Reverse Evolution of Proteases Studied by a Modified 3D-1D Method

Haretsugu HISHIGAKI gaki@cc.tuat.ac.jp
Tamio YASUKAWA yasukawa@cc.tuat.ac.jp

Faculty of Technology, Tokyo University of Agriculture and Technology
2-24-16 Naka-cho Koganei-city, Tokyo 184, Japan

Abstract

Molecular evolution processes of cysteine and serine proteases were followed in a reversed way by the combined use of the inverted Dayhoff matrix and the 3D-1D method.

1 Introduction

Brenner[1] suggest the possibility that all serine proteases, except subtilisin family, originate from homologous precursor cysteine proteases as a result of single base mutation of cysteine codons TGY (Y=T or C) to serine codons TCN or AGY (N=A,T,G or C). This point mutation might be accompanied with conformational changes, as well as appreciable changes in specificities of enzymatic activities. On the other hand, it is widely observed that folding patterns of proteins tend to be conserved so far as their specific activities are retained. We therefore suppose that serine proteases formed by this mutation conserve their specific folding patterns even after evolutions accompanied with large scale amino acid mutations.

Recently, Eisenberg's group[2][3] proposed the 3D-1D method to design polypeptide chains having the designated conformation. We have attempted to modify this method to follow the evolutionary processes of proteases in a reversed way, as well as the conformational transition between cysteine and serine proteases.

2 Method

2.1 Determination of the reference environment class sequences for serine and cysteine proteases

For the inverse protein folding, Bowie et al.[3] assigned the environmental class to each site of the protein with a given conformation.The compatibility of a certain amino acid sequence with this designated conformation can be estimated by the profile score.

For the present purpose, we determined the reference environment class sequence for both serine and cysteine proteases: namely, employing the crystal structure data for 11 serine proteases with 223-238 residues, environment class sequences were determined and, then, all 11 amino acid sequences were individually fitted to them and profile scores were calculated. By this procedure,the environment class sequences of tonin was chosen as a reference since it gave the least mean square deviations for other 10 serine proteases. As for cysteine protease, we adopted that of papain as a reference.

菱垣 晴次，安川 民男：東京農工大学工学部，〒184 東京都小金井市中町 2-24-16

2.2 Introduction of reverse mutations into amino acid sequence

As starting amino acid sequences too, we adopted tonin and papain respectively. Employing random numbers, ten sites were chosen in each trial cycle and, for respective sites, the choice of "deletion", "insertion" and "mutation" was made by another random number with tentative probabilities of 0.40, 0.40 and 0.20, respectively. When "insertion" was chosen,one amino acid residue was inserted before the chosen site with the probability of average occurrence of amino acid residues in proteins. When "mutation" was chosen, the chosen residue was mutated with the probability given by the inverted Dayhoff matrix; i.e., P(i,j)→P(j,i).

The choice of 10 sites was not homogeneously random: i.e., serine in the active site was out of the choice and the fit probability of secondary structure regions were reduced by a factor of $1/4 \sim 1/8$ following the structure dependent mutation probabilities found for serine proteases[4].

2.3 Structures of evolving sequences

The mutated amino acid residue sequence obtained by the above procedure was fitted by dynamic programming to the reference environment class sequences for serine and cysteine proteases and both profile scores were plotted against the number of trial cycle.

In the course of trial, the score deviates frequently from the initial value, which indicates that the 3 dimensional structure of newly generated sequence assumes folding pattern different from the reference system.

When the score obtained by serine protease environment class deviates largely for consecutive ten trials and, further, the score by cysteine protease environment class does not approach the value acceptable as cysteine protease,mutation trial was stopped and returned to the point where sequence assumed the normal serine fold pattern, and new trial of mutation was resumed.

3 Results

Profile score value oscillated around the initial values and the score for serine protease environment class sequence was frequently reduced by more than 20%, but simultaneous increase of the score for cysteine protease environment class was rarely observed,indicating that the definite transitions from typical serine protease fold to cysteine protease fold was not so easy to reproduce.More refined algorithm to introduce correlated mutations are now on progress and will be presented at workshop.

References

[1] S.Brenner, *Nature*, 334, 528 (1988).

[2] M.Gribskov, A.D.McLachlan and D.Eisenberg, *Pro. NAS. USA*, 84, 4355, (1987).

[3] J.U.Bowie, R.Luthy and D.Eisenberg, *SCIENCE*, 253, 164, (1991).

[4] T.Fukuhara, H.Hishigaki and T.Yasukawa, *Genome Workshop 3*: 171, (1992).