Volume 211, number 2, 185–189 FEB 04375 January 1987

Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility

Jiří Novotný and Robert E. Bruccoleri

Molecular & Cellular Research Laboratory, Massachusetts General Hospital & Harvard Medical School, Boston, MA 02114, USA

Received 10 November 1986

The relationship among accessibility to an enzyme, flexibility, and limited proteolysis was explored. Regions accessible to large probes, comparable in size to proteolytic enzymes, were computed in the crystallographic structures of thermolysin, trypsinogen and ribonuclease. Positions of these accessible regions were compared with sites of autolytic/proteolytic attacks, and with locations of flexible backbone segments. All the proteolytic sites were found to be exceptionally accessible. Most of them were also flexible, but at least one prominent site in trypsinogen appeared to be rigid. Thus, surface exposure seems to be more essential to proteolysis than flexibility.

Proteolysis; X-ray structure; Accessibility; Flexibility; Trypsinogen

1. INTRODUCTION

Limited autolysis and proteolysis has been a subject of considerable experimental work, reviewed most recently by Neurath [2] and Bennett and Huber [3]. Neurath [2] observed that the sites of limited proteolysis are always located at 'hinges and fringes', that is, at exposed polypeptide chain segments that either connect two compact, globular domains, or loop out from a compact fold of a domain. Bennett and Huber [3] discussed limited proteolysis in the light of the exceptional flexibility these sites seem to possess. They reached the following important conclusion:

'Crystallographic studies of chymotrypsinogen and trypsinogen show that the catalytic residues of the active sites of the zymogens have conformations very similar to those found in active proteases. In chymotrypsinogen the substrate binding site appears to have different conformation than in the active enzyme, while this region of trypsinogen is disordered to some degree... The function of the disorder in trypsinogen is thus to prevent the zymogen from functioning at inappropriate times.'

Fontana et al. [1] recently reported on their experiments which localized autolytic and proteolytic sites in thermolysin in the most mobile segments of the molecule. They concluded that:

'flexibility of the polypeptide chain of a globular protein at the site of proteolytic attack promotes optimal binding and proper interaction with the active site of the protease'.

Thus, although both Bennett and Huber [3] and Fontana et al. [1] ascribe functional importance to flexibility, they hold different views on whether mobile segments promote, or impede, proteolysis. In view of this apparent paradox, it becomes important to reevaluate the relationship among limited proteolysis, autolysis, temperature factors (the temperature factor, or the B value, is given by $B = 8/3\pi^2 \langle r^2 \rangle$, where $\langle r^2 \rangle^{1/2}$ is the root-mean-square atomic displacement from the crystal equilibrium position [4]) and surface exposure.

Correspondence address: J. Novotný, Molecular & Cellular Research Laboratory, Massachusetts General Hospital, MA 02114, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)

This report focuses on the question of whether it is the accessibility to an enzyme, rather than backbone flexibility, that determines locations of proteolytic sites. It has been known that surface exposure and *B* factor maxima are strongly correlated [3,5,6]. Hence, if either flexibility or surface accessibility determines proteolytic sites in native (folded) proteins, the other property (that is, surface accessibility; flexibility) will show a strong correlation, too, without being requisite to proteolysis.

Large probe accessibility computations [7,8] have recently been used to locate surface points available for contacts between two macromolecules (antibody and protein antigens [9–11]). In this report, accessibility to probes comparable in size to the relevant proteolytic enzymes is computed for thermolysin, trypsinogen and ribonuclease, and the results are compared with locations of proteolytic sites and relevant temperature factors in these proteins.

2. METHODS OF COMPUTATION

The atomic coordinates and temperature factors of thermolysin [12], trypsinogen [13,14], ribonuclease A [15,16] and subtilisin [17] were obtained from the Brookhaven Data Bank [18]. In order to determine optimal radii for the probes to be used in accessibility calculations, molecular

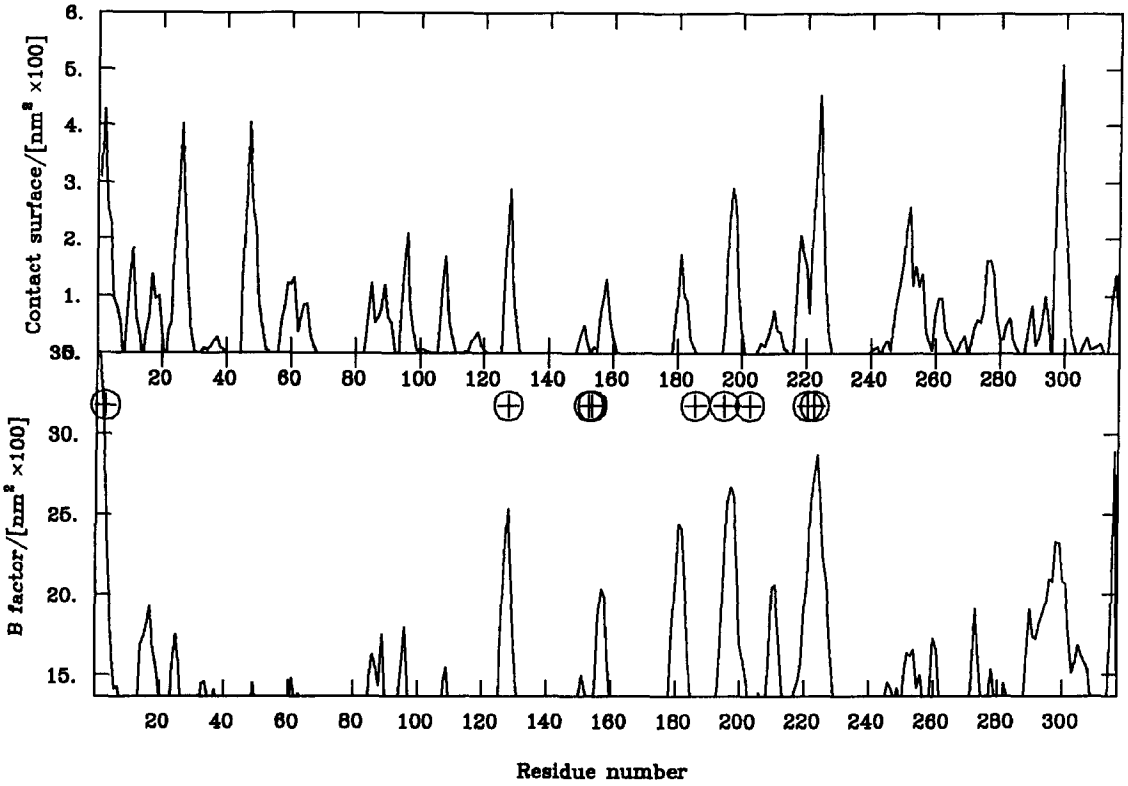

Fig.1. Profile of contact areas of thermolysin computed with a spherical probe 2 nm in radius. Contact areas (1 Å^2 equals 0.01 nm^2, see the upper panel) represent those parts of the van der Waals surface of a protein that come in direct contact with the sphere, when the sphere is rolled over the surface [8]. Residue contact areas shown were obtained as sums of computed atomic contact areas. The values were smoothed by a 7-point moving window algorithm [21] and plotted against the residue numbers. The above-average backbone *B* factors, as computed from the atomic *B* factors contained in the Brookhaven data bank [18], are given in the lower panel (the average backbone *B* value is 0.137 nm^2 or 13.7 Å^2). Peptide bonds known to be autolysed in native thermolysin [1] are indicated by crosses. Limited proteolysis of thermolysin with subtilisin [1] leads to the cleavage of three bonds. Two of them, Thr-224–Glu-225–Asp-226, coincide with one of the autolytic sites, whereas hydrolysis of the Thr-4–Ser-5 bond is unique to subtilisin. Location of the 4–5 proteolytic site is also indicated in the figure.

volumes, V, of subtilisin, trypsinogen and thermolysin were first computed from the coordinates using the Voronoi-Richards algorithm [19]. Radii, r, for the spherical enzyme-sized probes were then obtained using the formula $r = (3V/4\pi)^{1/3}$. For subtilisin-, trypsinogen- and thermolysin-like probes the radii were, respectively, 1.9, 1.8 and 2.0 nm. As the proteolytic process involves enzyme binding to peptide segments averaging 6–8 amino acid residues [20], residue sums of the computed atomic contact areas were smoothed with a 7-point moving window algorithm [21] before plotting the final large probe contact profiles. Average backbone B factors were computed as described [9].

3. RESULTS AND DISCUSSION

The results of computations, together with the locations of autolytic/proteolytic sites reported [1,22,23] are displayed in the figs 1–3. Both the large probe accessibility profiles and the autolytic/proteolytic points of these proteins can be seen to correlate very well with average backbone B factor maxima. Nevertheless, it is of particular note that the Arg-105–Val-106 bond of trypsinogen is not associated with any above-average maximum of backbone B factors in the structure solved by Kossiakoff et al. [14], as displayed in fig.3. The same is true for the structure solved independently by Fehlhammer et al. ([13], not shown). The fact that residues 104 to 107 form a well-defined secondary structure element (3_{10} helical turn [14]) is also consistent with this backbone segment not being exceptionally mobile. Yet, the Arg-105–Val-106 bond is the one most frequently cleaved in trypsinogen preparations. Higaki and Light [23] reported that the Val-neotrypsinogen represents up to 70% of commercial trypsinogen preparations, while the Ser-

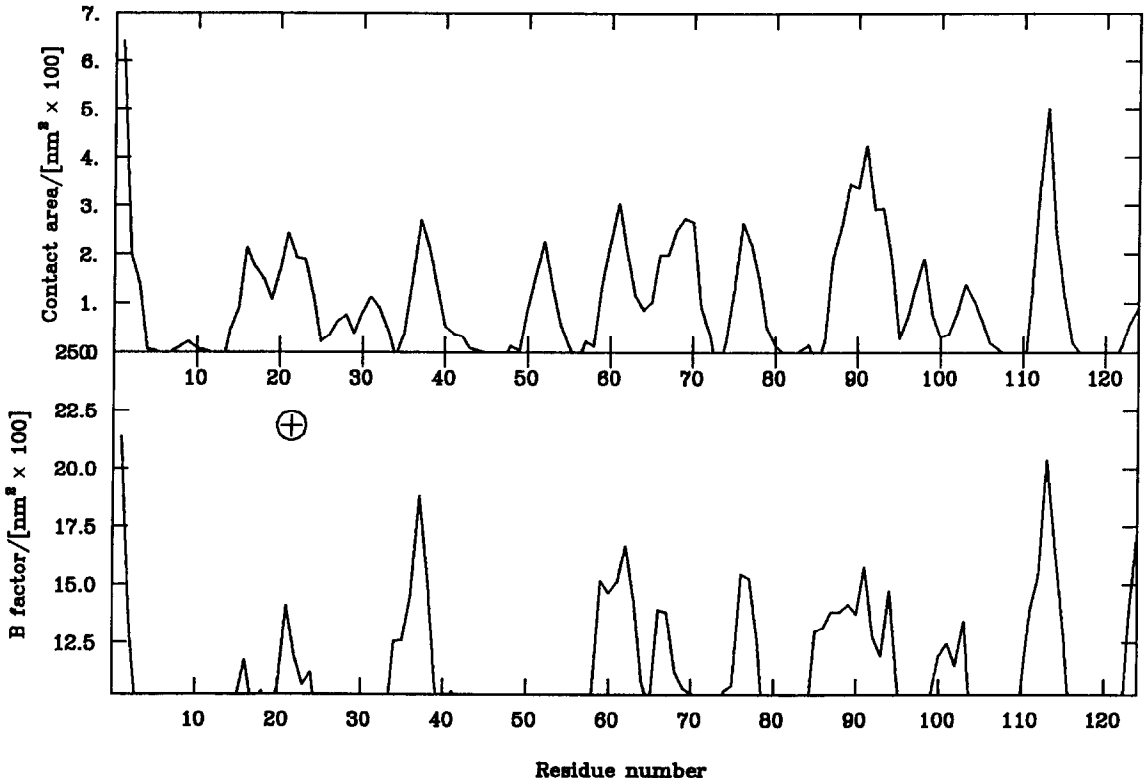

Fig.2. Profile of contact areas of ribonuclease A [15] computed with a subtilisin-sized spherical probe (radius = 1.9 nm). A plot of above-average backbone B factors [16] is also included in the lower panel. The average backbone B value is 0.103 nm^2 or 10.3 $Å^2$. Backbone B factor plots of the two ribonuclease A structures [15,16] are very similar and only one of them is shown in the figure. The location of the S-peptide cleavage point [22] is indicated by a cross.

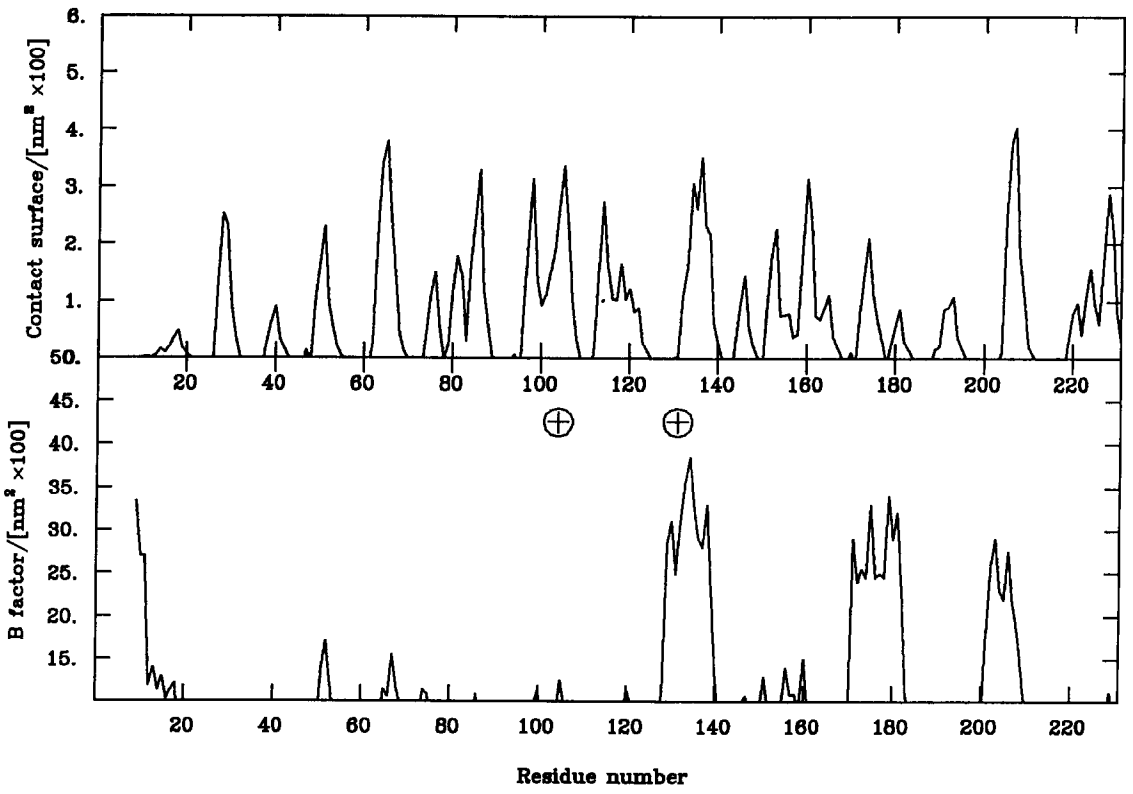

Fig.3. Profile of contact areas of trypsinogen computed with a trypsinogen-sized spherical probe (radius = 1.8 nm). A plot of above-average backbone *B* factors [14] is also included in the lower panel. The average backbone *B* value is 0.187 nm^2 or 18.2 $Å^2$. Backbone *B* factor plots of the two trypsinogen structures [13,14] are very similar and only one of them is shown in the figure. The locations of the two major autolytic sites [23] are indicated by crosses. Note that no crystallographic data exist for the N-terminal 'activation' heptapeptide, which is presumed to be completely disordered in both the crystal structures.

neotrypsinogen, resulting from autolysis of the flexible 'autocatalytic loop' at residues 131–132, is less abundant.

In summary, the data reported here, although not excluding the possible importance of segmental flexibility for enzyme-substrate binding, do nevertheless indicate that surface accessibility is as good a candidate for determining the sites of limited proteolysis as flexibility. The methodology used in this report may assist in designing experiments aimed at final clarification of the respective roles of surface exposure and mobility in proteolysis of native proteins.

ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research Grant no. N000140-86-K-0116, and the National Institutes of Health Grant no. HL-19259-11.

REFERENCES

[1] Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. and Zambonin, M. (1986) Biochemistry 25, 1847–1851.

[2] Neurath, H. (1980) in: Protein Folding (Jaenicke, R. ed.) pp.501–524, Elsevier/North Holland, Amsterdam, New York.

[3] Bennett, W.S. and Huber, R. (1984) CRC Crit. Rev. Biochem. 15, 291–384.
[4] Ladd, M.F.C. and Palmer, R.A. (1985) in: Structure Determination by X-Ray Crystallography, pp.165–166, Plenum, New York.
[5] Karplus, M. and McCammon, A.D. (1983) Annu. Rev. Biochem. 52, 263–300.
[6] Petsko, G.A. and Ringe, D. (1984) Annu. Rev. Biophys. Bioeng. 13, 331–372.
[7] Lee, B.K. and Richards, F.M. (1971) J. Mol. Biol. 55, 379–400.
[8] Richards, F.M. (1977) Annu. Rev. Biophys. Bioeng. 6, 151–176.
[9] Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R.E., Carlson, W.D., Fanning, D.W., Smith, J.A. and Rose, G.D. (1986) Proc. Natl. Acad. Sci. USA 83, 226–230.
[10] Novotný, J., Handschumacher, M. and Haber, E. (1986) J. Mol. Biol. 189, 715–721.
[11] Novotný, J. and Haber, E. (1986) Biochemistry 25, 6748–6754.
[12] Holmes, M.A. and Matthews, B.W. (1982) J. Mol. Biol. 160, 623–639.
[13] Fehlhammer, H., Bode, W. and Huber, R. (1977) J. Mol. Biol. 111, 415–438.
[14] Kossiakoff, A.A., Chambers, J.L., Kay, L.M. and Stroud, R.M. (1977) Biochemistry 16, 654–664.
[15] Borkakoti, N., Palmer, R.A., Haneef, I. and Moss, D.S. (1983) J. Mol. Biol. 169, 743–755.
[16] Wlodawer, A., Bott, R. and Sjolin, L. (1982) J. Biol. Chem. 257, 1325–1332.
[17] Wright, C.S., Alden, R.A. and Kraut, J. (1969) Nature 221, 235–242.
[18] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[19] Richards, F.M. (1974) J. Mol. Biol. 82, 1–14.
[20] Ottesen, M. (1967) Annu. Rev. Biochem. 36, 55–76.
[21] Novotný, J. and Auffray, C. (1984) Nucleic Acids Res. 12, 243–255.
[22] Richards, F.M. and Withayathil, P.J. (1959) J. Biol. Chem. 234, 371–379.
[23] Higaki, J.N. and Light, A. (1985) Anal. Biochem. 148, 111–120.