Protein Science (1994), *3*:757–768. Cambridge University Press. Printed in the USA.

Modeling studies of the change in conformation required for cleavage of limited proteolytic sites

S.J. HUBBARD,[1,2] F. EISENMENGER,[3,4] AND J.M. THORNTON[1]

[1] Department of Biochemistry and Molecular Biology, Biomolecular Structure and Modelling Unit, University College, Gower Street, London WC1E 6BT, United Kingdom
[2] European Molecular Biology Laboratory, Postfach 10 22 09, Meyerhofstrasse 1, 69012 Heidelberg, Germany
[3] Department of Crystallography, ICRF Unit of Structural Molecular Biology, Birkbeck College, Malet Street, London WC1E 7HX, United Kingdom
[4] Max Delbrück Centrum für Molekulare Medizin, Robert Roessle Strasse 10, O-1115 Berlin, Germany

(RECEIVED January 27, 1994; ACCEPTED March 14, 1994)

Abstract

Previous analyses of limited proteolytic sites within native, folded protein structures have shown that a significant conformational change is required in order to facilitate binding into the active site of the attacking proteinase. For the serine proteinases, the optimum conformation to match the proteinase binding-site geometry has been well characterized crystallographically by the conserved main-chain geometry of the reactive site loops of their protein inhibitors. A good substrate must adopt a conformation very similar to this "target" main-chain conformation prior to cleavage. Using a "loop-closure" modeling approach, we have tested the ability of a set of tryptic-limited proteolytic sites to achieve this target conformation and further tested their suitability for cleavage. The results show that in most cases, significant changes in the conformation of at least 12 residues are required. All the putative tryptic cleavage sites in 1 protein, elastase, were also modeled and tested to compare the results to the actual nicksite in that protein. These results strongly suggest that large local motions proximate to the scissile bond are required for proteolysis, and it is this ability to unfold locally without perturbing the overall protein conformation that is the prime determinant for limited proteolysis.

Keywords: flexibility; local unfolding; loop closure; nicksites; proteinase; recognition

Protein–protein molecular recognition is exemplified by the numerous serine proteinase–protein inhibitor complexes whose structures have been solved and deposited with the Brookhaven Protein Data Bank (Bernstein et al., 1977). The inhibitors possess a conserved main-chain geometry local to the scissile peptide when bound to the proteinase and when free in crystal structures (Laskowski & Kato, 1980; Read & James, 1986; Bode & Huber, 1992). Apart from this canonical recognition motif, there are common interactions observed in the complexes with the proteinases (Kraut, 1977; Read & James, 1986), including a short section of antiparallel β-sheet between inhibitor and enzyme, the oxyanion binding of the carbonyl oxygen by 2 main-chain amides of the enzyme, and the approach of the catalytic serine O_γ to the carbonyl carbon of the scissile peptide. The P_1 side chain (notation of Schechter & Berger, 1967) also binds into a primary specificity pocket. All these conserved interactions are consistent with the proposed catalytic mechanism and the binding of a good substrate. The attainment of a conformation very similar to that of the inhibitor recognition loop, in order to fulfill these binding requirements, must surely be a prerequisite for the cleavage of substrates such as limited proteolytic sites (or nicksites). This is the basic premise for making structural comparisons to inhibitors—that substrate loop binding is inhibitor-like. It has been established biochemically that inhibitors bind productively as substrates and are cleaved at the reactive site loop, although typically the binding is very tight and the cleavage very slow (Laskowski & Kato, 1980; Read & James, 1986). Indeed, BPTI is rapidly digested by a trypsin from another species, *Dermasterias imbicata* (starfish), as are several other trypsin inhibitors (Estell & Laskowski, 1980). The crystal structures of several tetrapeptide products bound to *Streptomyces griseus* proteinase A were observed to possess a very similar conformation to OMTKY3 bound to the same enzyme (James et al., 1980). A further compelling argument is the existence of recog-

Reprint requests to: S.J. Hubbard, European Molecular Biology Laboratory, Postfach 10 22 09, Meyerhofstrasse 1, 69012 Heidelberg, Germany; e-mail: hubbard@embl-heidelberg.de.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; Brcode, Brookhaven 4-letter code; OMTKY3, turkey ovomucoid third domain; RMSD, RMS deviation; CO, refers to the carbonyl carbon only. The prime notation (C′) is used in this paper to delineate altered positions of such atoms in the backbone (see Figs. 2, 3).

nition subsites leading to strong subsite specificity of some proteinases toward substrates and inhibitors alike, suggesting a common binding mode.

Nicksite loops do not, in general, possess a conformation in their crystal structures similar to the conserved inhibitor binding loop (Hubbard et al., 1991), and considerable structural change would be necessary to alter their conformations to that of the inhibitors. This is a very complex biological recognition problem involving the docking of 2 molecules, one of which requires a major conformational change. Ideally, we would like to conduct modeling experiments where the native nicksite-containing protein is run in a simulation (e.g., molecular dynamics), the nicksite loop is allowed to unfold until it adopts an inhibitor-like conformation and then is docked with the target enzyme. However, such a simulation would be prohibitively long and requires a highly effective potential and a very complicated protocol. We have attempted to reduce the problem to something more tractable by conducting loop-closure modeling experiments designed to test whether such a conformation can be reasonably introduced into protein nicksite regions without perturbing the rest of the protein. Resultant models were then tested for self-consistency and suitability for cleavage by means of a set of filters designed to eliminate closed-loop models that do not retain the inhibitor-like conformation and possess bad intra- and intermolecular contacts. The main aim is to assess the minimum amount of change necessary to achieve this goal, if indeed it is possible. Further to this, the ability of all possible tryptic sites (at lysines and arginines) in elastase were similarly modeled and tested to examine the ability of all such sites in 1 protein to adopt a cleavable conformation, and to find out why the true nicksite at arginine 125 is the preferred one.

Methods

Data

The 8 tryptic-limited proteolytic sites used in this study are listed in Table 1, taken from Hubbard et al. (1991). The trypsin and BPTI coordinates were taken from the Brookhaven entry 2ptc (Marquart et al., 1983). In addition to the nicksites, the experiments were also performed with elements of regular secondary structure generated using 20-residue segments of alanine built into right-handed α-helix and antiparallel β-strand using CHARMm (Brooks et al., 1983). The ϕ, ψ pairs used were $\phi = -57°$, $\psi = -47°$ for α-helix, and $\phi = -139°$, $\psi = 135°$ for antiparallel β-strand.

Overview of methods

Loop modeling in protein structures has been addressed using knowledge-based approaches (Blundell et al., 1988; Chothia et al., 1989), pattern matching of distance matrices (Jones & Thirup, 1986), ab initio approaches using conformational searching (Bruccoleri & Karplus, 1985, 1987; Fine et al., 1986; Moult & James, 1986), simulated annealing (Collura et al., 1993), scaling-relaxation techniques (Zheng et al., 1993), and combinations of several of these (Martin et al., 1989). For our problem, we need to introduce a predetermined conformation into a native loop without perturbing the rest of the structure. To accomplish this, we have combined a restricted conformational search with a loop-closure algorithm tested upon a set of known tryptic nicksites. Multiple conformations of each given loop were systematically generated, but the central 4 residues about the scissile bond were restricted to the inhibitor (BPTI) conformation, thus "breaking" the loop. Loops with close end-to-protein distances were then "closed" using a loop-closure algorithm, making minimal changes in the ϕ and ψ angles along the loop. The "closed" loops were then subjected to a series of filters to ensure that the inhibitor-like conformation had been retained and no bad intramolecular main-chain contacts had been introduced during the closure process. Further to this, it should be possible to "dock" the remodeled nicksite protein into the trypsin active site without generating many bad intermolecular contacts.

Generation of loops for attempted closure

For each nicksite and the 2 secondary structures, loops were generated and tested for 3 different loop lengths: 8, 10, and 12 residues centered about the scissile bond. In all cases, the P_2–P_2' segment was set to the corresponding ϕ, ψ values found in the BPTI binding loop. The remaining torsion angles were systematically set to 1 of 6 possible ϕ, ψ pairs as defined by Rooman et al. (1992) so that all possible combinations were generated. Specifically, the 6 ϕ, ψ pairs used were A (−64°, −42°), C (−87°, −4°), B (−123°, 129°), P (−70°, 138°), G (77°, 20°),

Table 1. *Tryptic limited proteolytic sites used in this study*

Protein	P_1 residue	Brcode	Resolution (Å)	References	
				Proteolytic	Structural
Staphylococcal nuclease	Lys 48 Lys 49	2sns	1.5	Taniuchi et al., 1967; Taniuchi & Anfinsen, 1968	Cotton et al., 1979
Ribonuclease	Lys 31 Arg 33	5rsa	2.0	Winchester et al., 1970	Wlodawer et al., 1982
Trypsinogen	Arg 117 Lys 145	1tgn	1.65	Bode et al., 1976	Bode et al., 1976; Kossiakoff et al., 1977
Elastase	Arg 125	3est	1.65	Ghelis et al., 1978	Meyer et al., 1988
Calmodulin	Lys 77	3cln	2.2	Drabikowski et al., 1982	Babu et al., 1988

and E (107°, −174°). These were chosen to represent the range of allowed dihedrals in conformational space. Thus, for the 8-residue loop tests, the conformations generated would be "AAFFFFAA," "CAFFFFAA," "BAFFFFAA" through to "EEFFFFEE," where F represents the "fixed" BPTI values, totaling 1,296 combinations. Similarly 46,656 and 1,679,616 loops were built for 10- and 12-residue loops. Thus, the closure distance, d_{close}, to close the loop is the distance between the old and new positions of the C^{α} of the last residue in the loop. Subsequent loop closures were attempted only if the distance, d_{close}, was less than a specified acceptance distance, d_{accept}, set to be 8.0, 6.0, and 3.2 Å for 8-, 10-, and 12-residue loops, respectively. If this criterion was satisfied, the closure was attempted. These values were chosen to allow the maximum number of closures to be attempted, but still allowing the algorithm to complete in a reasonable time. For those closures with large values of d_{close}, the closure typically fails due to introducing large changes in the loop, thus distorting the structure away from the BPTI target conformation. All side chains were ignored during the closure and then replaced with respect to the appropriate N-C^{α}-CO atoms upon completion.

Loop closure algorithm

The ring closure algorithm of Sklenar (Sklenar et al., 1986; Sklenar, 1989) was used, as implemented by one of us (F.E.). For loop closures where the number of torsion angles is exactly 6, the loop may be closed by the analytical procedure of Go and Scheraga (1970). However, we need to be able to close larger loops, and the Go and Scheraga algorithm is unable to do this. To solve the problem we introduce the side condition that the torsion-angle changes required to close the loop should be minimal. For a closure of length n residues (where $n > 4$), $2(n-1)$ torsion angles are allowed to vary as illustrated in Figure 1. The ϕ angle of residue 1 and the ψ angle of residue n, along with all ω torsion angles, are kept fixed. All ϕ and ψ angles in between are allowed to change. An example closure is shown in Figure 2, where $n = 5$. The original nicksite loop has a set of torsion angles ϕ_i, ψ_i ($i = 1, 5$), as shown in Figure 2A. If these angles are set to ϕ_i' ($i = 2, 5$) and ψ_i' ($i = 1, 4$), then the N,C^{α},CO triplet of the nth residue will be displaced from its original position (Fig. 2B). It must be generally possible, by finding a new set of torsion angles ϕ_i'' ($i = 2, 5$) and ψ_i'' ($i = 1, 4$), to bring the N,C^{α},CO triplet of main-chain atoms back to its original orientation, as shown in Figure 2C. If the torsion angle changes

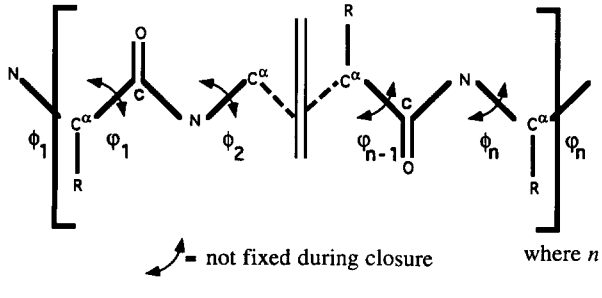

Fig. 1. Residues allowed to vary during the loop-closure process. These torsion angles are indicated by double-headed arrows across the rotable bond.

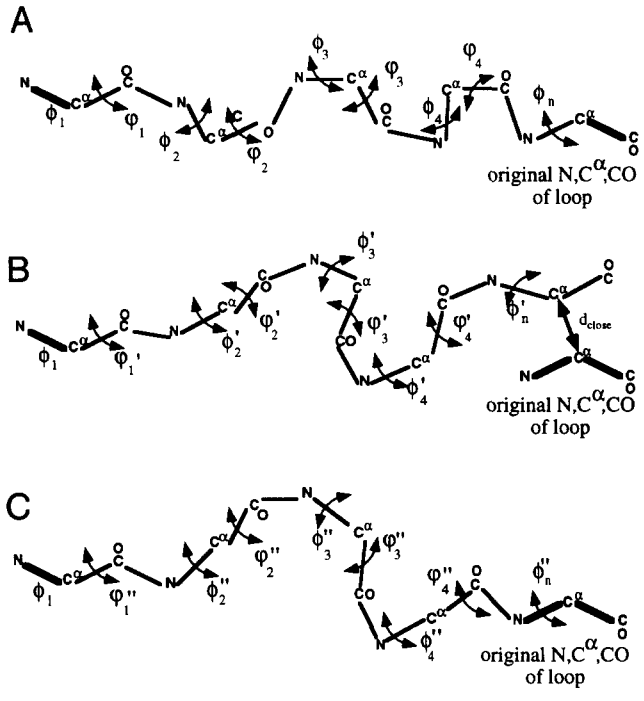

Fig. 2. Example closure for a loop of length 5 residues. Rotable bonds are indicated by double-headed arrows.

in going from Figure 2B to Figure 2C are small, then this new conformation will closely resemble that of the conformation set in Figure 2B. The problem for any given closure is to find the set of torsion angles ϕ_i'', n_i'' that satisfy the closure condition. To formulate this condition, a column vector $\vec{F}$ is considered. Figure 3 illustrates how the components of this vector are derived. The triplets of main-chain atoms N,C^{α},CO and N′,C^{α}′, CO′ represent the nth residue in the loop before and after the new torsion angles (ϕ_i', ψ_i') have been set. Let nitrogen atoms N and N′ represent the origins of 2 coordinate systems, and the vector joining them $\vec{R}$. The x-axes lie along the N-C^{α} bonds, the y-axes in the plane of the main-chain triplets, and the z-axes along the cross product of these two. There must exist a unit vector, $\vec{S}$, about which a rotation of θ relates the 2 axis systems. In-

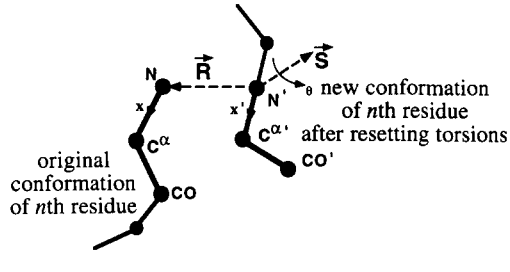

Fig. 3. Definition of the vectors $\vec{R}$ and $\vec{S}$. The vector $\vec{R}$ links the origins (the nitrogen atoms) of the 2 coordinate systems. After application of $\vec{R}$, there must be a vector $\vec{S}$ about which a rotation of θ will coincide the 2 coordinate systems. The x-axis lies along the N-C^{α} bond, the y-axis in the plane containing the main-chain triplets (N,C^{α},C′), and the z-axis along the cross product of the x- and y-axes. The vector $\vec{\sigma}$ is the simple product of θ and $\vec{S}$.

deed, when shifting the origin of the system N′,C$^{\alpha}$′,CO′ by $\vec{R}$ and rotating the coordinate axes around $\vec{S}$ by θ, the 2 systems coincide. If we define a vector, $\vec{\sigma}$, which is equal to $\theta\vec{S}$, then we may define $\vec{F}$ as consisting of the 3 components of $\vec{R}$ and the 3 components of $\vec{\sigma}$, giving 6 independent variables in total. Hence, when the loop is successfully closed:

$$F = \sqrt{R_x^2 + R_y^2 + R_z^2 + \sigma_x^2 + \sigma_y^2 + \sigma_z^2} = 0. \quad (1)$$

Equation 1 represents the closure condition that must be satisfied. If we now consider the column vector $\vec{F} = \{F_i, i = 1 \text{ to } 6\}$ composed of the components of $\vec{R}$ and $\vec{\sigma}$ as a nonlinear function of the $2(n-1)$ torsion angles given by $\vec{v} = \{v_j, j = 1 \text{ to } 2(n-1)\}$, then for a satisfactory closure the problem may be restated as

$$\vec{F}_i(\vec{v}) = 0 \qquad i = 1 \text{ to } 6. \quad (2)$$

For small deviations Δv from the initial set of starting torsion angles v_0 ($\equiv \phi_i', \psi_i'$), the components of $\vec{F}(\vec{v})$ can be expanded as a Taylor series, giving

$$F_i(\vec{v}) = F_i(\vec{v}_0) + \Delta\vec{v}^{\mathrm{T}}\left(\frac{\partial F_i}{\partial \vec{v}}\right)_{\vec{v}_0} + \ldots \qquad i = 1 \text{ to } 6. \quad (3)$$

The superscript T denotes a row vector instead of a column vector. Using only the linear terms from this expansion, and including Equation 2, we obtain

$$\underline{A}\Delta v = -\vec{F}(\vec{v}_0), \quad (4)$$

where $\underline{A}$ is the matrix with the elements

$$A_{i,j} = \left(\frac{\partial F_i}{\partial v_j}\right)_{\vec{v}_0} \qquad i = 1 \text{ to } 6 \quad \text{and} \quad j = 1 \text{ to } 2(n-1). \quad (5)$$

Following the algorithm of Sklenar et al. (1986), these partial derivatives with respect to each torsion angle v_j are calculated as follows:

$$\frac{\partial \vec{R}}{\partial v_j} = \vec{u} \times (\vec{R} - \vec{r}_j) \qquad i = 1, 2, 3 \quad \text{for} \quad \frac{\partial F_i}{\partial v_j} \quad (6)$$

and

$$\frac{\partial \vec{\sigma}}{\partial v_j} = \vec{u}_j \qquad i = 4, 5, 6 \quad \text{for} \quad \frac{\partial F_i}{\partial v_j}, \quad (7)$$

where $\vec{u}_j$ is the unit vector giving the direction of the jth single bond, and the vector $\vec{r}_j$ refers to any point on this bond.

Generally, there are an infinite number of solutions to this closure problem, as stated in Equation 4. According to Sklenar (1989), the problem is restricted by applying extra conditions to Δv to minimize the changes, both with respect to the torsion angles and with respect to the energy. These are expressed as:

$$\Delta\vec{v}^{\mathrm{T}}\Delta\vec{v} + w\Delta\vec{v}^{\mathrm{T}}\vec{g} \rightarrow \min, \quad (8)$$

where w is a weighting factor, and $\vec{g}$ represents the changes in energy of the system, as shown below:

$$g_j = \frac{\partial E}{\partial v_j}, \qquad j = 1 \text{ to } 2(n-1). \quad (9)$$

Here, we use a notional energy term based purely on geometry that restricts the algorithm from introducing large conformational changes in torsion angles and C$^{\alpha}$ positions. This term is given below:

$$E = c_1 \sum_{j=1}^{2(n-1)} (v_j^0 - v_j)^2 + c_2 \sum_{k=1}^{N_k} (\vec{P}_k^0 - \vec{P}_k)^2. \quad (10)$$

Hence the "energy" term $\vec{g}$ is calculated as a function of the amount of displacement incurred by each rotation about a bond j rather than a more complex (and more computationally expensive) potential, as used in molecular mechanics programs. Rotations about bonds near the N-terminus move more atoms than bonds near the C-terminal end of the loop and are more heavily constrained. Vector $\vec{g}$ is calculated from the derivatives of the "energy" term in Equation 10, where N_k is the number of C$^{\alpha}$ positions P_k altered by rotations around the jth single bond, and v_j^0 and P_k^0 denote the initial dihedral angles and C$^{\alpha}$ positions, respectively. The weighting factors c_1 and c_2 in Equation 10 were empirically found to be efficient if they are calculated as $c_1 = (|v_j^0 - v_j|)/30$ with the angles in degrees and $c_2 = 0.001(|\vec{P}_k^0 - \vec{P}_k|)$.

Introducing Equation 2 as side conditions by Lagrange multipliers to the partial derivatives, $\vec{\lambda} = \{\vec{\lambda}_j, j = 1 \text{ to } 2(n-1)\}$, together with the variation $\Delta\vec{v}$ gives:

$$\Delta\vec{v} + w\vec{g} + \underline{A}^{\mathrm{T}}\vec{\lambda} = 0. \quad (11)$$

Together with Equation 4, this gives:

$$(\underline{A}\underline{A}^{\mathrm{T}})\lambda^{\mathrm{T}} = \vec{F}(\vec{v}_0) - w\underline{A}\vec{g}. \quad (12)$$

For a given starting conformation, the set of Lagrange multipliers $\vec{\lambda}$ can be found from the system of linear equations given by Equation 12. Substituting these values back into Equation 11 gives the set of torsion angle changes Δv_j to close the loop. Because only the linear terms of the Taylor expansion are considered, the changes Δv do not, in general, lead to $F = 0$. Thus, the procedure is repeated up to a maximum of 20 steps to attempt to bring $F < 10^{-8}$. The weighting factor w is set to be the same as F throughout the closure so that the torsion angles become more constrained as the closure distance gets smaller.

Filtering and testing loop closure

The overall modeling process is subjected to 4 filtering stages described below.

1. The loop-closure algorithm, combined with the requirement that d_{close} is less than d_{accept}, forms the first filter, and unsuccessful closures were rejected at this stage. There are 4 possibilities for a closure to fail:

   i. Proline residue ϕ angles are unrestrained during the closure, and hence new torsion angles may be introduced outside of the usually accepted range of $-30°$ to $-90°$. Any closed loops with proline ϕ angles outside of this range are rejected.

ii. The closure algorithm fails to find a solution to one of the linear equations from Equation 12.
iii. More than 20 steps are made in attempting to bring $F < 10^{-8}$ and close the loop.
iv. More than 100 trials are made during 1 step in attempting to find a new set of torsion angles that satisfy the closure and weighting constraints. For each trial the weighting value w is halved, permitting larger Δv's.

In practice, closures very rarely fail due to ii, iii, or iv. If a loop contains a proline, then the closure algorithm often rejects closures due to failure of condition i.

2. The second filter rejects loops where the backbone atoms (N,C^{α},C') of the P_2–P_2' segment possess an RMSD from the equivalent BPTI segment greater than 0.5 Å after closure. All least-squares fitting was done using the method of McLachlan (1979). Our previous analysis (Hubbard et al., 1991) showed that the P_2–P_2' segments of known serine proteinase inhibitors typically differ by less than 0.5 Å, and this region is highly conserved in terms of structure throughout all known serine proteinase inhibitor binding loops.

3. Bad intramolecular contacts were introduced into many of the closed-loop models, particularly from the side chains that were replaced in their original conformations. However, we wish only to eliminate those models with poor main-chain geometry at this stage. Hence, a third filter rejects closures with bad intramolecular main-chain to main-chain atomic contacts, defined as less than 2.5 Å. Those closed loops with more than 2 such contacts were rejected (there are no such contacts in BPTI). Contacts were calculated from all atoms in a loop to all other atoms in the protein (excluding intraresidue and adjacent residue contacts).

4. Finally, the modeled nicksite proteins were "docked" into the trypsin active site by superposition upon the BPTI loop, using the resultant transformation matrices obtained when least-squares fitting the P_2–P_2' segment of the closed loop onto the corresponding segment of the trypsin-bound BPTI. Modeled complexes possessing more than 10 C^{α}–C^{α} intermolecular contacts less than 4.0 Å or more than 5 intermolecular contacts involving loop main-chain atoms less than 2.5 Å were rejected, thus eliminating untenable putative complexes due to bad intermolecular geometry.

All the successful loop closures for each nicksite were then clustered, to determine the number of unique solutions, by calculating the RMS difference in torsion angles along the loop (excluding P_2–P_2') for all pairs, and then grouping the loops using single-linkage clustering with a cut-off of 20°.

Further analyses of generated closed loop models

The atomic contacts made between the modeled nicksite proteins for the lysine 48 nicksite in staphylococcal nuclease (12-residue closures) and trypsin were examined in detail. Contacts were classified in terms of their hydropathic type (apolar, polar, or mixed less than 4.0 Å) and hydrogen bonding. For this purpose, all oxygen and nitrogen atoms were considered to be polar and all others apolar. Van der Waals contacts were defined as those whose distances were less than the sum of the van der Waals radii of the 2 contacting atom types. The radii used were those of Chothia (1976): tetrahedral carbons 1.80 Å, trigonal carbons 1.76 Å, tetrahedral nitrogens 1.65 Å, trigonal nitrogens 1.50 Å, oxygens 1.40 Å, and sulfurs 1.85 Å. To calculate hydrogen bonds, all polar hydrogens were built according to the definitions of Baker and Hubbard (1984), using a standard N···H bond length of 1.0 Å. A hydrogen bond was deemed to be made if the acceptor–donor distance was less than 3.6 Å and the hydrogen bond angle in the range 90–180° (angle at the hydrogen where defined, otherwise at the oxygen).

Putative elastase nicksites

The closure algorithm was applied to all possible tryptic sites (at lysines and arginines) in 1 protein, elastase, for 8-, 10-, and 12-residue loops about the putative scissile peptide, using the same protocol as before. Further to this, the relative solvent accessibility of each putative P_1 residue prior to modeling was calculated, using an implementation of the Lee and Richards (1971) algorithm. The same van der Waals radii as stated above and a probe size of 1.4 Å were used. Residue accessibilities were expressed as percentages of the accessible surface of that residue type in an extended Ala-X-Ala tripeptide (see Hubbard et al., 1991). The total change in absolute accessible surface of non-loop residues was also calculated by comparing the accessibility of the native protein with and without the loop section. This was done to assess the amount of surface that the loop is directly responsible for burying. Main-chain–main-chain hydrogen bonds made between the putative nicksite loop and the rest of the protein in its native conformation were also calculated.

Results

Loop closure and filtering

The numbers of successful closures for each nicksite and regular secondary structure through the loop closure protocol and subsequent passes through the filters are shown in Table 2. The algorithm was able to close the rebuilt nicksite loops in all cases, apart from the β-strand conformation, where no successful closures were produced for any of the 3 loop lengths. These results show that it was generally possible to model a conformation similar to an inhibitory one into a resident loop in a folded nicksite protein without deforming more than just the local structure. In most cases, loops were produced that could also pass all the filtering stages. Hence, these closed loop models possessed a loop that retained the BPTI inhibitor loop main-chain conformation, did not introduce bad main-chain contacts into the protein, and could successfully be "docked" with trypsin without producing large numbers of bad intermolecular contacts. For most nicksites this was achievable using 10 or more residues, although it proved impossible in all instances using only 8. Clearly, some sites were more easily modeled than others because they ultimately produced more unique solutions. Using loops of length 12 almost always produced more successful closures than with 10. In the 1 exception (the trypsinogen site at lysine 145), most closures failed to pass the fourth filter, and the only successful closure was obtained using a 10-residue loop, which only passed this filter by 1 contact. Generally, however, longer loops produced more successful closures.

An example closure is shown in Figure 4A. The original nicksite loop of staphylococcal nuclease is shown (dark shading) along with the remodeled loop (white shading), and the P_1 residue is indicated. The loop is brought into a more open position

Table 2. *Loop closure data for tryptic limited proteolytic sites*[a]

Brcode	P_1 residue	Loop length	Loop end-end distance (Å)	(a)	(b)	(c)	(d)	(e)
2sns	Lys 48	8	9.5	30	0	0	0	
		10	9.5	642	188	152	67	26
		12	8.9	4,734	2,800	1,483	747	222
2sns	Lys 49	8	14.6	0	0	0	0	
		10	13.0	996	424	341	164	45
		12	5.6	4,908	1,943	1,541	883	174
5rsa	Lys 31	8	10.5	42	0	0	0	
		10	9.7	708	194	171	78	19
		12	14.5	4,386	1,554	412	197	59
5rsa	Arg 33	8	9.3	24	0	0	0	
		10	15.4	132	3	1	0	
		12	15.3	3,456	1,890	1,669	568	103
1tgn	Arg 117	8	11.5	24	0	0	0	
		10	16.9	1,800	909	178	11	6
		12	18.2	3,342	1,923	899	164	49
1tgn	Lys 145	8	11.4	114	29	0	0	
		10	10.3	588	165	51	1	1
		12	11.1	4,284	2,234	362	0	
3est	Arg 125	8	15.4	282	22	20	0	
		10	17.9	1,512	558	327	4	1
		12	24.7	5,004	3,185	730	169	48
3cln	Lys 77	8	11.3	60	0	0	0	
		10	15.2	1,050	76	76	34	8
		12	16.5	4,032	1,729	1,505	357	123
α-Helix	Ala 10	8	10.8	0	0	0	0	
		10	14.5	486	69	69	68	15
		12	16.8	4,956	2,717	2,405	1,395	337
β-Strand	Ala 10	8	24.2	36	0	0	0	
		10	31.1	0	0	0	0	
		12	38.0	0	0	0	0	

[a] (a) Number of loop closures attempted (filter 1). (b) Number of loops actually closed with "BPTI-like" conformation (filter 2). (c) Number of loops that passed the self-contacts filter (filter 3). (d) Number of loops that passed the "docking" filter (filter 4). (e) Number of unique loops after clustering.

by the closure algorithm. In Figure 4B, the interactions of the modeled structure with trypsin are shown. The loop fits well into the trypsin active site and no steric hindrance from more distant parts of the nicksite protein are introduced. In Figure 4C, the P_6–P_6' segment of the closed loop is shown superposed with BPTI, using the corresponding P_2–P_2' segments from the modeled loop and BPTI. The main-chain conformation of the modeled loop superposes with that of BPTI extremely well, although the P_1 lysine residues have different but equally accessible positions. Two further example closures are shown in Figure 5. In Figure 5A and Kinemage 1, 1 successful closure for the trypsinogen arginine 117 nicksite is shown. In its native conformation (the darker loop), the loop containing this site lies flat against the protein and would be unable to enter the proteinase active site. The modeled loop conformation is much more accessible.

The ribonuclease sites produced a number of successful closures, despite the P_1 residues being relatively inaccessible and situated in a helix. Previous work has suggested that accessibility is the prime determinant for limited proteolytic sites (Novotny & Bruccoleri, 1986), yet there are other putative nicksites that are more accessible in ribonuclease which are not preferentially cut by trypsin. An example closure for ribonuclease is shown in Figure 5B and Kinemage 2. The ribonuclease sites are situated in or close to a short helical segment from residues 25–32. In the modeled protein, the helix is unfolded and the loop is now protruding away from the protein bulk. In calmodulin, the nicksite lies in the center of the helical segment connecting the 2 domains (Babu et al., 1988). Although satisfactory closures were found for this nicksite, the structure of calmodulin is different when bound to a peptide ligand (Ikura et al., 1992; Meador et al., 1992) where the 2 lobes are brought closer together and the connecting helical hinge is uncoiled to permit this. This would increase the susceptibility of the loop to limited proteolysis because the intermolecular steric hindrance between enzyme and nicksite protein would be reduced.

The closures for the 2 secondary structure types are very different. Satisfactory closures were found for α-helix for both 10- and 12-residue closures. In contrast, no single β-strand closures were made, let alone passed the filtering tests. This can be explained on simple geometric grounds. As shown in Table 2, the end-end distance for 12 residues of the β-strand is 38.0 Å. The C^α–C^α distance from P_2–P_2' in BPTI is 8.9 Å; therefore, the remaining 8 residues in the loop must span 38.0 − 8.9 = 29.1 Å, or 3.64 Å per residue. The corresponding distances for 10- and

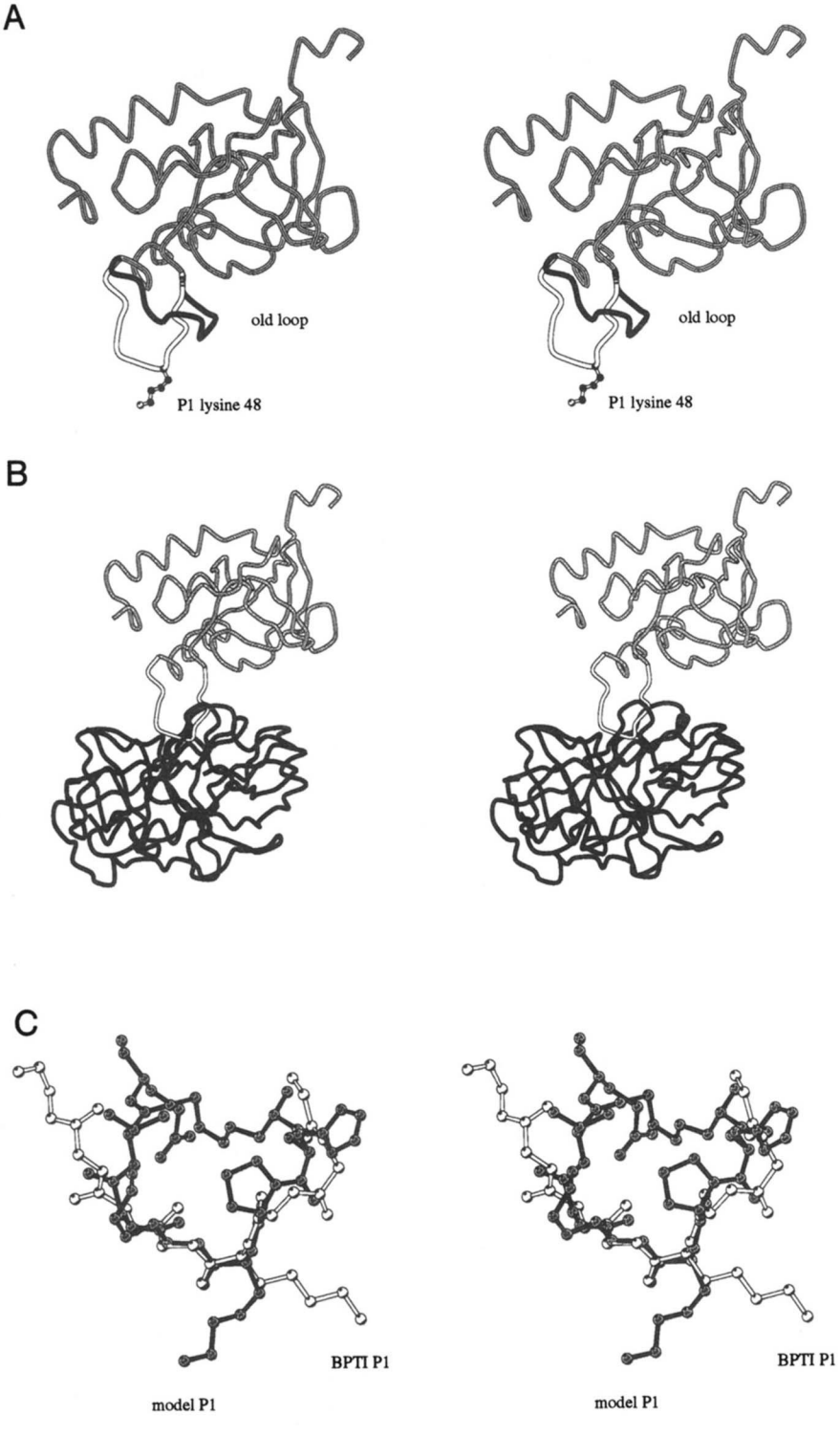

Fig. 4. Stereo plots of an example staphylococcal nuclease loop closure for the lysine 48 nicksite. **A:** C^{α} trace of staphylococcal nuclease with the original P_6–P_6' loop conformation shown in black and the modeled nicksite conformation shown in white. **B:** The modeled nicksite protein is shown "docked" into the trypsin active site (trypsin shown in black), with the nicksite loop again shown in white. **C:** Superposition of the BPTI (gray) and the modeled P_6–P_6' nicksite loop (white) achieved by superposing the P_2–P_2' main-chain atoms. For the sake of clarity, only the main-chain and P_1 side-chain atoms are shown for the nicksite loop. Plots were produced using the program MOLSCRIPT (Kraulis, 1991).

8-residue loops are 3.7 and 3.83 Å, respectively. Given that the maximum C^{α}–C^{α} distance for the extended conformation is close to 3.7 Å, it would be virtually impossible to close any loops for the β-strand, which is indeed the case. Applying the same calculation for all the other nicksites, the average C^{α}–C^{α} distance to close the loop is always under 2.0 Å, and closure is therefore quite possible.

Proteinase/modeled-substrate interactions for the 2sns Lys 48 nicksite

The interactions observed between trypsin and the "docked" 12-residue closed-loop nicksites for the staphylococcal nuclease lysine 48 site are summarized in Table 3 and compared to those observed in the BPTI/trypsin (Brcode 2ptc) complex itself. The

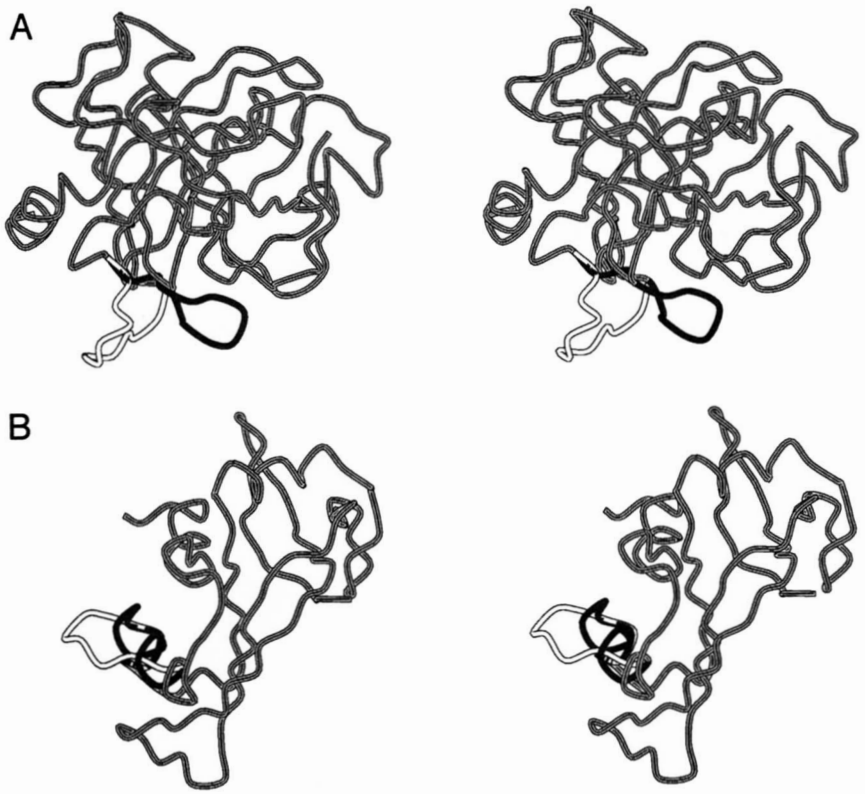

Fig. 5. Stereo plots of further loop-closure examples. C^{α} traces are shown with the original P_6–P_6' loop conformation in black and the modeled loop conformation in white. **A:** Trypsinogen with Arg 117 at P_1. **B:** Ribonuclease with Lys 31 at P_1. Plots were produced using the program MOLSCRIPT (Kraulis, 1991).

mean contact values over the 747 successful closures show that, on average, the models still make many contacts with the enzyme upon "docking." However, the majority of these contacts (about 67%) are formed from side chains of the remodeled loops, which are not considered in the modeling process. Considering contacts made by main-chain groups only, the mean value for all 747 loops falls from 375 to 125, compared to 73 of the 125 made by BPTI to trypsin in the 2ptc complex. In both the modeled and real complexes, the majority of the contacts formed are mixed, with the fewest being polar. However, the main-chain hydrogen bonding between the loop protein and trypsin is good, with a mean of 7 being formed over the 747 loop models. This is comparable with BPTI–trypsin, where 5 are formed.

One example closure-complex (number 19), with relatively few bad intermolecular contacts with the enzyme after "docking," was examined in more detail. The prime source of the bad contacts was the P_1 lysine side chain itself, which lies in a different orientation compared to BPTI (see Fig. 4C). Upon adjustment of the side-chain χ angles to the same as observed in the BPTI P_1 side chain, the number of contacts falls to values more in line with those observed in the BPTI/trypsin complex, and the number of interacting trypsin residues increases to 19

Table 3. *Intermolecular contacts in modeled complexes and BPTI–trypsin*[a]

	Contacts less than 4.0 Å					No. of residues involved	
Protein	Total	Nonpolar	Mixed	Polar	VDW	(a)	(b)
2sns Lys 48 models							
Mean values	375	133	183	60	211	15	26
2sns Lys 48 model 19							
Original P_1 angles	136	43	68	25	77	7	16
Altered P_1 angles	102	25	54	23	44	7	19
2ptc							
BPTI–trypsin complex	125	26	70	29	30	13	19

[a] The 2sns Lys 48 models refer to all the successful 12-residue loop models. (a) Number of residues from putative substrate or inhibitor interacting with enzyme. (b) Number of residues from enzyme interacting with substrate/inhibitor. VDW, number of van der Waals contacts, defined as contacts less than the sum of the contributing atomic radii.

Table 4. *Detailed intermolecular contacts in modeled and real complex*

Enzyme		Inhibitor/substrate	Contact distance (Å) 2ptc	2sns model
Phe E 41 O	→	(P_2') Arg I 17 N	2.83	2.70
Gly E 193 N	→	(P_1) Lys I 15 O	2.76	2.75
Asp E 194 N	→	(P_1) Lys I 15 O	3.06	3.01
Ser E 195 N	→	(P_1) Lys I 15 O	2.81	2.77
Gly E 214 O	→	(P_1) Lys I 15 N	3.56	3.37
Gly E 216 N	→	(P_3) Pro I 13 O	3.22	4.13
Ser E 195 OG	→	(P_1) Lys I 15 C	2.68	2.77

(see Table 3). As shown in Table 4, the conserved recognition interactions observed in inhibitor–proteinase complexes are also made in the modeled complex.

Putative elastase nicksites

The results of the loop-closure modeling experiments for all putative tryptic nicksites in elastase are shown in Table 5. In the majority of cases, no tenable solutions were found for any of the 3 loop lengths; good solutions were only found for 5 of the 15 potential sites examined. Of the 5 successful sites, the true nicksite at Arg 125 produced the most successful models with 48 unique solutions found for the 12-residue closures. Most solutions failed at the "docking" stage, where it was impossible to fit the closed loop into the trypsin active site without introducing large numbers of bad intermolecular contacts. Many of the putative nicksite segments are somewhat buried, so that although the inhibitor conformation could be successfully modeled, the loop was not sufficiently protruding from the protein environment to facilitate "docking" with trypsin. All the sites that produced successful models possessed relative accessibilities in excess of 50% for the P_1 residue. However, accessibility in itself is not enough to determine the suitability for cleavage, because Arg 217A is fully accessible yet produces no successful "docked" models. Apart from the Arg 223 site, the loops are all responsible for burying approximately the same amount of accessible surface. The true nicksite Arg 125 buried the fourth least of all the sites. However, the true site makes by far the fewest intramolecular main-chain hydrogen bonds compared to the others. Indeed, only 3 of the 6 main-chain hydrogen bonds made by the 12-residue loop are "non-local" hydrogen bonds formed from the loop to sequentially distant regions of the protein. The equivalent numbers for Arg 24, Arg 61, Arg 145, and Arg 223 are much higher, being 7, 6, 7, and 10 main-chain hydrogen bonds, respectively.

Discussion

The loop-closure algorithm was able to successfully model a fixed conformation into a protein nicksite region, without disturbing the rest of the protein structure. By attempting many such closures, for almost all of the nicksites assayed, a loop can be found that satisfies a number of criteria: namely, that the closed loop still closely resembles the proteinase inhibitor binding loop; that once closed, the loop does not introduce a large number of bad contacts within the nicksite protein; and that it can be docked (by superposition) into the proteinase active site to produce a reasonable starting point for further modeling. There remain some limitations to this modeling approach, however. The search of conformational space is quite crude; the values chosen were restricted only to those from "allowable" regions of conformational space. However, the dihedral angles of the residues were unrestricted (apart from proline) during the closure and could take up values quite distant from the 6 ϕ, ψ pairs set prior to the closure step. Secondly, motions in the rest of the protein were not considered, and it is conceivable that successful, "dockable" closures could be modeled using less local change if gross motions such as hinge-bending were permitted in distal regions of the protein. However, such modeling is beyond the scope of this work, and the larger changes involved might very well be energetically more expensive.

The modeling of regular secondary structural elements, α-helices and antiparallel β-strands, showed that only helices could produce an inhibitor-like conformation. The extended conformation of β-strands is less malleable, it would be impossible to introduce the BPTI conformation into β-strand without bringing the start and end points of the "loop" closer together. However, this is a condition imposed upon the system by our algorithm, i.e., that the end points are fixed. In vivo, the protein structure would not be expected to be so rigid. Nevertheless, it is clear that such movements would have to distort the strand arrangement of the protein in question. It might be expected, therefore, that limited proteolytic sites would only be possible in edge strands. It has also been suggested elsewhere that an extended conformation is unlikely to be suitable for a limited proteolytic site (Fontana, 1989).

It still remains difficult, in light of these results, to precisely quantify the extent of the flexibility required by limited proteolytic segments in order to be cleaved. No loops could be closed and subsequently modeled into plausible inhibitory conformations (P_2–P_2') using only 8 residues; 10 residues was the minimum number required, but generally superior closures in terms of similarity to BPTI were produced using larger segments. Also, it should be noted that it remains unclear to what degree a proteolytic substrate must mimic the inhibitor conformation. Although a certain number of interactions would be expected to be conserved (P_1 pocket interactions, oxyanion binding, antiparallel β-sheet hydrogen bonding), it is difficult to translate this into the number of residues that must occupy similar conformations. We have taken a 4-residue segment (P_2–P_2') as a template for structural comparisons because it is the longest segment that is highly structurally conserved within the inhibitor families. The inhibitor reactive-site loop segments begin to diverge in structure from P_3 outward (Hubbard et al., 1991; Bode & Huber, 1992).

It seems reasonable that proteins will have evolved in such a way to avoid large-scale cleavage by proteinases, hence the general absence of long, flexible, exposed loops that could be easily degraded. Similarly, no nicksites with conformations close to the requisite cleavage conformation are observed. The modeling experiments on all putative elastase nicksites point to local unfolding to allow subsequent docking as the prime determinant for limited proteolysis. The scissile peptide bond must be presented to the proteinase without interference from the rest of the protein. The loop that is best able to do this will be the most

Table 5. *Closure data for putative elastase nicksites*

Loop P_1 residue	Loop length	Number of unique closures[a]	Accessibility of P_1 residue (%)	"Loop" surface area lost[b] (Å^2)	Hydrogen bonds lost[c]
Arg 24	8	0	54	470	6
	10	1		567	9
	12	15		671	11
Arg 36	8	0	52	427	10
	10	0		519	12
	12	0		554	14
Arg 48	8	0	40	668	12
	10	0		730	14
	12	0		825	17
Arg 61	8	0	76	354	7
	10	0		556	9
	12	0		649	11
Arg 65A	8	0	21	728	10
	10	0		803	10
	12	5		878	11
Lys 87	8	0	47	532	8
	10	0		684	10
	12	0		796	10
Arg 107	8	0	27	621	12
	10	0		793	14
	12	0		912	14
Arg 125[d]	**8**	**0**	**87**	**473**	**4**
	10	**1**		**544**	**5**
	12	**48**		**631**	**6**
Arg 145	8	0	78	280	7
	10	0		455	8
	12	15		475	10
Lys 177	8	0	37	537	7
	10	0		575	10
	12	0		682	12
Arg 188A	8	0	32	466	6
	10	0		514	7
	12	0		608	9
Arg 217A	8	0	100	466	4
	10	0		514	6
	12	0		608	10
Arg 223	8	0	53	552	10
	10	1		673	11
	12	7		795	15
Lys 224	8	0	23	558	12
	10	0		720	15
	12	0		918	17
Arg 230	8	0	13	604	16
	10	0		766	20
	12	0		859	22

[a] Unique closures refers to the number of distinct loops passing all filtering tests after clustering.
[b] "Loop" surface lost refers to the amount of surface-accessible area uncovered upon removal of the loop.
[c] The number of hydrogen bonds lost on removal of the loop refers only to main-chain···main-chain hydrogen bonds.
[d] The results for the actual nicksite (Arg 125) are shown in boldface.

susceptible to proteolysis. This explains the correlation with accessibility, flexibility, and protrusion observed previously with limited proteolytic sites (Novotny & Bruccoleri, 1986; Vita et al., 1988; Fontana, 1989; Hubbard et al., 1991). Loops that are heavily tied down by hydrogen bonding and extensive van der Waals contacts would find it harder to unfold. Accessibility and flexibility are necessary but not sufficient conditions to determine nicksites (Hubbard, 1992) because a loop may be accessible and flexible but still unable to adopt a cleavable conformation without a great deal of conformational change in the protein's structure. Hence, local unfolding must be the key factor in determining limited proteolysis. This hypothesis is supported by recent experimental work on the proteolytic stability of a neutral proteinase (Eijsink et al., 1992; Vriend & Eijsink, 1993; Hardy et al., 1994) where the rate-limiting step in inactivation by autolysis of the protein is reported to be the local unfolding of a surface loop. Mutants of this protein, which stabilize the cleaved loop, stabilize the protein against denaturation.

These results describe a method to predict limited proteolytic sites based on the hypothesis that nicksites correspond to segments that can unfold and become accessible to proteinases. The modeling procedure described herein also provides a method to expose local unfolded conformations. This could prove useful for the design of degradation-resistant proteins in which the target loops are identified and mutated by protein engineering.

Acknowledgments

S.J.H. was supported in part by a SERC/CASE award with Pfizer UK, and F.E. was supported in part by an ICRF fellowship. We thank one of the referees for pointing out the simple geometric grounds for the impossibility of β-strand closures. We also thank Simon Campbell at Pfizer UK for useful discussions.

References

Babu YS, Bugg CE, Cook WJ. 1988. Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol 204*:191–204.

Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Progr Biophys Mol Biol 44*:97–179.

Bernstein FC, Koetzle TF, William GJB, Meyer EF Jr, Brice MD, Rodgers GR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:532–542.

Blundell T, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, Overington J, Singh DA, Sibanda BL, Sutcliffe M. 1988. Knowledge-based protein modelling and design. *Eur J Biochem 172*:513–520.

Bode W, Fehlhammer H, Huber R. 1976. Crystal structure of bovine trypsinogen at 1.8 Å resolution. *J Mol Biol 106*:325–335.

Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interactions with proteinases. *Eur J Biochem 204*:433–451.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMm: A program for macromolecular energy minimization and dynamics calculations. *J Comput Chem 4*:187–217.

Bruccoleri RE, Karplus M. 1985. Chain closure with bond angle variations. *Macromolecules 18*:2767–2773.

Bruccoleri RE, Karplus M. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers 26*:136–168.

Chothia C. 1976. The nature of the accessible and buried surface in proteins. *J Mol Biol 105*:1–14.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davis D, Tulip WR, Colman PM, Spinelli S, Alzari PM, Poljak RJ. 1989. Conformations of immunoglobulin hypervariable regions. *Nature 347*:882–883.

Collura V, Higo J, Garnier J. 1993. Modelling of protein loops by simulated annealing. *Protein Sci 2*:1502–1510.

Cotton FA, Hazen EE Jr, Legg MJ. 1979. Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate calcium ion complex at 1.5 Å resolution. *Proc Natl Acad Sci USA 76*:2551–2555.

Drabikowski W, Brzeska H, Venyaminov SY. 1982. Tryptic fragments of calmodulin. *J Mol Biol 257*:11584–11590.

Eijsink VGH, Vriend G, van der Vinne B, Hazes B, van den Burg B, Venema G. 1992. Effects of changing the interaction between subdomains on the thermostability of *Bacillus* neutral proteinase. *Proteins Struct Funct Genet 14*:224–236.

Estell DA, Laskowski M Jr. 1980. *Dermasterias imbricata* trypsin 1: An enzyme which rapidly hydrolyzes the reactive-site peptide bonds of protein trypsin inhibitors. *Biochemistry 19*:124–131.

Fine RM, Wang H, Shenkin PS, Yarmush DL, Levinthal C. 1986. Predicting antibody hypervariable loop conformations: Minimization and molecular dynamics studies of MCP603 from many randomly generated loop conformations. *Proteins Struct Funct Genet 1*:342–362.

Fontana A. 1989. Limited proteolysis of globular proteins occurs at exposed and flexible loops. In: Kotyk A, Skoda J, Paces V, Kostka V, eds. *Highlights of modern biochemistry*. Zeist, The Netherlands: VSP International Science Publishers. pp 1711–1726.

Ghelis C, Tempete-Gaillourdet M, Yon JM. 1978. The folding of pancreatic elastase: Independent domain refolding and inter-domain interaction. *Biochem Biophys Res Commun 84*:31–36.

Go N, Scheraga HA. 1970. Ring closure and local conformational deformation of chain molecules. *Macromolecules 3*:178–187.

Hardy F, Vriend G, van der Vinne B, Frigerio F, Grandi G, Venema G, Eijsink VGH. 1994. The effect of engineering surface loops on the thermal stability of *Bacillus subtilis* neutral protease. *Protein Eng 7*:425–430.

Hubbard SJ. 1992. Analysis of protein-protein molecular recognition [thesis]. London: University of London.

Hubbard SJ, Campbell SF, Thornton JM. 1991. Molecular recognition: Conformational analysis of limited proteolytic sites and serine proteinase inhibitors. *J Mol Biol 220*:507–530.

Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science 256*:632–638.

James MNG, Sielecki AR, Brayer GD, Delbaere LTJ, Bauer CA. 1980. Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. *J Mol Biol 144*:43–88.

Jones TA, Thirup S. 1986. Using known substructures in protein model building and crystallography. *EMBO J 5*:819–822.

Kossiakoff AA, Chambers JL, Kay LM, Stroud RM. 1977. Structure of bovine trypsinogen at 1.9 Å resolution. *Biochemistry 16*:654–664.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Kraut J. 1977. Serine protease: Structure and mechanism of catalysis. *Annu Rev Biochem 46*:331–358.

Laskowski M Jr, Kato I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem 49*:593–626.

Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol 55*:379–400.

Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B 39*:480–490.

Martin ACR, Cheetham JC, Rees AR. 1989. Modelling antibody hypervariable loops: A combined algorithm. *Proc Natl Acad Sci USA 86*: 9268–9272.

McLachlan AD. 1979. Gene duplications in the structural evolution of chymotrypsin. *J Mol Biol 128*:49–79.

Meador WE, Means AR, Quiocho FA. 1992. Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science 257*:1251–1255.

Meyer E, Cole G, Radahakrishnan R, Epp O. 1988. Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Crystallogr B 44*:26–38.

Moult J, James MNG. 1986. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins Struct Funct Genet 1*:146–163.

Novotny J, Bruccoleri RE. 1986. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett 211*: 185–189.

Read RJ, James MNG. 1986. Introduction to the proteinase inhibitors: X-ray crystallography. In: Barrett AJ, Salvesen G, eds. *Proteinase inhibitors*. Amsterdam: Elsevier. pp 301–336.

Rooman MJ, Kocher JPA, Wodak SJ. 1992. Extracting information on folding from the amino acids sequence. *Biochemistry 31*:10226–10238.

Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157–162.

Sklenar H. 1989. [thesis]. Berlin: Academy of Science of the GDR.
Sklenar H, Lavery R, Pullman B. 1986. The flexibility of the nucleic acids I: "SIR", a novel approach to the variation of polymer geometry in constrained systems. *J Biomol Struct Dyn 3*:967–987.
Taniuchi H, Anfinsen CB. 1968. Steps in the formation of active derivatives of staphylococcal nuclease during trypsin digestion. *J Biol Chem 243*:4778–4786.
Taniuchi H, Anfinsen CB, Sodja A. 1967. Nuclease-T: An active derivative of staphylococcal nuclease composed of two noncovalently bonded peptide fragments. *Proc Natl Acad Sci USA 58*:1235–1242.
Vita C, Dalzoppo D, Fontana A. 1988. Limited proteolysis of globular proteins: Molecular aspects deduced from studies on thermolysin. In: Chaiken IM, Chiancone E, Fontana A, Neri P, eds. *Macromolecular recognition: Principles and biotechnological applications*. Clifton, New Jersey: Humana Press.
Vriend G, Eijsink V. 1993. Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases. *J Computer-Aided Mol Des 7*:367–396.
Winchester BG, Mathias AP, Rabin BR. 1970. Study of the thermal denaturation of ribonuclease A by differential thermal analysis and susceptibility to proteolysis. *Biochem J 117*:299–307.
Wlodawer A, Bott R, Sjolin L. 1982. The refined crystal structure of ribonuclease A at 2.0 Å resolution. *J Biol Chem 257*:1325–1332.
Zheng Q, Rosenfeld R, Vajda S, DeLisi C. 1993. Determining protein loop conformation using scaling-relaxation techniques. *Protein Sci 2*:1242–1248.

Forthcoming Papers

Thermodynamics of staphylococcal nuclease denaturation. I. The acid-denatured state

J.H. Carra, E.A. Anderson, and P.L. Privalov

Thermodynamics of staphylococcal nuclease denaturation. II. The A-state

J.H. Carra, E.A. Anderson, and P.L. Privalov

Receptor-binding properties of four-helix-bundle growth factors deduced from electrostatic analysis

E. Demchuk, T. Mueller, H. Oschkinat, W. Sebald, and R.C. Wade

Folding of the multidomain human immunodeficiency virus type-I integrase

D.P. Grandgenett and G. Goodarzi

Crystallization and preliminary structural studies of neurotrophin-3

J.A. Kelly, E. Singer, T.D. Osslund, and T.O. Yeates

Kringle-kringle interactions in multimer kringle structures

K. Padmanabhan, T.-P Wu, K.G. Ravichandran, and A. Tulinsky

Association of the catalytic subunit of aspartate transcarbamoylase with a zinc-containing polypeptide fragment of the regulatory chain leads to increases in thermal stability

C.B. Peterson, B.-B. Zhou, D. Hsieh, A.N.H. Creager, and H.K. Schachman

A 70-amino acid zinc-binding polypeptide fragment from the regulatory chain of aspartate transcarbamoylase causes marked changes in the kinetic mechanism of the catalytic trimer

B.-B. Zhou, G.L. Waldrop, L. Lum, and H.K. Schachman

Identification of glutamate 344 as the catalytic residue in the active site of pig heart CoA transferase

J.-C. Rochet and W.A. Bridger