Mapping active sites of blood coagulation serine proteases—activated protein C and thrombin—on simple graphics models

Kazunori Toma

Computer Science Department, Asahi Chemical Industry Co., Ltd., Fuji, Shizuoka, Japan

Koji Suzuki

Department of Laboratory Medicine, Mie University School of Medicine, Edobashi, Tsu, Mie, Japan

Employing the known three-dimensional (3D) structure of trypsin, we constructed simple graphics models of human-activated protein C and thrombin catalytic domains. Considering the structural analysis of bovine trypsin and pancreatic trypsin inhibitor complex, the difference of active-site amino acid sequences of human protein C inhibitor and antithrombin III and their inhibitory selectivity toward activated protein C and thrombin, we estimated the enzymatic subsites of activated protein C and thrombin and mapped them on the graphics models. Predicted favorable contacts can explain substrate selectivity of the enzymes. In this study, we used two types of modified ALPHA representations extensively. Since almost no report on the 3D structure of a blood coagulation factor has appeared and even an extensive molecular mechanics or dynamics calculation cannot produce satisfying results, simple graphics representation has several advantages.

Keywords: computer graphics, α-carbon model, blood coagulation factor, enzymatic subsite, protein C, thrombin, protein C inhibitor, antithrombin III

INTRODUCTION

Blood coagulation involves a complicated combination of proteolytic reactions. Several blood coagulation factors are classified into the serine protease superfamily, since they show considerable amino acid sequence homology with trypsin, chymotrypsin, etc.[1] The differences in their catalytic selectivity toward substrates and inhibitors mainly control stepwise reactions of the blood coagulation cascade.

Although no report has appeared concerning the three-dimensional (3D) structure elucidation of a blood coagulation factor except that of prothrombin fragment-1,[2,3] 3D structures of the proteases would be helpful to understand the blood coagulation mechanism in molecular detail. It is, however, possible to make a model structure from known 3D structures of superfamily proteins.[4] Using known 3D structures of serine protease superfamily members, Greer[5,6] and Furie *et al.*[7] had constructed graphics models of blood coagulation factors and discussed some important factors that might determine their selectivity toward substrates. Their methods are widely accepted as standard strategies for model building of homologous proteins, but they are rather complicated and computationally time consuming, if the conclusion they draw is considered. As a result of a model-building study of human protein C inhibitor (PCI), we made an all-atom model of the interaction site of human-activated protein C (APC) and PCI and discussed some potentially important residues for specific interactions.[8]

With any model-building methods presently known, it may be difficult to predict precise atomic positions and, therefore, atomic interactions of proteins. However, even simple graphics representation that displays each amino acid residue as a sphere is considered to be enough

Address reprint requests to Kazunori Toma at the Computer Science Department, Asahi Chemical Industry Co., Ltd., 2-1 Samejima, Fuji, Shizuoka 416, Japan.
Received 10 January 1989; accepted 14 February 1989

to deduce some important subsite residues that may determine the substrate selectivity of enzymes.

In this paper, we examine the difference of substrate selectivity between APC and thrombin to investigate such a possibility. It must be an interesting example also from a biochemical point of view, since APC and thrombin are two of the important enzymes in the blood coagulation system. Their selectivity is distinctive in their behavior toward antithrombin III (ATIII), which is a specific inhibitor of thrombin but not so against APC.[9] Their α-carbon models were constructed from the structure of trypsin, and their enzymatic subsites (S3–S3′) were mapped on the models by considering active-site sequences of PCI and ATIII.

METHODS

The amino acid sequences of human APC heavy chain and human thrombin heavy chain were taken from the protein sequence database of Protein Identification Resource.[10] The reactive-site amino acid sequences of PCI and ATIII were taken from the literature.[9] Both sequence and 3D coordinates of bovine trypsin were taken from the data of its complex with bovine pancreatic trypsin inhibitor (BPTI) in the Protein Data Bank (entry name 2PTC).[11,12] Sequences were manually aligned with reference to alignments reported by Greer[5] and Furie *et al.*,[7] special attention being paid to the structurally conserved regions. Chymotrypsinogen numbering[12] is adopted throughout the following discussion.

Graphics models were made on the raster display of the Evans and Sutherland Picture System 340 using modified versions of ALPHA[13] and PIMIG.[14] Two additional visual dimensions were introduced to the amino acid information color-coded CPK-like representation in which each sphere corresponded to each amino acid residue.

One of the additional visual dimensions was sphere size. In this study, each sphere size was set to represent the relative approximate bulkiness of an amino acid side chain. Twenty amino acids were classified into four groups from small to large: Ala, Cys, Gly, Pro, Ser, Thr and Val; Asn, Asp, Ile and Leu; Gln, Glu, His and Met; Arg, Lys, Phe, Trp and Tyr.

The other visual dimension was simultaneous projection. In this study, it was used to compare the same characteristics of different proteins on the same picture. This representation could also be used to compare different characteristics of one protein. In the color plates, Arg, Asn, Asp, Gln, Glu, His, Lys, Ser and Thr were considered hydrophilic and shown in blue; Ile, Leu, Met, Phe, Trp, Tyr and Val were considered hydrophobic and shown in red; and others were considered neutral and shown in white. Arg, His and Lys were categorized as basic amino acids and shown in blue; Asp and Glu were acidic and shown in red; and other amino acids were neutral and shown in white.

RESULTS AND DISCUSSION

Amino acid sequences of trypsin, APC heavy chain and thrombin heavy chain are aligned as shown in Figure 1. Based on trypsin, about 30% of residues are identical among the three proteins. Important residues for catalytic reaction, such as His57, Asp102 and Ser195 (shown by triangles in Figure 1), are strictly conserved among them. Three disulfide bonds (42–58, 168–182, 191–220) are also reserved. It is noteworthy that these three disulfide bonds have smaller residue intervals among the original six disulfide bonds of trypsin. One remaining cysteine of APC or thrombin, which corresponds to Ser122 of trypsin and is marked with an asterisk in Figure 1, is linked to one of the cysteine residues of the light chain. Five insertions are necessary for APC and eight for thrombin based on trypsin. In the following discussion, they are termed (a) through (h), as shown in Figure 1.

The most obvious feature of the model structure is that the active center and Cys122 exist on the opposite side of the molecule and represent two faces of the proteins. Hence, color plates are shown in two directions (i.e., the active center side or front side and the domain interface side or backside) in the following discussion.

Insertion positions corresponding to the alignment result are displayed on the model structure with various colors in Color Plate 1. Positions are shown by both sides of the insertions, which are indicated with vertical bars in Figure 1. In the case of insertion (e), three other residues are also colored. All insertions except (d) exist on the surface of the molecule, and even (d) is close to the molecular surface.

Since insertions (a), (b) and (d) exist around the active center, they may affect the substrate selectivity. Being remote from the active center, other insertions may not influence the substrate selectivity directly. Insertions (e)

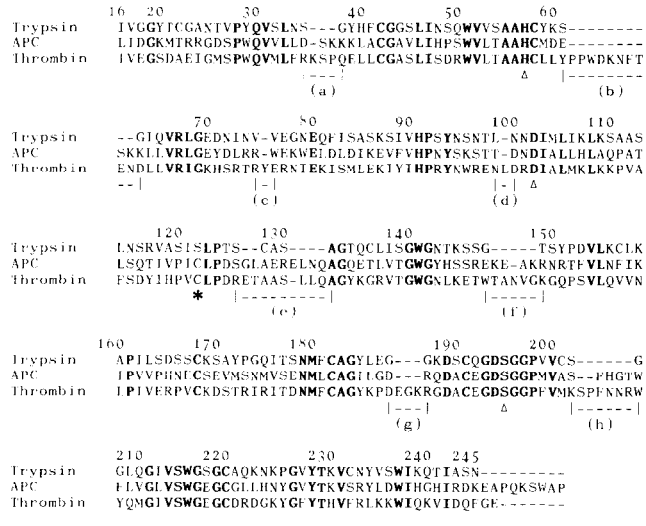

Figure 1. Amino acid sequence alignment of bovine trypsin, human APC heavy chain and human thrombin heavy chain. Amino acid residues are shown in one-letter codes, and identical residues among three proteins are shown with boldface letters. Catalytic triad residues are marked with triangles. Cys122 of APC and thrombin is marked with an asterisk. Insertions are shown with horizontal bars and termed (a) through (h) along the sequence

and (h) lay on the backside of the proteins and near the Cys122 that links heavy and light chains of APC and thrombin, and they may be involved in the interface contact of two domains.

Conserved residues are shown with green spheres in Color Plate 2. Though conserved residues tend to gather in the core of the models (pictures not shown), as is well known among globular proteins, the active site is the only part where green spheres considerably surface on the molecular models. This distinctive difference between the active center side and the domain interface side is demonstrated in the color plate.

Color Plate 3 compares acidic and basic residue distribution, and Color Plate 4 shows hydrophobic and hydrophilic residue distribution of the three proteins. In the color plates, the active center sides of the three proteins are shown simultaneously. On the same sphere, top color shows the residue characteristics of trypsin, lower right does those of APC, and lower left does those of thrombin. These models can be considered as spatial alignments of the three proteins, models that compare residue characteristics at their 3D positions. Trypsin has fewer characteristic residues than the other two proteins around the active center, which may be why trypsin does not show preferences other than P1 position Arg and Lys. Considerable differences can be seen between thrombin and APC active sites, which may explain the differences in their selectivity toward substrates.

The reactive site sequences of PCI and ATIII are

	P3	P2	P1	P1′	P2′	P3′
PCI	Thr	Phe	Arg	Ser	Ala	Arg
ATIII	Ala	Gly	Arg	Ser	Leu	Asn

and it must be stressed that P1 and P1′ are common among the two inhibitors.[9] However, PCI inhibits APC and thrombin, while ATIII is ineffective toward APC.

Based on the structure of trypsin-BPTI complex, S2 through S2′ can be marked. Since the interactive site conformation of BPTI and that of PCI or ATIII may differ depending on the distance from the active center, potential S3, S2, S2′ and S3′ sites are tentatively located on a wider range. Though it is possible that insertions (a) and (b) play parts in the S2′ and S3′, the specificity difference between APC and thrombin is assumed to be explained without insertion loops. Both APC and thrombin have conserved Ser214 to Gly216, which may be the main part of S2. Trp215 is invariantly conserved in a biochemical domain present in all eukaryotic serine proteases, located in an α-helix region facing hydrophobic residues.[12] Considering all of the above, active sites S3 through S3′ are mapped on the model structures as follows.

Color Plate 5 shows the active center side of an APC model with sphere sizes relative to amino acid side chains. Color Plate 5a shows enzymatic subsites with various colors, in which S3 is made up of Met175 and Glu217 and is shown in green, S2 of Trp215 in sky blue, S1 of Asp189 to Glu192 and Ser195 in blue, S1′ of His57 in red, S2′ of Lys39 to Ala41 in pink and S3′ of Asp60 and Glu61 in yellow. Color Plate 5b shows the distribution of acidic and basic residues, and Color Plate 5c shows the distribution of hydrophobic and hydrophilic residues.

Color Plate 6 shows the active center side of a thrombin model with sphere sizes relative to amino acid side chains. Color Plate 6a shows enzymatic subsites. Color codes are the same as those of APC. S3 site is formed with Arg175 and Glu217, S2 with Trp215, S1 with Asp189 to Glu192 and Ser195, S1′ with His57, S2′ with Glu39 to Leu41 and S3′ with Leu60 and Tyr61. Color Plates 6b and 6c correspond to Color Plates 5b and 5c, respectively.

Favorable aromatic contact at P2 and ion pair contact at P3′ in APC–PCI interaction, hydrophobic contact at P2 position in thrombin–ATIII interaction and aromatic contact at P2 position in thrombin–PCI interaction may be important, but the lack of favorable interactions at P2, P2′ and P3′ may make ATIII ineffective against APC. Common Glu217 at S3 seems to be of no importance in the explanation of their selectivity, but this residue may act as a real S3 subsite in the case of some synthetic peptide substrates,[15-17] since free N-terminal peptides are generally good substrates for both thrombin and APC. Although it becomes harder to estimate active sites prior to S3 and they must be much less important in the reaction with natural protein substrates, many of them might be traced on the models for each peptide sequence. Insertion (d) may make the catalytic cleft narrow and make thrombin S2 specific to small or aromatic amino acids.

CONCLUDING REMARKS

We have discussed molecular modeling and the prediction of subsite-forming residues of APC and thrombin. As more proteins have been isolated and more genes have been cloned, a great amount of information is now available concerning blood coagulation factors. A simple graphics tool is useful in model-building studies of blood coagulation factors. The important aspects of this modeling procedure are ease of model handling and the final visual representation. Even if the 3D structure of a homologous protein is elucidated, it would be unwise to incorporate molecular mechanics or dynamics calculation and to spend an enormous amount of CPU time modeling insertions and deletions in the present state of reliability. There is a chance, however, that development of molecular dynamics programs and computer hardware will meet the present need in the future.

It is impossible to discuss the catalytic reaction itself, which requires detailed atomic coordinates even more precise than that of an X-ray determined structure, and it is out of the scope of this paper. But we considered it possible to investigate enzymatic subsites with simple graphics models, since subsites must be more operational than the active center. Though these models must be considered only as first approximations, they would provide some working hypotheses for protein engineering experiments, such as changing enzymatic selectivity and inhibitor specificity, even before the real 3D structure of the enzyme is elucidated. If simplicity and ease of the modeling procedure are taken into account, this type

of approach would be convenient for planning experiments.

ACKNOWLEDGEMENT

This work was carried out in part through Special Coordination Funds of the Science and Technology Agency of the Japanese Government.

REFERENCES

1 Rosenberg, R.D. and Rosenberg, J.S. Natural anticoagulant mechanisms. *J. Clin. Invest.* 1984, **74**, 1–6
2 Tulinsky, A., Park, C.H. and Skrzypczak-Jankun, E. Structure of prothrombin fragment 1 refined at 2.8 Å resolution. *J. Mol. Biol.* 1988, **202**, 885–901
3 Harlos, K., Holland, S.K., Boys, C.W.G., Burgess, A.I., Esnouf, M.P. and Blake, C.C.F. Vitamin K-dependent blood coagulation proteins form heterodimers. *Nature* 1988, **330**, 82–84
4 Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* 1987, **326**, 347–352
5 Greer, J. Comparative model-building of the mammalian serine proteases. *J. Mol. Biol.* 1981, **153**, 1027–1042
6 Greer, J. Model of a specific interaction. *J. Mol. Biol.* 1981, **153**, 1043–1053
7 Furie, B., Bing, D.H., Feldmann, R.J., Robinson, D.J., Burnier, J.P. and Furie, B.C. Computer-generated models of blood coagulation factor Xa, factor IXa, and thrombin based upon structural homology with other serine proteases. *J. Biol. Chem.* 1982, **257**, 3875–3882
8 Toma, K., Yamamoto, S., Deyashiki, Y. and Suzuki, K. Three-dimensional structure of protein C inhibitor predicted from structure of α1-antitrypsin with computer graphics. *Protein Engineering* 1987, **1**, 471–475
9 Suzuki, K., Deyashiki, Y., Nishioka, J., Kurachi, K., Akira, M., Yamamoto, S. and Hashimoto, S. Characterization of a cDNA for human protein C inhibitor. *J. Biol. Chem.* 1987, **262**, 611–616
10 Protein Identification Resource, National Biomedical Research Foundation, Washington, DC, USA
11 Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, New York, USA
12 Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complex with inhibitors. *Acta Cryst.* 1983, **B39**, 480–490
13 Toma, K. Simple protein model building tool. *J. Mol. Graphics* 1987, **5**, 101–102
14 Toma, K. Investigation of protein inter- and intramolecular interactions with a simple graphics tool. *J. Mol. Graphics* 1987, **5**, 149–151
15 Stone, S.R. and Hofsteenge, J. Specificity of activated human protein C. *Biochem. J.* 1985, **230**, 497–502
16 Cho, K., Tanaka, T., Cook, R.R., Kisiel, W., Fujikawa, K., Kurachi, K. and Powers, J.C. Active-site mapping of bovine and human blood coagulation serine proteases using synthetic peptide 4-nitroanilide and thio ester substrates. *Biochemistry* 1984, **23**, 644–650
17 Suzuki, K., Nishioka, J. and Deyashiki, Y. Unpublished results