Modeling of Protein Structure by Hidden Markov Models

Kiyoshi Asai[1] asai@etl.go.jp
Kentaro Onizuka[2] onizuka@mrit.mei.co.jp
Masayuki Akahoshi[3] akahoshi@icot.or.jp
Hidetoshi Tanaka[3] htanaka@icot.or.jp
Katunobu Itou[1] kito@etl.go.jp

Electrotechnical Laboratory (ETL), 1-1-4 Umezono, Tsukuba, 305 Japan
Matsushita Research Institute, 3-10-1 Higashimita, Tama-ku, Kawasaki, 214 Japan
ICOT, Mita Kokusai Bldg. 21F, 1-4-28 Mita, Minato-ku, Tokyo, 108 Japan

Prediction of protein structures from amino-acid sequences is one of the most important, but difficult problem in computational biology. The authors have been working on the expressions of protein structures because we need expressions of the targets in order to predict them. Now the authors are trying to classify the protein structures by the expression, Multi Scale Structure Description (MSSD). In this paper, labeling of local structures and classification of local structure packings are performed by Hidden Markov Models (HMMs) using MSSD.

HMMs have been used for protein researches for local structure prediction [Asai91, Asai93A, Asai93B], for sequence alignment [Haussler93], for protein classification [Tanaka93], and for motif extraction [Fujiwara94]. Most of them used 20 amino acids as the discrete output symbols of distributions in HMMs. In this paper, however, HMMs have continuous output distributions of MSSD-parameters of protein structures. MSSD is a robust and residue length free parameterization of protein structures using 3D coordinates of α carbons [Onizuka94]. For the network shape determination, iterative duplication method [Fujiwara94] and successive state splitting (SSS) algorithm [Tanaka93] have been used for protein HMMs. We used modified iterative duplication method, where negligible links are deleted and states of the largest output variances are duplicated.

For the purpose of local structure clustering and the automatic labeling of the structures, the parameters of 5-residue MSSD are used. Both continuous structures and short ranged structures are categorized naturally as the hidden states in HMMs, which nearly correspond to the secondary structures, "α helix" and "β strand" and many types of "turns" and "coils". After the automatic labeling, the packings of the local structures are detected and parameterized by MSSD, which can treat the structures of variable lengths. The HMM of the MSSD parameters of structure packings are constructed and trained just as same as local structure HMM.

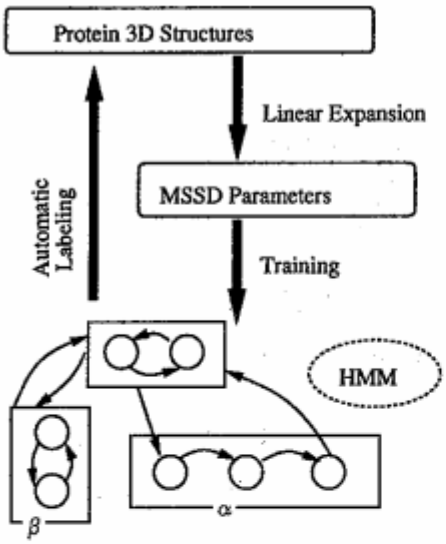

Fig.1: Labeling of Structures by HMMs

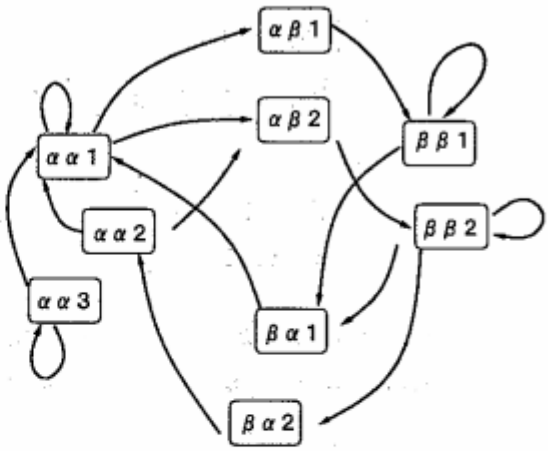

Fig.2: Transitions of Packing Patterns

By estimation of the hidden state transitions using Viterbi algorithm, the protein structures are aligned to the HMMs. The labeling of the local structures and the structure packings are easy translations of this alignment. At the same time, the HMMs extract the rules between the local structures and between the structure packings as the matrices of the transition probabilities. These rules are important for modeling the protein structures including the higher level structures. We need these models and rules for the structure prediction of protein.

References

[Asai91] Asai,K.; Hayamizu,S. and Handa,K.: "Genetic Information Processing by Stochastic Model: HMM for Secondary Structure Prediction of Protein," *Proc. Genome Info. Workshop 2, pp.144-147. (in Japanese, 1991).*

[Asai93A] Asai,K.; Hayamizu,S. and Handa,K.: "Secondary Structure Prediction by Hidden Markov Model," *CABIOS Vol.9, No.2, pp.141-146, 1993.*

[Asai93B] Asai,K.; Hayamizu,S. and Onizuka,K.: "HMM with Protein Structure Grammar," *Proc. HICSS26, vol.1, pp.783-791.(1993).*

[Fujiwara94] Fujiwara,Y.; Asogawa,M. and Konagaya,A.: "Stochastic Motif Extraction Using Hidden Markov Model," *Proc. ISMB94, pp.121-129.(1994).*

[Haussler93] Haussler,D.;Krogh,A.;Mian,I.S. and Sjölander,K.: "Protein modeling using hidden Markov Models: Analysis of globins," *Proc. HICSS26, vol.1, pp.792-802. (1993).*

[Onizuka94] Onizuka,K.; Akahoshi,M.; Ishikawa,M. and Asai,K.: "The Multi-Scale 3D-1D Compatibility Scoring for Inverse Protein Folding Problem," *Proc. of ISMB-94 pp.314-321. (1994).*

[Tanaka93] Tanaka,H.; Onizuka,K. and Asai,K.: "Classification of Proteins via Successive State Splitting of Hidden Markov Network," *Proc. W26 in IJCAI93. (1993).*