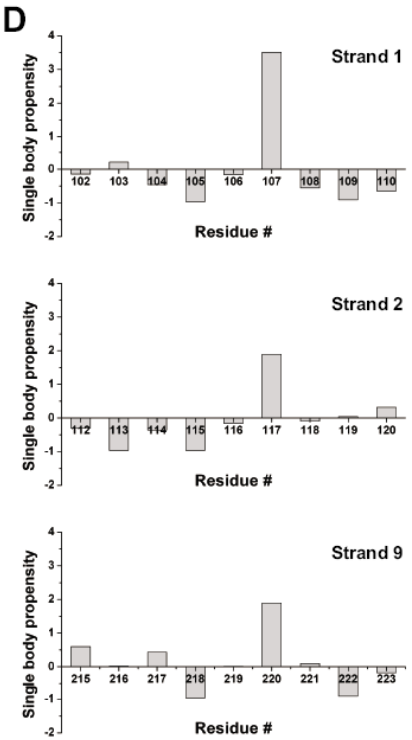
(Naveed et al, *J Mol Biol* , 2012 419:89-101)

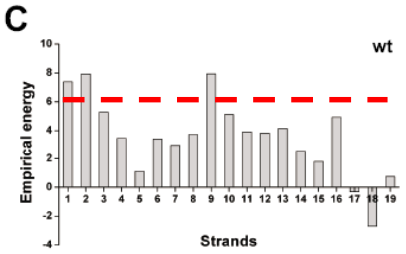
Altering Protein-Protein Interactions in Eukaryotic Mitochondria: Human Tom40

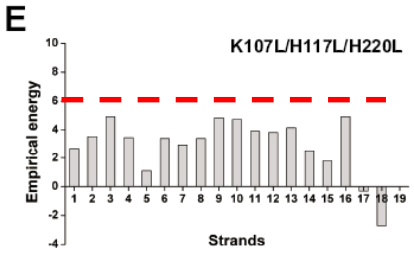
(with *Dennis Gessman* and *Stephan Nussberger*) (Gessman, et al, *J. Mol. Biol.* , 2011:413:150-161)

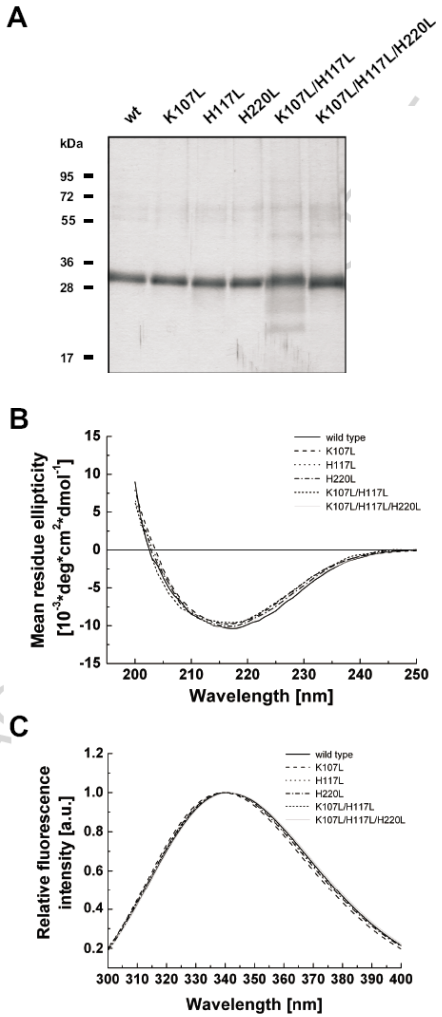


(Predicted structure.
Experimentally unknown)

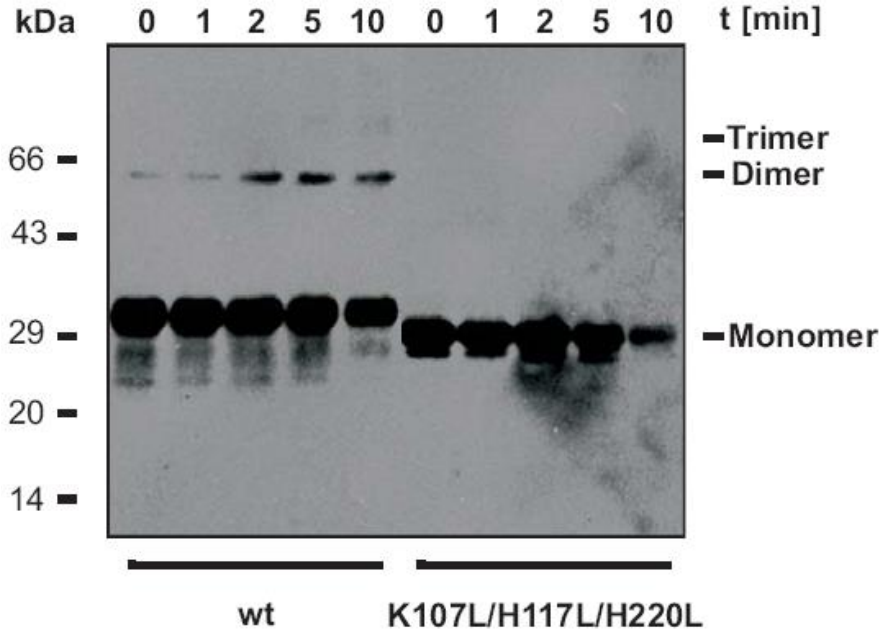


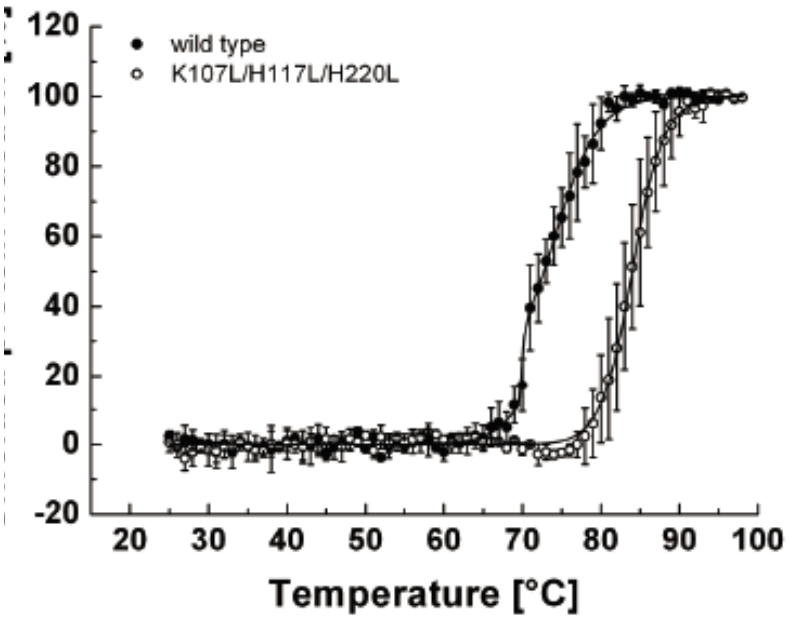




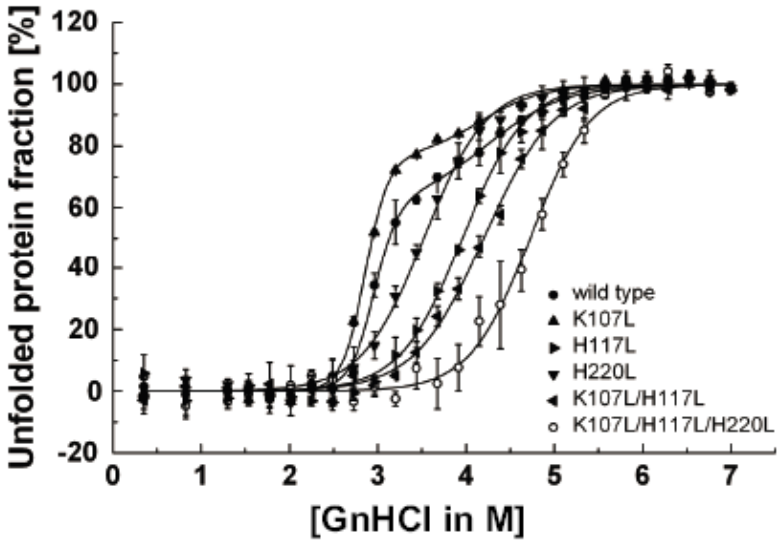


Predicting protein-protein interactions in eukaryotic mitochondria (humanTom40)





C

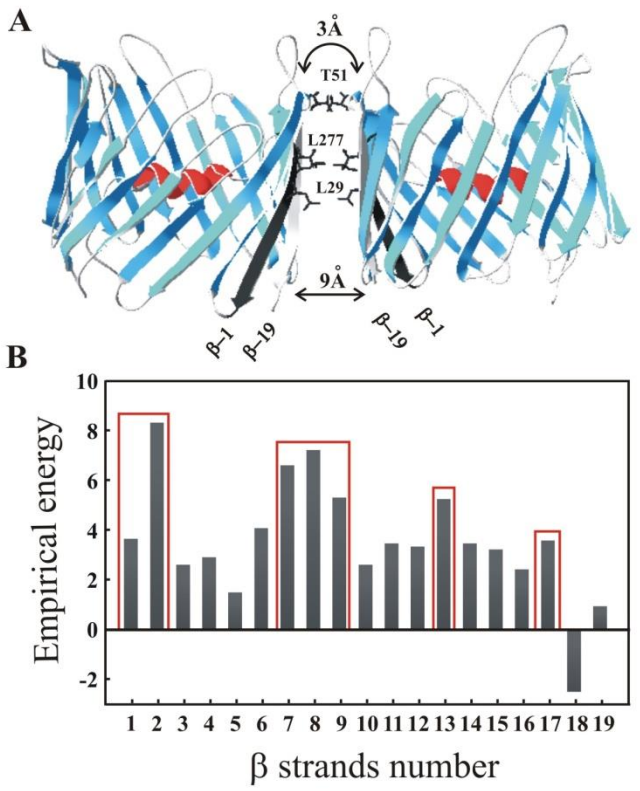


- Mutant protein form monomers only
- Protein stabilized:
  Unfolding Tm increased by 11°C.
- Complete denaturation of the mutant protein occurred at 6.3 M GnHCl compared to 4.5 M GnHcl for Wild type.

(Gessmann, et al, *J. Mol. Biol.* , 2011:413:150-161)

VDAC Study

- Voltage dependent anion channel
  - Key player in apoptosis induction
  - Mechanism: through oligomerization? where?
- Our predictions:
  - Discrete sites of PPI
  - Mutations suggested
- Experimental verification:
  - Mutants made
  - Cross-linking
  - Gel analysis



(with Guela and Shoshan-Barmatz, Ben-Gurion U)

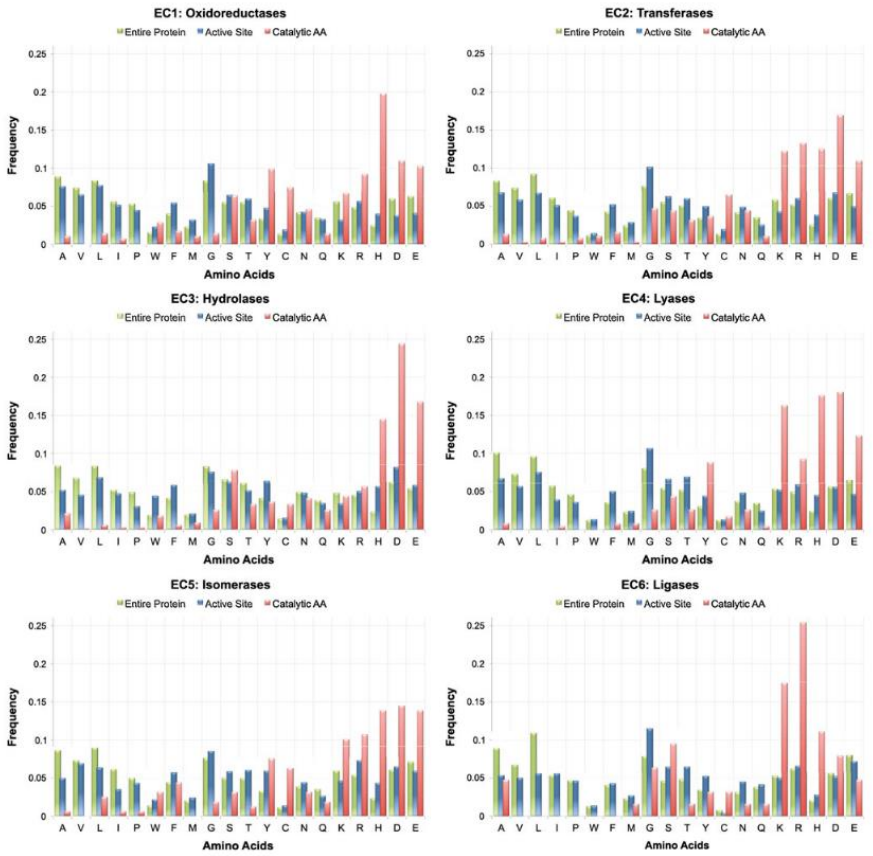
(Guela et al, *J Biol Chem* , 2012, 287(3):2179-90)

2. Predicting Beta Barrel Membrane Protein TM Structures from Sequences

(Naveed et al, *J Am Chem Soc*, 2012**,** 134:1775-1781
Tang et al, *PLoS Comput Biol*, 2014, DOI: 10.1371/journal.pcbi.1003539)

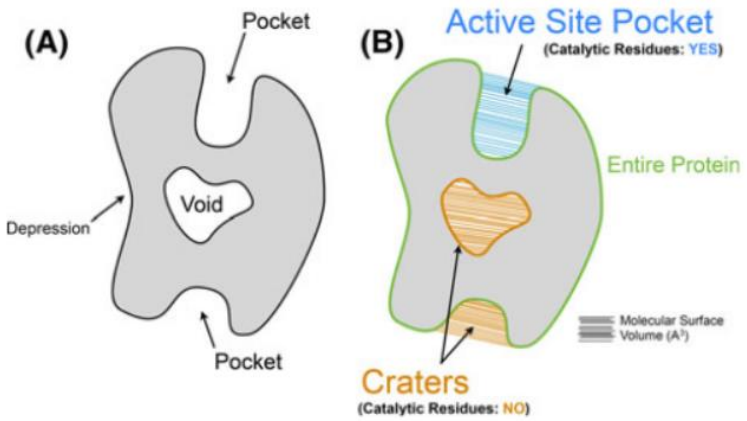
Structures of Proteins Are Important

- Lead to new insight
- Enzyme active sites are unusually crowded with charges:



(Jimenez-Morales, Liang and Eisenberg, 2012, *Eur J Biophys*, 41:449–460,)

(Jimenez-Morales, Liang and Eisenberg, 2012, *Eur J Biophys*, 41:449–460,)



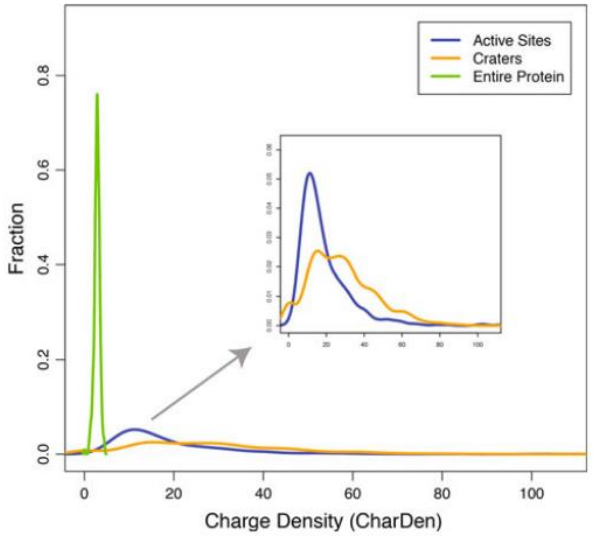


Fig. 5 Density estimation of the fraction of proteins with a given charge density (CharDen). Catalytic active site, craters and the entire protein CharDen

- 4 acid and 5 base side chains in an avg of 1,000 A^3
- 8.3M of acid and 10.6M of basic side chains

- Reactants and side chains are crowded in a mixture of ionic liquid
  - Not ideal infinitely dilute solutions

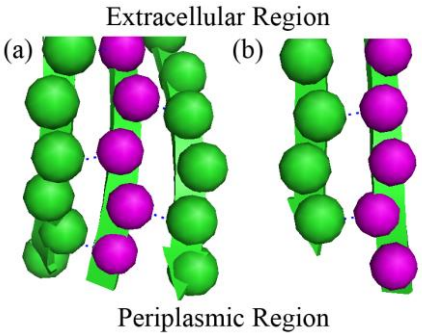
- Given a sequence of a beta barrel membrane protein, can we predict its 3D structure?
  - Prediction of secondary structure from sequence is quite reliable
    - 96% for number of strands
    - 87% for residues in each strand

Ou et al. *J Comput Chem.* 2010 Jan 15;31(1):217-23.

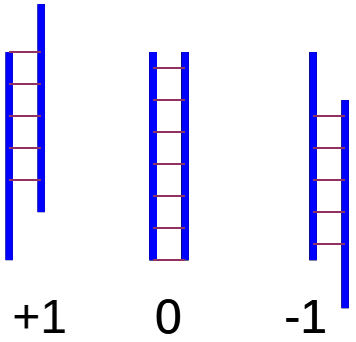
  - Challenges:
    - 170-700 residues, beyond existing methods
      - Only template based methods have been successful, TMBpro

Empirical Pair Propensity

- One-body region-specific propensity
- Two-body spatial contact propensities



- Types of contacts
  - Strong H-bond
  - Side chain (vdw)
  - Weak H-bond



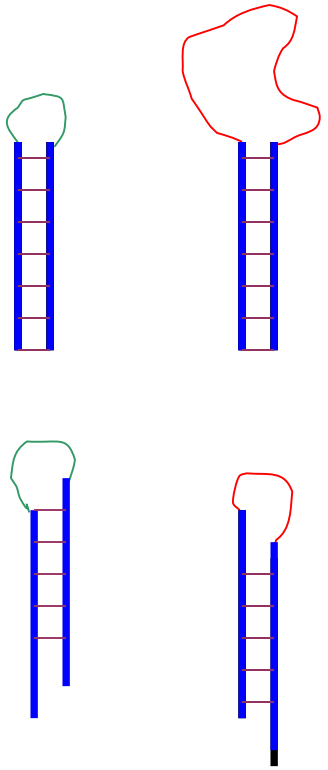
Jackups, R. Jr & Liang, J. (2005). *J. Mol. Biol. 354, 979-993*
Jackups, R. Jr & Liang, J. (2009). IEEE/ACM Trans Comput Bio & Bioinfo

Loop Entropy and Asymmetry

- Effects of loop entropy
  - Average length of loop is *nref* = 8.5

$$L.E. = \ln\left(\frac{nref + \Delta L}{nref}\right)$$

- Conformation sampling
  - Up and down: from –L to +L
  - L: strand length



Chou *et al* (1982) J. Mol. Biol. 162, 89-112
Chou *et al* (1983) Biochemistry, 22, 6213-6221
Chou *et al* (1983) J. Mol. Biol. 168, 389-407
Wang *et al* (1996) J. Mol. Biol. 262, 283-293

Predict Registration

- Energy function

$$E(i,d_i) = \alpha[\sum_{k_i}\sum_{k_{i-1}} E_H(k_i,k_{i-1};d_i)] + \beta[\sum_{k_i}\sum_{k_{i-1}} E_{SC}(k_i,k_{i-1};d_i)] +$$
$$\gamma[\sum_{k_i}\sum_{k_{i-1}} E_{WH}(k_i,k_{i-1};d_i)] + \delta[\ln[\frac{n_{ref}+\Delta L}{n_{ref}}]] + \epsilon[LH]$$

- Overall accuracy
  - By identifying the configuration of lowest energy
  - Correct hydrogen bond b/w the preplasmic residues

	Small	Med	Large	All
TMBpro	0.52	0.43	0.50	0.48
Ours	0.71	0.73	0.76	0.73

(small: <14 strands; med: between 14 and 22; large: >22 strands)

	Total	TP
1BXW	8	4
1QJ8	8	6
1P4T	8	5
2F1T	8	6
1THQ	8	5
1K24	10	8
1I78	10	7
1QD6	12	8
1UYN	12	10
1T16	14	11
2F1C	14	9

	Total	TP
2POR	16	12
1PRN	16	13
2OMF	16	13
1E54	16	12
2O4V	16	11
2MPR	18	14
1A0S	18	10
1FEP	22	14
2FCP	22	17
1KMO	22	18
1NQE	22	16
1XKW	22	19

Coiled Coil Model: Our Work

- Unit Speed curve
- Frenet Formula's
- Reparametrization

$$f(s) = (a\cos(\frac{s}{c}), a\sin(\frac{s}{c}), b\frac{s}{c})$$

a = radius of the barrel

b = height of one beta turn

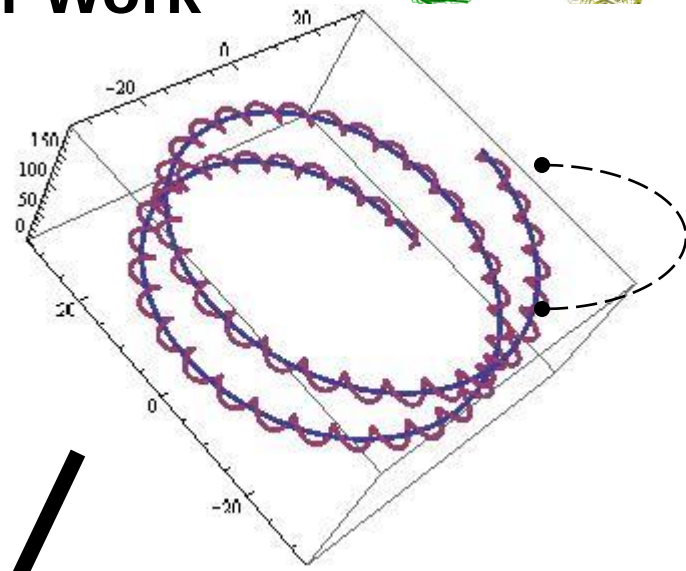
$c = \sqrt{a^2+b^2}$

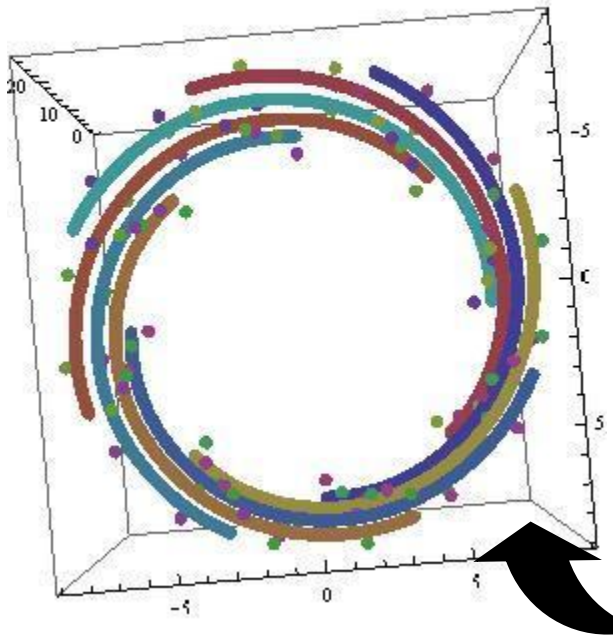
Minor helix can be written as

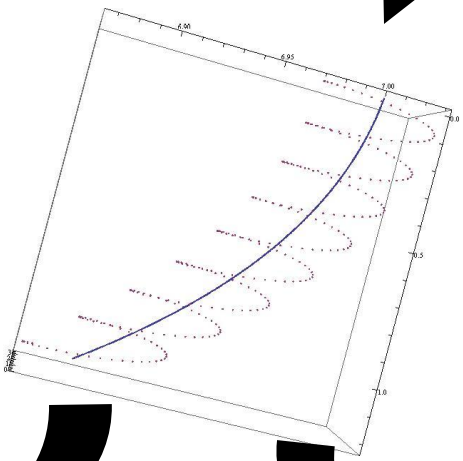
$r(s) = \lambda(s)\alpha(s) + \mu(s)\beta(s) + \nu(s)\gamma(s)$

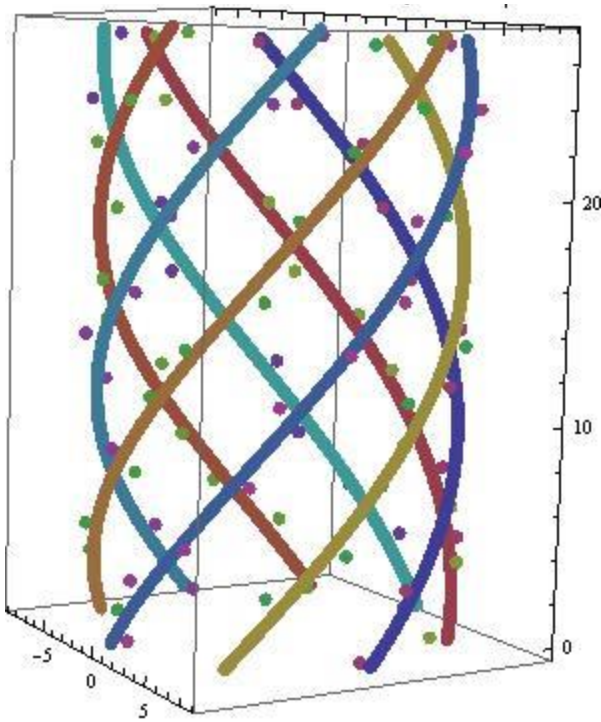
Solve for $\lambda(s), \mu(s)$ and $\nu(s)$ using frenet frame properties.

$$f(s) = \frac{b^2}{c^2}s\alpha(s) + (-a)\beta(s) + \frac{ab}{c^2}s\gamma(s)$$

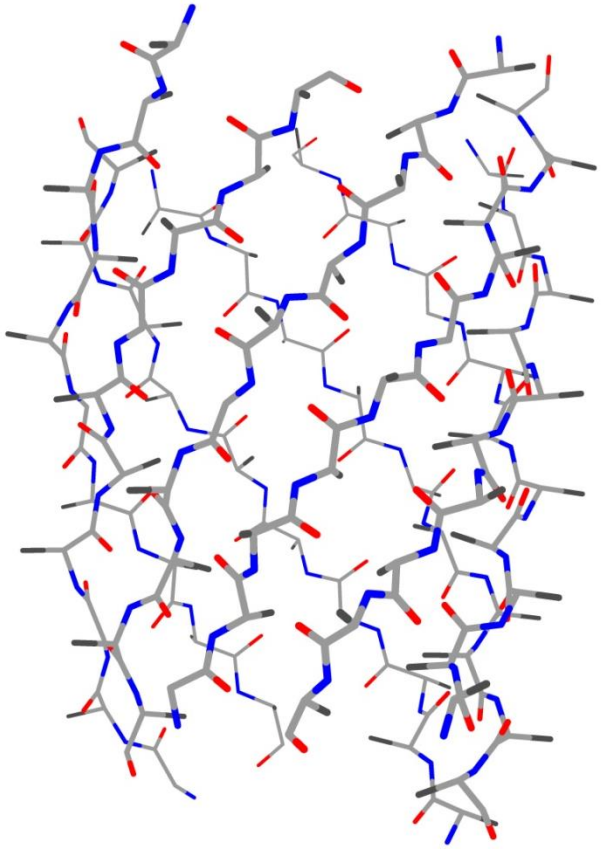


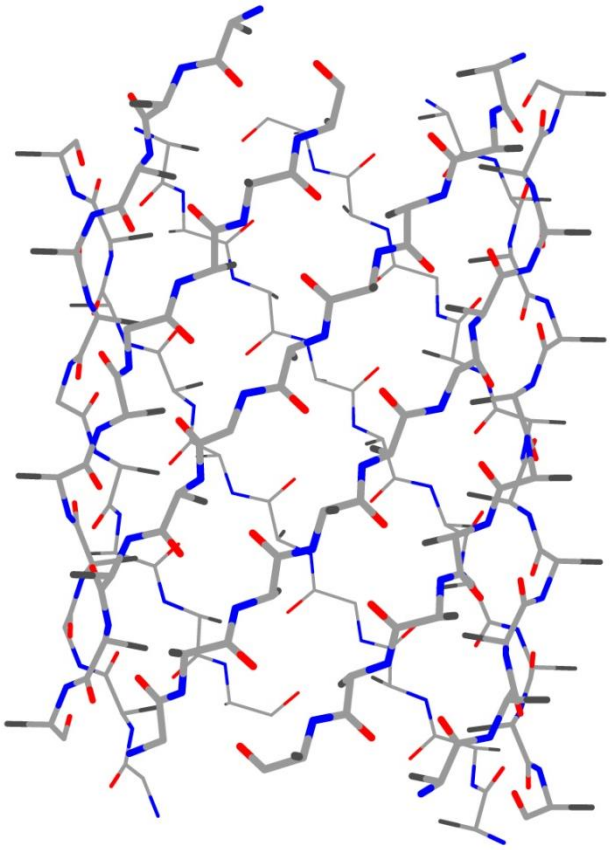




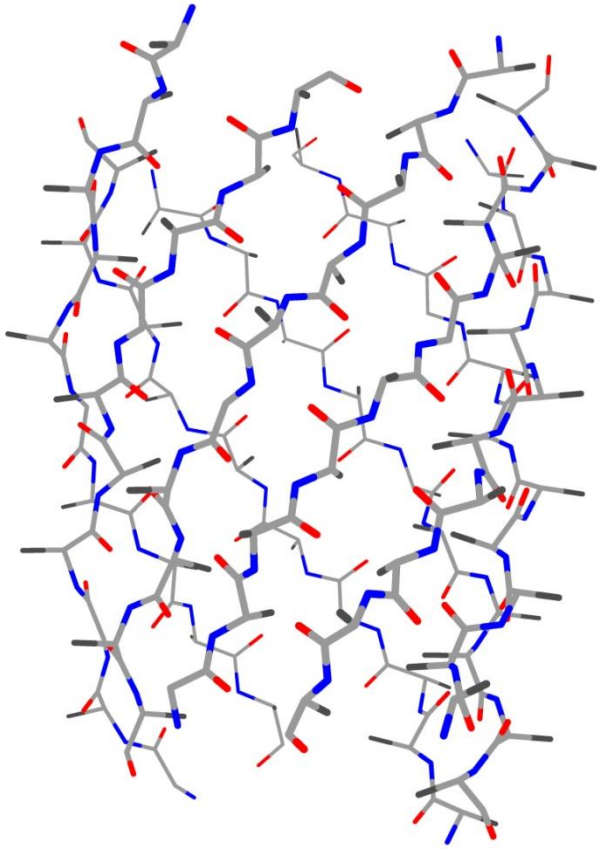


Example: NspA



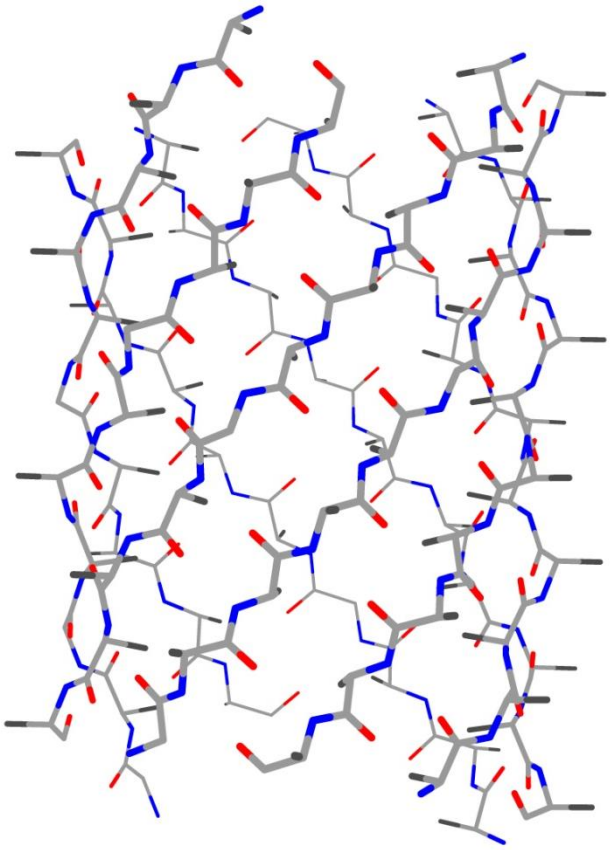


Example: NspA



True structure of
NspA (PDB `1p4t`)

RMSD = 2.35
for 391 atoms

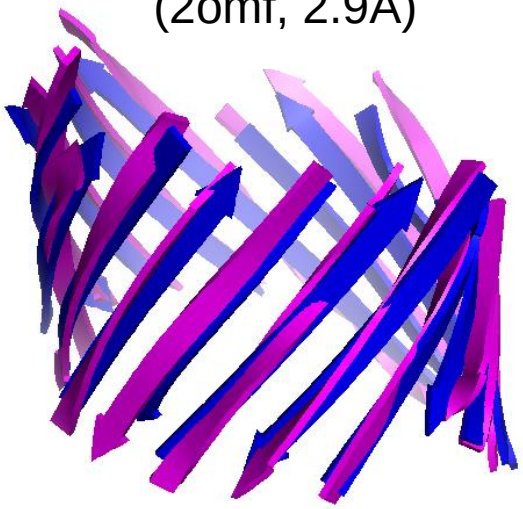


Predicted structure

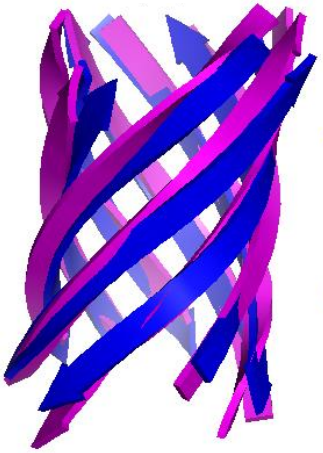
Predicted TM Structures

- Predicted: magenta
- Pdb: blue

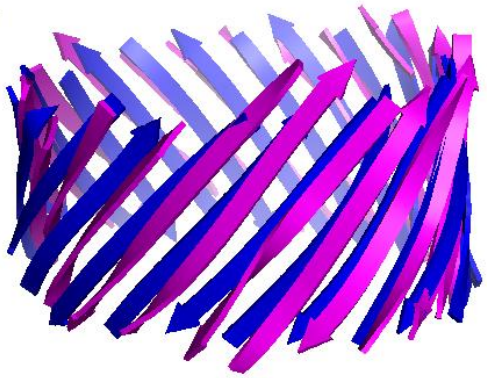
Omp F
(2omf, 2.9A)



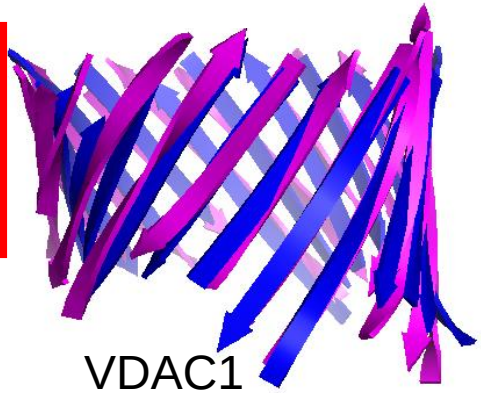
Omp A
(1bxw, 2.2A)



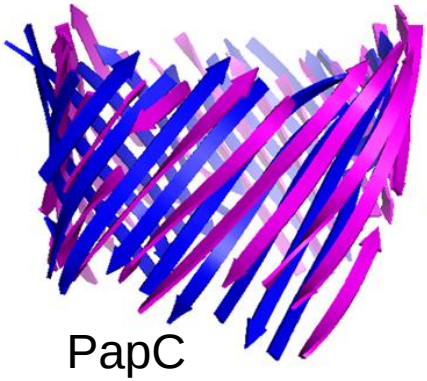
BtuB
(1nqe, 3.5A)



VDAC and PapC:
19 and 24 strands
never seen before



VDAC1
(3emn, 3.7A)



PapC
(3emn, 7.4A)

PapC


Results Using Sequence only

Protein	#Strands	bRMSD	aRMSD
OmpA/1bxw	8	2.23	4.72
OmpX/1qj8	8	3.34	4.93
OmpF/2omf	16	2.87	5.15
FadL/1t16	14	3.69	5.23
NspA/1p4t	8	3.57	5.27
BtuB/1nqe	22	3.50	5.29
Porin/1prn	16	3.34	5.53
FhuA/2fcp	22	3.76	5.57
FepA/1fep	22	4.04	5.59
Porin/2por	16	4.21	5.61
OmpWt/2f1t	8	3.58	5.74

Protein	#Strands	bRMSD	aRMSD
OpcA/1k24	10	4.15	5.83
NalP/1uyn	12	4.34	5.89
PorinP/2o4v	16	3.93	5.96
OmpG/2f1c	12	4.04	6.07
PagP/1thq	8	4.51	6.13
FptA/1xkw	22	4.56	6.29
Omp32/1e54	16	4.64	6.41
OmpT/1i78	10	5.09	6.71
LamB/2mpr	18	5.00	6.76
FecA/1kmo	22	5.30	6.80
OmpLA/1qd6	12	6.13	7.67
Scr Y/1a0s	18	8.37	9.89

bRMSD: backbone RMSD; aRMSD: atomic RMSD

Methods

	RMSD (A) of TM Domain			
	Small	Medium	Large	All
TMBpro Server	6.02	6.28	11.78	7.35
Our Method	5.83	6.47	5.90	6.05
Our Method (backbone)	4.06	4.62	4.23	4.27

TMBpro: requires template. Will not work for VDAC and PapC.

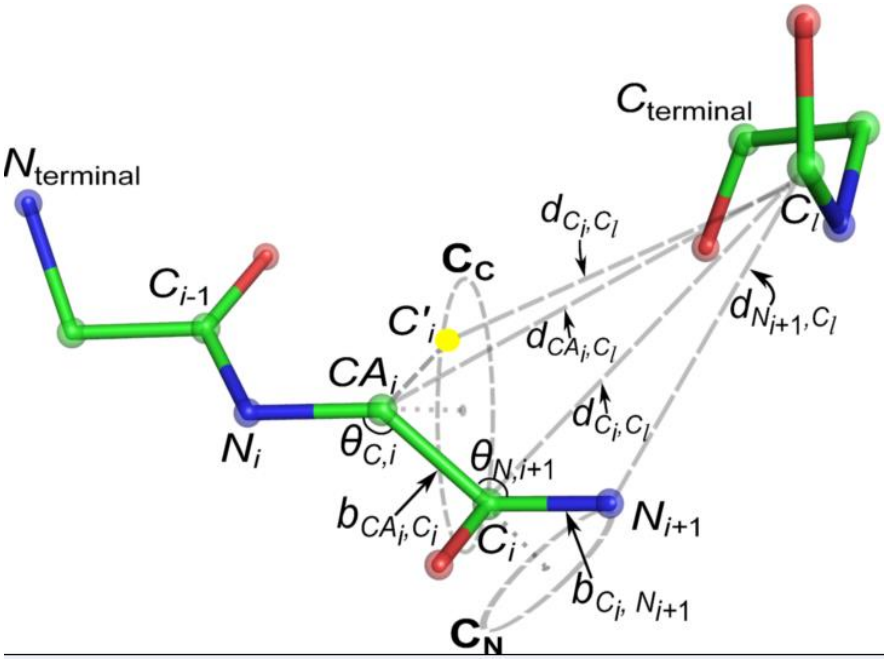
Structural Organizational Principle

- Models built from bacterial can be applied to eukaryotic mitochondria

- Structural and physicochemical constraints are more fundamental than evolutionary signal

Predicting Loops: Loop Distance-guided Sequential Monte Carlo Process (DiSGRO)

- For one residue, sample *m* (φ, ψ) trials from a conditional empirical distance distribution π_d.
- Samples *n* out of *m* trials from backbone torsion angle distribution π_τ.
- Select one trial according to probability:



$$x_{i+1} \sim p_{i+1}(x_{i+1} = s | x_1 \ldots x_i) \cdot \pi_\tau \cdot \pi_d$$

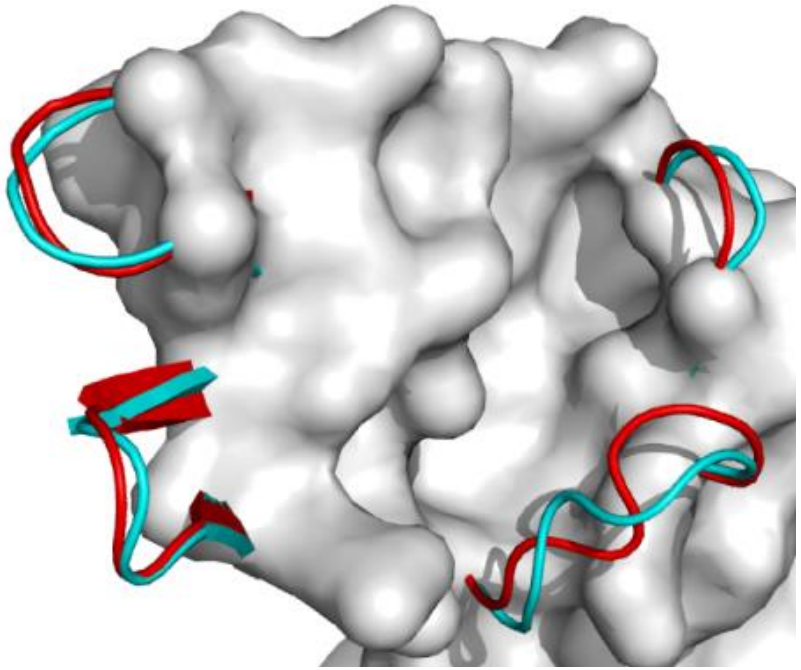
$$p_{i+1}(x_{i+1} = s | x_1 \ldots x_i) = e^{-E(s)}$$

$E(s)$: an incremental energy function accounts for the interaction of the trial *s* of x_{i+1} with the remaining part of the protein

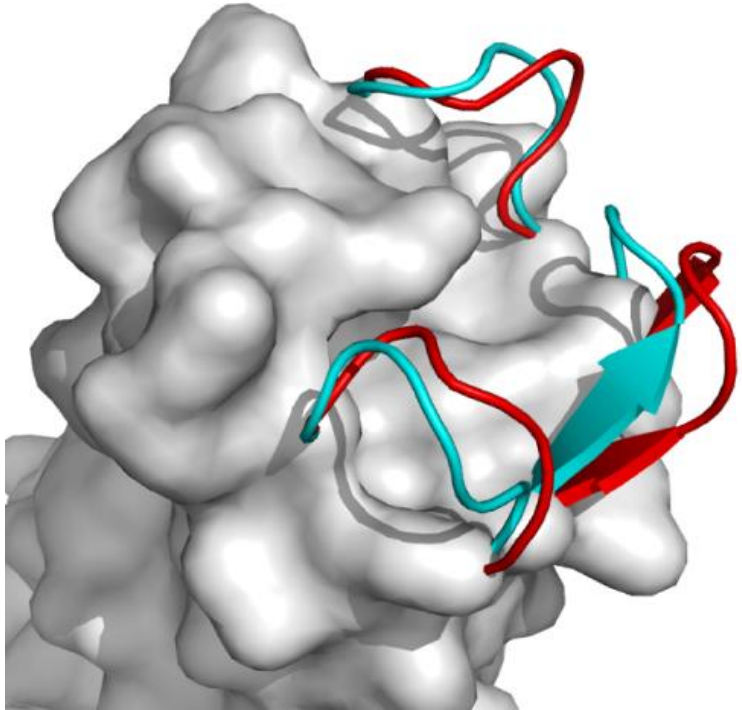
(Ke Tang, Jinfeng Zhang, and JL, *PLoS Comput Biol*, 2014)

Predicting Loops in OMPs

- OmpX (`1qj8`)
  - Cyan: native loops.
  - Red: minimal energy loops by mDiSGro.



- Extracellular:
  - 8res + 7res + 5res + 4res = 24 res
  - RMSD: 1.26 Å
  - Interacting loop pairs

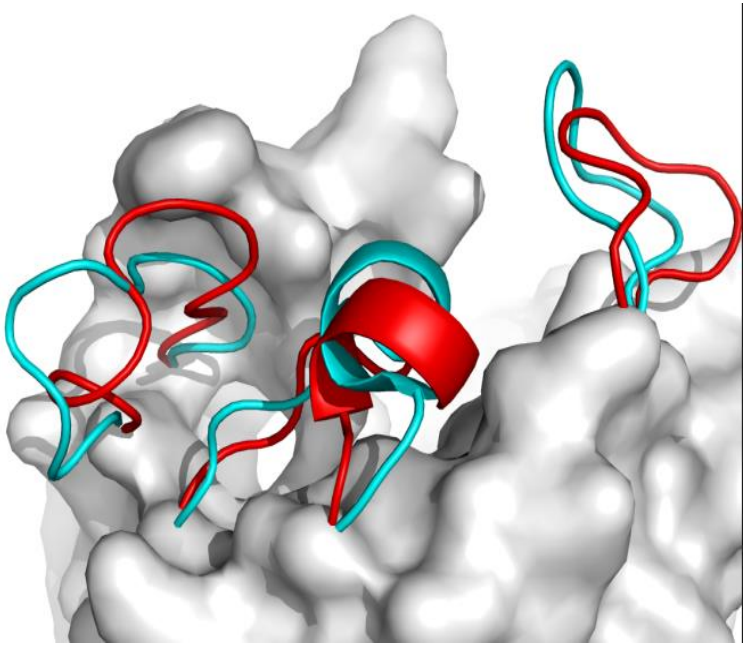


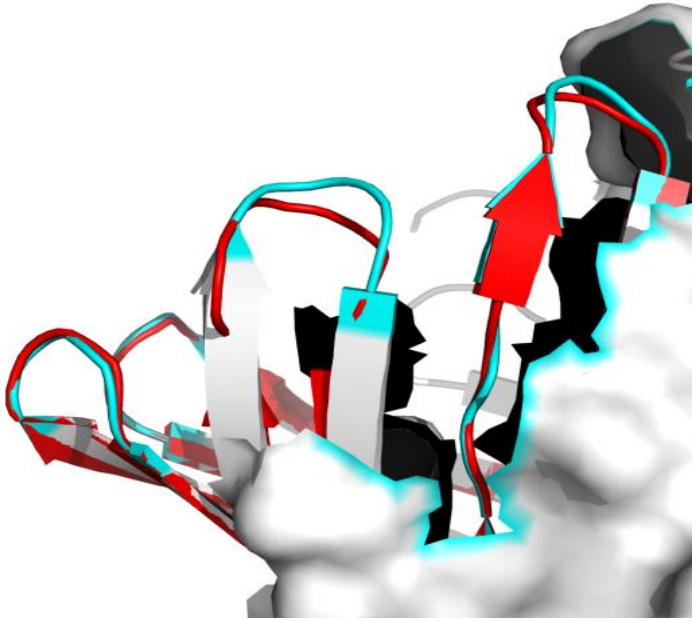
- Periplasmic:
  - 7res + 7res + 7res = 21 res
  - RMSD: 1.71 Å

(Ke Tang, et al, *manuscript*)

Predicting Loops in OMPs

- OmpG (pdb: `2x9k`)
  - Cyan: Native loops.
  - Red: min energy loop by mDiSGro.
- Extracellular
  - 12res + 12res + 10res = 34 res
  - RMSD: 2.62 Å
  - Interacting loops
- Periplasmic
  - 9res + 4res + 4res + 4res = 21 res
  - RMSD: 0.93 Å
  - Interacting loops

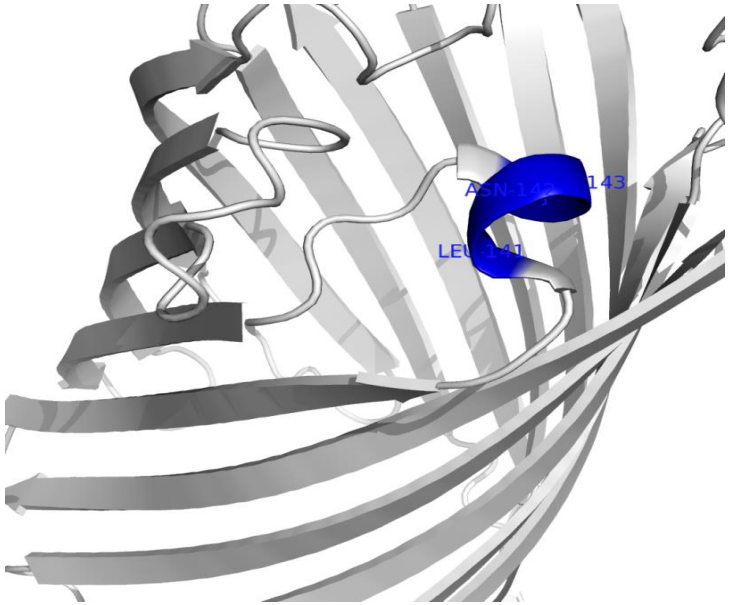




(Ke Tang, et al, *manuscript*)

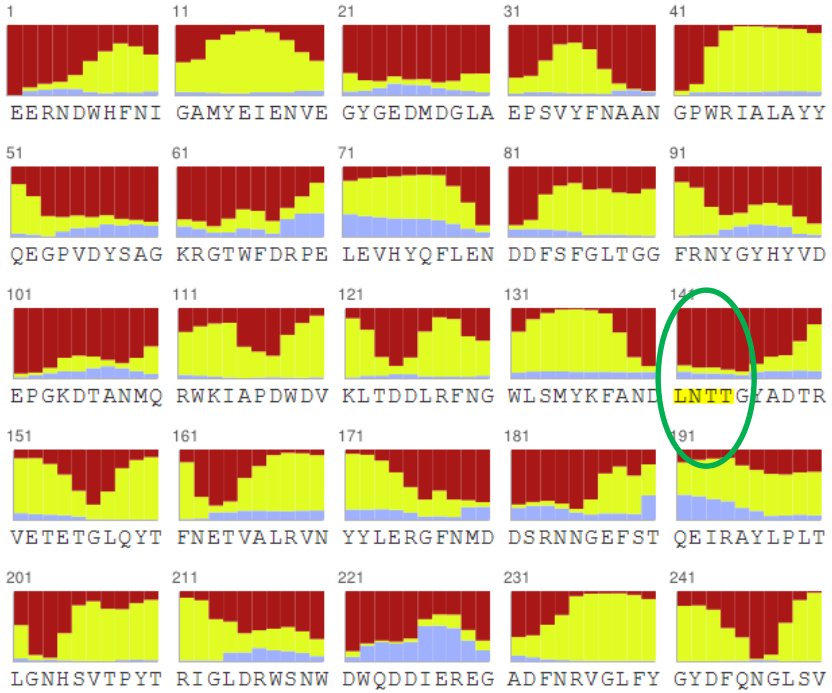
de novo Prediction of Small Secondary Structures in Loops

- Secondary structure prediction methods
  - Cannot predict short SSE in loop of OmpG



GOR V prediction: coil

- Helix: blue
- Beta: yellow
- Coil: red



(T.Z. Sen *et al*, *Bioinformatics*, 2005)

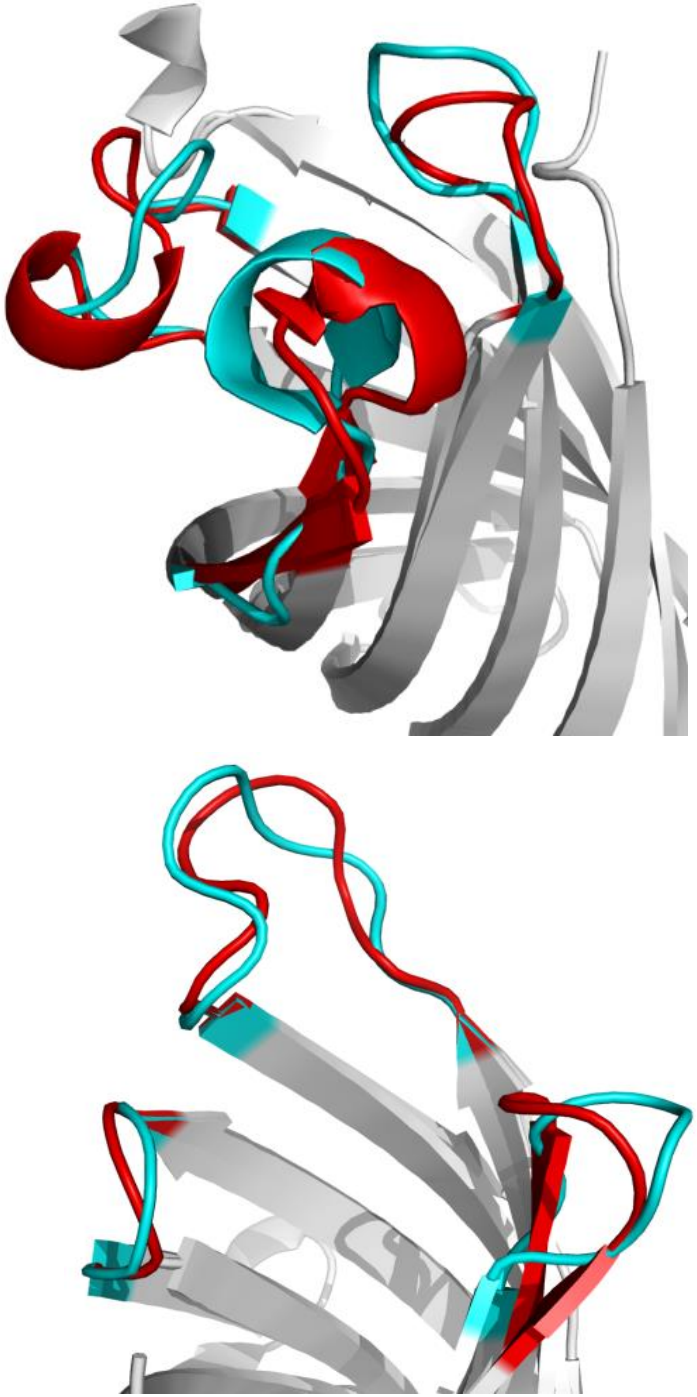
Predicting Loops in OMPs

- PagP (pdb: `1thq`)
  - Cyan: Native loops.
  - Red: min energy loop by mDiSGro.

- Extracellular
  - 11res + 10res + 7res = 29 residues
  - RMSD: 2.08 Å
  - Triplet interacting loops

- Periplasmic:
  - 9res + 7res + 4res = 20 residues
  - RMSD: 1.83 Å
  - Interacting loops

(Ke Tang, et al, manuscript)



Comparative Model:
Scoring Matrix from Residue Substitution Matrix

- Existing approach:
  - PAM and BLOSUM: for soluble proteins
  - PHAT and SLIM matrice: for alpha helical proteins.

- Our approach for beta-barrels:
  - Evolutionary pattern:
    Explicit phylogenetic tree.
  - Model:
    Continuous time Markov process.
  - Evolution of only residues located in the TM region.
  - Bayesian Markov chain Monte Carlo.
    (Yan-Yuan Tseng and JL, 2006, *Mol Biol Evo*)

(David Jimenez-Morales and JL, PLoS ONE, 2011)

- Application:
  - Highly confident homology model

Workflow

blastp

nr-ncbi

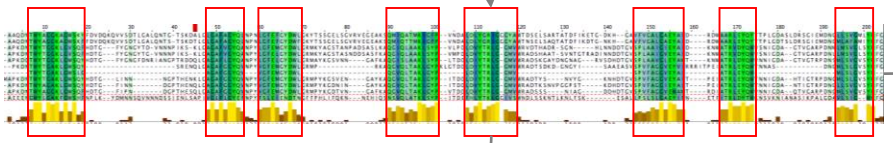
```
>1BXW:A|PDBID|CHAIN|SEQUENCE
MAPKDNTWYTGAKLGWSQYHDTGLINNNGPTHENKLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQ
L
TAKLGYPITDDLDIYTRLGGMVWRADTYSNVYGKNHDTGVSPVFAGGVEYAITPEIATRLEYQWTNNIGDAHTIGTRPD
N
GMLSLGVSYRFG
```

Distance tree of results NEW Related Structures

Sequences producing significant alignments:	Score (Bits)	E Value	
gb|AAS47896.1| sucrose-specific outer membrane porin [Escherichi	188	9e-47	
sp|P22340|SCRY_SALTY Sucrose porin precursor >emb|CAA40656.1|...	188	1e-46	
pdb|1A0T|P Chain P, Sucrose-Specific Porin, With Bound Sucros...	182	5e-45	S
pdb|1OH2|Q Chain Q, Sucrose-Specific Porin, With Bound Sucrose M	171	2e-41	S
ref|YP_001334195.1| sucrose porin [Klebsiella pneumoniae subs...	166	5e-40	G
ref|ZP_00725659.1| COG4580: Maltoporin (phage lambda and malt...	146	6e-34	
ref|YP_001439577.1| hypothetical protein ESA_03526 [Enterobac...	144	2e-33	G
emb|CAG25845.1| sucrose porin precursor [Escherichia coli]	143	3e-33	
ref|YP_001176747.1| porin, LamB type [Enterobacter sp. 638] >...	141	2e-32	G
ref|YP_001454488.1| hypothetical protein CKO_02946 [Citrobact...	140	3e-32	G
emb|CAC14599.1| Porin [Erwinia amylovora]	140	3e-32	
ref|ZP_00833400.1| COG4580: Maltoporin (phage lambda and malt...	134	2e-30	
ref|ZP_00828494.1| COG4580: Maltoporin (phage lambda and malt...	133	3e-30	

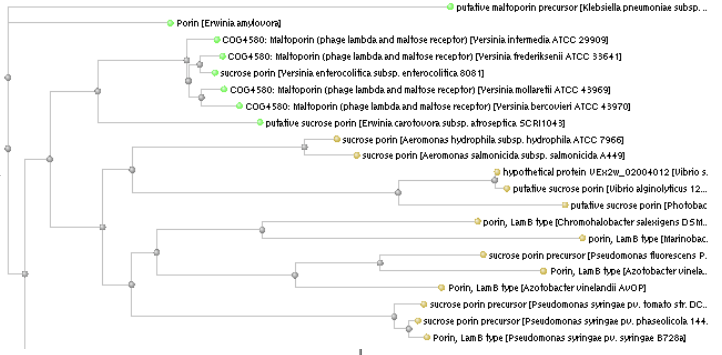
clustalw

* Select sequences between **90 to 30%** identity and perform a MSA.

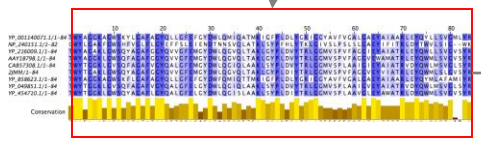


protml

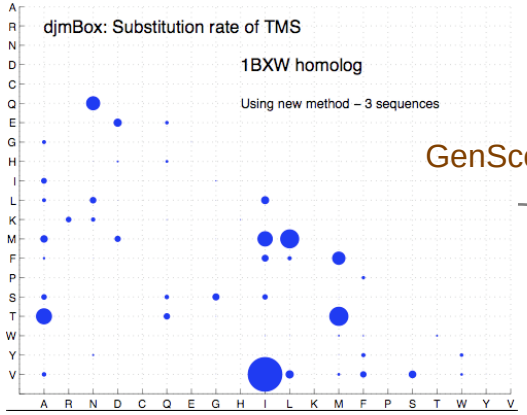
phylogenetic tree



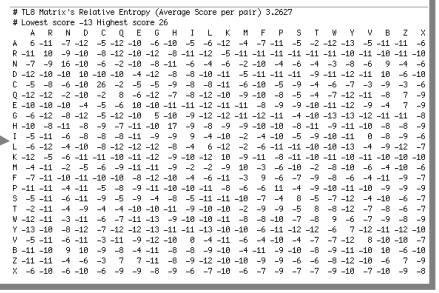
whitebox



codeml



GenScoringMatrix



(David Jimenez-Morales and JL, PLoS ONE, 2011, 6(11):e26400)

Comparative Models of Beta Barrel Membrane Proteins

- Estimated number of proteins whose TM regions can be reliably modeled:
  - About 120 - 180 for 17 templates!

(David Jimenez-Morales and JL, PLoS ONE, 2011)

Summary

- Empirical potential function and reduced state model
  - Combinatorial analysis for weak signal
  - Reduced state model and conformation enumeration

  - Mechanism for stabilization
  - Oligomerization state and PPI prediction
  - Engineering different oligomeric states and resistance to unfolding

- Computational transfer free energy
  - General scales and depth dependent profiles
  - General cooperativity
  - Important insight and predictions

- Template-free TM 3D structure prediction
  - TM barrel structures
  - Loop conformations

Collaborators

- **Andrew Binkowski** (Argonne National Lab)
- **Jeffrey Yan-Yuan Tseng** (Wayne State U)
- **Youfang Cao (**UIC)
- **Dennis Gessmann (UIC)**
- **Gamze Gursoy** (UIC)
- Sema Kachalo (now Intel graphics unit)
- Ronald Jackups, Jr (Wash U, Assist Prof)
- Yingzi Li (SJTU)
- **Meishan Lin** (UIC)
- **Hammad Naveed** (UIC/now Toyota Institute)
- **Ke Tang** (UIC)
- **Anna Terebus** (UIC)
- Wei Tian (UIC)
- **Yun Xu** (UIC)
- **Jieling Zhao** (UIC)
- Jinfeng Zhang (FSU, Associate Prof)

- David Stone (UIC)
- **Luisa Di Pietro (UIC)**
- **Linda Kenney (UIC)**
- **Amy Kenter (UIC)**
- John Marko (NWU)
- Lisa Xu (SJTU)

Papers: http://gila.bioe.uic.edu/lab/ (left column)

Acknowledgement

- NSF DBI
- NIH NIGMS
- CBC Catalyst