

[Chem. Pharm. Bull.]
36(12)4652—4658(1988)

Computer Screening and Visualization of Hydrophobic Core of Protein

YOJI UMEZAWA^{*,a} and HIDEAKI UMEYAMA^b

*Institute of Microbial Chemistry,^a 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141,
Japan and School of Pharmaceutical Sciences,^b Kitasato University,
Shirokane, Minato-ku, Tokyo 108, Japan*

(Received May 9, 1988)

We have succeeded in developing a computer program (SCROIL) for screening the hydrophobic core residues from the three-dimensional structure of a protein and visualizing the hydrophobic core residues by computer graphics. Minimal spanning trees are described in connection with the side chains of the hydrophobic core residues in these displays.

Keywords—hydrophobic core; computer screening; visualization; protein; minimal spanning tree

The three-dimensional structure of a protein in aqueous solution generally has hydrophobic residues located inside the molecule and hydrophilic residues placed at the protein surface. The hydrophobic amino acid residues energetically stabilize the protein structure as the core residues in the hydrophobic interactions.^{1,2)} Therefore, it is supposed that screening of the hydrophobic core residues could be useful in elucidating the folding process and the structural dynamics of proteins. Moreover, extraction of the hydrophobic core residues which are involved in the structurally conserved region³⁾ from the three-dimensional structure of a protein is very important in connection with amino acid sequence alignments for other structurally homologous but unknown proteins.

Methods

It is difficult to determine quantitatively the involvement of an amino acid residue in the hydrophobic core of a protein, since few physical data are available on the hydrophobic core residues. We have developed a method to determine hydrophobic residues constituting the hydrophobic core, which are located inside of the protein and are connected with other hydrophobic residues, by computer graphics.

We assumed an imaginary distance which is called the hydrophobic distance (HPOD). We used hydrophobic weights calculated from values of π (PAI)⁴⁾ or hydropathy index (HI)⁵⁾ to represent the hydrophilic and hydrophobic properties of each of the 20 amino acid side-chains. Values of PAI are 0.40 (Ala), 0.82 (Cys), -0.16 (Asp), -0.98 (Glu), 1.63 (Phe), 0.0 (Gly), -0.40 (His), 1.55 (Ile), -1.14 (Lys), 1.64 (Leu), 1.42 (Met), -0.16 (Asn), 0.77 (Pro), 0.10 (Gln), -0.90 (Arg), -0.08 (Ser), 0.33 (Thr), 1.18 (Val), 1.41 (Trp) and 0.88 (Tyr). Values of HI are 1.8 (Ala), 2.5 (Cys), -3.5 (Asp), -3.5 (Glu), 2.8 (Phe), -0.4 (Gly), -3.2 (His), 4.5 (Ile), -3.9 (Lys), 3.8 (Leu), 1.9 (Met), -3.5 (Asn), -1.6 (Pro), -3.5 (Gln), -4.5 (Arg), -0.8 (Ser), -0.7 (Thr), 4.2 (Val), -0.9 (Trp) and -1.3 (Tyr). A residue is determined to be a hydrophobic residue if the value of PAI or HI is positive (hydrophobic residues are Ala, Cys, Gln, Ile, Leu, Met, Phe, Pro, Thr, Trp, Tyr and Val in the case of PAI, and are Ala, Cys, Ile, Leu, Met, Phe and Val in the case of HI). The actual distance between the side-chains of two amino acid residues was added by the second term on the right-hand side obtained as a function of PAI or HI values as shown in the following Eq. 1.

$$\text{HPOD}(i, j) = \text{DIST}(i, j) + \text{CONST} \times (2.0 \times V_{\max} - V(i) - V(j)) \quad (1)$$

“DIST” is the minimal distance of the van der Waals surface of a heavy atom between the side-chains of two amino acid residues (i) and (j). V is the value of PAI or HI for each amino acid residues. V_{\max} is the maximal value among them. V_{\max} values are 1.64 (Leu) and 4.5 (Ile) for PAI and HI, respectively. Some values of CONST, which is a scaling factor, were assumed to scale the hydrophobic weight to the distance dimension (1.00, 1.50, 1.75 and 2.0 in the

case of PAI, 0.25, 0.50, 0.75 and 1.00 in the case of HI). HPOD values were calculated for all the pairs of residues. A pair of residues having a DIST value of more than 2.8 Å was neglected as a hydrophobic interaction pair, because the hydrophobic interaction is thought to be weak at the DIST distance of 2.8 Å, which is the diameter of a water molecule.^{6,7)} We extracted the hydrophobic core residues by applying the minimal spanning tree (MST)⁸⁾ of HPOD, because the MST method should be appropriate to determine the division value (DIVIS) of HPOD for screening of hydrophobic core residues and to classify the screened residues. The screened pairs were placed in order from the minimal HPOD value to search for MST. According to the definition of MST,⁸⁾ a pair was left out if the path made a circuit in the connection graph. The DIVIS of HPOD was determined so as to extract the hydrophobic core residues involved in structurally conserved regions³⁾ and selected from the three-dimensional structure of trypsin by computer graphics.

Results and Discussion

CONST and DIVIS were determined on the basis of serine protease family proteins (bovine trypsin (1TPP), porcine elastase (1EST) and bovine chymotrypsin (2CHA)).⁹⁾ The values of CONST and DIVIS are 1.75 and 2.25 (condition 1), respectively, in the case of PAI values, and are 0.75 and 2.85 (condition 2), respectively, in the case of HI values (Table I). However, in the case of PAI, since it was difficult to screen Ala and Thr which should mostly be included in the hydrophobic core, the PAI values of 0.45 (Ala) and 0.33 (Thr) were increased to 0.90 (Ala) and 0.80 (Thr).

In the case of HI in condition 2, the hydrophobic core residues of trypsin extracted by our SCROIL program are shown in Table II. The column of NBRC is the number of branches at the node of the residue in MST, showing the state of hydrophobic connection as simplified by MST. The NRED column is the number of the hydrophobic core residues placed less than 2.8 Å in DIST from the residue. It is considered that this number shows the local importance of the residues in the hydrophobic core.

The hydrophobic core residues extracted from trypsin, elastase and chymotrypsin by SCROIL in condition 2 are shown in Fig. 1. The amino acid sequence alignments were made by comparison of the three-dimensional structures. The top row is the standard chymotrypsin numbering. The hydrophobic residues using HI (marked ●) in trypsin indicate the hydrophobic core residues involved in structurally conserved regions and extracted from the three-dimensional structure of trypsin by computer graphics. The amino acid residues marked ▼ and ▽ are extracts by SCROIL using HI values. In the case of amino acid residues under the mark ▼, the NRED numbers of the marked residues are more than two. It is thought that the marked residues are very important to construct the three-dimensional structure of the protein because they are located in the center of the hydrophobic core. Practically all of the ▼ marked residues were in good agreement with ● marked residues which were the hydrophobic core residues of trypsin extracted by computer graphics. The ▽ mark means that the NRED numbers of the ▽ marked residue are less than three. It is not considered that the ▽ marked residues are in the center of the hydrophobic core; they could be located at the surface of the protein. Therefore most ▽ marked residues are substituted by hydrophilic residues in other proteins (for example, 75 (V), 76 (V), 129 (C), 154 (V) and 232 (C) in trypsin, 62 (L), 129 (I), 167 (I) and 221 (V) in elastase, 23 (V), 71 (F) and 80 (I) in chymotrypsin), and some of them are found in an addition or a deletion region at the protein surface (for example, 185 (L) and 221A (A) in trypsin, 126 (A), 203 (V) and 219A (L) in elastase).

TABLE I. The Determined Values of CONST and DIVIS Using PAI or HI Values

Condition	Parameter	CONST	DIVIS
1	PAI	1.75	2.25
2	HI	0.75	2.85

TABLE II. Hydrophobic Core Residues of Trypsin (1TPP) Extracted by SCROIL in Condition 2

Group	NCHT ^{a)}	Residue	NRED ^{b)}	NBRC ^{c)}	Group	NCHT ^{a)}	Residue	NRED ^{b)}	NBRC ^{c)}
1	16	Ile	4	3	1	137	Leu	7	2
1	17	Val	4	3	1	138	Ile	7	4
1	22	Cys	4	1	1	154	Val	1	1
1	27	Val	4	3	1	155	Leu	4	2
1	31	Val	5	2	1	157	Cys	5	2
1	33	Leu	9	2	1	158	Leu	5	1
1	41	Phe	4	1	1	160	Ala	3	1
1	42	Cys	4	2	1	162	Ile	6	4
1	46	Leu	10	2	1	163	Leu	4	3
1	47	Ile	8	3	1	168	Cys	6	2
1	52	Val	8	3	1	176	Ile	4	2
1	53	Val	10	4	1	180	Met	4	1
1	55	Ala	5	1	1	181	Phe	7	1
1	56	Ala	4	1	1	182	Cys	5	2
1	58	Cys	5	1	1	185	Leu	1	1
1	63	Ile	6	2	1	191	Cys	2	2
1	65	Val	10	3	1	199	Val	8	2
1	67	Leu	9	2	1	200	Val	5	2
1	83	Ile	4	1	1	201	Cys	6	1
1	85	Ala	4	1	1	209	Leu	9	1
1	89	Ile	3	2	1	212	Ile	6	4
1	90	Val	3	2	1	213	Val	4	2
1	103	Ile	7	2	1	220	Cys	3	1
1	104	Met	10	1	1	221A	Ala	2	1
1	105	Leu	8	2	1	227	Val	4	2
1	106	Ile	8	6	1	231	Val	7	1
1	108	Leu	8	2	1	235	Val	3	1
1	112	Ala	6	1	1	238	Ile	9	3
1	114	Leu	4	1	1	242	Ile	3	2
1	118	Val	4	2	2	128	Cys	1	1
1	121	Ile	6	2	2	232	Cys	1	1
1	123	Leu	7	2	3	75	Val	1	1
1	129	Ala	3	1	3	76	Val	1	1
1	136	Cys	8	3					

a) Identity of the amino acid residue in trypsin (1TPP) by standard chymotrypsin numbering. b) Number of hydrophobic core residues placed less than 2.8 Å in DIST from the residues. c) Number of branches of the minimal spanning tree at the node of the amino acid residues.

The hydrophobic core screening using HI values (condition 2) was successful for serine protease and many other proteins. A similar screening of the hydrophobic core was performed by using PAI values (condition 1). Good results were obtained for other proteins in addition to serine protease, although a small revision was necessary to the DIVIS values.

Visualization of Hydrophobic Core of Protein

We tried to visualize a hydrophobic core in a three-dimensional protein structure in condition 2, by means of a drawing of MST of HPOD in which the location of the β -carbon in each hydrophobic core residue was used as a node of MST. The hydrophobic core extracted from trypsin by SCROIL is easily observed with α -carbon plots showing the protein structure in a stereoscopic view (Fig. 2(a)). In the figure, the actual length of the branch in MST is different from the HPOD value because the location of the β -carbon is used as a node of MST to visualize the hydrophobic connection.

The stereo view of hydrophobic cores screened from γ H-crystallin (1GCR)⁹⁾ by SCROIL in condition 2 is shown in Fig. 2(b). There are three hydrophobic cores. The yellow

[illegible]

Fig. 1. Hydrophobic Core Residues Screened from Trypsin (1TPP), Elastase (1EST) and Chymotrypsin (2CHA) by SCROIL in Condition 2

The hydrophobic residues under the mark **●** in trypsin indicate the hydrophobic core residues extracted from the three-dimensional structural structure of trypsin by computer graphics. The amino acid residues under the marks **▼** and **▽** are extracts by SCROIL in condition 2. The **▼** and **▽** marks mean that the numbers of the hydrophobic core residues placed less than 2.8 Å in DIST from the marked residue are more than two and less than three, respectively.

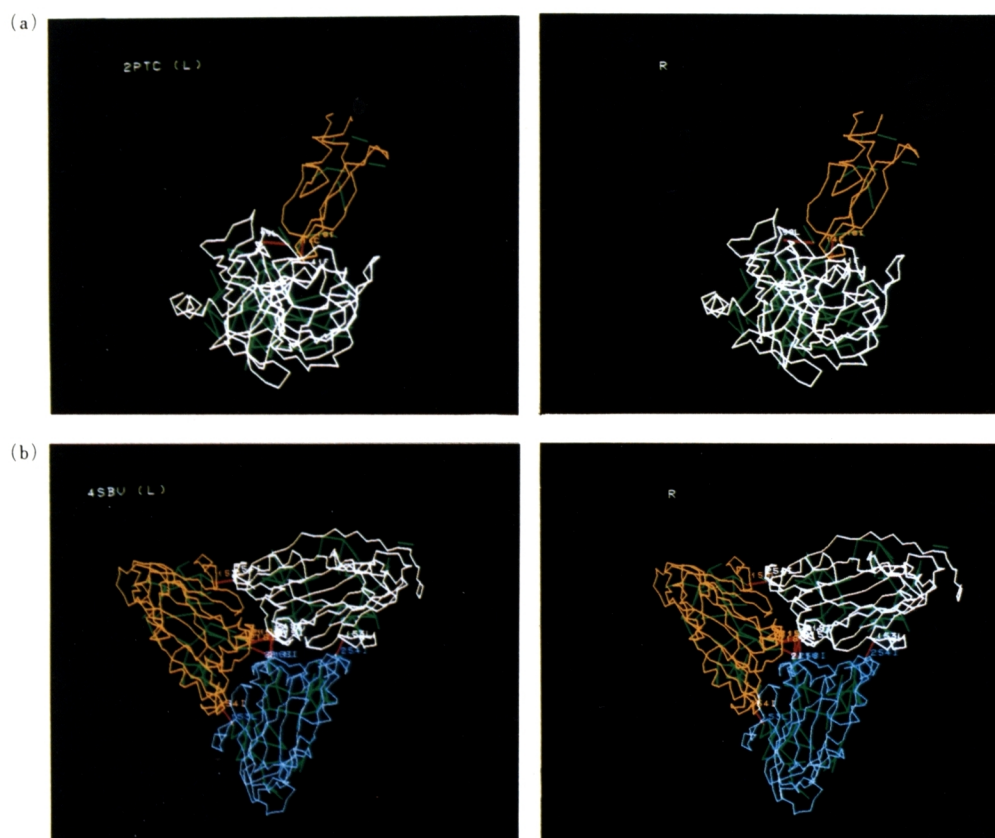


Fig. 3. Hydrophobic Cores of Complex and Oligomer

The green trees are the intramolecular hydrophobic connections, and the red trees are the intermolecular connections.

(a) A stereo view of the hydrophobic cores extracted from the complex (2PTC) of trypsin and BPTI in condition 2 with α -carbon plots of the proteins. α -Carbon plots of trypsin, white; BPTI, yellow.

(b) A stereo view of the hydrophobic cores extracted from three protein units of southern bean mosaic viral coat protein (4SBV) in condition 2 with α -carbon plots of proteins. α -Carbon plots, white, blue and yellow.

hydrophobic core of cytochrome c (1CCR)⁹⁾ was used as an example. There are two hydrophobic cores (green and red) in the protein. The hydrophobic connections of three residues, 17(I), 76(L) and 102 (L), are observed in the green core. We consider that the hydrophobic connections of these residues are those in secondary structures, because the three residues are involved in three α -helices (10—22, 68—78 and 95—111) separately. Similar hydrophobic connections in two α -helices (78—83 and 57—63) are observed in the other (red) core.

We tried to screen the hydrophobic core from the complex and oligomer. In Fig. 3, the green trees are the intramolecular hydrophobic connections, and the red trees are the intermolecular connections.

The hydrophobic cores of the complex (2PTC)⁹⁾ of trypsin and bovine pancreatic trypsin inhibitor (BPTI) are shown in Fig. 3(a). Two intermolecular hydrophobic connections, 41 (F) in trypsin & 18 (I) in BPTI and 99 (L) in trypsin & 14 (C) in BPTI, which are colored red, are observed. It is suggested that the location of the 99 (L) is important for the binding site, because the position, even though it is located at the surface of proteins, is hydrophobically

conserved in serine proteases.

The particle of southern bean mosaic virus has a shell of 180 protein subunits arranged with icosahedral symmetry around a core of ribonucleic acid. Each icosahedral unit contains three protein subunits. The hydrophobic cores extracted from the three protein subunits (4SBV)⁹⁾ are shown in Fig. 3(b). The red-colored hydrophobic connections are observed at the same position in the center (211 (V), 215 (I) and 218 (I)) and around the proteins symmetrically (153 (L) and 254 (I)). It is surmised that these intermolecular connections hydrophobically stabilize the structure of the three protein subunits together with other intermolecular interactions.

Conclusions

Screening of the hydrophobic core of a protein could be useful in connection with amino acid sequence alignments for structurally homologous but unknown proteins. Moreover, information obtained from the observation of the hydrophobic cores in the three-dimensional structure of a protein, such as the shapes of the hydrophobic connections in the protein and in secondary structures and the hydrophobic binding sites of a complex or oligomer, could be useful to elucidate the folding process of proteins.

References and Notes

- 1) W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
- 2) R. B. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972).
- 3) J. Greer, *J. Mol. Biol.*, **153**, 1027 (1981).
- 4) V. Pliska, M. Schmidt and J.-L. Fauchere, *J. Chromatogr.*, **216**, 79 (1981).
- 5) J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, **157**, 105 (1982).
- 6) K. Akahane and H. Umeyama, *Chem. Pharm. Bull.*, **34**, 3492 (1986).
- 7) C. Pangali, M. Rao and B. J. Berne, *J. Chem. Phys.*, **71**, 2975 (1979).
- 8) Minimal spanning tree means that the total value of imaginary distance of the branch is minimal in the tree. The tree is a connection graph without circuits.
- 9) The data on the three-dimensional structure were obtained from the Protein Data Bank (PDB).