



S2-2

Predicting Contact-dependent Secondary Structure Propensity (CSSP); relevance to amyloidogenic sequences

Sukjoon Yoon¹, William J. Welsh² and Heeyoung Jung¹

¹Department of Biological Sciences, Research Center for Women's Diseases (RCWD), Sookmyung Women's University, Seoul 140-742, Korea, ²Department of Pharmacology, Robert Wood Johnson Medical School and the Informatics Institute of UMDNJ, Piscataway, NJ 08854, USA

The preponderance of evidence implicates protein misfolding in many unrelated human diseases. In all cases, normal correctly folded proteins transform from their proper native structure into an abnormal β -rich structure known as amyloid fibril. We have previously demonstrated that calculation of contact-dependent secondary structure propensity (CSSP) was highly sensitive in detecting non-native β -strand propensities in the core sequences of known amyloidogenic proteins. Here we describe an improved CSSP method based on dual artificial neural networks that rapidly and accurately estimate the potential for the non-native secondary structure formation in local regions of protein sequences. Long-range interaction patterns in diverse secondary structures were quantified by potential energy calculations and decomposition on a pairwise per-residue basis. The calculated energy parameters and 7-residue sequence information were used as inputs for artificial neural networks (ANNs) to predict sequence potential for secondary structure conversion. The trained single ANN using the $>(i, i\pm 4)$ interaction energy parameter exhibited 74% accuracy in predicting the secondary structure of test sequences in their native energy state, while the dual ANN-based predictor using $(i, i\pm 4)$ and $>(i, i\pm 4)$ interaction energies showed 83% prediction accuracy. Analysis of 1,692 non-homologous protein domains using the dual network reveals that 13% of helical residues had higher beta propensity than the average beta propensity in natively beta residues. Furthermore, we identified a large number of amyloidogenic subsequence patterns using the dual ANN-based CSSP method. The present method provides a simple and accurate tool for predicting sequence potential for secondary structure conversions without using 3D structural information. It may find utility in many medically relevant applications, such as the engineering of protein sequences and the discovery of therapeutic agents that specifically target these sequences for the prevention and treatment of amyloid diseases. A user friendly web-interface (<http://cssp2.sookmyung.ac.kr>) was constructed to provide the present energy-based CSSP methods.

