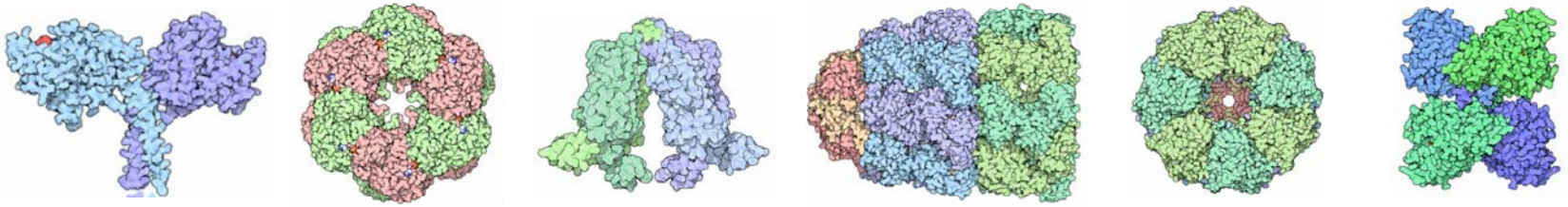
Refinement of reduced protein models with all-atom force fields



Liliana Wróblewska

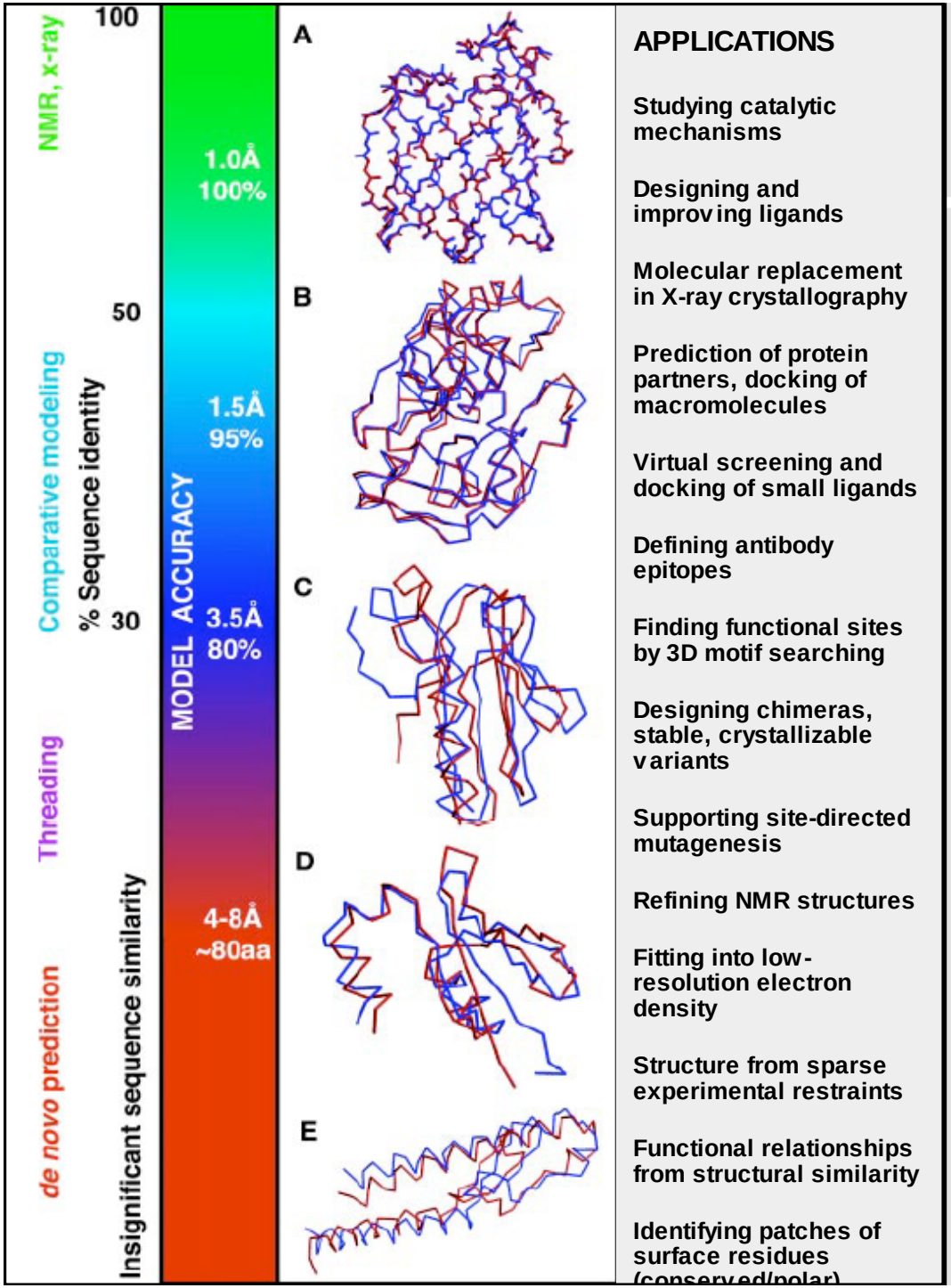
Jeffrey Skolnick's lab

School of Biology

Ph.D. defense 10/30/07

Outline

- General and Specific Aims
- Background and Significance
- Results
- Summary
- Future Work



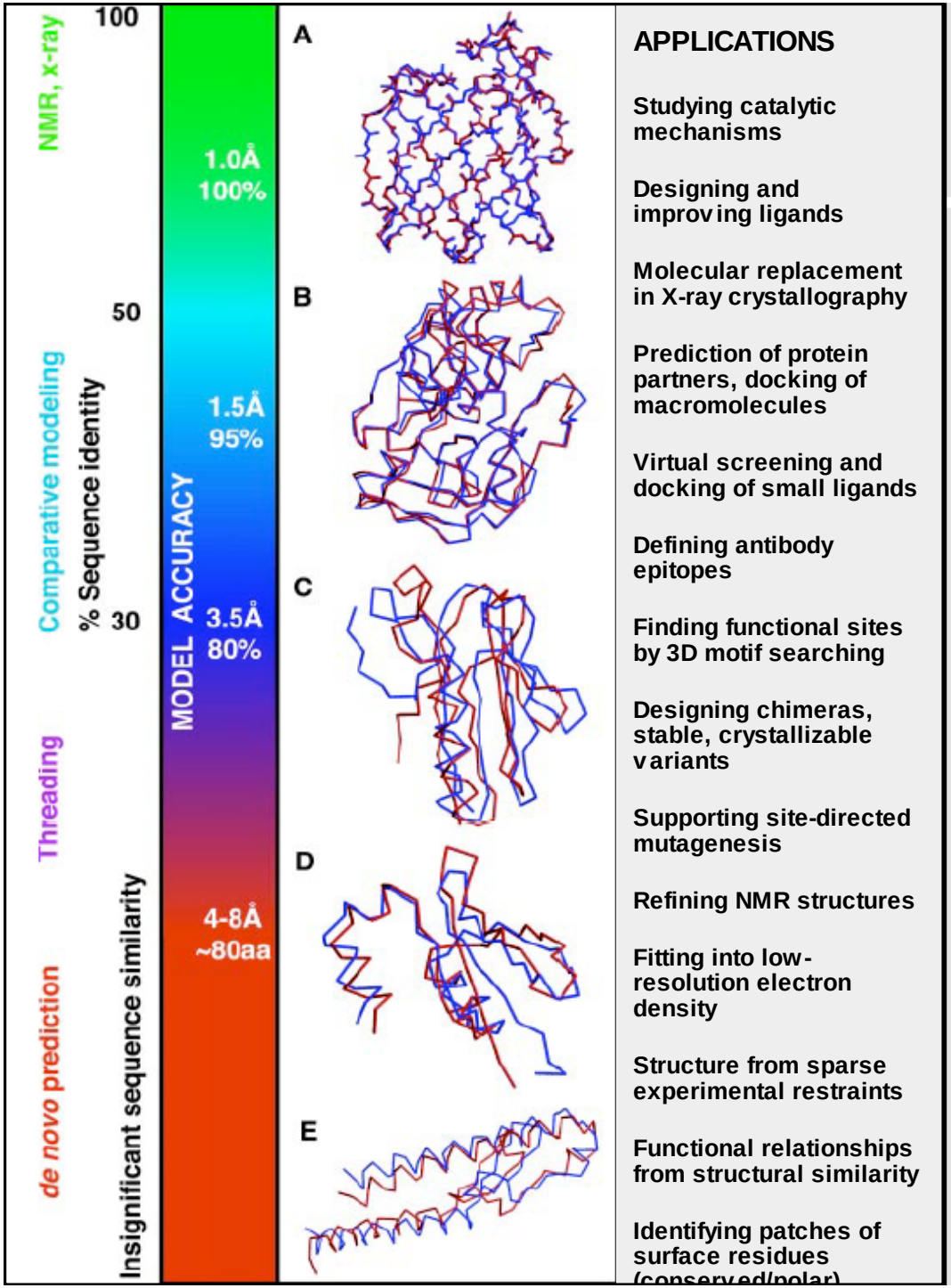
General Aim

refinement of protein models to higher resolution (from ~2-6.5Å backbone RMSD to less than 2 Å)

→

development of a systematic approach

D.Baker, A.Sali, Science, 2001, 294: 93-6



General Aim

refinement of low resolution predicted protein models to higher resolution (from ~2-6.5Å backbone RMSD to less then 2 Å)

→

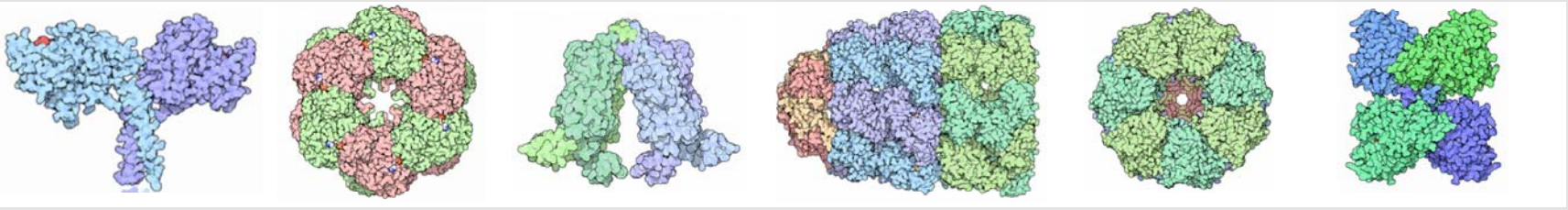
development of a systematic approach

D.Baker, A.Sali, Science, 2001, 294: 93-6

Specific Aims

- An assessment of ability of the existing all-atom force fields to refine protein models
- Identification of the key problems in the all-atom force fields
- Optimization of the all-atom force field components
- Modification of the energy function
- Refinement using optimized force field

Background and Significance



Protein structure prediction – the rationale

"The amino acid sequences of polypeptide chains (…) only make functional sense when they are in the three dimensional arrangement that characterizes them in the native protein structure"

- C.B.Anfinsen, 1972, Nobel Lecture

Protein structure prediction – the foundation

THERMODYNAMIC HYPOTHESIS:

"The three-dimensional structure of a native protein in its normal physiological milieu (…) is the one in which the Gibbs free energy of the whole system is lowest"

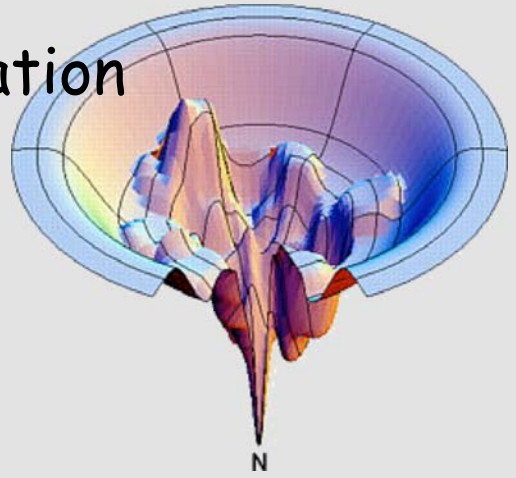
- C.B.Anfinsen, 1972, Nobel Lecture

Protein structure prediction

Ideal protocol:

- Generate all possible conformations of a polypeptide chain
- Calculate the free energy for each model
- Choose the lowest energy conformation

→ native structure (N)



Ken A. Dill and Hue Sun Chan, *Nature Structural Biology*, January 1997 4(1)

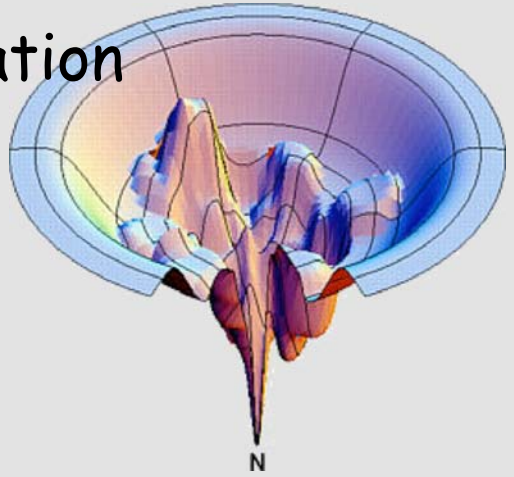
Protein structure prediction

Ideal protocol:

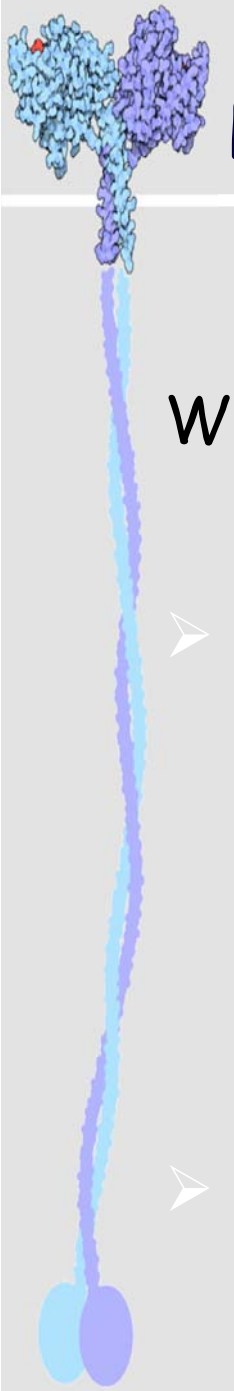
- Generate all possible conformations of a polypeptide chain
- Calculate the free energy for each model
- Choose the lowest energy conformation

→ native structure (N)

IMPOSSIBLE



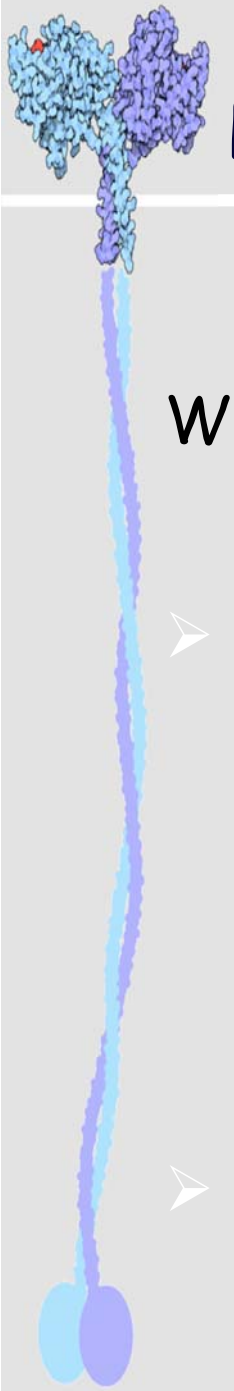
Ken A. Dill and Hue Sun Chan, *Nature Structural Biology*, January 1997 4(1)



Protein structure prediction

Why impossible?

- Generation of all possible conformations of a polypeptide chain
  - 100 amino acid protein, 2 configurations per amino acid:
  - 2^{100} (~ 10^{30}) conformations
  - 1000 conformations/1s ~ 10^{20} years
- Calculation of the free energy for each model
  - high accuracy quantum methods can handle on the order of tens of atoms

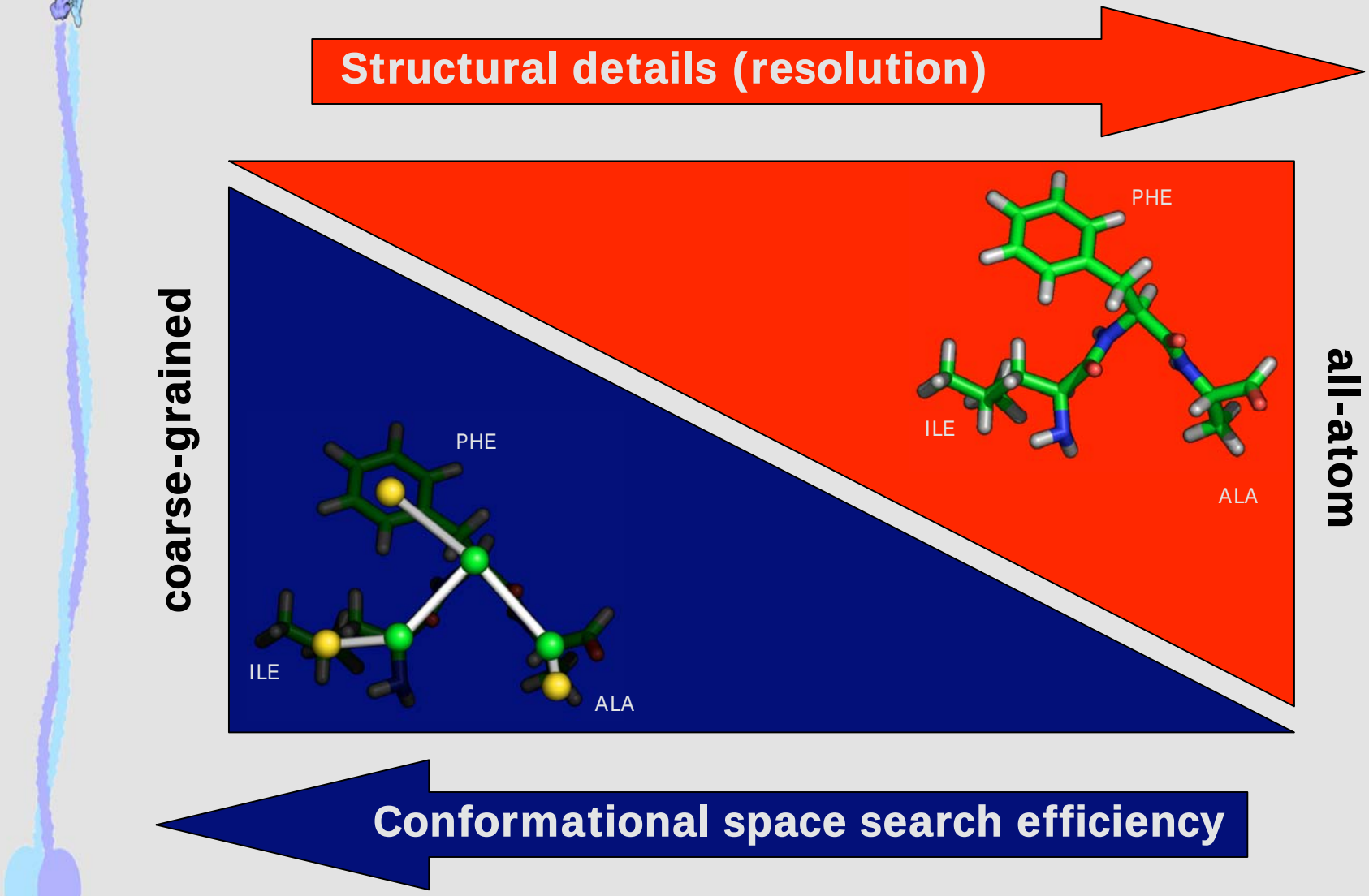


Protein structure prediction

Why impossible?

- Generation of all possible conformations of a polypeptide chain
  - 100 amino acid protein, 2 configurations per amino acid:
  - 2^{100} (~ 10^{30}) conformations
  - 1000 conformations/1s ~ 10^{20} years
    - SAMPLING PROBLEM
- Calculation of the free energy for each model
  - high accuracy quantum methods can handle on the order of tens of atoms

The trade-off



Hierarchical approach

Conformational space search efficiency

- Reduction of the number of interaction centers (coarse-grained models) for GLOBAL SEARCH

Structural details (resolution)

- Reconstruction of the atomic details for REFINEMENT

Hierarchical approach

- Reduction of the number of interaction centers (coarse-grained models) for GLOBAL SEARCH

• Explored extensively - the performance of current state-of-the-art methods for structure prediction: 70% of sequences <200aa have structures predicted with accuracy of < 6.5 Å (TASSER)

- Reconstruction of the atomic details for REFINEMENT

TASSER: Y.Zhang, J.Skolnick, PNAS, 2004, 101: 7594-7599

Hierarchical approach

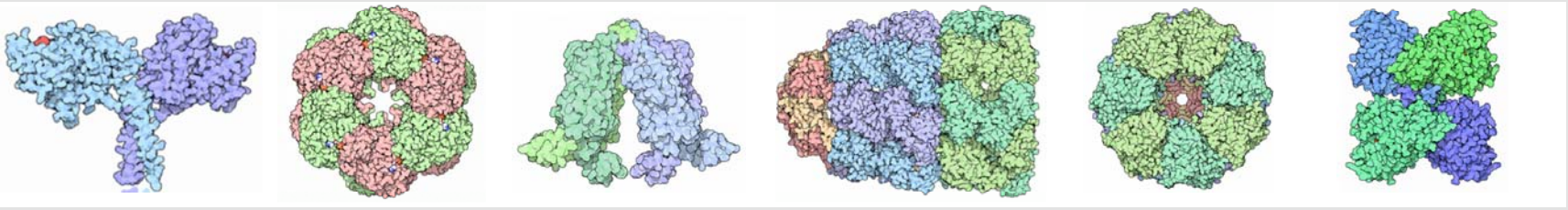
- Reduction of the number of interaction centers (coarse-grained models) for GLOBAL SEARCH

• Explored extensively - the performance of current state-of-the-art methods for structure prediction: 70% of sequences <200aa have structures predicted with accuracy of < 6.5 Å (TASSER)

- Reconstruction of the atomic details for REFINEMENT

• There is no method for systematic all-atom refinement of protein models

Results



Part I

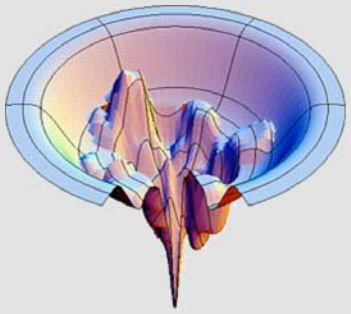
Can contemporary all-atom potentials refine protein models?

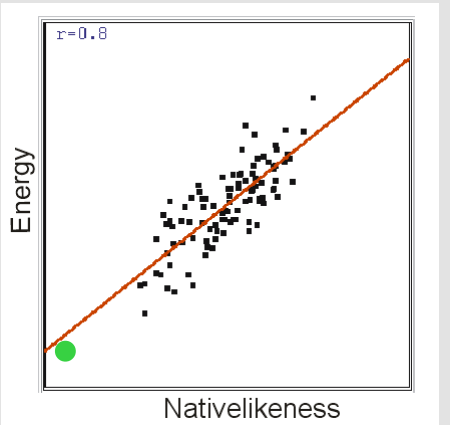
L. Wróblewska and J. Skolnick, J. Comput. Chem. (2007) 28:2059-2066

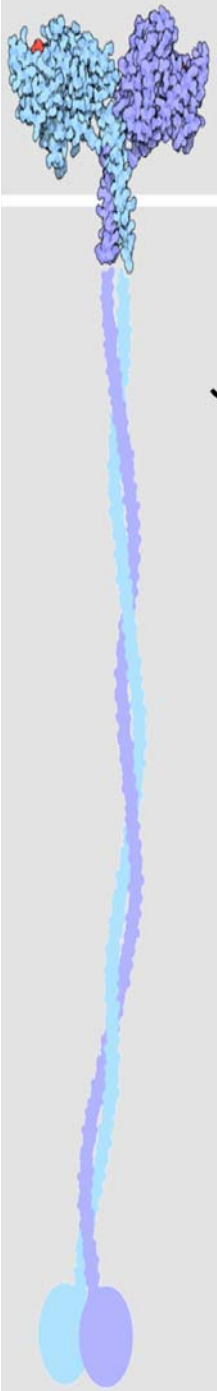
I. Desired characteristics of a force field

1. Native structure is the lowest energy conformation of a given protein sequence

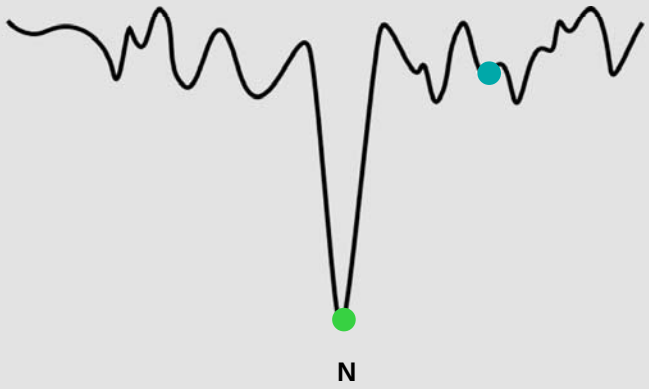
2. Energy correlates with native-similarity (the lower the energy, the closer a particular conformation is to the native structure)

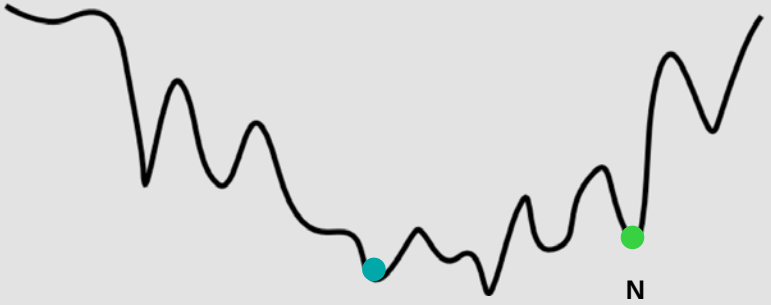


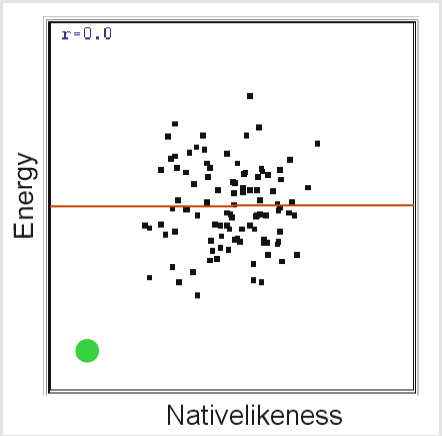


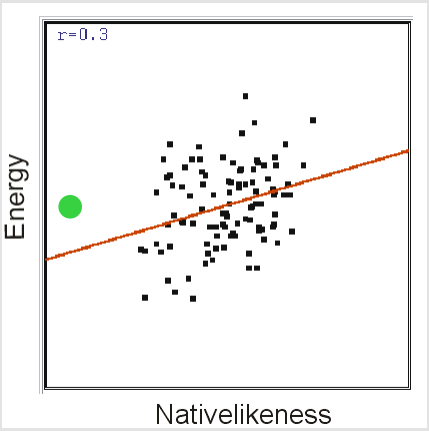


I. Typical energy landscapes

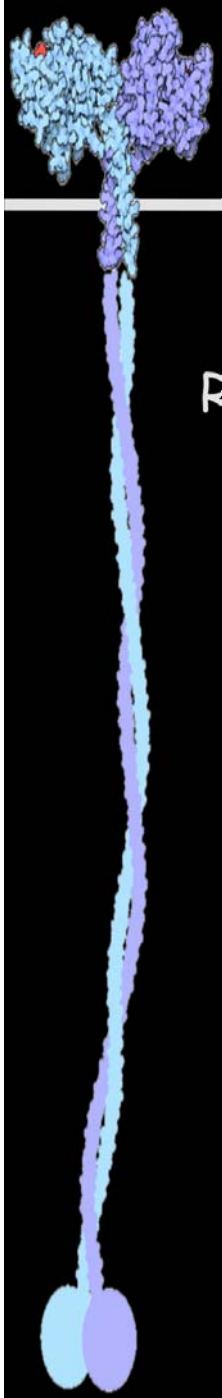






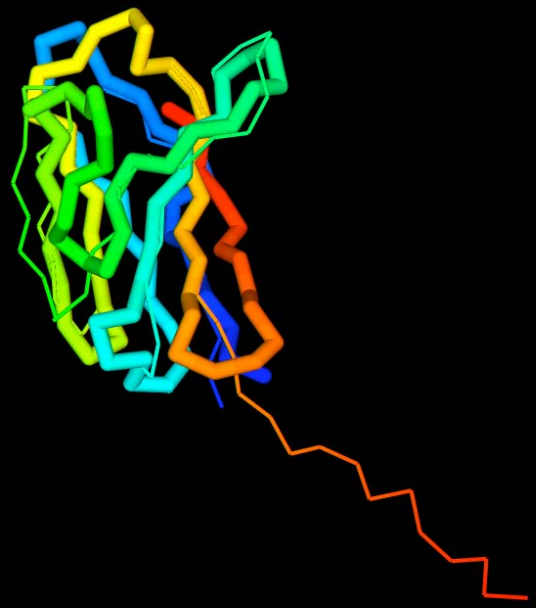


BAD SHAPE



I. Structure similarity metrics

	RMSD	TM-score
Range:	0-up	1-0
	chain length dependent	chain length independent



Thick tube: native
Thin tube: decoy
RMSD: 10.91 Å, TM-score: 0.72

Thick tube: native
Thin tube: decoy
RMSD: 10.30 Å, TM-score: 0.31

I. AMBER force field

E = E (bond) +	harmonic bond stretching
E (angle) +	harmonic angle bending
E (dihedral) +	cosine-like dihedral angle term
E (van der Waals) +	6-12 Lennard Jones potential
E (electrostatics) +	coulombic potential
E (solvation)	free energy of solvation consisting of polar contribution modeled by Generalized Born approximation and non-polar surface area dependent component (SA)

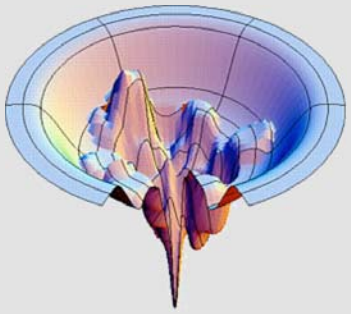
(GB)

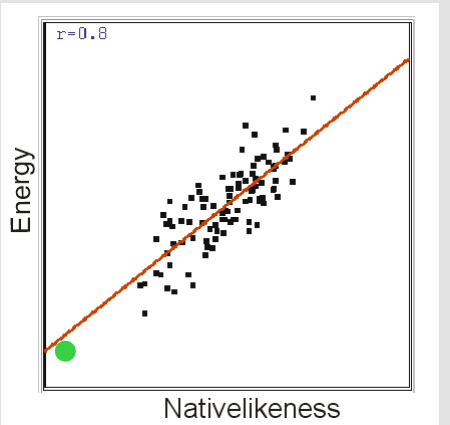
AMBER: D.A.Case *et.al.*, J. Computat. Chem. 2005, **26:** 1668-1688.

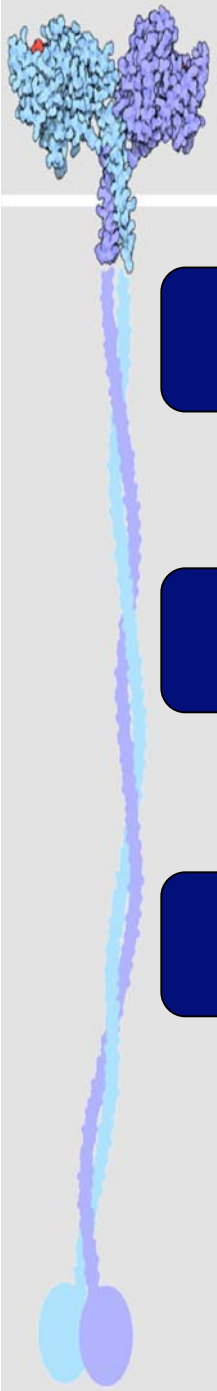
I. Desired characteristics of a force field

1. Native structure is the lowest energy conformation of a given protein sequence

2. Energy correlates with native-similarity (the lower the energy, the closer a particular conformation is to the native structure)







I. Assessment of AMBER force field

TASSER decoys
in reduced representation

↓

All-atom reconstruction

↓

Energy minimization
with AMBER

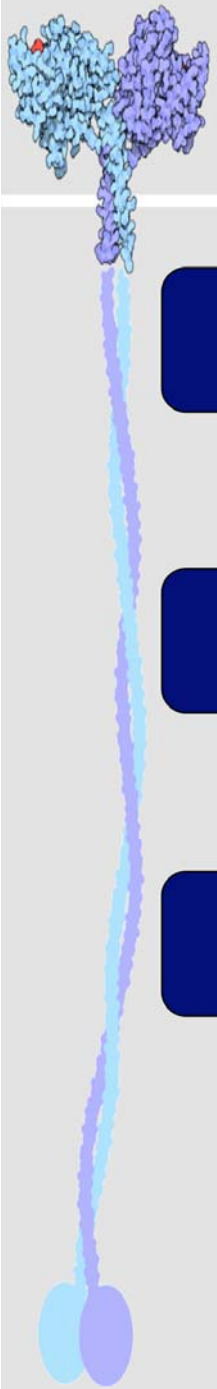
- 100 nonhomologous proteins
- 50-200 amino acids
- 1000 decoys per protein

I. Assessment of AMBER force field

TASSER decoys
in reduced representation

All-atom reconstruction

Energy minimization
with AMBER

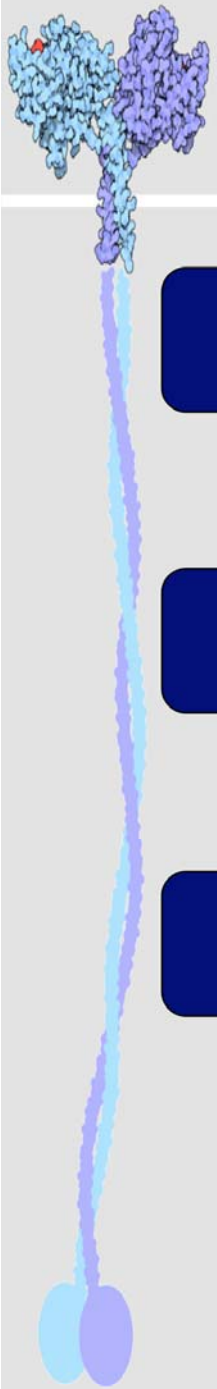


I. Assessment of AMBER force field

TASSER decoys
in reduced representation

All-atom reconstruction

Energy minimization
with AMBER

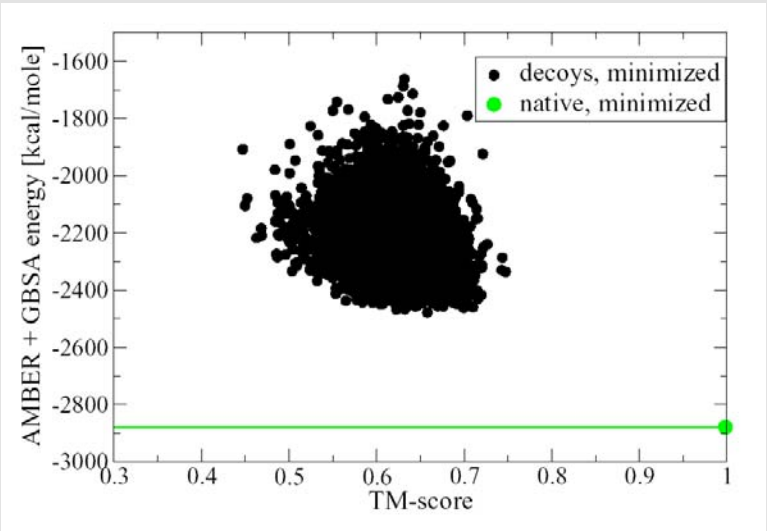


I. Assessment of AMBER force field

TASSER decoys
in reduced representation

All-atom reconstruction

Energy minimization
with AMBER



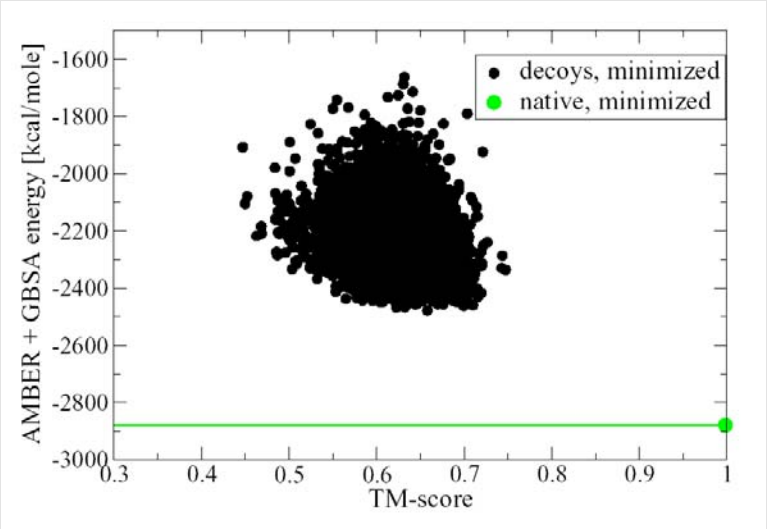
I. Assessment of AMBER force field

TASSER decoys
in reduced representation

All-atom reconstruction

Energy minimization
with AMBER

AMBER conformational search
1000 decoys per protein



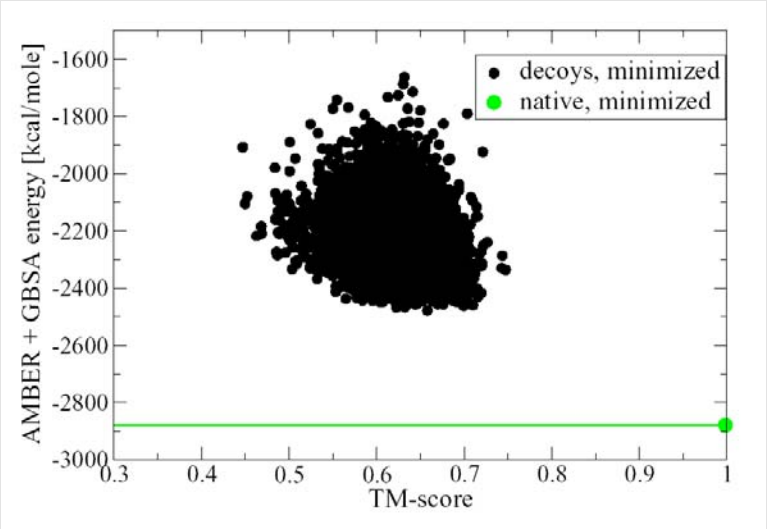
I. Assessment of AMBER force field

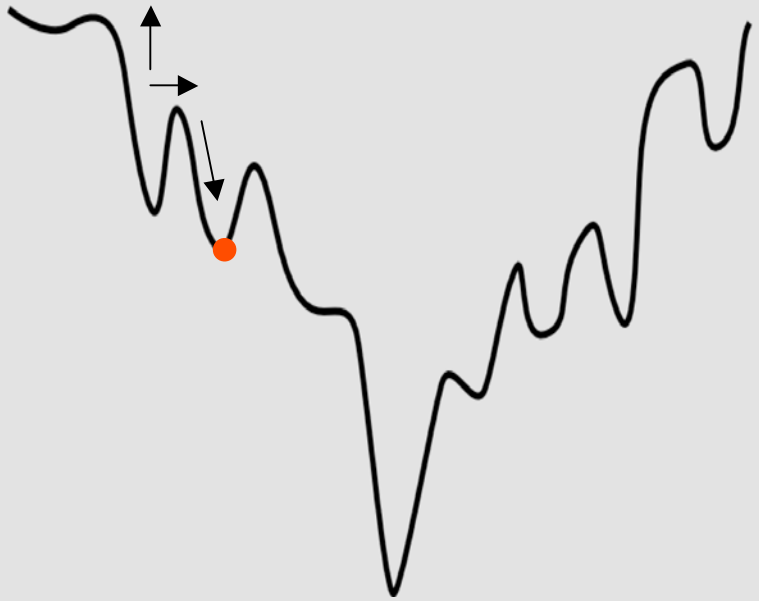
TASSER decoys
in reduced representation

All-atom reconstruction

Energy minimization
with AMBER

AMBER conformational search
1000 decoys per protein





I. Assessment of AMBER force field

TASSER decoys
in reduced representation

↓

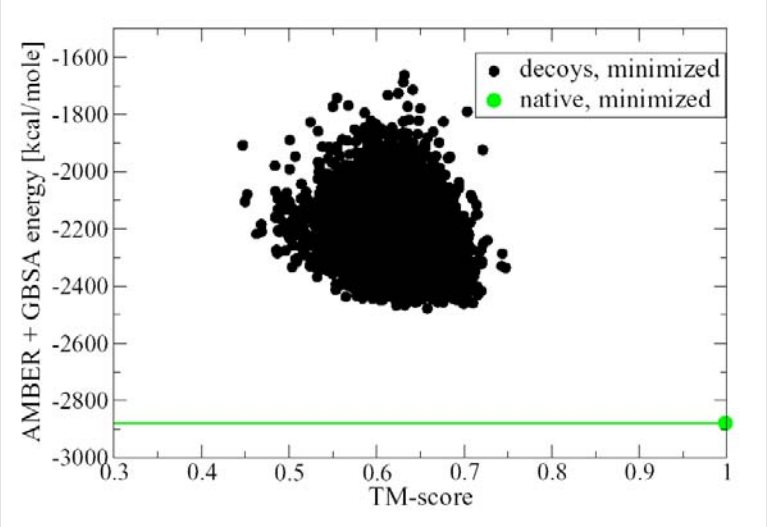
All-atom reconstruction

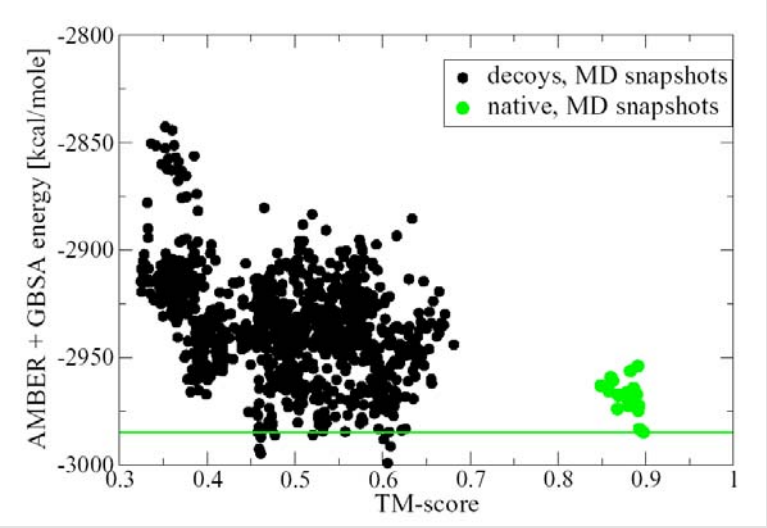
↓

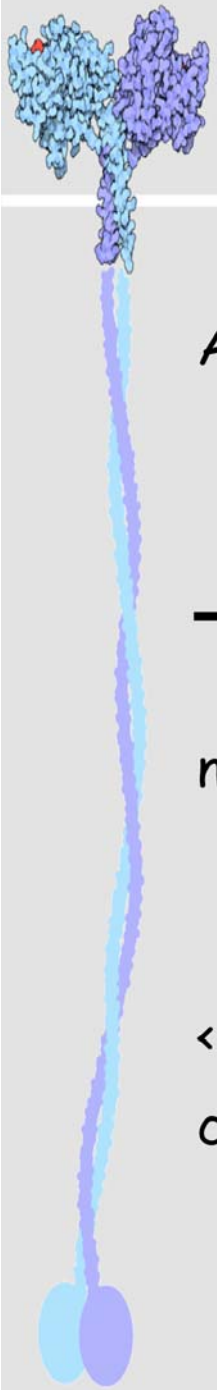
Energy minimization
with AMBER

↓

AMBER conformational search
1000 decoys per protein







I. Assessment of AMBER force field

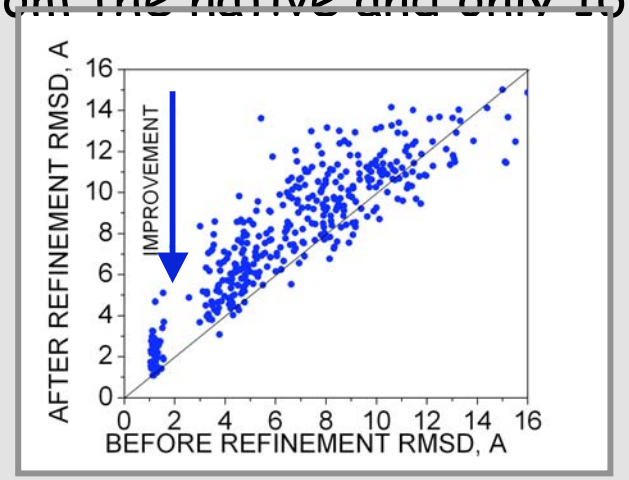
Average results:

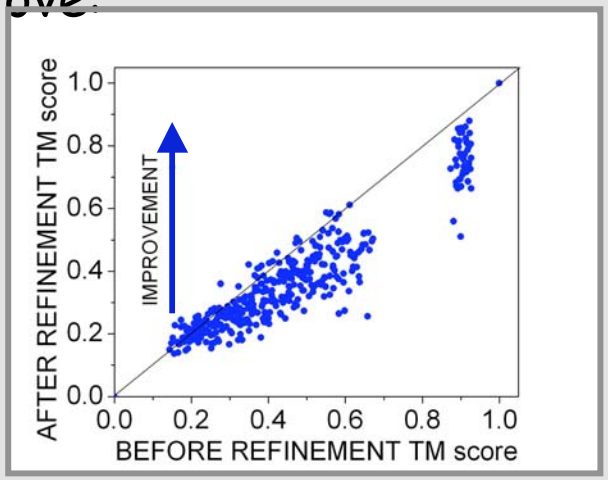
	FF99	FF03
native ranking #1	20%	48%
<energy – TM-score correlation coeff.>	0.1	0.25

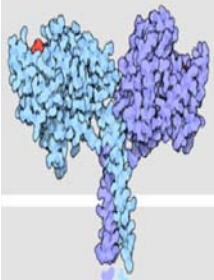
I. Assessment of AMBER force field

- AMBER ff03 performs better than ff99
- Neither version is capable of model refinement

during search models 84% of decoys drift farther away from the native and only 16% improve:







I. Identification of the key problems

- The energy components that mostly disfavor native-like structures:
  - dihedral angle energy
  - electrostatics of bonded atoms (1-4)

- The energy components that display no native-likeness specificity:
  - electrostatics
  - generalized Born solvation

} large contributions

- Hydrogen bond energy seems to be underestimated

(numerous hydrogen bonds are distorted during simulations)

I. Conclusions

- The original AMBER potential is not capable of protein model refinement (flat energy landscape)

Part II

Development of a physics-based force field for the scoring and refinement of protein models

L. Wróblewska A. Jagielska and J. Skolnick, Biophys. J. (submitted)

II. Optimization of AMBER energy components

- Data set: 58 proteins, diverse set of ~20,000 decoys per protein generated by AMBER molecular dynamics and atomic-TASSER Monte Carlo search

- Methodology: optimizing w_i in: $E = \sum w_i e_i$ to:
  - maximize energy - TMscore correlation coefficient
  - maximize native - decoy energy gap (Z-score)

II. Optimization of AMBER energy components

	FF03	FF03 OPT	
		training set	testing set
<energy – TM-score CC>	0.25	0.63	0.61
CC > 0.60	12%	47%	49%
TM-score > 0.90	22%	93%	84%
RMSD < 2.0 A	48%	93%	88%
<Z-SCORE>	0.16	2.65	2.18

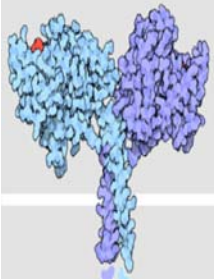
II. Modification of the energy components

- New component for hydrogen bond energy added

$$E_{HB} = q_1 q_2 (1/r_{NO} + 1/r_{CH} - 1/r_{OH} - 1/r_{CN}) \cdot f \qquad \text{(DSSP)}$$

	FF03	FF03 OPT	FF03/HB OPT
<energy – TM-score CC>	0.28	0.62	0.65
CC > 0.60	12%	48%	64%
TM-score > 0.90	22%	86%	90%
RMSD < 2.0 A	48%	89%	91%
<Z-SCORE>	0.16	2.30	2.29

DSSP: W. Kabsch, C. Sander, Biopolymers, 1983, 22: 2577-2637.



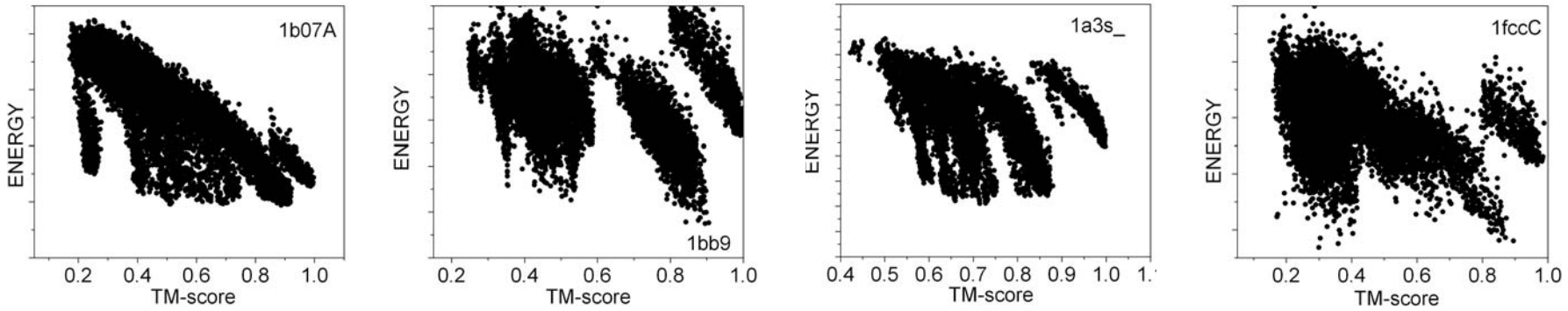
II. Modification of the energy components

- Electrostatics and Generalized Born solvation parts omitted

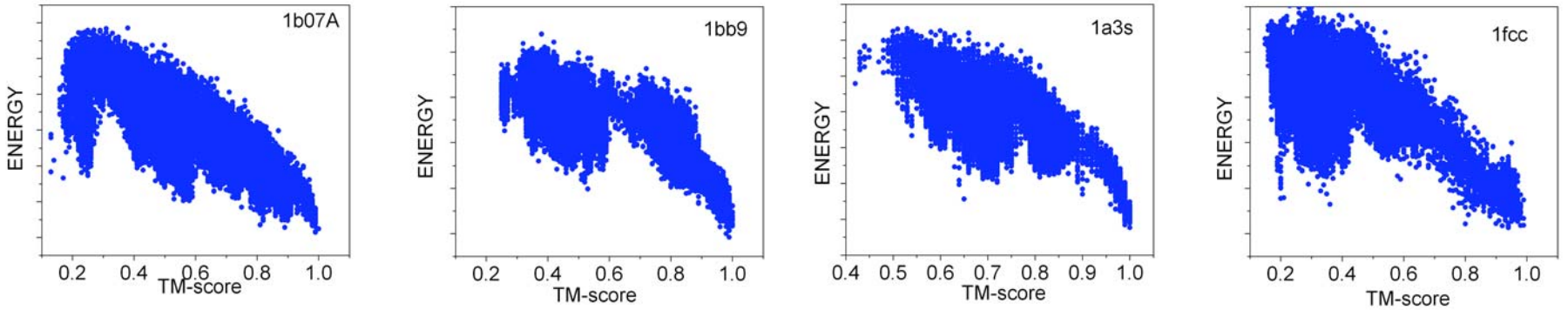
	FF03	FF03 OPT	FF03/HB OPT	FF03/HB Reduced/OPT
<energy – TM-score CC>	0.28	0.62	0.65	0.58
CC > 0.60	12%	48%	64%	43%
TM-score > 0.90	22%	86%	90%	71%
RMSD < 2.0 A	48%	89%	91%	84%
<Z-SCORE>	0.16	2.30	2.29	1.56

II. The optimized force field

Before optimization



After optimization (FF03/HB)



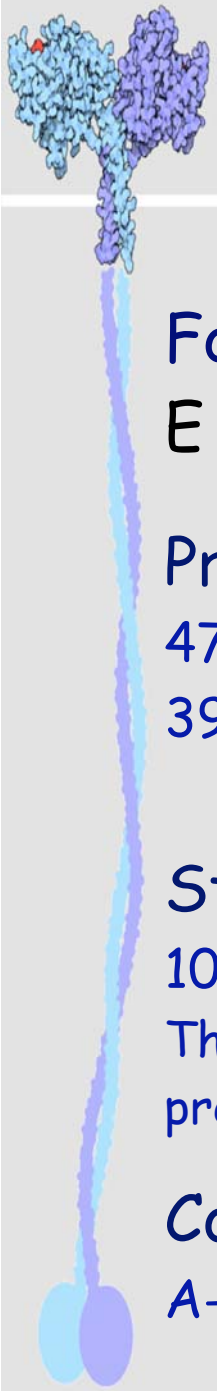
II. Conclusions

- Optimization of weights of AMBER FF03 energy significantly improved
  - native scoring
  - energy – native-similarity correlation coefficient
- Explicit hydrogen bond potential further improved the optimized force field
- Reduced optimized potential, without time consuming electrostatics and GB solvation is still much better than the original FF03 force field

Part III

Systematic refinement of reduced protein models with all-atom force field

A. Jagielska, L. Wróblewska and J. Skolnick (in prep.)



III. Refinement protocol

Force field:

$$E = w_1 E_{dihedral} + w_2 E_{VdWaals} + w_3 E_{SA} + w_4 E_{HB}$$

Protein set

47 proteins 54-123 amino acids (a, b, a/b)

39 proteins were not used in the optimization of the potential

Starting decoys

100 decoys per protein, that span 0-8 Å RMSD to the native

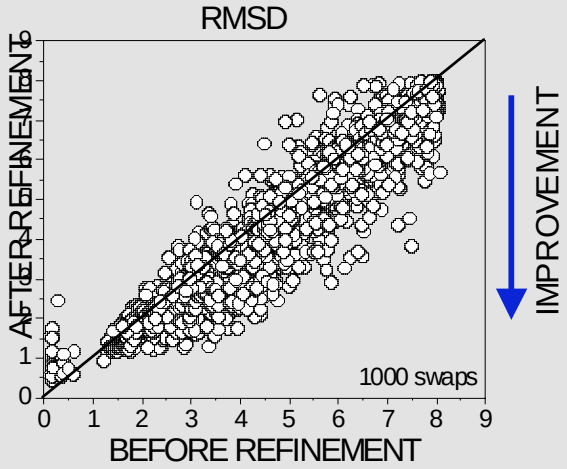
The decoys were randomly chosen from the decoy clouds obtained previously using TASSER/ff03-MD/ff03-ATASSER

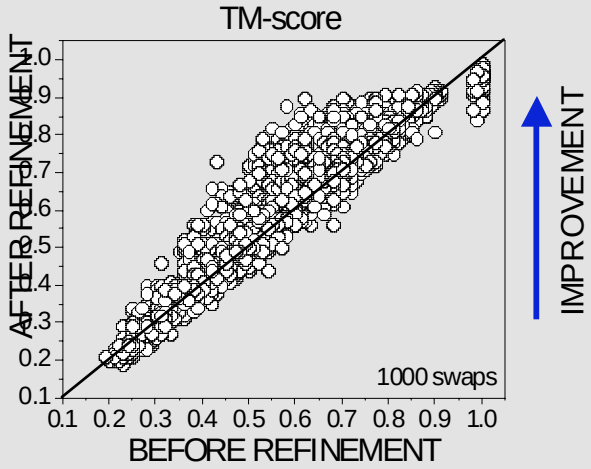
Conformational Search

A-TASSER Monte Carlo Replica Exchange

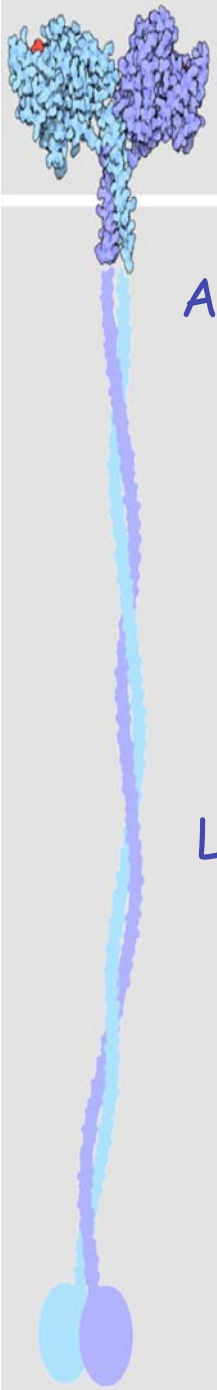
III. Refinement

refined model = lowest energy model from the refinement run



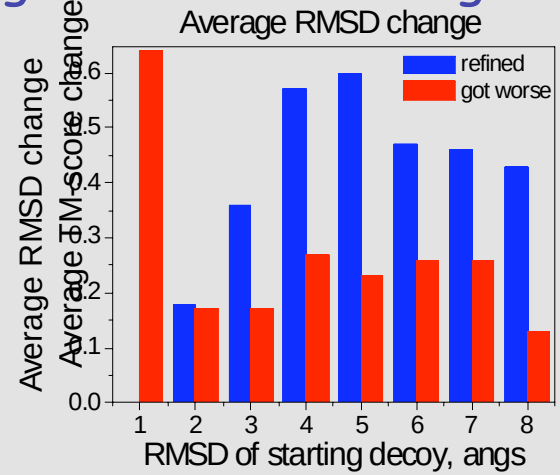


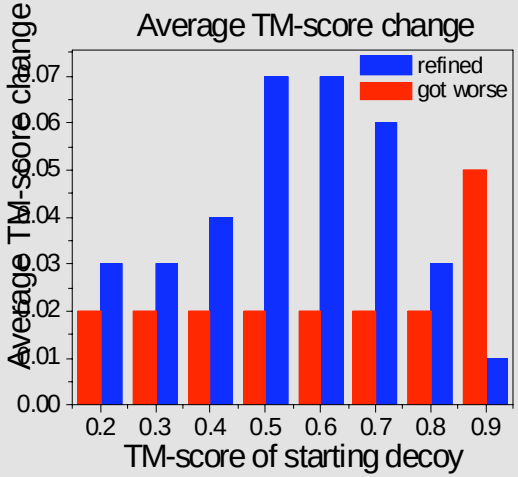
	RMSD	TM-score
Improvement	70%	70%
Deterioration	28%	18%
no change	2%	12%



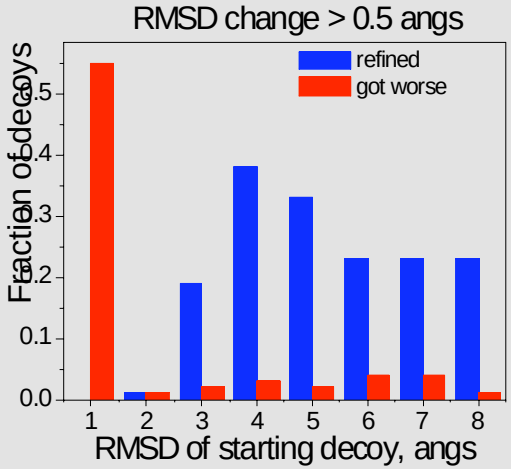
III. Refinement

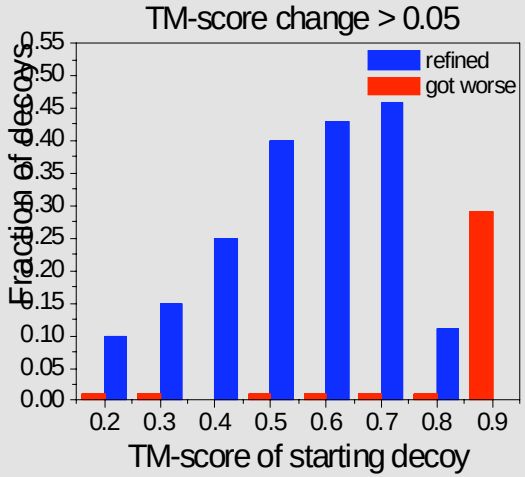
Average structural changes



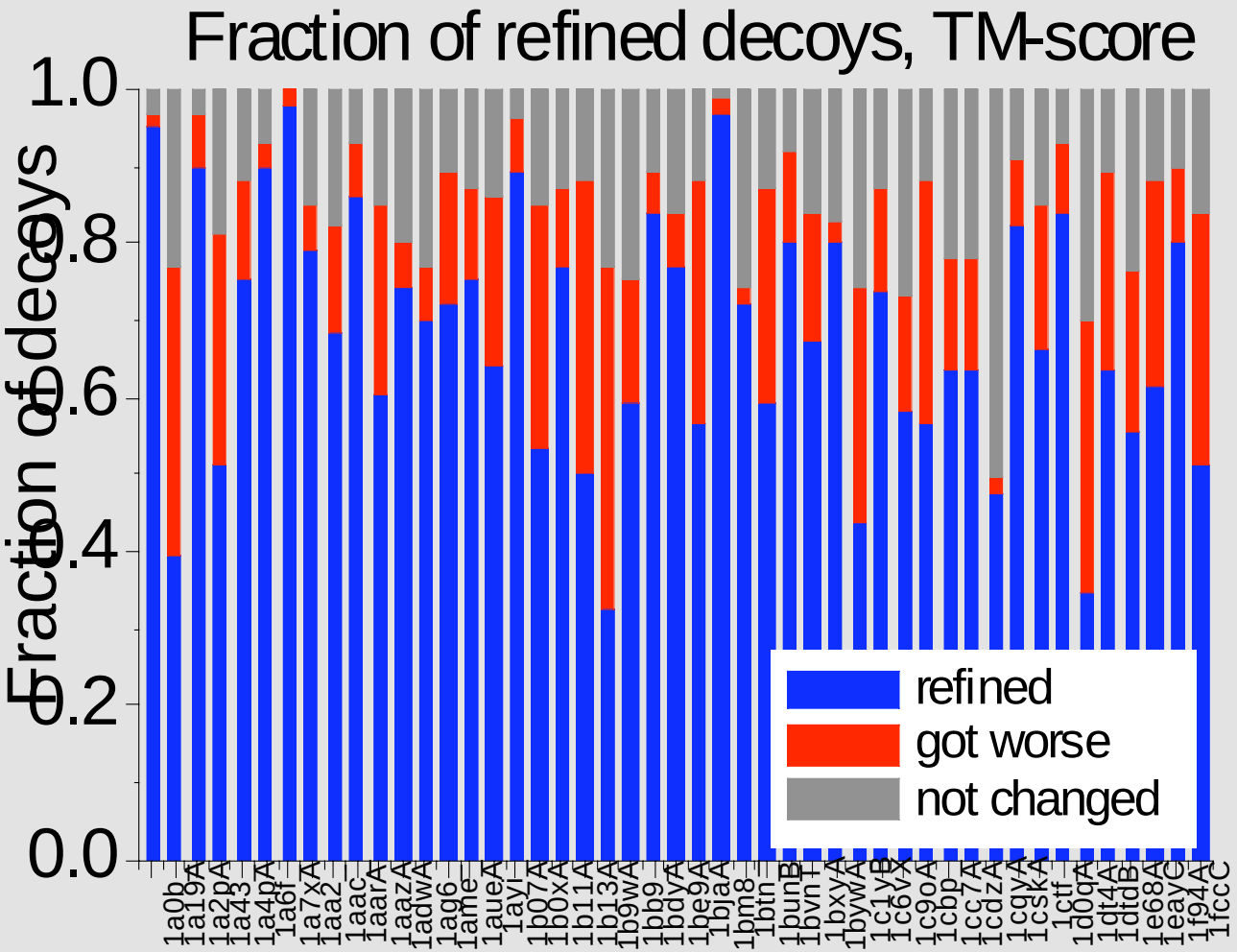


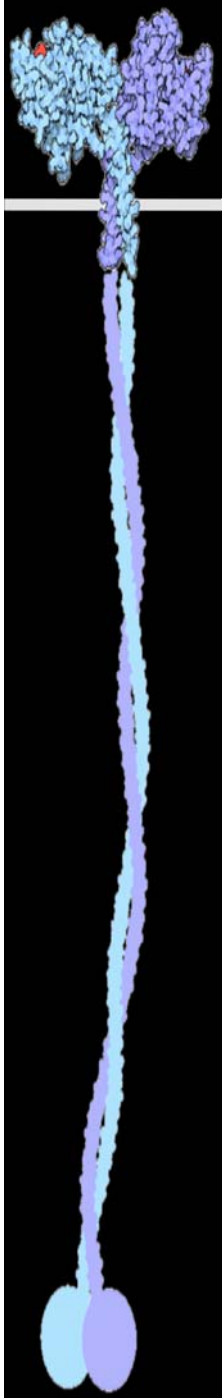
Large structural changes



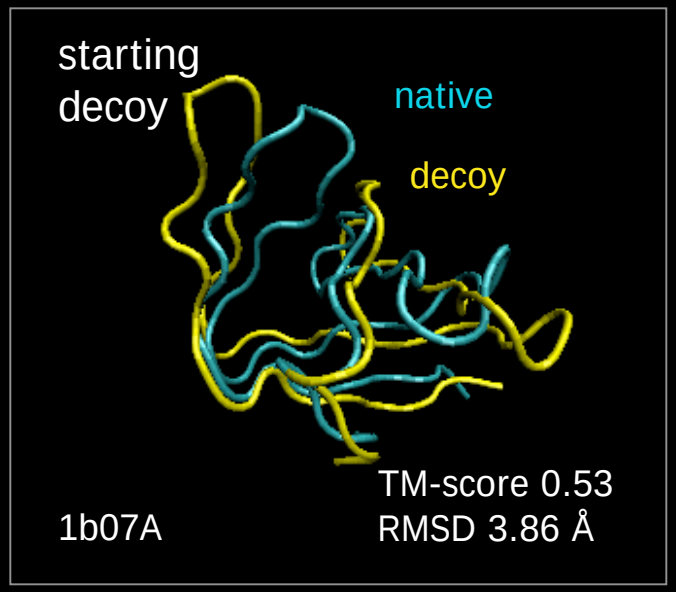


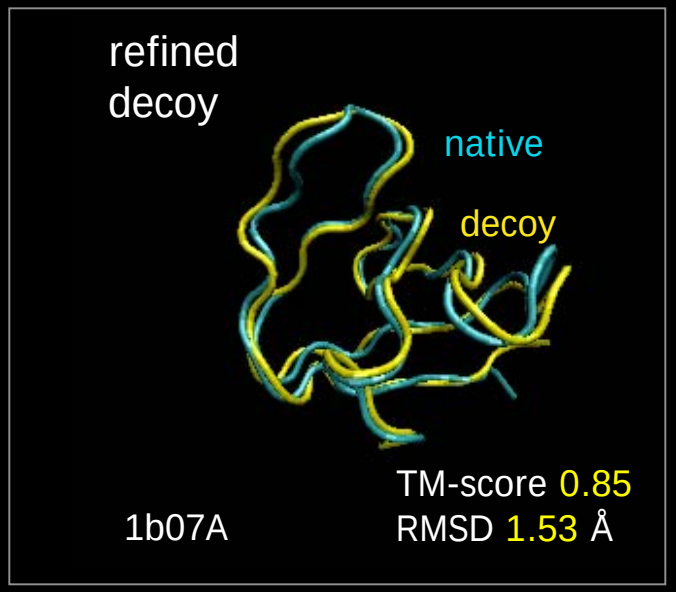
III. Refinement



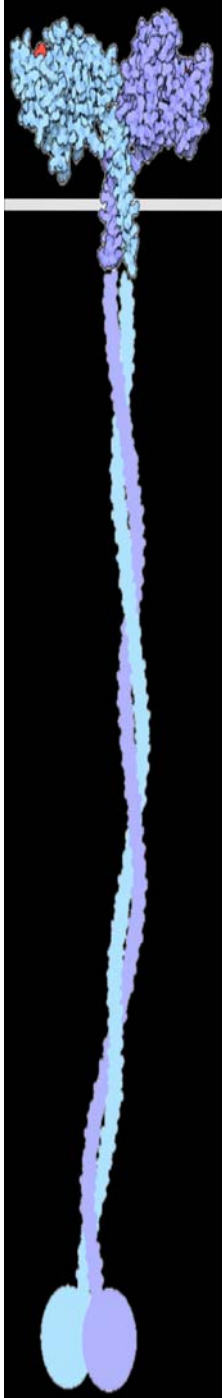


III. Refinement

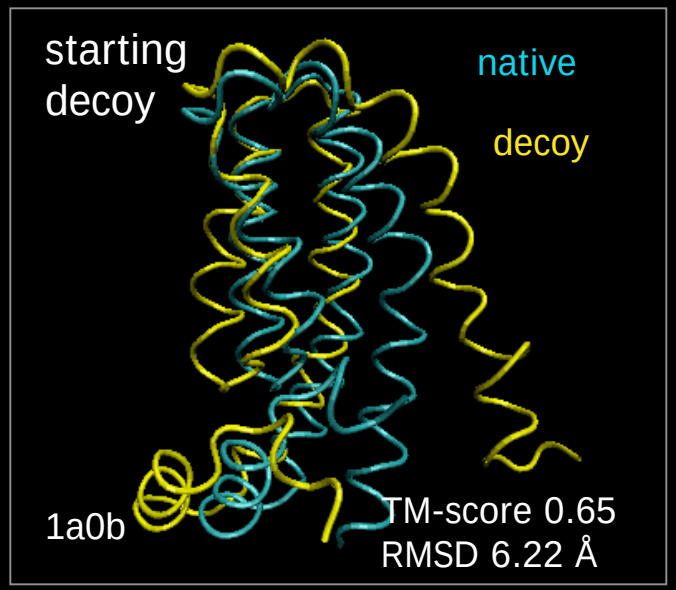


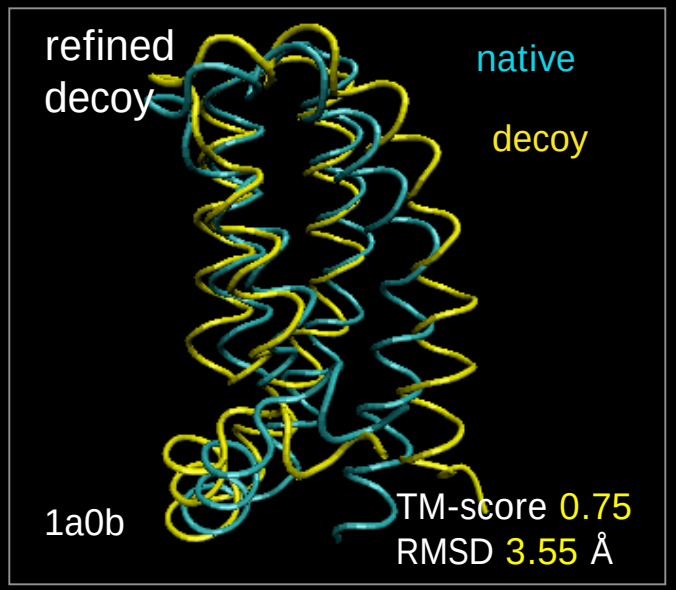


the lowest energy structure is shown



III. Refinement

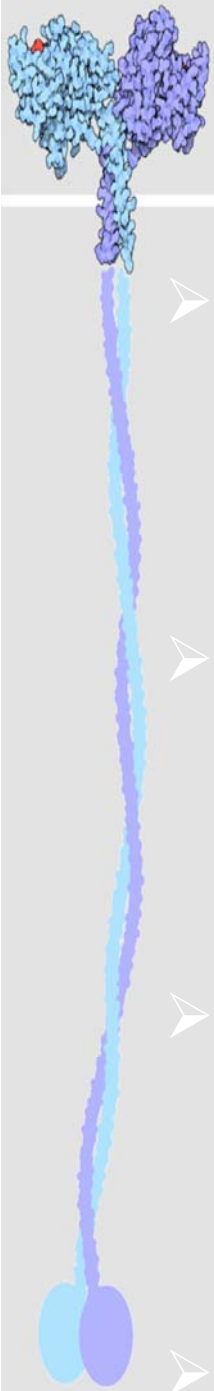




the lowest energy structure is shown

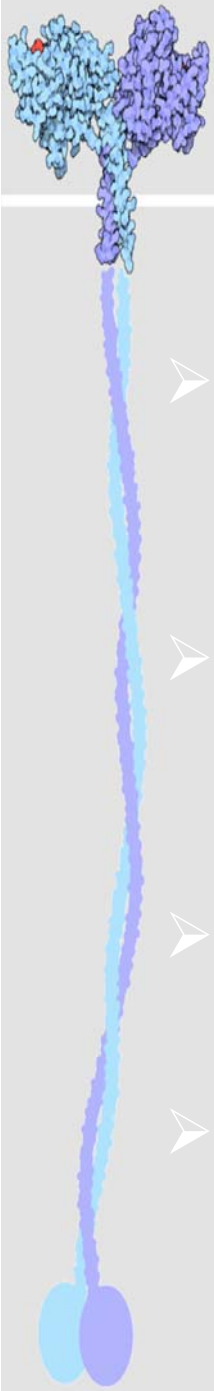
III. Conclusions

- The optimized force field allows for a systematic refinement of protein models
- For the first time such unrestrained, systematic all-atom refinement has been achieved



Summary

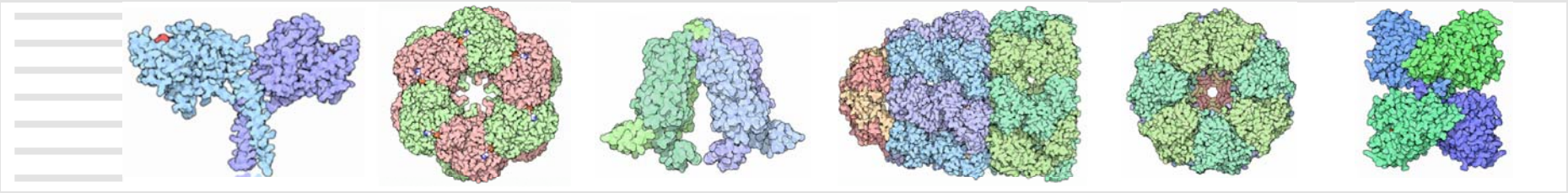
- Extensive testing of AMBER potential allowed for identification of the key problems in terms of refinement purposes
- Optimization of the AMBER energy contributions improved both: correlation between energy and TM-score, and scoring of native structures
- With the addition of the hydrogen bond energy and further optimization the improvement was much more pronounced
- In our set of representative decoy structures, the



Future Work

- Incorporation of the refinement procedure in the automated protein structure prediction pipeline
- Development of a confidence score for model selection
- Improvement of the dihedral angle energy term
- Reiteration of the optimization procedure with the new decoy structures obtained during refinement

Acknowledgements



Jeffrey Skolnick

Facundo M. Fernandez
King Jordan
John McDonald
David C. Sherrill

Laura Cook
Kathryn Macken
I[illegible]ntoro

Skolnick group:

Piotr Rotkiewicz
Anna Jagielska
Marcos Betancourt
Daisuke Kihara
Yang Zhang
Wei Li
Weidong Tian
Andras Szilagyi
Eckart Bindewald
Vera Grimm
Olaf Zimmermann
Dukka K.C.
Jason Rappleye

Adrian Arakaki
Jessica Gilmore
Ying Huang
Jake Boggan
Hongyi Zhou
Seung Yup Lee
Ryang Guk Kim
Jose Borreguero
Bartosz Ilkowski
Michal Brylinski
Shashi Pandit
Dmitry Ivankov
[illegible]

Thank You!