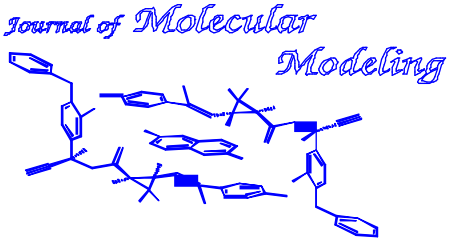
J. Mol. Model. **1997**, *3*, 31 – 47



A 'Fragment Fitting Approach' to Model Disulfide Loops by Utilizing Homologous Peptide Fragments from Unrelated Proteins of Known Structures: Application to the V3 Loop of the HIV-1 Envelope Glycoprotein gp120

Asim Kumar Debnath

Lindsley F. Kimball Research Institute, The New York Blood Center, 310 E 67th Street, New York, NY, 10021, USA; Tel: (212) 570-3373; Fax: (212) 570-3299 (adebnath@server.nybc.org)

Received: 13 November 1996 / Accepted: 6 January 1997 / Published: 20 January 1997

Abstract

The V3 loop of the human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein gp120 has gained considerable attention for developing subunit vaccines against HIV-1 and also as a target to develop anti-HIV-1 drugs. These endeavors would be significantly enhanced by understanding the structural aspects of this loop. The structure of the full-length V3 loop has not been defined yet. Therefore, a novel modeling technique, termed 'Fragment Fitting Approach' (FFA), was developed to model the V3 loop. This technique utilizes fragments (≥ 6 residue long) with local sequence and secondary structure similarity from unrelated proteins with known x-ray crystallographic structure and concatenating the fragments to build the model. A systematic search method was devised to identify the fragments using the combined criteria of sequence and secondary structure identity and/or similarity, predicted by a combination of methods. FFA requires partial three-dimensional coordinates of the target sequence to be modelled to get the overall coordinate path correct. The method was validated with nine disulfide-bonded loops from the Protein data bank. The modelled structures conform well with the corresponding x-ray crystallographic structures. As the models were built using the x-ray coordinates with reasonable resolution (≤ 3 Å), they are expected to have stereochemically correct structures. The modelled V3 loop structure might assist in structure-based drug design of anti-HIV-1 agents targeted to this loop.

Keywords: Fragment Fitting Approach (FFA), HIV-1, V3 loop, homology, molecular modeling

Introduction

Determination of three-dimensional structures of proteins and peptides is essential for understanding the molecular interactions of proteins with another protein (e.g. antigen-antibody reaction) or with small molecules (e.g. drug-receptor interactions). The advancement of gene sequencing techniques has made possible an exponential growth of gene sequencing data. As of September 1996, the GenBank had 921,500 sequences, while the Brookhaven Protein data bank (PDB) had only 4761 entries of 3D-coordinates of proteins/peptides. This discrepancy is due to the time consuming nature of the x-ray diffraction and nmr experiments and is compounded by the difficulty of getting crystals appropriate for x-ray diffraction studies. Therefore, theoretical prediction

methods might be useful in decreasing this gap. The most widely and accepted method for predicting the 3D-structure of proteins is homology modeling. In this method, closely related proteins with known three dimensional structures are used to model a protein having a primary amino acid sequence homology of >50% [1], with these proteins. Unfortunately, less than 20% of proteins have homologous proteins with known 3D-structures [2,3]. To gain insight into the structures of most (~80%) proteins is difficult. It has been recognized that only limited numbers of structural motifs (α-helix, β-sheet, reverse turns, etc.) exist and that these motifs recur in different proteins with limited number of protein folds [4,5]. The local similarities of structural patterns were first exploited by Jones and Thirup [6] for their protein crystallography work by selecting fragments from unrelated proteins with known structures to fit the electron density map of polypeptide chains. Several groups later utilized local fragments to model proteins [7,8]. Recently, Levitt has applied this concept to the 'Segment Match Modeling (SMM)' method to model several proteins with known 3D-structures to facilitate direct comparisons [9]. The initial success of these researchers and the recent success of predicting the secondary structure by >72% accuracy [10,11] motivated us to exploit fragments having not only sequence homology but also secondary structure homology to model small disulfide bonded loops. 4652 disulfide loops were found in the July 1995 PDB adopted in the Iditis Data V14.0 of which 452 loops are 25 to 40 residue long.

Attempts to gain better understanding of the structure of the third variable domain, designated as V3 loop, of the human immunodeficiency virus (HIV-1) gp120 envelope glycoprotein, a target for anti-HIV drug discovery [12-14], was the principal reason for us to model disulfide bonded loops in proteins. This loop, bonded by a pair of invariant cysteines, has been shown to play an important role in virus entry into CD4+ cells and in the tropism and infectivity of HIV-1 [15]. A principal neutralizing domain (PND) is located in the V3 loop region and antibodies that recognize the PND blocked virus infectivity and syncytium formation [16-18].

A number of laboratories have initiated a considerable effort to obtain structural information on the PND region by nuclear magnetic resonance (nmr) and circular dichroism (cd) [19-24]. The resulting reports describe the presence of populations of random structures in water. No attempts were made to generate tertiary structures from these experimental results except by combination of 2D nmr and Monte Carlo simulated annealing techniques [24].

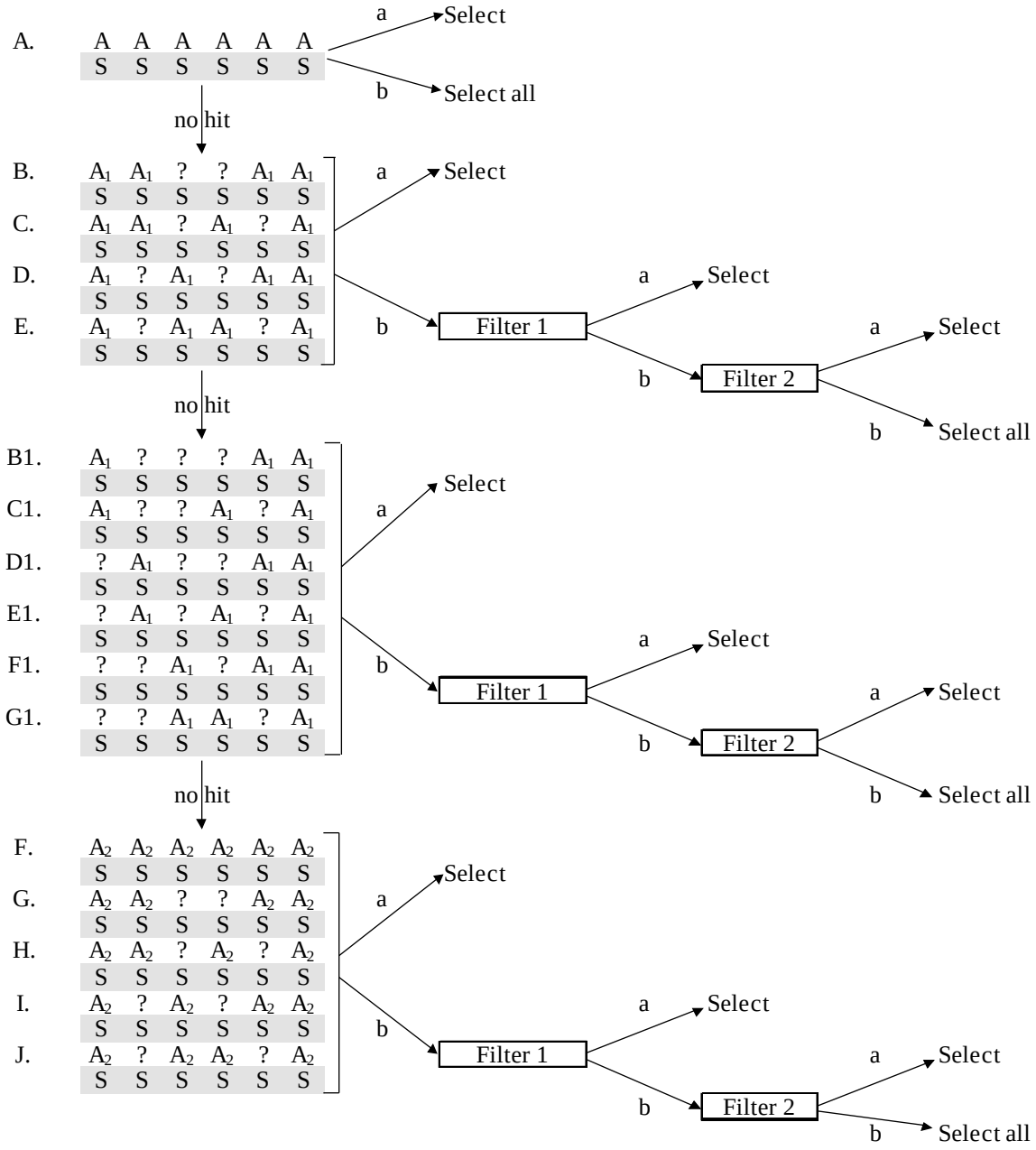
Recently, two crystal structures of HIV-1 MN V3 loop peptides complexed with Fab fragments of anti-V3 antibodies were reported [25,26]. In both cases, only short peptides from the V3 loop were used. In the first case, though a 14 residue long peptide cyclized by two cysteines (CKRIHIGPGRAFYTTC) was used, the coordinates for only nine residues (CKRIHIGPG) are available due to the disorder in the electron density map of the adjoining segments. Similarly, in the second case, a 24 residue long peptide (YNKRKRIHIGPGRAFYTTKNIIGC) was used, but coordinates for only 10 residues (HIGPGRAFYT) are available.

To the best of our knowledge, there is no primary amino acid sequence homology between the V3 loop and any protein for which the three-dimensional structure is known. We report here a method, designated as 'Fragment Fitting Approach', to generate the structure of the entire V3 loop by combining crystal structure coordinates of shorter V3 loop peptides with those of peptide fragments from unrelated proteins. This method may have general application to predict 3D-structures of small peptides using the steadily growing structural information from the PDB.

General Method for Model Building

Fragment Search

Proteins available in the July 1995 update of PDB adopted in the program Iditis v3.0/Iditis Data v14.0 [27] were used to search for fragments. The search was initiated using a six residue long fragment beginning at the N-terminal end of the disulfide loop to be modelled. A combination of criteria were used for search: (a) sequence homology of ≥ 50% and (b) secondary structure similarities as predicted by PHD [10] and NNSSP [11] methods for disulfide loops with unknown 3D-structures and by the method of Kabsch and Sander [28] for disulfide loops with known 3D-structures. The term sequence homology in our study was used to indicate either identical amino acids or a combination of identical and/or similar amino acids as defined by Vogt et al. [29]. Amino acids with substitution weights≥ 5 in the amino acid exchange matrix [29] were considered homologous to amino acid residues in the target sequence. The search procedure is shown schematically in Figure 1. In the first run (A), the fragment with six identical amino acids and the predicted secondary structure was searched for using the search option in Iditis program. If there was a single hit, it was selected as the fragment for that region of the disulfide loop to be modelled. If there were more than one hit (identical hexapeptides from different proteins) all of them were selected to build the model separately using one fragment at a time for that particular region. If there was no hit, steps B-E were followed. In each case, four out of six residues as well as two out of three residues at the N- and C-terminals were kept as identical or similar amino acids from the 'amino acid exchange matrix' [29] with three highest substitution weights [e.g., for cysteine(C) - C, S, A and V were selected; for alanine(A)- A, C, S, T and G were selected] to perform the search. The latter constraints were imposed to obtain fragments with the closest amino acids in the anchor region for superimposition in the following steps. Again, if searches B-E resulted in one hit, it was selected as the fragment for that region as described before. If there were more than one hit some 'filters' were used. The filters were used as a screening device to select the most closest fragments to that of the target sequence



a,	One hit
b,	More than one hit
Filter 1,	Fragments with the highest number of identical amino acids and secondary structure elements
Filter 2,	Fragments with the highest number of identical amino acids and secondary structure elements plus the highest number of amino acids with the substitution weights ≥ 5 in the amino acid exchange matrix in Ref. 29.

Figure 1. *Fragment search and selection rules. In A, amino acids A represent amino acids identical to those of the target sequence; in B - E and B1 - G1, A_1 represents identical amino acids and/or amino acids with the top three substitution weights each ≥ 5 in the amino acid exchange matrix [29]; in F– J, A_2 represents identical and/or any amino acids with substitution weights ≥ 5 in the exchange matrix in Ref. 29. S, in the shaded area in this figure, represents the secondary structure used for the search.*

in terms of sequence identity and/or similarity keeping the secondary structural elements fixed. These screening devices also kept the number of possible hits to a manageable number for model building. The first filter screened out fragments with the maximum number of identical amino acids. If there was only one hit, it was selected. If there were more than one hit, they were further subjected to a pass through another filter. This filter selected fragments with the maximum

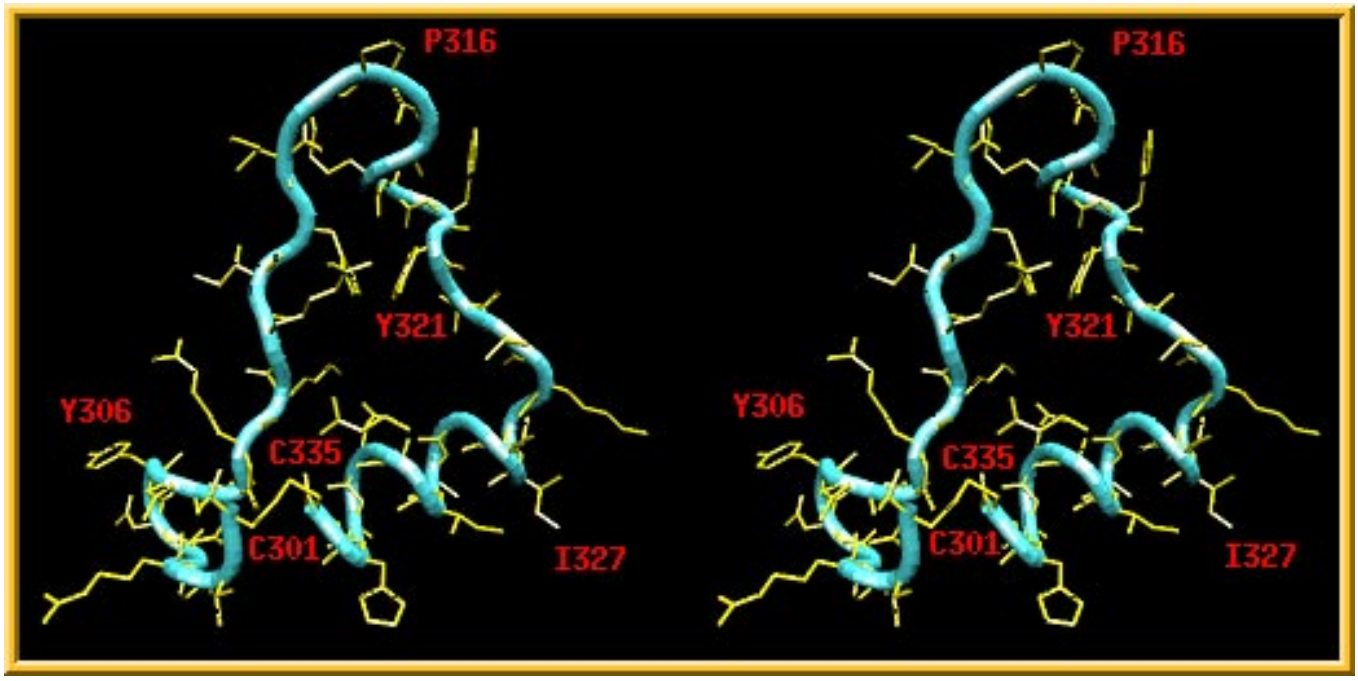


Figure 2. *Stereo representation of the C-α tracing (cyan) of the modelled V3 loop structure of HIV-1 MN gp120. The side chains are shown in yellow. The amino acid numbering is shown in one letter code. The figure was generated by VMD-1.1 [38] and ImageMagick-3.7.7 [39].*

number of identical plus similar residues. If there was only one hit, it was selected. If there was more than one hit, all of them were selected as possible fragments for that particular region of the peptide to be modelled and were used one at a time to derive multiple models (see Table 4 and 7). If there was no hit, the stringency of search was reduced and steps B1-G1 were followed using the same sequences of search, filters and selection procedures as described before. If no hits resulted from the entire search, the stringency of search was further reduced and the search steps F-J were used sequentially as before using all identical and/or similar residues with substitution weights ≥ 5 [e.g.,for serine(S)- S, T, P, A, G, N, D, E, Q, H, R, K were used]. After getting a hit or multiple hits in each of these steps, the fragment(s) was further extended by few amino acids and inspected for any possible sequence and secondary structural homology with the corresponding region of the disulfide loop to be modelled. If the extended fragment had homology as mentioned before, it was always selected preferentially to any six residue long fragments. To illustrate the above search techniques the disulfide loop in Lysozyme (178L, residues 127-154, in Table-2) was selected as an example. The fragment search on residues 127-132 yielded no hit in steps A-E, whereas 72 hits were obtained in the next steps B1-G1. The filters screened out all but one hit i.e., VEAAVN. Similarly, the next search on residues 130-135 yielded no hit in steps A-E but 28 hits in steps B1-G1. Only one hit emerged as the best fragment in terms of sequence and secondary structure homology after extending the fragment length by two more residues, and this hit was selected. The next search on residues 135-140 resulted in only one hit in steps F-G but no hit in either A-E or B1-G1 steps. The search on residues 143-148 resulted in 40 hits in steps B1-G1. Again only one fragment with a two residue extension resulted in the best match with the target sequence and was selected. The last search on residues 148-153 resulted in one hit in steps A-E. This example demonstrates the possibility of getting multiple hits and how the filters can help in selecting the best fragments. When multiple models are generated as a result of multiple hits, the selection of the model(s) has to be based on the evaluation of the qualities of models by a combination of programs e.g., Procheck [30], ProsaII [31], Whatif-Check [32], etc. In the next search, the fragment from the target disulfide loop was selected based on the previously selected fragments. In every case, a three residue overlap of the N-terminus of the new fragment with the C-terminus of the preceding fragment was used (see Tables). The search continued till the last fragment(s) was found to model the C-terminal region of the target disulfide loop. Due to the limited availability of the 3D coordinates, protein structures with up to 3.0 Å resolution were considered for the fragment search. FFA requires partial three-dimensional coordinates of the target sequence to be modelled to get the overall coordinate path correct.

To confirm the validity of FFA, nine additional disulfide loops from proteins with defined 3D-structures were modelled as described for the V3 loop. In most cases the fragment of the central portion (six-ten residues long) of the disulfide-loop was taken from the same protein in the PDB. In some cases, a random coil region, the most difficult to model, was selected instead.

Alignment of peptide fragments, extraction of the coordinates from the fragments to the target sequence and energy refinement of the modelled structure

The first fragment(s) selected, consisting of at least six residues, started with the N-terminus of the target peptide. The second fragment was then aligned on the three end residues at the C-terminal end of the first fragment as shown in Tables 1-10 and superimposed on the C-α atoms. The Protein Design Module within Quanta 4.0 [33] was used for alignment and superimposition. These alignments and superimpositions were continued with all the selected fragments until the C-terminal end of the modelled loop was reached. The coordinates of the main-chain atoms from the fragments shown by underline in Tables were copied onto the target sequence of the disulfide loops. The fragments which were not copied were used as linker to join two fragments. A four-residue section at each fragment interface (i.e., where two copied portions of the fragments met) was regularized by using 50 steps of steepest descent and 200 steps of adopted basis Newton-Raphson (ABNR) method using CHARMm minimization option adopted within Quanta 4.0 [33] to connect the copied fragments. The coordinates of the section selected from the same protein were always kept constrained. The modelled structures were then minimized by 500 steps of steepest descent and 2000 steps of adopted basis Newton-Raphson (ABNR) method. Again the coordinates of the section taken from the same protein were always kept constrained and a disulfide constraint was also used to form the disulfide loop.

Results and Discussion

Application of the 'Fragment Fitting Approach' to model the tertiary structure of the HIV-1 MN gp120 V3 loop

The disulfide bonded 35-residue V3 loop from the HIV-1 MN gp120 envelope glycoprotein was constructed using eight peptide fragments from the July 1995 update of the Protein data bank (PDB) as implemented in the Iditis v3.0/Iditis Data v14.0 program. Six of these peptide fragments were from unrelated proteins (Table 1) and two fragments corresponded to the V3 loop peptides, the coordinates of which (in the form of complexes with Fab fragments of anti-V3 loop antibodies) have recently been published [25,26]. As described in the method section, combined criteria of sequence identity and/or similarity along with secondary structure information were used to search for fragments of unrelated proteins with already defined 3D-structures to construct the model tertiary structure of the HIV-1 MN V3 loop. This procedure was adopted to exclude identical fragments with different secondary structure [34,35]. The secondary structure of the V3 loop was taken from the recently published predicted secondary structure of HIV-1 gp120 [36]. This approach incorporated two of the most widely used and accurate (>72% accuracy) methods for secondary structure prediction, PHD (Profile Network System Heidelberg) [10] and NNSSP (Nearest-Neighbor Secondary Structure Prediction) [11]. This three-state prediction method includes H, helix (α, 3_{10}, π); E (extended β-strand); and blank space (turn, bend, coil, loop, β-bridge and non-periodic structures). U, was assigned by this method when conflicting predictions were assigned for helix and β-strand. The secondary structure assignment method in the Iditis program used to search the fragments was as follows: alpha helix (H, h); extended β-strand (E, e); 3_{10} helix (G, g); turn (T, t); bridge (B); bend (S) and coil (space). While doing the search, these assignments along with the assignments in PHD and NNSSP methods were synchronized to keep uniformity in the search method. The fragment search was initiated beginning with the N-terminal residue of the V3 loop and up to the sixth residue. In the first search, the secondary structure of the first residue was assigned as an extended β-strand (E), the second to sixth residues could be either bend (S) or end of turn (t) or turn (T) or bridge (B) or coil (blank space). This yielded no hit, so the next search was done assigning the secondary structure of the first residue to be any secondary structure while the assignment of second to sixth residues were as described above. This decision was based on the fact that only the first residue was predicted to be in an extended β-strand, and changing it to any secondary structure may yield hits. Only one hit was obtained (see Table 1). A similar approach was followed for all other fragment searches for the V3 loop. It is to be pointed out that the secondary structure assignments for two peptide fragments from the V3 loop, the crystal coordinates of which were determined, i.e., KRIHIGPG and HIGPGRAFYT may not be correct, since the Kabsch and Sander method utilizes hydrogen bond information to assign the secondary structures. In case of 8-10 residue long peptide fragments, the prediction is bound to miss the hydrogen bonding effects of neighboring amino acids which are not available for calculation. It was well established by both nmr and x-ray crystallography that the tip of the V3 loop region (GPGR) has a type II β-turn but the assignment clearly missed that. These assignments are presented in the Table 1 to clarify these points. The sequence alignment, extraction of coordinates to the V3 loop sequence, regularizing the junction points of two fragments and the energy refinements were done as described above. The model could not be evaluated for any rms deviation from the reference coordinates as there is no full length V3 loop structure reported yet. The modelled structure is expected to be stereochemically correct as the fragments used to construct the structures were taken from well defined x-ray crystal structures. To verify that, the structure was checked by the PROCHECK program [30]. The stereochemical qualities of the predicted structure of the V3 loop at 2.0 Å conform well when compared to a set of well-defined (≤ 2.0 Å) crystal structures from PDB. The predicted structure is shown in Figure 2. The C-α tracing is shown in cyan while the side chains are shown in yellow. The amino acids are indicated in one letter code.

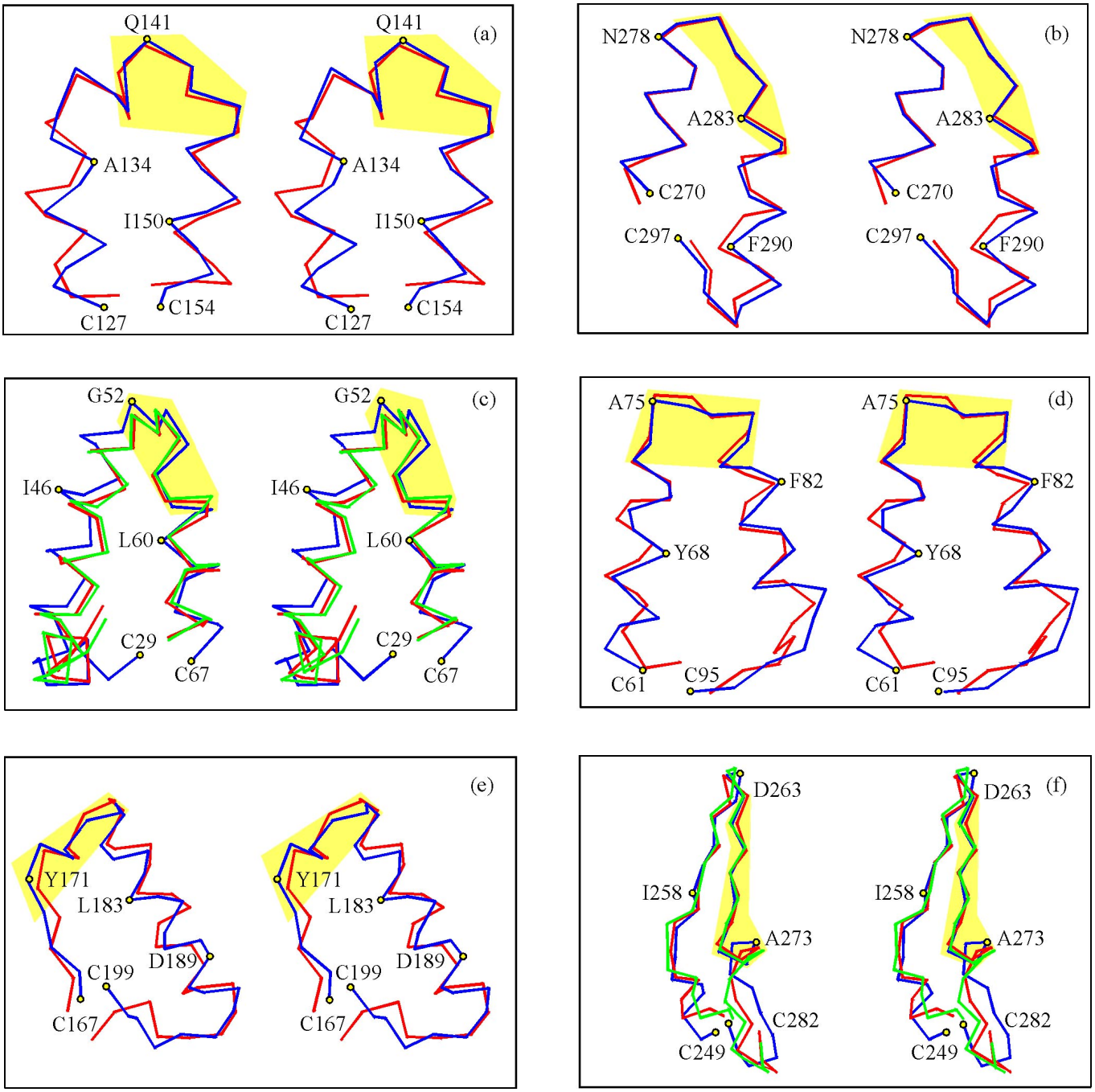
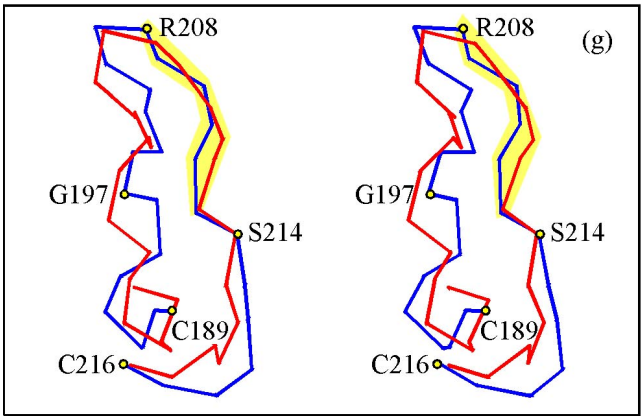
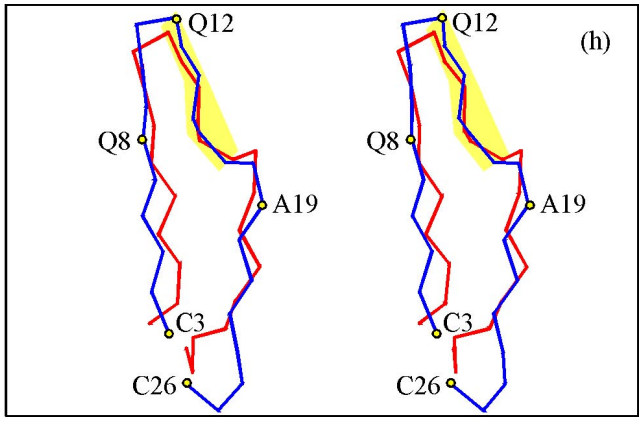


Figure 3 (continuous next page). *Stereo representation of the C-α tracings of nine disulfide loop models created by FFA (red and green) superimposed on the corresponding x-ray crystal structures (blue):*
[a] 178L, residues 127-154;
[b] 1ezm, residues 270-297;
[c] 1hyp, residues 29-67;
[d] 1poc, residues 61-95;
[e] 2exo, residues 167-199;
[f] 3psg, residues 249-282;

Validation of the FFA method

The fragment fitting method was evaluated using nine additional disulfide loop peptides 28 to 39 residues long (Tables 2-10) from the Protein data bank. These loops were selected at random. Attempts to model three more disulfide loops (1BP2-61-91; 2CGA-191-220 and 7LPR-191-224) failed since some of the fragments are not available in the PDB. The Kabsch and Sander method of secondary structure prediction as implemented in the Iditis V3.0 [27] program was chosen to search for fragments. It is the objective to show that if the secondary structure prediction accuracy increases, the fragment search method yields a better hit when com-





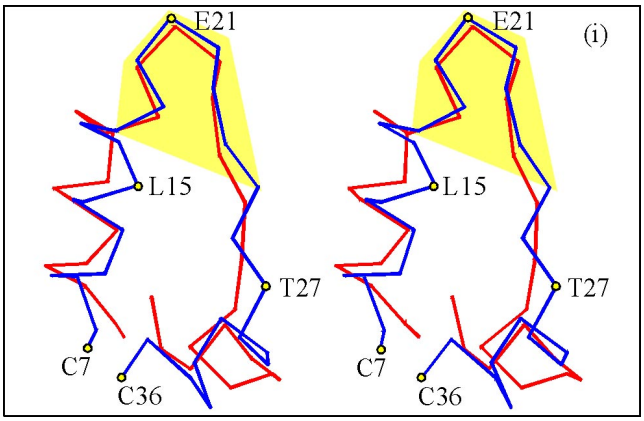


Figure 3 (continued).
[g] 4sgb, residues 189-216;
[h] 5azu, residues 3-26;
[i] 6ins, residues 7-36.
The figure was generated by Quanta 4.0 [33]. Segments in each loop taken from the corresponding x-ray crystal structure in PDB is demarcated by yellow background (see also Table 2-10).

bined with sequence identity and/or similarity as described in the Method section. There were two instances when the search technique yielded more than one fragment even after using the filters (see Table 4 and 7). In both cases, one fragment was used at a time in conjunction with all other fragments to built the model. In other words, two models were built in each case. They conform well with each other [red and green tracings in Figure 3(c) and 3(f)]. Figure 3(a)-(i) shows the stereo diagrams of the C-α tracing of all the disulfide loops modelled in addition to the V3 loop. The blue tracing are the x-ray crystal structures while the modelled structures are red and green. The figures clearly show that FFA yielded tertiary structures which conform well with the crystal structures. The rms (Å) deviations between the model and the x-ray crystallographic structures of C-α atoms, main-chain atoms and all atoms are listed in Table 11. Results of the modeling suggest that disulfide loops having fewer turns, bends, random coil, β-bridge and non-periodic structures are relatively simple to model and that fragments in the helix regions are, in general, available in the PDB.

Conclusion

A novel modeling technique, a fragment fitting approach (FFA), was developed to predict the tertiary structures of disulfide-bonded loops in proteins when homology modeling could not be used because of the absence of three-dimensional coordinates of homologous proteins or peptides. The FFA modeling method utilizes sequence identity and/or similarity and secondary structure prediction methods for fragment search and selection. This method was applied to model the third hypervariable loop, known as V3 loop, of the HIV-1 MN gp120 envelope glycoprotein. The application of the FFA method to model nine additional disulfide loops validated the overall technique of search, alignment, and other methods incorporated to built the models.

The main drawback of this technique is the insufficient availability of appropriate fragments in the protein data bank and the limited (~72%) accuracy of secondary structure predictions by the combined PHD and NNSSP methods incorporated into the FFA. Therefore, the models predicted by this method should not be expected to be more than 72% accurate when compared with the x-ray crystal structures although more favorable results were obtained in nine model systems (Table 11). The success of these models mostly lies on the availability of the homologous fragments with the correct secondary structures. The modelled structures with this accuracy are good starting points when there is no other structural information available. This method may be applied to model hypervariable loops of antibody molecules.

Acknowledgement. This study was supported by Grant CA43315 from the National Institute of Health and by Philip Morris Companies, Inc. I am indebted to Dr. A. Robert Neurath for his encouragement to undertake this project and for reading the manuscript. I thank J. Ng Pack for help in preparing the Figures.

References

1. Johnson, M.S.; Srinivasan, N.; Sowdhamini, R.; Blundell, T.L. *Critical Rev. Biochem. Mol. Biol.* **1994**, *29*, 1 and references cited therein.
2. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
3. Schneider, R.; Sander, C. *Proteins* **1991**, *9*, 56.
4. Orengo, C.A.; Jones, D.T.; Thornton, J.M. *Nature* **1994**, *372*, 631.
5. Russell, R.B.; Copley, R.R.; Barton, G.J. *J. Mol. Biol.* **1996**, *259*, 349.
6. Jones, T.A.; Thirup, S. *EMBO J.* **1986**, *5*, 819.
7. Finzel, B.C.; Kimatian, S.; Ohlendorf, D.H.; Wendoloskii, J.J.; Levitt, M.; Salemme, F.R. In, Crystallographic and Modeling Methods in Molecular Design, Bugg, C.E.; Ealick, S.E., eds; 1989, pp175-188, Springer-Verlag, Berlin.
8. Blundell, T.L.; Sibanda, B.L.; Sternberg, M.J.E.; Thornton, J.M. *Nature* **1987**, *326*, 347.
9. Levitt, M. *J. Mol. Biol.* **1992**, *226*, 507.
10. Rost, B.; Sander, C. *Proteins* **1994**, *19*, 55.
11. Salamov, A.A.; Solovyev, V.V. *J. Mol. Biol.* **1995**, *247*, 11.
12. Neurath, A.R.; Strick, N.; Haberfield, P.; Jiang, S. *Antiviral Chem. Chemother.* **1992**, *3*, 55.
13. Debnath, A.K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P.; Neurath, A.R. *J. Med. Chem.* **1994**, *37*, 1099.
14. Neurath, A.R.; Strick, N.; Debnath, A.K. *J. Mol. Recog.* **1995**, *8*, 345.
15. Moore, J.P.; Nara, P.L. *AIDS* **1991**, *5*, S21.
16. Javaherian, K.; Langlois, A.J.; McDanal, C.; Ross, K.L.; Eckler, L.I.; Jellis, C.L.; Profy, A.T.; Rusche, J.R.; Bolognesi, D.P.; Putney, S.D.; Mathews, T.J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6768.
17. Palker, T.J.; Clark, M.E.; Langlois, A.J.; Mathews, T.J.; Weinhold, K.J.; Randall, R.R.; Bolognesi, D.P.; Haynes, B.F. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1932.
18. Kenealy, W.R.; Mathews, T.J.; Ganfield, M.-C.; Langlois, A.J.; Waselefsky, D.M.; Petteway, S.R.,Jr. *AIDS Res. Hum. Retroviruses* **1989**, *5*, 173.
19. Chandrasekhar, K.; Profy, A.T.; Dyson, H.J. *Biochemistry* **1991**, *30*, 9187.
20. Zvi, A.; Hiller, R.; Anglister, J. *Biochemistry* **1992**, *31*, 6972.
21. Laczko, I.; Hollosi, M.; Urge, L.; Ugen, K.E.; Weiner, D.B.; Mantsch, H.H.; Thurin, J.; Otvos, L. Jr. *Biochemistry* **1992**, *31*, 4282.
22. de Lorimer, R.; Moody, M.A.; Haynes, B.F.; Spicer, L.D. *Biochemistry* **1994**, *33*, 2055.
23. Vranken, W.F.; Budesinsky, M.; Fant, F.; Boulez, K.; Borremans, F.A.M *FEBS Letters* **1995**, *374*, 117.
24. Catasti, P.; Fontenot, J.D.; Bradbury, E.M.; Gupta, G. *J. Biol. Chem.* **1995**, *270*, 2224.
25. Rini, J.M.; Stanfield, R.L.; Stura, E.A.; Salinas, P.A.; Profy, A.T.; Wilson, I.A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6325.
26. Ghiara, J.B.; Stura, E.A.; Stanfield, R.L.; Profy, A.T.; Wilson., I.A. *Science* **1994**, *264*, 82.
27. Iditis v3.0/Iditis Data v14.0: Oxford Molecular, Oxford Science Park, Oxford, U.K.
28. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
29. Vogt, G.; Etzold, T.; Argos, P. *J. Mol. Biol.* **1995**, *249*, 816.
30. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. *J. Appl. Cryst.* **1993**, *26*, 283. (Procheck v3.4.2 was used)
31. Sippl, M. J. *Proteins* **1993**, *17*, 355.
32. Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52.
33. Quanta v4.0: Molecular simulations, Inc./Biosym San Diego, California, U.S.A.
34. Kabsch, W.; Sander, C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075.
35. Cohen, B.I.; Presnell, S.R.; Cohen, F.E. *Protein Science* **1993**, *2*, 2134.
36. Hansen, J.E.; Lund, O.; Nielsen, J.O.; Brunak, S.; Hansen, J-E.S. *Proteins* **1996**, *25*, 1.
37. Meyers, G.; Wain-Hobson, S.; Henderson, L.E.; Korber, B.; Jean, K.-T.; Pavlakis, G.N.: A compilation and analysis of nucleic acid and amino acid sequences: Los Alamos, MN, U.S.A.
38. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
39. ImageMagick-3.7.7 (available free of charge from ftp://ftp.wizard.dupont.com/pub/ImageMagick).

Table 1. *Alignment of peptide fragments on the target sequence of the HIV-1 MN V3 loop.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			301 310 320 330 335
			C T R P N Y N K R K R I H I G P G R A F Y T T K N I I G T I R Q A H C
			E E E E E E E E E E H H H H H U H H H H H H
1sgt	1.7	100	C A R P G Y
			S S t T T t
2tmd	2.4	83.3	P Y Y N K R
			T T T t
1ikb	2.7	75	N T G K E I V I
			T T t e E E E E
1ggi	2.8	100	K R I H I G P G
			t
1acy	3.0	100	H I G P G R A F Y T
			t T g G G G g
1ril	2.8	66.7	A C T T N N
			E E E h H H
1hex	2.5	100	T G N I F G D I
			E h H H H H H H
1gdh	2.4	75	G S I G Q A L A
			h H H H H H H H

[a] Percent homology (identical or residues with substitution weights ≥ 5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28] except the target sequence (top) of HIV-1 MN V3 loop, the prediction of which was taken from Ref. 36. The blank in all the secondary structure prediction represents coil; H, helix; E, β-sheet; U, unknown. Sequence numbering is according to Meyers et al. [37].

Table 2. *Alignment of peptide fragments on the target sequence of Lysozyme (178L), residues 127-154.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			127 136 146 154
178L	2.7		C E A A V N L A K S R W Y N Q T P N R A K R V I T T F C H H H H H H H H h h H H H H H h H H H H H H H H H H H H
1hrs	2.6	100	V E A A V N H H H H H H
2lhb	2.0	62.5	A D E L K K S A H H H H H h h H
1crl	2.1	83.3	Q S K V F E h h H H H H
178L	2.7	100	W Y N Q T P N R H H H H h H H H
1nfp	1.6	62.5	I N F D K R V I H H H H H H H H
1ilk	1.8	57.1	R V K T F F Q H H H H H H H

Table 3. *Alignment of peptide fragments on the target sequence of Elastase (1ezm), residues 270-297.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			270 279 289 297
1ezm	1.5		C G V I R S A Q N R N Y S A A D V T R A F S T V G V T C H H H H H H H H H h T t h H H H H H H H H H H h T t
1pfk	2.4	85.7	R G V V R S A H H H H H H H
1anh	2.4	87.5	E Q A Q A R G Y H H H H H h T t
1ezm	1.5	100	R N Y S A A D h T t h H H H
1cc5	2.5	71.4	A A W K T R A H H H H H H H
1did	2.5	85.7	T S A F F T V H H H H H H H
1azm	2.4	100	T A V G S S C H H h T t

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 4. *Alignment of peptide fragments on the target sequence of Hydrophobic Protein from Soybean (1hyp), residues 29-67.*

Source of Fragment (pdb Code) [e]	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
		29 38 48 58 67
1hyp		C A L I G G L G D I E A I V C L C I Q L R A L G I L N L N R N L Q L I L N S C
		H H H H H h t h H H H H H H H H H H H H H H h T t S h H H H H H H H H H H h
2mta [c]	83.3	S T L Y G G
		H H H H H h
1pfk [d]	83.3	A A I A G G
		H H H H H h
1pxt	50	I G R F P E
		H H h t h H
1edb	66.7	L T E A E A
		t h H H H H
1asy	83.3	I Q A G V C
		H H H H H H
1fps	62.5	A V G W C I E L
		H H H H H H H H
1csc	75	I W S R A L G F
		H H H H H h T t
1hyp	100	L G I L N L N R
		h T t S h H H H
1cew	83.3	L Q R A L Q
		H H H H H H
3ink	88.9	D L Q M I L N G I
		H H H H H H H H H

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

[c] When this fragment was used, d was not used to derive the model.

[d] When this fragment was used, c was not used to derive the model.

[e] X-ray crystal structure resolutions (Å) of the proteins from top to bottom in the table are: 1.8, 2.4, 2.4, 2.8, 2.0, 3.0, 2.6, 1.7, 1.8, 2.0 and 2.5.

Table 5. *Alignment of peptide fragments on the target sequence of Phospholipase A2 (1poc), residues 61-95.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			61 70 80 90 95
1poc	2.0		C D C D D K F Y D C L K N S A D T I S S Y F V G K M Y F N L I D T K C
			H H H H H H H H H H H H h t S h H H H H H H H H H H H H h t E
1pii	2.0	83.3	A D V V D K
			H H H H H H
3hhr	2.8	88.9	D S N V Y D L L K
			H H H H H H H H H
1oxy	2.4	83.3	V F W E A G
			H H H h t S
1poc	2.0	100	N S A D T I S S Y
			h t S h H H H H H
1cpc	1.7	66.7	S S V A V G
			H H H H H H
1mio	3.0	66.7	P V G A M Y
			H H H H H H
1htl	2.5	83.3	R Y Y R N L
			H H H H H h
1ltd	2.6	83.3	L K F S N T
			H H h

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 6. *Alignment of peptide fragments on the target sequence of Exo-1,4-beta-D-Glycanase (2exo), residues 167-199.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			167 176 186 196 199
2exo	1.8		C I N D Y N V E G I N A K S N S L Y D L V K D F K A R G V P L D C
			E E E E S S S S S h H H H H H H H H H H H H H H H h t e E
1afn	2.6	66.7	V H S Q A N
			E E E E S S
2exo	1.8	100	D Y N V E G I N A
			E S S S S S h H
1oya	2.0	83.3	M D A E Q E
			S h H H H H
2bmh	2.0	70	L S F A L Y F L V K
			H H H H H H H H H H
2lbp	2.4	100	L V K D L K A
			H H H H H H H
1lnc	1.8	83.3	F D A V G V
			H H H H h t
1aco	2.0	71.4	E G W P L D I
			H h t e E

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 7. *Alignment of peptide fragments on the target sequence of Pepsinogen (3psg), residues 249-282.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			249 258 268 278 282
3psg	1.6		C S S I D S L P D I V F T I D G V Q Y P L S P S A Y I L Q D D D S C
			G G G G G G g E E E E E T T E E E E E h H H H H E E E e S S e E
1ofv [c]	1.7	100	A S D L N A
			G G G G G G
1acm [d]	2.8	100	A S D L H N
			G G G G G G
1cca	1.8	66.7	E D T T P D
			G G G g
1fdl	2.5	50	T P P S V F
			B E E E
1dhr	2.3	71.4	Q V V T T D G
			E E E E E T T
1dr1	2.2	83.3	E D G I Q Y
			E T T E E E
3psg	1.6	100	V Q Y P L S P S
			E E E E E h H H
1hah	2.3	100	A A H C L L
			h H H H H E
2ach	2.7	83.3	S L E F R E
			H h E E E E
1asy	3.0	71.4	L R Q Q A S L
			E E e S S e E

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

[c] When this fragment was used, d was not used to derive the model.

[d] When this fragment was used, c was not used to derive the model.

Table 8. *Alignment of peptide fragments on the target sequence of Serine Proteinase B Complex with potato inhibitor (4sgb), residues 189-216.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			189 198 208 216
4sgb	2.1		C A E P G D S G G P L Y S G T R A I G L T S G G S G N C
			t T T t T T e E E E E T e E E E E E E E E E E E E T
1gof	1.7	100	S A Q S G N
			t T T t
1ton	1.8	88.9	A G D S G G P L I
			T T t T T e E E E
2por	1.8	83.3	T Y S A G A
			E E E E T e
2aai	2.5	83.3	S N G K C L
			E T e E E E
4sgb	2.1	100	R A I G L T
			E E E E E E
1hcg	2.2	83.3	G I V S W G
			E E E E E E
1xnd	1.8	57	F S G S Y N P
			E E E E E E E

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 9. *Alignment of peptide fragments on the target sequence of Azurin (5azu), residues 3-26.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			3 12 22 26
5azu	1.9		C S V D I Q G N D Q M Q F N T N A I T V D K S C
			e E E E E E e t T T t B S e E E E E t T T t
1atn	2.8	50	D I V L I Q
			e E E E E E
1aak	2.4	50	V I F G P D
			E E E e t T
7pcy	1.8	83.3	G G D D G S
			e t T T t
5azu	1.9	100	Q M Q F N T
			T t B S
7pcy	1.8	75	F V P N N I T V G
			E E S e E E E E E
1cse	1.2	100	A V D S N S
			E E t T T t

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 10. *Alignment of peptide fragments on the target sequence of Insulin (6ins), residues 7-36.*

Source of Fragment (pdb Code)	X-ray Crystal Resolution (Å)	Percent Homology [a]	Sequence of Target Peptide and the Fragments [b]
			7 16 26 36
6ins	2.0		C G S H L V E A L Y L V C G E R G F F Y T P K G I V E Q C C
			t h H H H H H H H H H H H h T t E E E t T h H H H H H H h
6rlx	1.5	100	C G R E L V R A
			h H H H H H H H
1coy	1.8	83.3	V A A L R L
			H H H H H H
1anw	2.5	100	L L L L C G
			H H H H H h
6ins	2.0	100	V C G E R G F F
			H H h T t E E
1ant	3.0	88.9	N I F L S P L S I
			E E E t T h H H
1anx	2.5	62.5	S G N L E Q L L
			h H H H H H H H

[a] Percent homology (identical or residues with substitution weights ≥5 in Ref. 29) of the fragments compared to the target sequence.

[b] Underlines indicate the portion of the main chain coordinates copied. Shaded areas show secondary structures predicted by the method of Kabsch and Sander [28]. Sequence numbering is from the crystal structure.

Table 11. *RMS (Å) deviations of the modelled disulfide loop structures from the corresponding crystal structures.*

Name of the protein: residue number		rms (Å) c-α	main-chain only	all atoms
1. 178L:127-154		1.57	1.65	1.63
2. 1ezm:270-297		0.73	0.87	0.87
3. 1hyp:29-67	c	2.02	2.06	2.41
	d	1.93	1.89	2.36
4. 1poc:61-96		2.82	2.83	2.83
5. 2exo:167-199		1.89	1.98	1.98
6. 3psg:249-282	c	2.58	2.52	2.47
	d	2.74	2.77	2.70
7. 4sgb:191-220		2.04	2.03	2.12
8. 5azu:3-26		2.61	2.71	2.71
9. 6ins:7-36		2.77	2.73	2.77

c and d, as described in Tables 4 and 7.