Prediction of potential protein-protein interaction sites from amino acid sequence
Identification of a fibrin polymerization site

R. Manjunatha Kini*, Herbert J. Evans

Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0614, USA

Received 5 March 1996

Abstract Identification of a protein-protein interaction site is an important step that has significant potential to clarify structure-function relationships of proteins and drug design. We propose here a unique predictive method to identify protein-protein interaction sites based on the observation that proline is the most common residue found in the flanking segments of interaction sites [Kini, R.M. and Evans, H.J. (1995) Biochem. Biophys. Res. Commun. 212, 1115–1124]. Accordingly, the interaction sites of proteins might be predicted directly from the amino acid sequence based on the presence of proline brackets. Using this strategy, we have predicted a polymerization site in the epitope of the Aα-chain of fibrinogen recognized by a monoclonal antibody, 9E9 which inhibits fibrin polymerization [Cierniewski, C.S. and Budzynski, A.Z. (1992) Biochemistry 31, 4248–4253]. The synthetic peptide comprising this predicted site inhibited the coagulation of human blood and allosterically interfered in fibrin polymerization. This is the first known allosteric polymerization site of fibrinogen. Thus the results validate the predicted site and the method for prediction. This unique predictive method should help in identifying the interaction sites of many proteins.

Key words: Molecular recognition; Bioactive peptide; Proline; Fibrinogen; Structure-function relationship; Protein engineering

1. Introduction

Identifying protein-protein interaction sites, and thereby understanding structure-function relationships of proteins, is one of the most difficult and time-consuming challenges in protein chemistry. A prior knowledge of the location of the interaction site would be very useful in these studies. Without such knowledge, the studies are often tedious and cumbersome. Several methods using structural homology of the protein with other proteins, molecular model building and molecular docking using computer graphics, and theoretical deductive methods are being used to obtain an initial indication of interaction sites. Earlier, we showed that proline brackets often enclose protein-protein interaction sites [1]. These proline residues act as distinct barriers between the interaction sites and their neighboring segments, and may 'bracket' (or protect) the integrity and conformation of the site [1]. Proline brackets also help in the presentation of the interaction site to its complementary protein and thus facilitate protein-protein interactions [1,2]. In this report, we describe a corollary to our hypothesis, that protein-protein interaction sites can often be identified directly from the amino acid sequence of the protein, when proline residues bracket the sites. Accordingly, a short 3–7 residue segment enclosed between proline residues would be identified as a putative protein-protein interaction site. Here, we describe the identification of a fibrin polymerization site using this corollary.

Fibrin monomers, formed by proteolytic cleavage of fibrinogen by thrombin, self-assemble into multistranded fibers by a polymerization process forming a 'soluble' clot. These polymers are then covalently linked by the transglutaminase activity of factor XIIIa to form an 'insoluble' clot [3]. According to the prevailing model, the release of fibrinopeptides A and B from the amino-terminal end by thrombin cleavage exposes some polymerization sites that are complementary to preexisting sites in the carboxy-terminal domain [4]. Some polymerization sites have been tentatively identified. Recently, Cierniewski and Budzynski [5] developed three monoclonal antibodies that recognized the Aα-chain of fibrinogen and inhibited fibrin polymerization. They identified the epitopes in fibrinogen recognized by these antibodies and concluded that these sites probably play an important role in polymerization. Since the interaction of antibodies could interfere at remote sites, the results only show the possibility of polymerization sites in these epitopes. The author did not identify specific sites in these epitopes. Based on the presence of proline brackets in one of these epitopes, we predicted that the intervening segment is an interaction site which plays a significant role in clot formation. The synthetic peptide interferes in blood coagulation and fibrin polymerization supporting the predicted site.

2. Materials and methods

2.1. Materials

t-Butyloxycarbonyl (*t*-BOC) amino acids and trifluoroacetic acid were obtained from Advanced Chemtech, Inc., and other reagents were obtained from Fisher Scientific. Human fibrinogen (plasminogen-free, approx. 95% clottable protein) was obtained from Sigma.

2.2. Peptide synthesis

We synthesized the peptide HPGIAEFPSRA (peptide PSP-9E9, polymerization site peptide based on the epitope recognized by 9E9 monoclonal antibody) by solid-phase peptide synthesis, using *t*-BOC chemistry on Merrifield resin [6], employing a Milligen/Biosearch Model 9600 peptide synthesizer. The peptide was purified by a reverse-phase HPLC system to more than 95% purity, and the yield was about 90%. The amino acid composition after 24 h hydrolysis

*Corresponding author (present address). Bioscience Centre, Faculty of Science, National University of Singapore, Lower Kent Ridge Road, Singapore 0511, Republic of Singapore. Fax: (65) 779-2486. E-mail: bsckinim@nus.sg

SSDI S0014-5793(96)00327-4

in 6 N HCl at 110°C showed the expected ratios of amino acids. The structure was also confirmed by determining the mass by fast atom bombardment mass spectra on a Finnegan TSQ 70 mass spectrometer. The mass by mass spectrometry was 1182, whereas the calculated molecular weight for the peptide is 1181.3.

2.3. Coagulation assays

Various in vitro coagulation assays were used to determine the effect of PSP-9E9 on the coagulation of human plasma. Blood from healthy volunteers was drawn and anticoagulated with 0.11 M trisodium citrate (1:9; v/v). The coagulation times of human plasma were measured at 37°C using a BBL fibrometer. The effect of the peptide was determined by measurement of prothrombin time, Stypven time, thrombin time and activated partial thromboplastin time assays [7–9]. The first three clotting assays were performed essentially as described earlier [10].

2.4. Fibrin polymerization

Fibrin monomer was prepared from human fibrinogen according to the method of Belitser et al. [11]. Fibrin polymerization was assayed by a stopped-flow method. Each reaction mixture contained equal volumes of 170 μg/ml of fibrin monomer in 0.02 M acetic acid and 0.05 M Tris-HCl buffer, pH 7.3, containing 0.1 M NaCl. The initial rate of polymerization was measured as the increase in absorbance at 350 nm. The effect of PSP-9E9 on the rate of polymerization was determined by including various amounts of peptide in 0.05 M sodium phosphate buffer, pH 7.3.

2.5. Fluorescence measurements

All fluorescence measurements were made at room temperature in a Shimadzu RF-5000 recording spectrofluorophotometer at 280 nm excitation. The effect of various concentrations of PSP-9E9 on the intrinsic fluorescence of fibrin monomers was measured in 0.02 M acetic acid.

2.6. Fibrin fiber mass/length ratios

Mass/length ratios of fibrin fibers were determined according to the method of Carr and Gabriel [12]. Fibrin gels were formed at 0.1 M ionic strength in 0.05 M Tris, pH 7.4, containing 5 mM $CaCl_2$. The final concentrations of human fibrinogen and bovine thrombin in the reaction mixture were 1 mg/ml and 1.25 units/ml, respectively. After 20 h, the gels were scanned from 400 to 800 nm and the mass/length ratios were calculated.

3. Results and discussion

3.1. Prediction of a fibrin polymerization site

In our attempts to understand structure-function relationships of proteins, we have used both theoretical and experimental approaches [1,2,13–17]. Our recent studies of protein-protein interaction sites indicated that proline residues are found in the flanking segments of interaction sites at 2.5 times higher probability than expected by random distribution [1]. We exploited this finding by developing a predictive method to identify unknown protein-protein interaction sites based on the presence of proline residues, as shown in Fig. 1. When a short 3–7 residue segment is bracketed by proline residues, the intervening segment would be identified as a potential protein-protein interaction site, as a corollary to our survey (Fig. 1A,B). Cierniewski and Budzynski [5] identified a 34-mer segment in the Aα-chain of fibrinogen as an epitope recognized by a monoclonal antibody (9E9) that inhibits fibrin polymerization (Fig. 1C). This epitope contains two proline residues. Based on our predictive method (Fig. 1a) we identified the segment between these proline residues, GIAEF, as a potential fibrin polymerization site. To provide experimental evidence for the prediction, we synthesized peptide HPGIAEFPSRA (named PSP-9E9, see section 2) with the predicted site, and studied its effect on blood coagulation and fibrin polymerization.

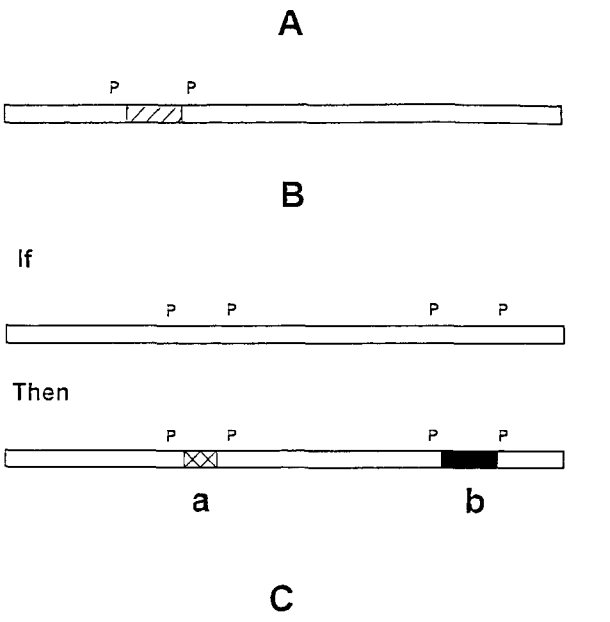

Fig. 1. Predictability of protein-protein interaction sites based on the presence of proline residues. (A) Proline residues are the most commonly found amino acid residues in the flanking segments of continuous protein-protein interaction sites [1] (shown as hatched area). (B) A corollary to the above finding is that if proline residues are separated by a short segment of the sequence, then the intervening segments are potential interaction sites (shown as cross hatched and solid areas). Two independent interaction sites (a,b) are shown as an example of a protein containing two sites. (C) Step-by-step description of prediction of fibrin polymerization site. In step 1, proline residues found in the sequence of the epitope recognized by monoclonal antibody 9E9 are highlighted. In step 2, the short intervening segment between two prolines (underlined) is identified as potential polymerization site. Step 3, gives the sequence of the synthetic peptide, PSP-9E9.

3.2. Experimental evidence for the predicted site

We studied the effect of PSP-9E9 on the coagulation of human plasma. The clotting time increased in all four assay systems with increasing peptide concentration (Fig. 2). The dose required for doubling the coagulation time ranged between 2.1 and 4.1 mM. Thus although it inhibited coagulation, the peptide is not a potent inhibitor. Since it inhibited coagulation initiated by thrombin, it interferes at either the cleavage of fibrinogen to fibrin monomer, or in fibrin polymerization.

We examined the effect of PSP-9E9 on fibrin polymerization. The initial rate of fibrin polymerization was not affected at low concentrations of the peptide, but at higher concentrations of the peptide, the rate was significantly enhanced (Fig. 3). The rate of polymerization reached about 20-fold greater than the control rates at 1.5 mM peptide. Thus the peptide appears to have an allosteric effect on fibrin polymerization. This is the first known allosteric site in fibrin polymerization. This suggests that interaction of the peptide with fibrin mono-

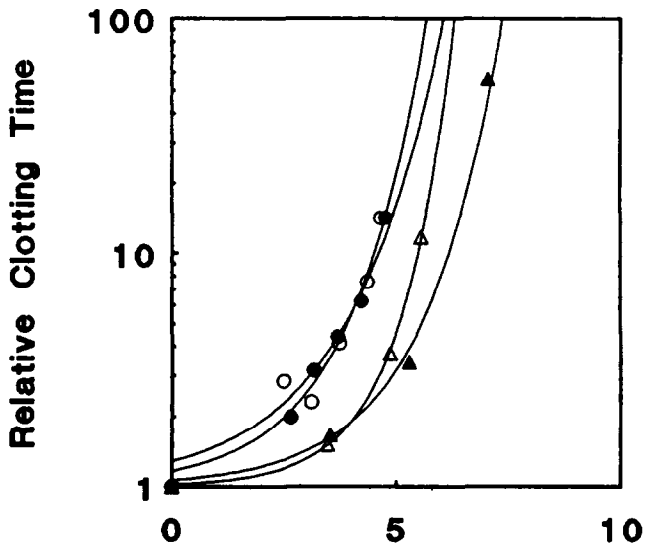

Fig. 2. Effect of PSP-9E9 on coagulation of human plasma. Various amounts of the peptide were incubated with plasma for two min before initiating the clotting. (○) Activated partial thromboplastin time; (●) prothrombin time; (△) Stypven time; and (▲) thrombin time. Each data point is an average of at least three independent experiments. Relative clotting times were determined by comparing the clotting times in the presence of the peptide to control clotting times.

mer probably induces subtle conformational changes that enhance the rate of polymerization. Therefore, we examined the effect of the peptide on the intrinsic fluorescence of fibrin monomers. As shown in Fig. 4, the intrinsic fluorescence was somewhat quenched at higher concentrations of the peptide. This is consistent with a conformational change in fibrin monomers upon binding of the peptide. However, PSP-9E9 did not affect the intrinsic fluorescence of fibrinogen even at high concentrations of the peptide (data not shown). The dose-response of fluorescence (Fig. 4, inset) paralleled that of the initial rates of polymerization (Fig. 3). Thus, from these studies, we conclude that PSP-9E9 interacts with fibrin monomers and allosterically modifies fibrin polymerization. Control peptides with unrelated sequences, for example, IARGDMNA and ADEFGHIKLMNP, did not show any effect on either fibrin polymerization or the intrinsic fluorescence (data not shown).

Finally, we examined the effect of PSP-9E9 on μ values (mass/length ratio) of the fibrin polymers formed (Fig. 5). At lower concentrations of peptide, μ decreased (Fig. 5, inset) indicating a probable decrease in the thickness of the fibers. In contrast, at higher concentrations, there was an increase in the μ values, and thus probably the thickness of the fibrin fibers. Whether the decrease in μ values cause decreases in the initial rate of fibrin polymerization is unclear. It is also not clear at this time why high concentrations of the peptide are required to exhibit these effects. It is important to note that protein-protein interaction in fibrin polymerization is based on multipoint contact. This new site may be one of the low affinity sites in these interactions. Further studies are required to clar-

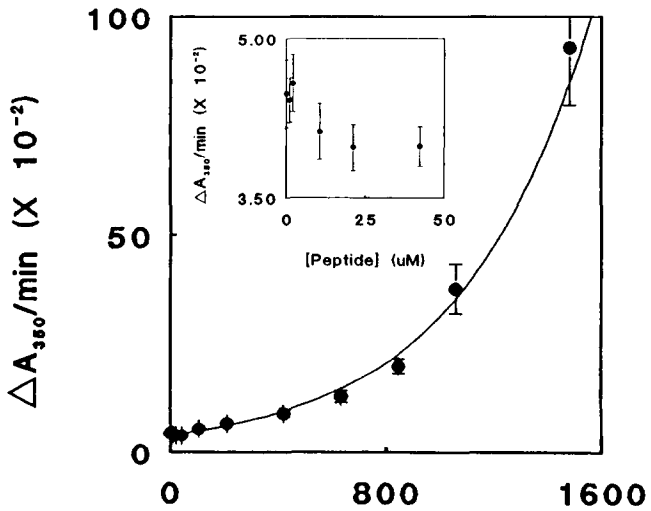

Fig. 3. Effect of PSP-9E9 on initial rate of fibrin polymerization. Initial rates were calculated between 0.252 and 4 s. Each data point is an average of eight determinations. The inset shows a decrease in the initial rate at lower concentrations of the peptide.

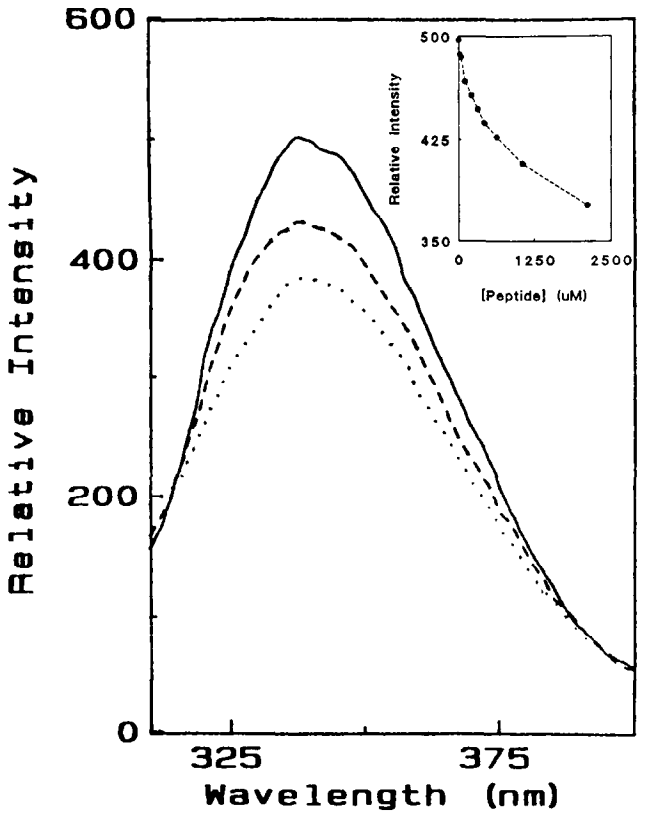

Fig. 4. Effect of PSP-9E9 on the intrinsic fluorescence emission of fibrin monomers. Fluorescence emission spectra of fibrin monomers with or without peptide. The emission spectra were measured upon excitation at 280 nm. Various amounts of the peptide were incubated with fibrin monomers for one min before fluorescence measurements were made at room temperature. Typical of two independent measurements. Solid line, no peptide; broken line, 0.5 mM peptide; and dotted line, 2 mM peptide. Inset: Effect of peptide on emission intensity of fibrin monomers. Excitation, 280 nm; emission, 345 nm. Each data point is an average of two experiments.

ify the mechanistic role of this site in fibrin polymerization. However, the results demonstrate that PSP-9E9 interacts with fibrin monomers and allosterically modifies fibrin polymerization, and thus the site is involved in protein-protein interaction. Thus, the current studies provide experimental evidence that protein-protein interaction sites can be predicted based on the presence of proline brackets, a corollary of our hypothesis [1].

3.3. Prediction of other interaction sites

We have presented here experimental evidence to support the prediction of a fibrin polymerization site. To test the capability of this new predictive method, we have examined other structurally, functionally and phylogenetically unrelated proteins, and have collected experimental evidence for their predicted sites. Although we have data from systematic and detailed studies of these examples, we have included only the summary of some of our results to avoid laborious, confusing and lengthy descriptions as each example is from a distinctly different field. For example, a new class of serine proteinase inhibitor was recently isolated from mustard seed [18]. We identified the reactive site of this inhibitor; the synthetic peptide, based on the predicted site between residues 18 and 24, acts like a substrate-analog inhibitor of trypsin (Kini and Evans, unpublished observations). We have also identified the interaction site of a calcium channel-blocking toxin from black mamba snake venom [19]. As expected, the synthetic peptide based on the predicted site shows negative inotropic effects on isolated rat hearts. The peptide also specifically blocks L-type calcium channels in isolated cardiac muscle cells as determined by patch clamp techniques (R.M. Kini et al., unpublished observations). Thus, we have predicted protein-protein interaction sites in three independent, unrelated examples with a success rate of 100% (three out of three). Consequently, the reliability of this approach is excellent. Hence, we have gathered strong experimental evidence for this novel predictive method to identify protein-protein interaction sites. Obviously no single method can guarantee to identify all protein-protein interaction sites. However, the described method can help identify a significant number of continuous interaction sites. This possibility is supported by the presence of proline residues in the flanking segments of a large number of protein-protein interaction sites [1] and their ability to enhance protein-protein interaction and potency of bioactive peptides [1,2]. Although the success rate will no doubt drop significantly, a success rate of even 20–30% is a significant step toward in identifying interaction sites from amino acid sequences, considering the present rate of 0%.

3.4. Novel method for prediction of interaction sites

In this method, short segments of 3–7 residues flanked by proline brackets are predicted as potential interaction sites. The method also considers 10-residue segments on either side of a single proline residue, with preference shown for segments containing a cysteine residue, which could form a boundary on the other side of the interaction site [1]. The number of such predicted sites will be small, because of the frequency of occurrence of proline residues ($\sim$5%) [20,21]. Any other information available can be used either to rule

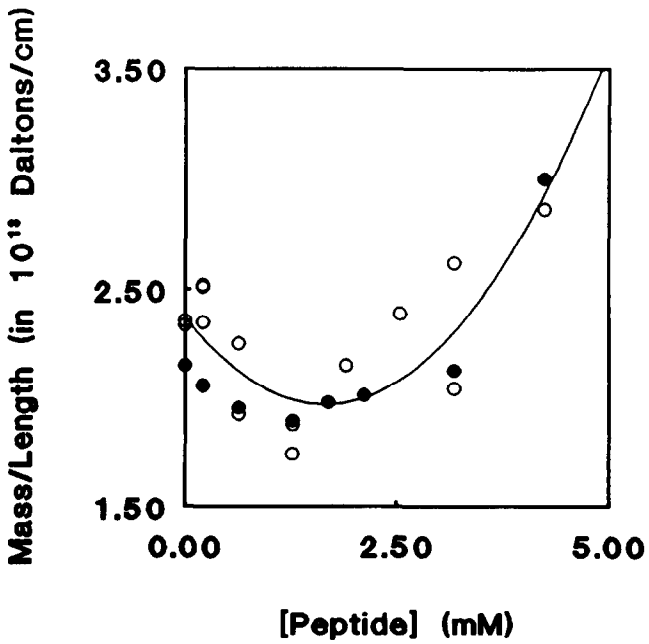

Fig. 5. Effect of PSP-9E9 on the mass/length ratio (μ) of fibrin polymers. The two different symbols used indicate the results of independent experiments.

out or to support the prediction of an interaction site. We used the experimental evidence presented by Cierniewski and Budzynski [5] of the epitope of monoclonal antibodies to predict a fibrin polymerization site. The presence of lysine or arginine within the site of a trypsin inhibitor, or the location of the interaction site on the external loop of a receptor protein, would also aid in the prediction. The systematic use of information obtained by homology studies can also direct the prediction. If available, a three-dimensional model of the protein would indicate the accessibility of the sites for interaction. Hydrophilicity [22,23], flexibility [24] and surface accessibility [25] may also assist in the prediction. Additional information is helpful, but not essential for the prediction of interaction sites. To our knowledge, this is the first method which allows the prediction of protein-protein interaction sites directly from the amino acid sequence.

The predicted site(s) can be critically tested by two distinct approaches depending on the facilities and the expertise available: (1) direct synthesis of peptides; or (2) site-directed mutagenesis. The structure of the peptides should include proline residues. When cysteine residues are used in the predictive step, replacement of these residues by proline should help in obtaining peptides with higher potency and specificity. These peptides can be directly tested by an assay of their biological effects to determine the success of the prediction. Although protein-protein interactions are important determinants of the biological activity of a protein, the contributions of other factors, such as enzymatic activity and role of other domains, should also be considered. Therefore, specific binding studies and binding competition between the peptide and the native protein should also be considered for testing the validity of predicted site(s). In site-directed mutagenesis experiments, the replacement of specific proline residues may not give informative results. Some proteins may retain their native folding even after replacement of proline residues because of other factors. Therefore, it would be prudent to mutate residues in the interaction site (the intervening segment) rather than the proline residues. Negative results in both these approaches only mean that the predicted site is not involved in the protein-protein interaction of the test system, and it may be necessary to examine the role of the site in other systems. Such studies may in fact reveal other unknown functions of the protein. After identifying the interaction site, further studies would also be required to identify the minimal recognition sequence.

In recent years, the theory of molecular recognition code has been used to identify or develop peptides that bind to a specific site. According to this theory, peptides encoded by the complementary sequences of nucleic acids interact with each other with high affinity and specificity [26,27]. The nucleic acid sequence of the protein is a mandatory requirement. The method also requires the prior knowledge of the interaction site of the protein. Without this information, designing the specific binding peptide or identifying the complementary site in its protein partner is difficult. The knowledge of topographic accessibility of a site may help to alleviate this problem to a small extent. Sometimes, the complementary strand may have a stop codon. Besides the universality of the theory is still controversial [28].

3.5. Conclusions and future prospects

We have developed a novel predictive method to identify protein-protein interaction sites based on the presence of proline brackets in the amino acid sequence. Our approach to prediction of interaction sites has several attractive features: (a) It is simple and straightforward. It reduces blind search for interaction sites; (b) it is based on an entirely new concept, the recognition of flanking segments, rather than the interaction sites themselves. The method is open to further improvement and fine tuning. Further studies of the flanking segments may help us improve the predictability using this strategy; (c) the only requisite for prediction is the complete amino acid sequence of the protein; (d) very little time is required to identify and test the predicted site(s) by this approach. In many cases, the method could reduce the time required to solve structure-function relationships from several years to a few short weeks. Thus, the method should have a considerable impact on the economics of research and development of novel bioactive peptides. Even if the method fails to identify the site, only a little time would be wasted; (e) the method appears to be robust, and can be readily applied to a large number of proteins. Thus the predictive method opens several avenues in protein chemistry and protein engineering, particularly for solving structure-function relationships of proteins; and (f) based on its excellent reliability, we believe that this approach is superior to all existing methods and approaches. Since the database covers only continuous interaction sites [1], the predictive method may or may not predict discontinuous sites. Finally, this corollary reinforces our hypothesis dealing with the importance of proline brackets [1,2].

Acknowledgements: We thank Mrs. Nancy Haar for excellent technical assistance, and Drs. H.T. Wright and K.R. Shelton for comments on the manuscript. This work was supported by a grant-in-Aid from the American Heart Association, Virginia Affiliate. Some parts of this paper are the subject of pending patents.

References

[1] Kini, R.M. and Evans, H.J. (1995) Biochem. Biophys. Res. Commun. 212, 1115–1124.
[2] Kini, R.M. and Evans, H.J. (1995) FEBS Lett. 375, 15–17.
[3] Mosesson, M.W. (1992) Semin. Hematol. 29, 177–188.
[4] Olexa, S.A. and Budzynski, A.Z. (1980) Proc. Natl. Acad. Sci. USA 77, 1374–1378.
[5] Cierniewski, C.S. and Budzynski, A.Z. (1992) Biochemistry 31, 4248–4253.
[6] Stewart, J.M. and Young, J.D. (1985) Solid Phase Peptide Synthesis, 2nd edn., Pierce Chemical Co., Rockford, IL.
[7] Quick, A.J. (1966) in: Hemorrhagic Diseases and Thrombosis, 2nd edn. (Quick, A.J. ed.) pp. 391–395, Lea and Febiger, Philadelphia, PA.
[8] Williams, W.J. and Esnouf, M.P. (1962) Biochem. J. 84, 52–62.
[9] Basu, D., Gallus, A., Hirsh, J. and Cade, J. (1972) N. Engl. J. Med. 287, 324–327.
[10] Stefansson, S., Kini, R.M. and Evans, H.J. (1989) Thromb. Res. 55, 481–491.
[11] Belitser, V.A., Varetskaja, T.V. and Malneva, G.V. (1968) Biochim. Biophys. Acta 154, 367–375.
[12] Carr, M.E. and Gabriel, D.A. (1980) Macromolecules 13, 1473–1477.
[13] Kini, R.M. and Iwanaga, S. (1986) Toxicon 24, 527–541.
[14] Kini, R.M. and Iwanaga, S. (1986) Toxicon 24, 895–905.
[15] Kini, R.M. and Evans, H.J. (1987) J. Biol. Chem. 262, 14402–14407.
[16] Kini, R.M. and Evans, H.J. (1989) Int. J. Peptide Protein Res. 34, 277–286.
[17] Kini, R.M. and Evans, H.J. (1989) Biochemistry 28, 9209–9215.
[18] Menegatti, E., Tedeschi, G., Ronchi, S., Bortolotti, F., Ascenzi,

P., Thomas, R.N., Bolognesi, M. and Palmieri, S. (1992) FEBS Lett. 301, 10–14.
[19] De Weille, J.R., Schweitz, H., May, P., Tartar, A. and Lazdunski, M. (1991) Proc. Natl. Acad. Sci. USA 88, 2437–2440.
[20] Robinson, A.B. and Robinson, L.R. (1991) Proc. Natl. Acad. Sci. USA 88, 8880–8884.
[21] Cserzo, M. and Simon, I. (1989) Int. J. Peptide Protein Res. 34, 184–195.
[22] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[23] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[24] Karplus, M. and Schultz, G.E. (1985) Naturwissenschaften 72, 212–213.
[25] Janin, J., Wodak, S.J., Levitt, M. and Maigret, M. (1978) J. Mol. Biol. 125, 357–386.
[26] Blalock, J.E. and Smith, E.M. (1984) Biochem. Biophys. Res. Commun. 121, 203–207.
[27] Bost, K.L., Smith, E.M. and Blalock, J.E. (1985) Proc. Natl. Acad. Sci. USA 82, 1372–1375.
[28] Beattie, J. and Flint, D.J. (1982) Biochem. J. 283, 473–478.