A Fast Method for Large-Scale De Novo Peptide and Miniprotein Structure Prediction

Maupetit, J., Derreumax, P., Tufféry, P.
J Comput Chem 31: 726-738, 2010

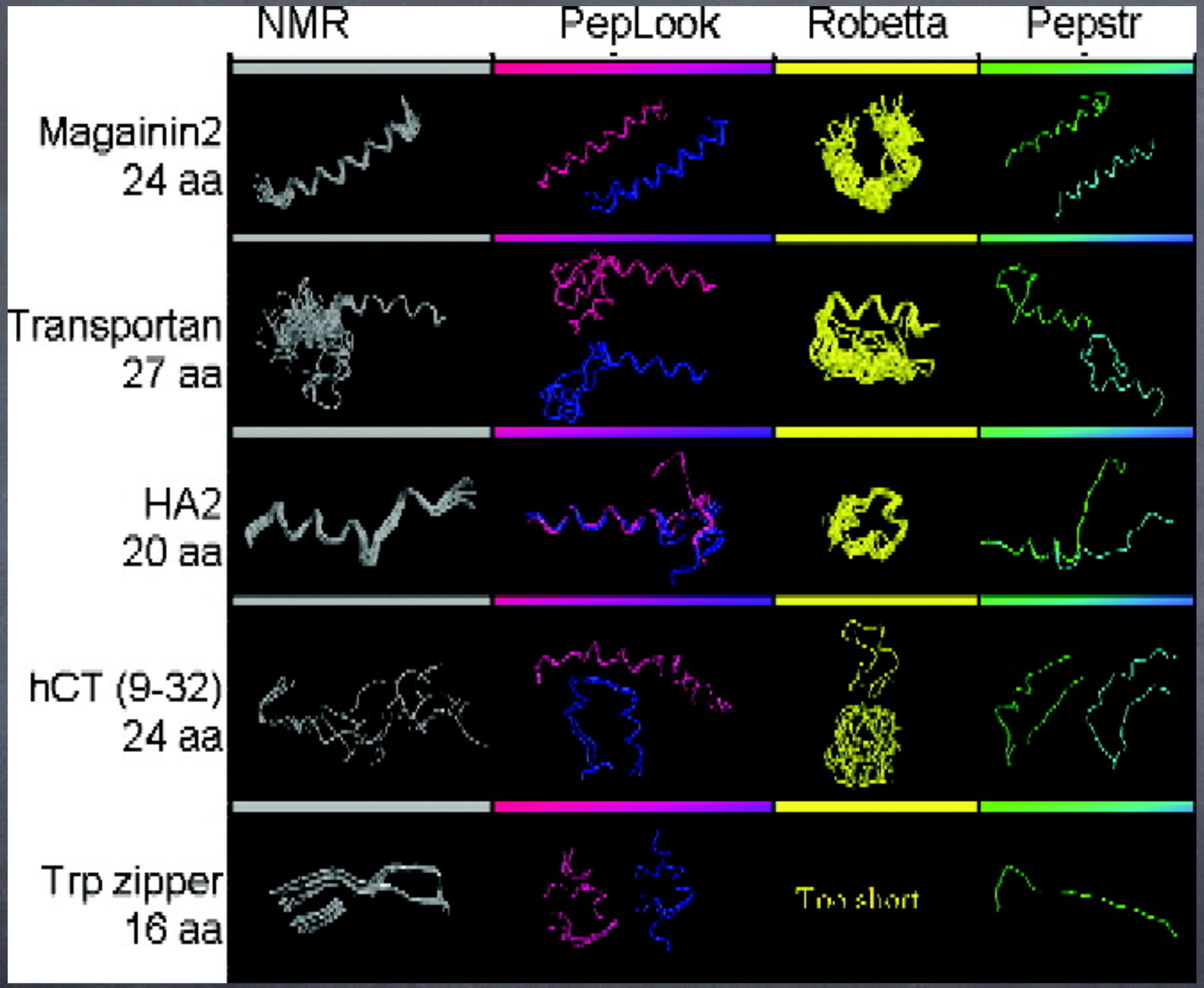
Programs for peptide structure prediction

- **Geocore** - a growing chain algorithm (Φ/Ψ choices and a sum of hydrophobic and hydrogen-bond interactions)

- **PepStr** - secondary structure and β-turn prediction with an energy MD-based refinement

- **Peplook** - Boltzmann-stochastic algorithm with 64 Φ/Ψ backbone combinations

- **Generalized pattern search (GPS)** - secondary structure prediction and an all-atom energy model



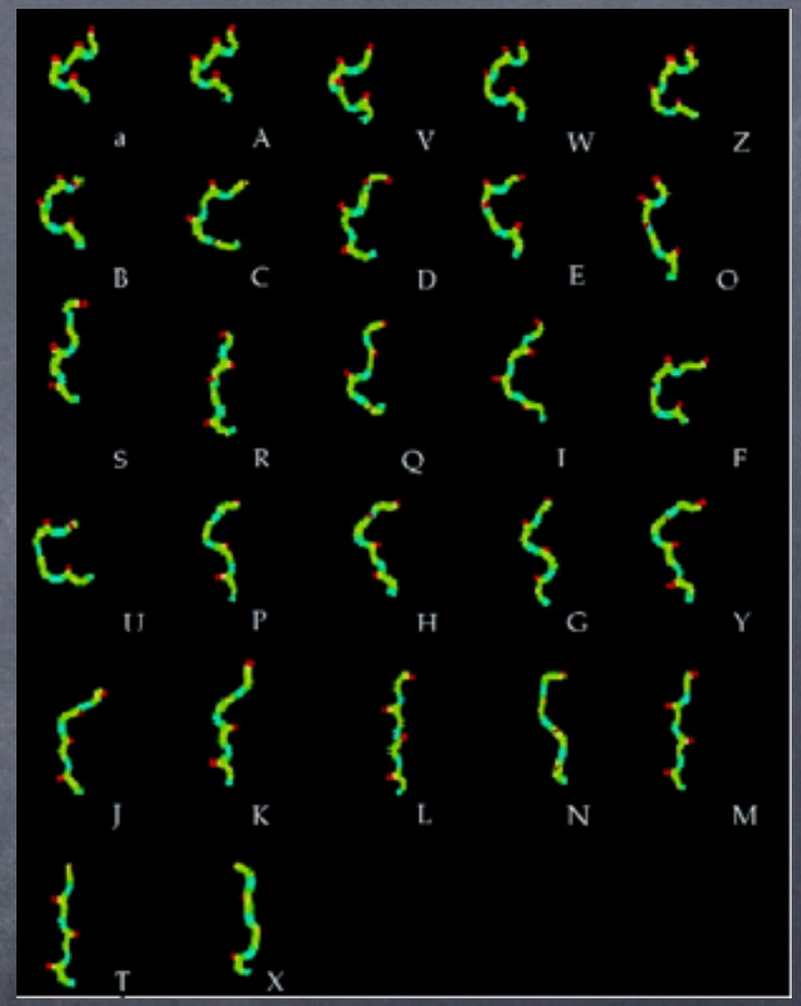
Prediction of peptide structure: How far are we?
Thomas et al., 2006

PEP-FOLD basis

- Structural alphabet (SA)
- Support vector machines (SVM)
- Forward-backward (FB) algorithm
- Stochastic greedy algorithm

Structural alphabet

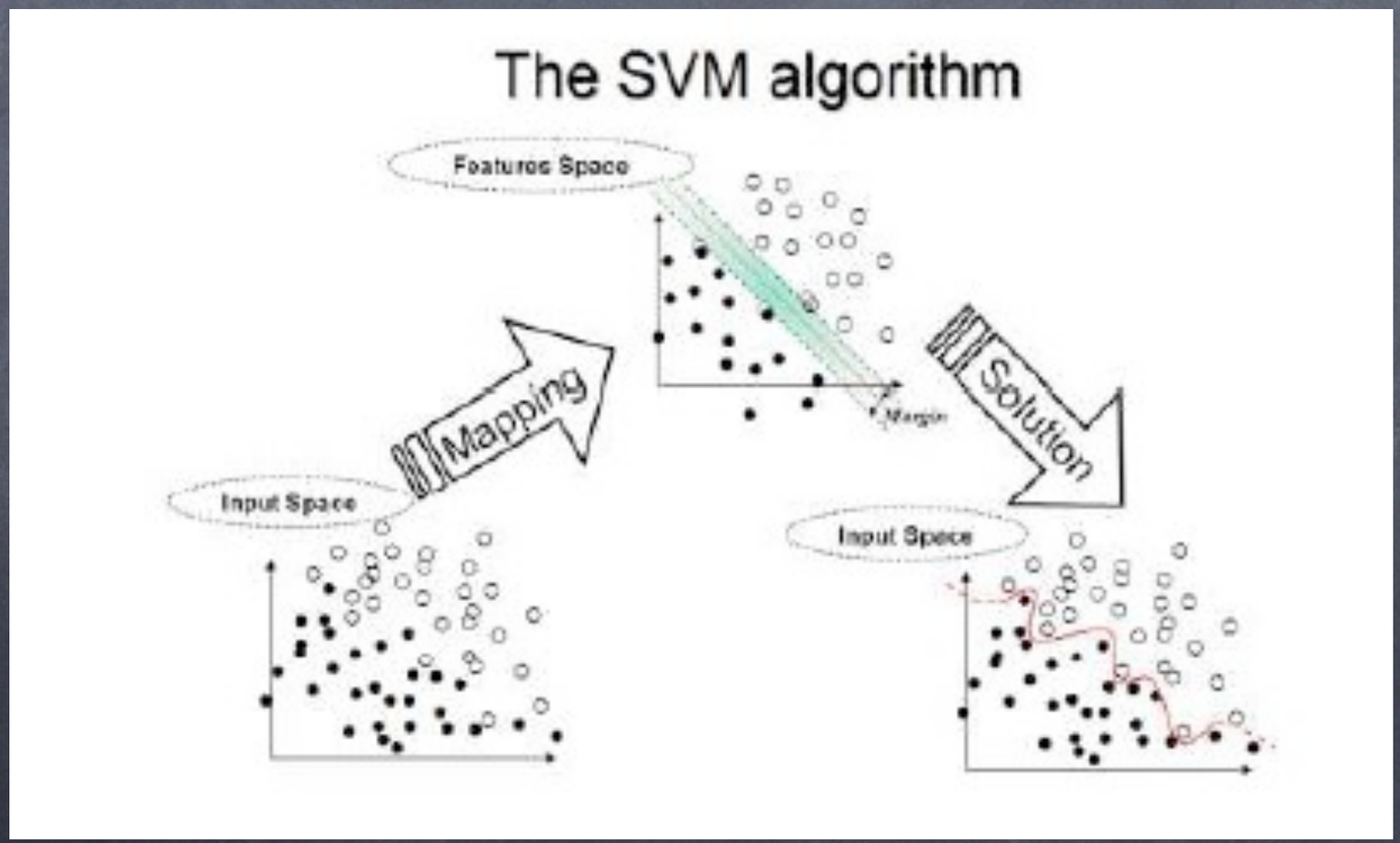
- proteins as series of overlapping four residues were used for HMM training
- 12 states SA from a dataset of 100 proteins, on the basis of a minimal representation of the states of 3%
- HMM_1 (Markovian dependence between the states) strategy resulted in 27 SA



A Hidden Markov Model Derived Structural Alphabet for Proteins
A.C Camproux, R Gautiera and P Tufféry, 2004

Support Vector Machines

- performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories



http://www.dtreg.com/svm.htm

Forward-backward algorithm

- an algorithm for computing the probability of a particular observation sequence in the context of hidden Markov models
- computes a set of forward probabilities, a set of backward probabilities and then the probability of being in each state at a specific time during the observation sequence
- finds the most likely state for a hidden Markov model at any time

Peptide Test Sets

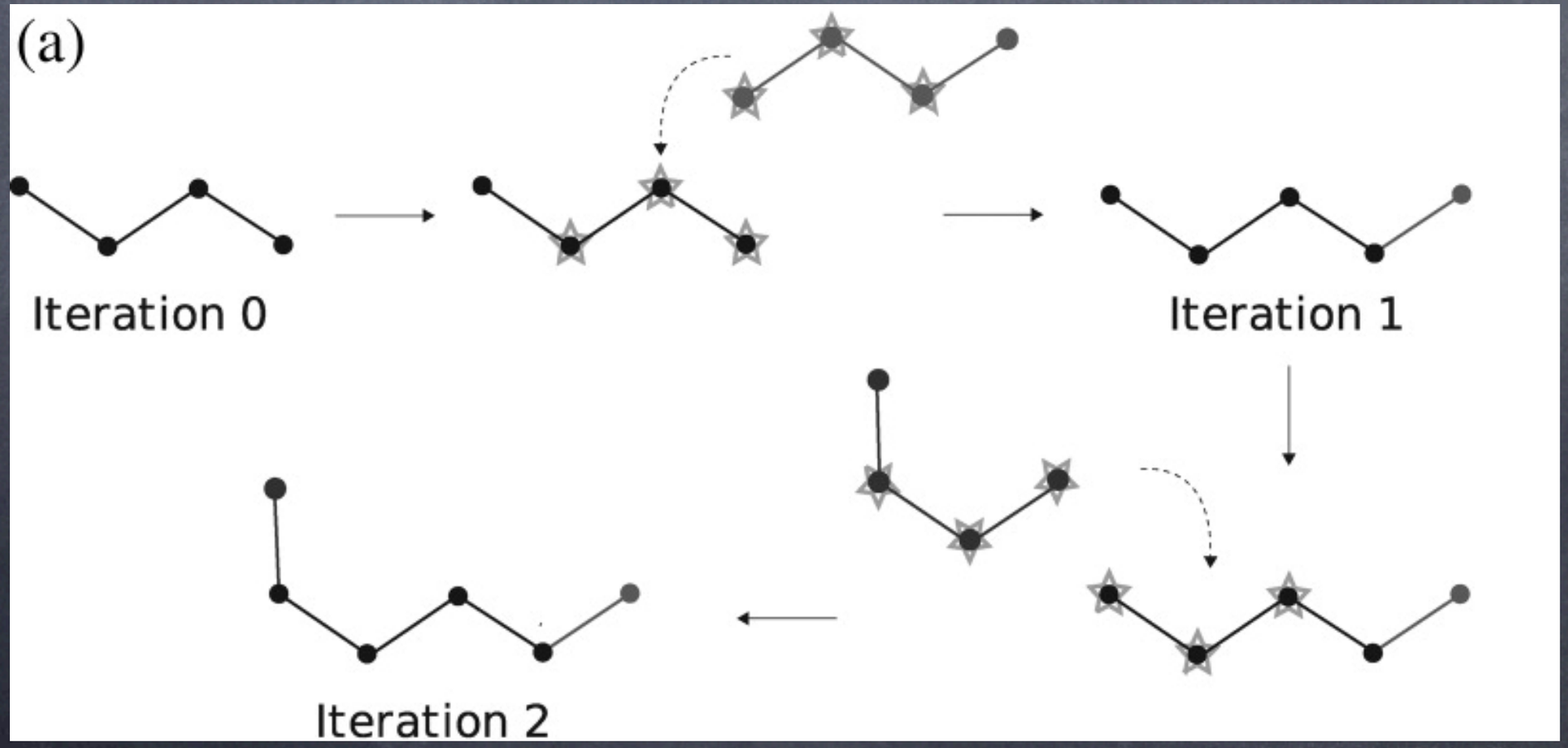
- **PepStr set** - 42 linear, bioactive peptides with 9-20 amino acids, free of S-S bridges, known structure from NMR
- **Pep-Fold set** - 23 PDB (http://www.rcsb.org/pdb) sturctures of 10-50 aa solved by NMR. 10 small peptides 10-23 aa and 13 miniproteins 27-49 aa

SA Letter Encoding from Structure and SA Letter Prediction from Sequence

- Forward-backward algorithm to encode protein structures as a series of SA letters
- SA predictors learnt from proteins and applied to peptides
- Nonredundant protein set of 3672 chains from PDB, solved by X-ray diffraction
- 860 000 4-residue fragments
- Learning set of 182 000 fragments, others used as validation set

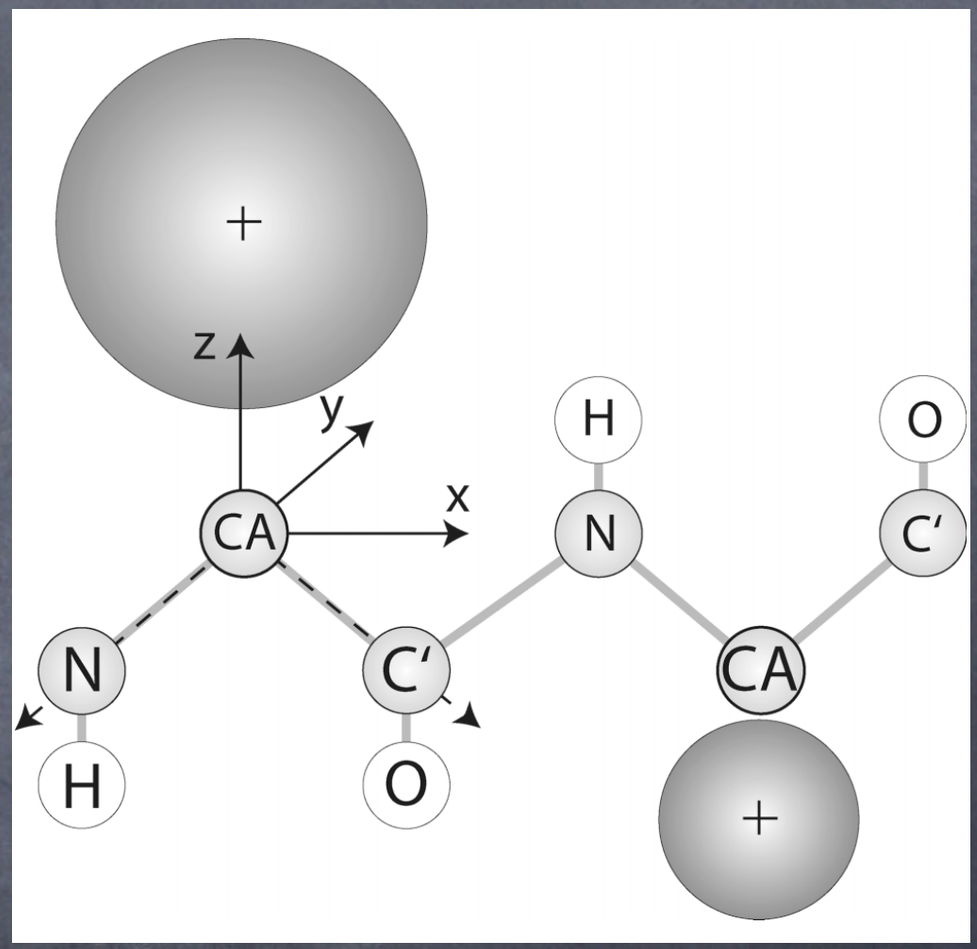
Greedy Zip operator

- Enhanced version of greedy algorithm, that starts 3D structure building from SA profile at any position of the structure, alternatively adding residues at each side of the growing structure.



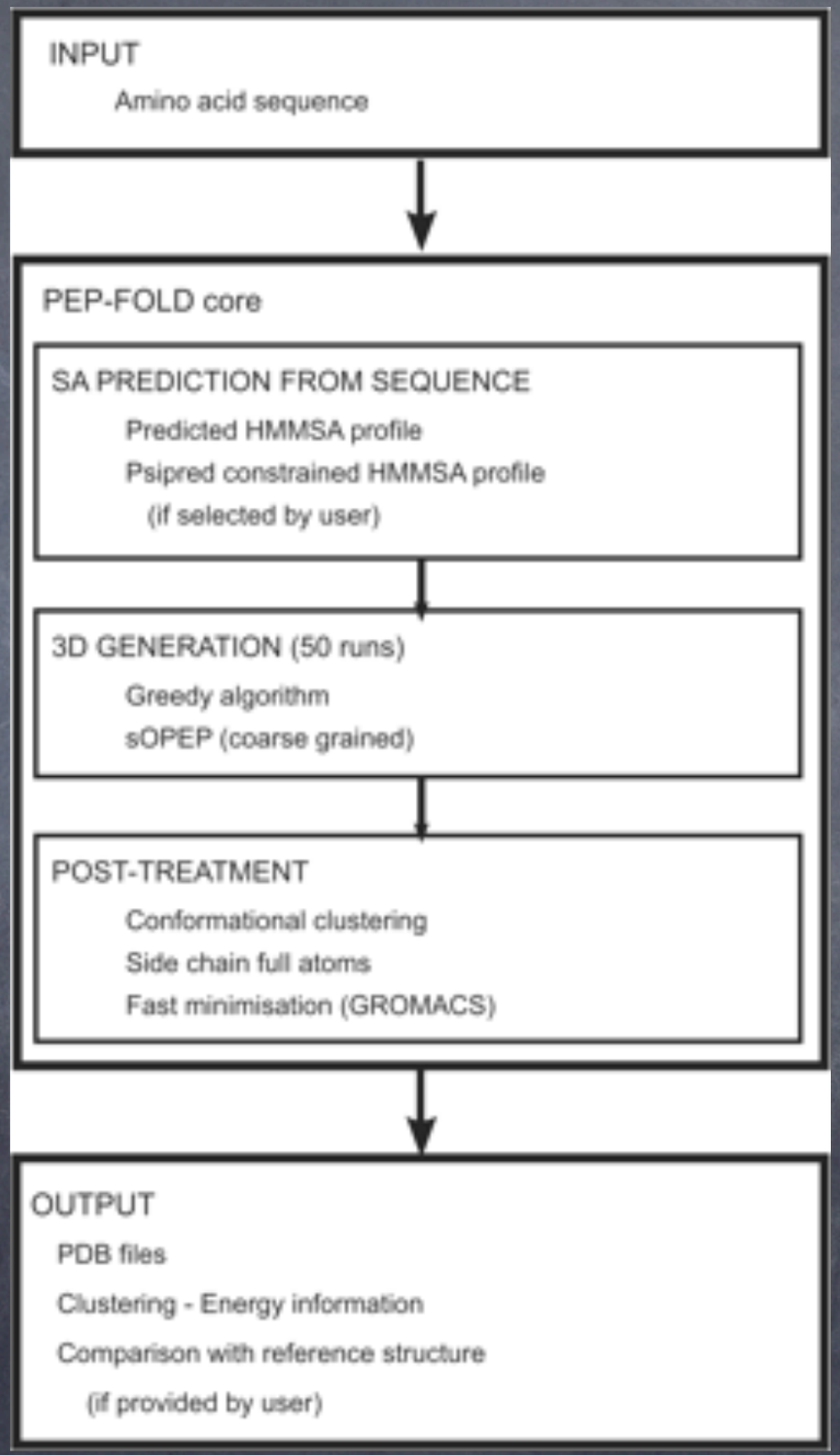
Coarse-Grained potential

- Amino acid side-chains are represented by one centroid in OPEP force field. Their positions, defined with respect to the backbone heavy atoms (N, CA, C'), and their van der Waals radii vary from one residue to another



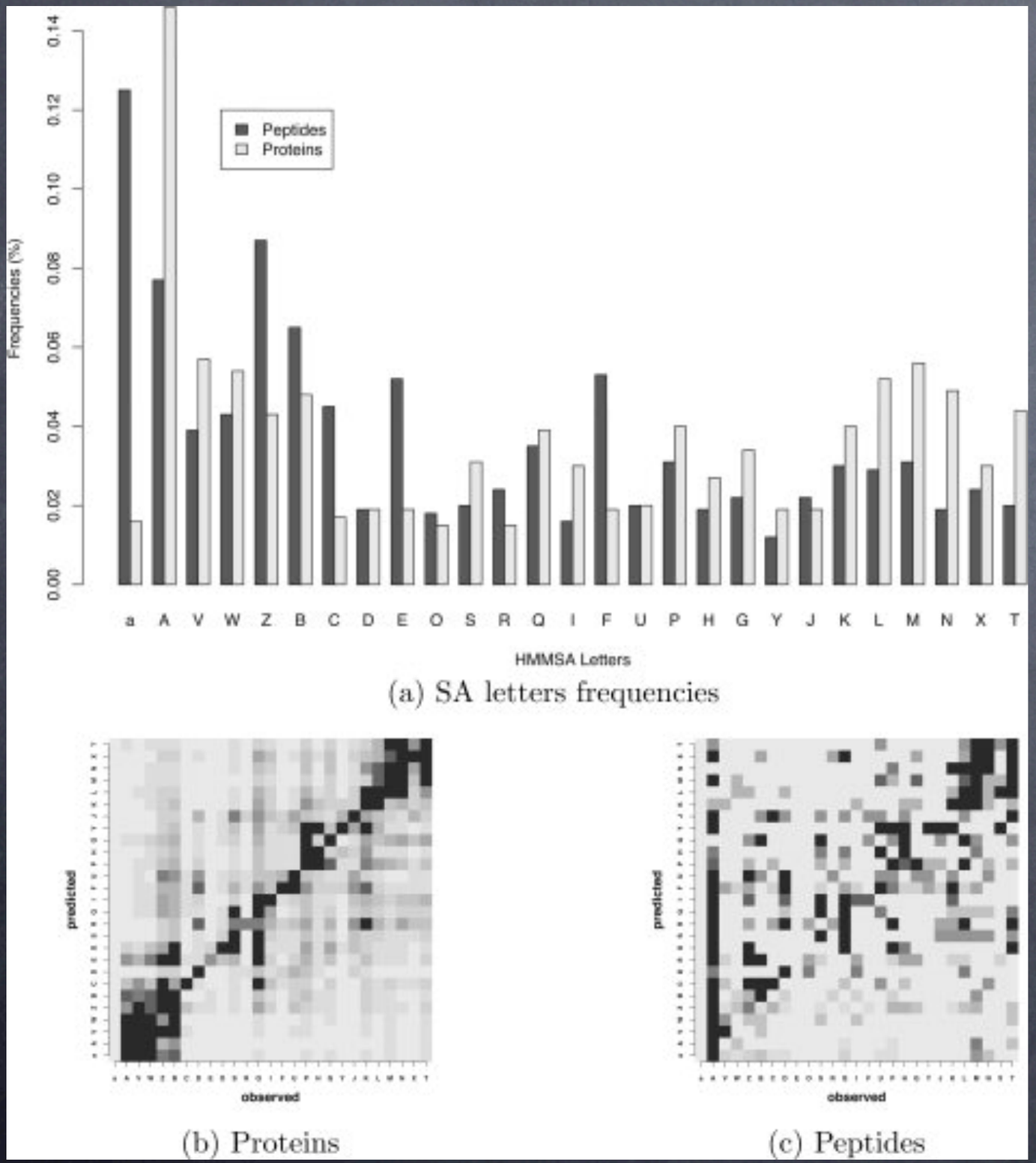
- sOPEP is a simplified OPEP version adapted to a greedy algorithm

PEP-FOLD flowchart



RESULTS

HMM-SA Letter Prediction



(a) SA letters frequencies

(b) Proteins

(c) Peptides

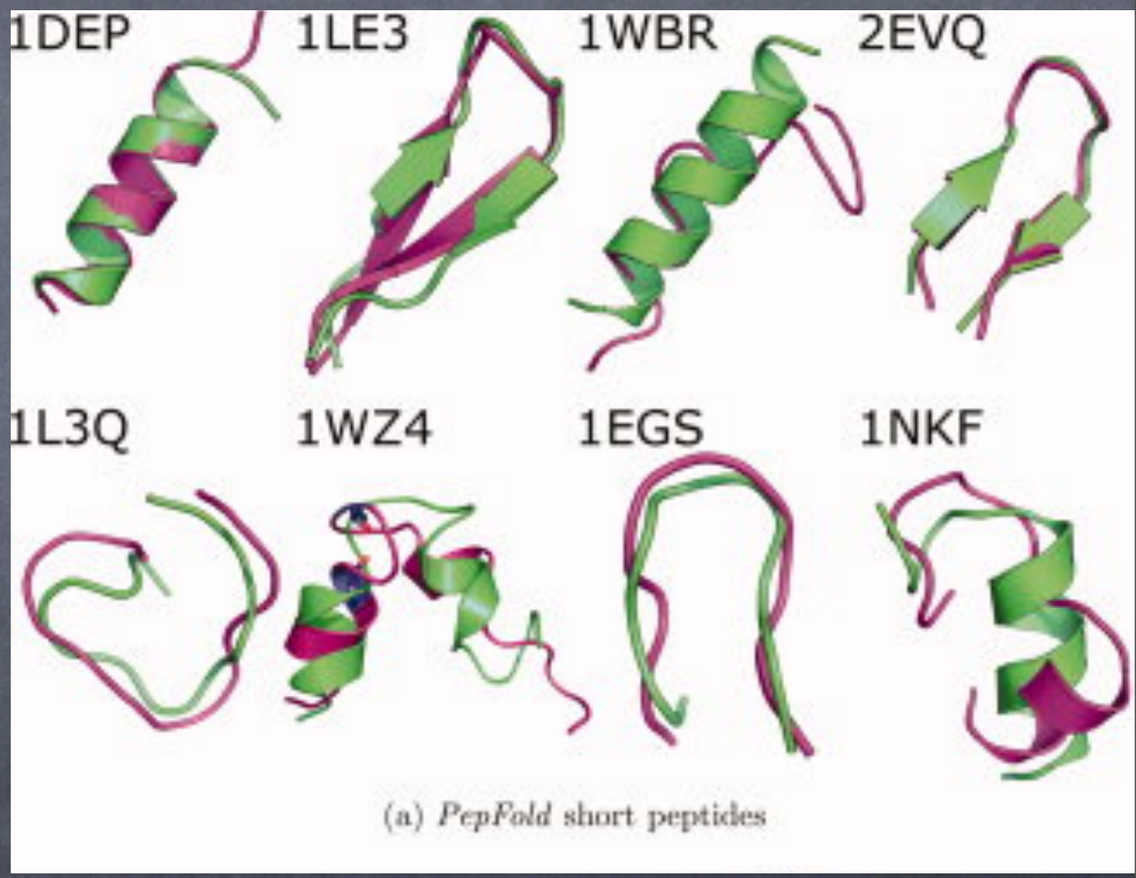
- Variation in distribution of SA letters [a,A] (α-helix cores), [Z] (α-helix extremities), [E] (3.10 helix or typeI turn), and [F] (fuzziest letter)
- Indication, that peptide conformations are less regular than protein structures

sOPEP Force Field Effectivness

- sOPEP can identify a native or native-like structure as the lowest energy conformation for 22 among 29 targets
- The reconstructions differ from experiment by 1.7 Å for short peptides and 2.9 Å for miniproteins
- Using sOPEP generates 3D models near the experimental conformations

De Novo Prediction on the PepFold set - peptides

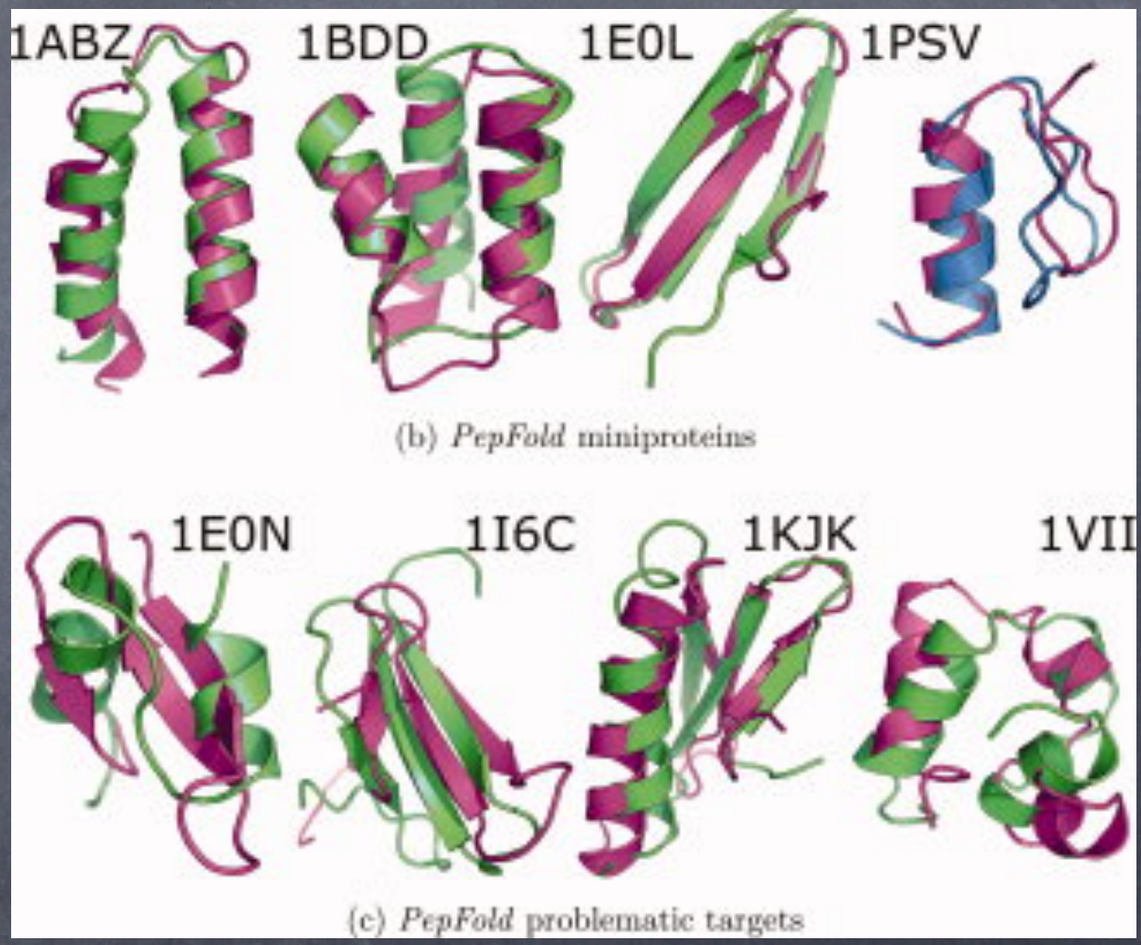
- For each target 50 independent runs launched
- Average number of clusters for peptides 5.1 (1-12), most populated and best clusters match for 5 out of 10; lowest energy and best match for 5 of 10, 4 match for all.
- PEP-FOLD generates native-like conformations for 10 short peptides

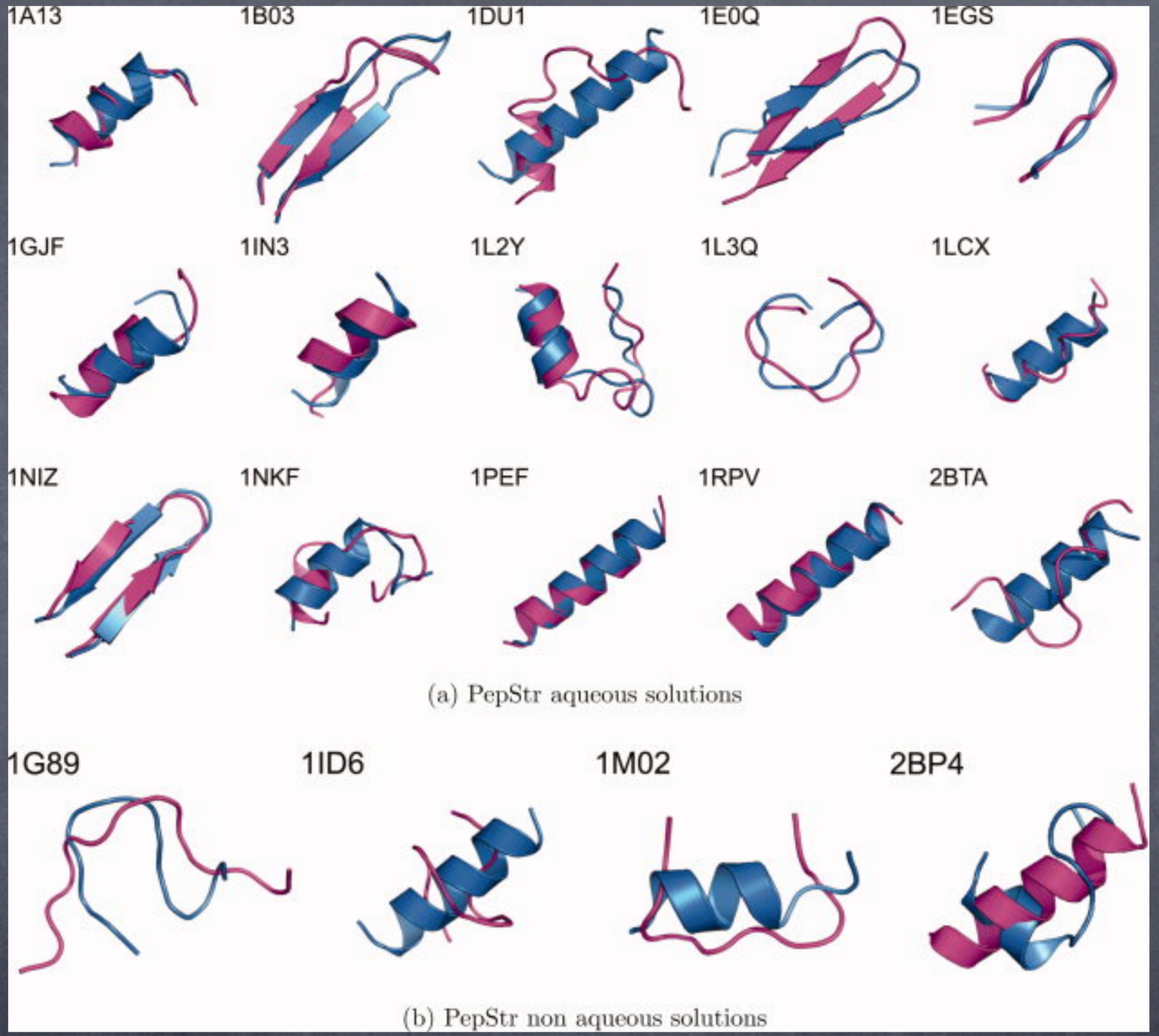


(a) *PepFold* short peptides

De Novo Prediction on the PepFold set - miniproteins

- sOPEP is not optimal for recognizing near-native from higher RMSd states
- PEP-FOLD generates near-native conformations for 9 among 13 targets

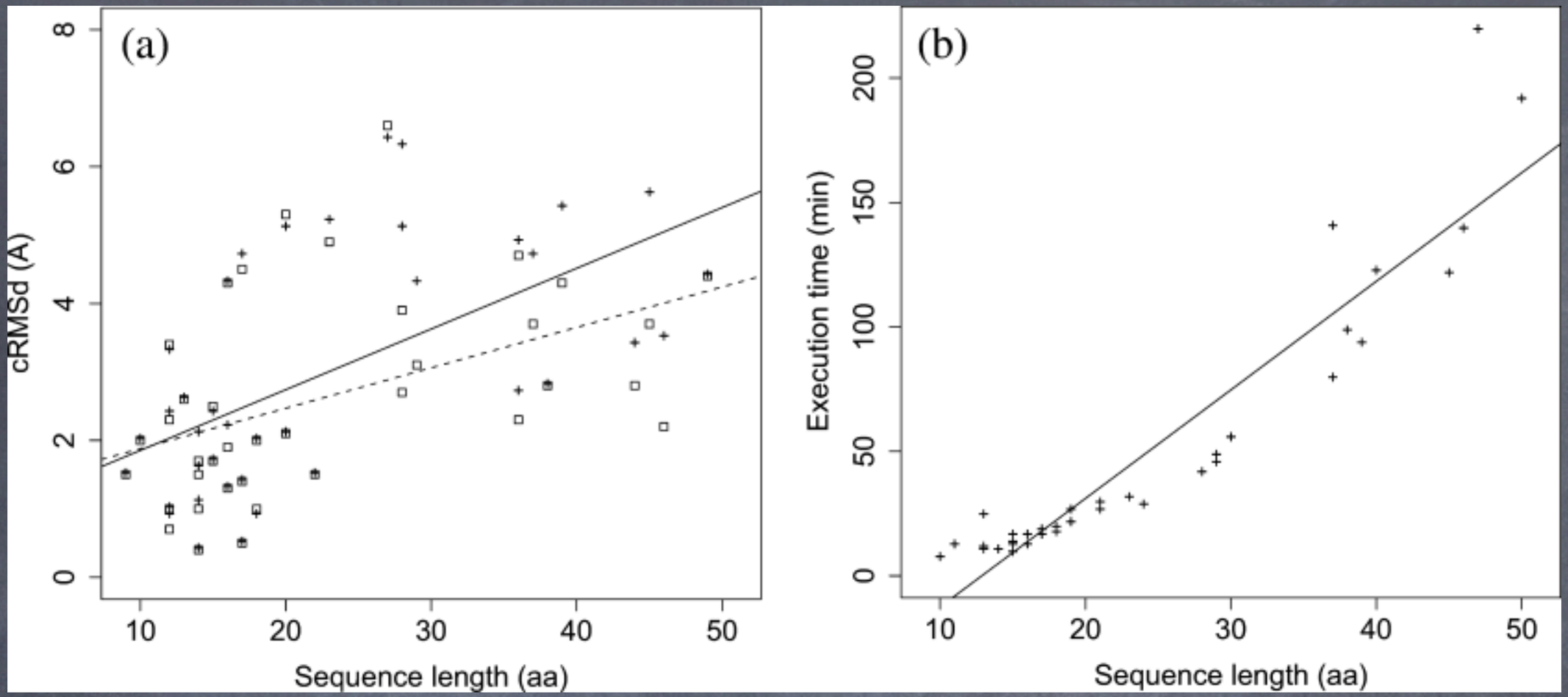




Best predicted models for PepStr set. For all targets, the native structure (magenta) is superimposed on the best predicted model (blue) for 15 peptides in aqueous solution set (a) and four problematics targets for peptides in nonaqueous solutions (b). Only the centroids of the best cluster are represented.

Discussion

- PepStr approximated the structure of PepStr set by 4 Å, GPS by 3.2 Å, PEP-FOLD by 2.7-3.0 Å
- The 8 peptides for which they could not find good enough model they find good excuses



- a) reports the RC-cRMSd of the lowest energy conformation and of the best cluster with respect to NMR structure
- b) shows execution times. Using 10 CPU Intel Xeon 2.8 GHz 50 simulations take 30 min and 90 min for 20-residue and 35-residue targets, respectively

CONCLUSIONS

- A new approach for de novo structure prediction of peptides from amino acid sequences.
- PEP-FOLD does not rely on any secondary structure information
- The speed of the algorithm and present results open the door to large-scale prediction of peptide structure and peptide engineering

Future Plans

- Option for structure prediction in non-aqueous environment
- Structure prediction of linear and cyclic peptides combining L- and D-amino acids
- Structure prediction of peptides with disulfide bridges
- Improvements for mini-protein structure prediction

PEP-FOLD server

De novo peptide structure prediction

PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino acid sequences. This method, based on structural alphabet SA letters to describe the conformations of four consecutive residues, couples the predicted series of SA letters to a greedy algorithm and a coarse-grained force field.

Access the PEP-FOLD server @ the RPBS Mobyle Portal or the PEP-FOLD standalone server.

[Restricted to peptide sizes from 9 to 25 residues]

Mobyle server | Standalone server

Please cite the following reference:

Maupetit J, Derreumaux P, Tufféry P.
PEP-FOLD: an online resource for de novo peptide structure prediction.
Nucleic Acids Res. 2009. doi:10.1093/nar/gkp323

History

2009, sep 17 - Bugfix: in some cases, secondary structure prediction constraints were not efficient.

http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD/