Ensembles of Secondary Structure Predictors for Sequence-based β-residue Pair Prediction

Kanaka Durga Kedarisetti, Scott Dick, and Lukasz Kurgan

Department of Electrical and Computer Engineering, University of Alberta, Canada T6G 2V4

INTRODUCTION

β-sheets are secondary structures that tend to increase the overall stability of a protein. Prediction of these structures is an important step in tertiary structure prediction, elucidation of folding pathways, and rational drug design [1]. As one example, research shows that formation of β-sheets is a key part of the conformational transition from the cellular (PrP^{c}) to scrapie (PrP^{Sc}) form of the prion protein, which cause various TSE diseases (BSE, CWD, Scrapie and CJD) [2]. These diseases greatly impact international trade; Canada alone has lost over $7 billion due to trade restrictions triggered after a few BSE-infected cattle were discovered [3]. Our long term research goal is to provide an accurate prediction of the location of β-sheets in PrP^{Sc}.

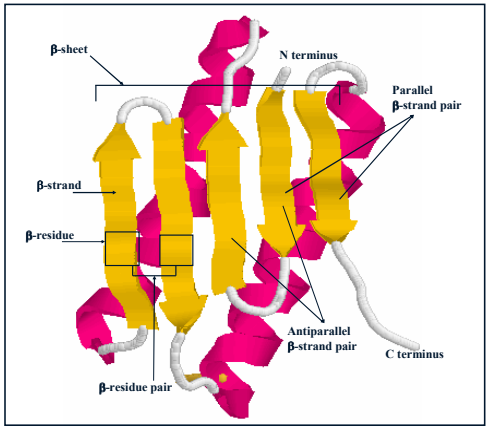

Figure 1. Topology of β-sheets

β-sheets involve the formation of backbone hydrogen bonds between adjacent β-strands; however, the amino acids involved in a bond may be far apart in the primary sequence. Figure 1 depicts the basic topology of β-sheets.

MOTIVATION

Prediction of β-sheets from the primary amino acid (AA) sequence is a challenging task due to the long range interactions that are involved in the hydrogen bonds; in addition, complementarity between the AAs in a β-residue pair is relatively weak [4]. A literature review reveals that:

I. Systematic analysis and prediction of β-sheet topology has not yet been fully explored.

II. To date, only one method, BETAPRO [8], predicts β-sheet topology from the known secondary structure (SS).

III. The long-range interactions in β-sheets have no equivalent in the other SS classes (helices and coils); this may be a reason why the β-sheet class is not predicted as accurately as the others.

IV. Research shows favorable side-chain interactions between β-residue pairs that contribute to stability of β-sheets [5]. This implies that accurate prediction of β-sheets from the primary AA sequence should be feasible.

METHOD

β-sheets are made up of a number of β-residue pairs, which are the amino acids that form hydrogen bonds. We will attack the problem of predicting which amino acids will form β-residue pairs based on the primary amino acid sequence. The current goal is to develop a sequence-based prediction of β-residue pairs that form intra-chain β-sheets, which is an intermediate step to predicting the β-sheets.

We have developed a meta-predictor that employs five modern sequence-based SS predictors (JNET, PROTEUS, PSIPRED, TRANSEC, and YASPIN) [6,7] and combines these predictions to improve the quality of β-residue prediction. We then employ the BEATPRO method to identify β-residue pairings. The overall design in shown in Figure 2.

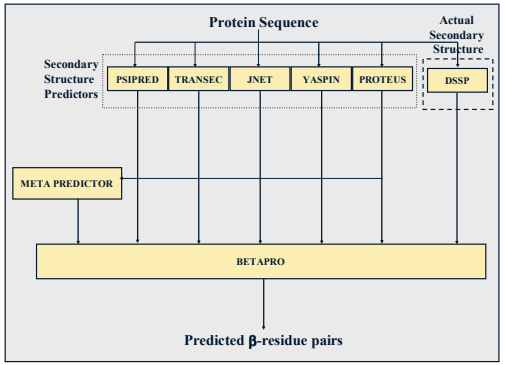

Figure 2. Design of the proposed system for prediction of β-residue pairs

We perform a two-fold cross-validation on a dataset of 872 non-redundant protein chains that were deposited in the PDB after January 2007, and are composed of at least 10% β-residues. This dataset is also non-redundant with respect to PDB entries from before January 2007, which eliminates any evaluation bias in SS prediction methods using structural templates.

RESULTS

We found that BETAPRO achieves sensitivity/specificity of 39.6/39.8% for β-residue pairs when using the actual SS derived with the DSSP. The sensitivity/specificity measures drop to between 19.2/19.3% (PSIPRED) and 7.7/8.5% (TRANSEC) when using SS predictions from individual methods.

We observed that improved BETAPRO predictions are obtained for SS predictions characterized by higher specificity for β-residues and acceptable levels of sensitivity. We developed a decision-table based meta-predictor for SS (using the five SS predictors), which obtains higher specificity (81.5% vs. 78.6%) and higher sensitivity (74% vs. 73.5%) for β-strand residues when compared with the individual SS predictors. Using these SS predictions, BETAPRO obtains improved sensitivity (19.4% vs. 19.2%) and specificity (19.5% vs. 19.3%) for prediction of β-residue pairs compared to using the PSIPRED predictions. Detailed results are given in Tables 1 and 2.

Table 1. Sensitivity and specificity of β-residue predictions

SS Method	Sensitivity	Specificity
PROTEUS	81.22	72.05
YASPIN	72.80	69.19
PSIPRED	73.45	78.59
TRANSEC	40.85	63.02
JNET	59.15	79.03
META PREDICTOR	**73.95**	**81.85**

Table 2: Sensitivity and specificity of β-residue pair prediction

SS Method	Sensitivity	Specificity
DSSP	39.59	39.78
PROTEUS	17.77	17.89
YASPIN	14.53	14.58
PSIPRED	19.20	19.32
TRANSEC	7.70	8.48
JNET	15.90	16.08
META PREDICTOR	**19.37**	**19.51**

CONCLUSIONS

We proposed a novel sequence-based method to improve β-residue pair prediction using a consensus of SS predictions to predict β-residues. Our study constitutes a step toward designing a predictor that provides accurate prediction of β-residue pairs, which in turn will contribute to prediction of intra-chain β-sheets.

In future work the proposed method will be extended to consider information about local and long distance interactions that affect β-residue pair prediction. We will also apply this method to predicting the (unknown) SS of the PrP^{Sc} protein. This method may also be useful in predicting external β-strand pairs and β-contacts.

ACKNOWLEDGEMENTS & REFERENCES

The authors would like to thank ISMB 2008 for the opportunity to present these findings, and the authors of the BETAPRO method for providing their software. We gratefully acknowledge financial support from NSERC Canada, Alberta Ingenuity, and ISBM for providing the travel fellowship.

[1] Rost B, Liu J, Przybylski D, Nair R, Wrzeszczynski KO, Bigelow H, and Ofran Y, Prediction of protein structure through evolution. In Gasteiger J and Engel T, (eds), *Handbook of Chemoinformatics - From Data to Knowledge*, 1789–1811, 2003.

[2] Stöhr J, Weinmann N, Wille H, Kaimann T, Nagel-Steger L, Birkmann E, Panza G, Prusiner SB, Eigen M, and Riesner D (2008), Mechanisms of prion protein assembly into amyloid, *Proc Natl Acad Sci U S A.* 2008, 105(7):2409–14.

[3] Forge F and Fréchette JD, Mad Cow Disease And Canada's Cattle Industry, http://www.parl.gc.ca/information/library/prbpubs/prb0301-e.htm, 2008.

[4] Mandel-Gutfreund Y, Zaremba SM, and Gregoret LM, Contributions of residue pairing to beta-sheet formation: conservation and covariation of amino acid residue pairs on antiparallel beta-strands, *J. Mol. Biol.* 2001, 305:1145–59.

[5] Smith CK and Regan L, Guidelines for protein design: The energetics of β sheet side chain interations, *Science* 1995, 270:980–2.

[6] Montgomerie S, Sundararaj S, Gallin WJ, and Wishart DS, Improving the accuracy of protein secondary structure prediction using structural alignment, *BMC Bioinformatics* 2006, 7:301.

[7] Lin K, Simossis VA, Taylor WR, and Heringa J, A Simple and Fast Secondary Structure Prediction Algorithm using Hidden Neural Networks. *Bioinformatics* 2005, 21(2):152–9.

[8] Cheng J and Baldi P, Three–stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms, *Bioinformatics* 2005, Suppl 1:175-84.