Probing intermolecular backbone H-bonding in serine proteinase–protein inhibitor complexes

Wuyuan Lu[1]*, Michael Randal[2], Anthony Kossiakoff[2]* and Stephen BH Kent[1]

Background: Intermolecular backbone H-bonding (N–H···O=C) is a common occurrence at the interface of protein–protein complexes. For instance, the amide NH groups of most residues in the binding loop of eglin c, a potent serine proteinase inhibitor from the leech *Hirudo medicinalis*, are H-bonded to the carbonyl groups of residues in the target enzyme molecules such as chymotrypsin, elastase and subtilisins. We sought to understand the energetic significance of these highly conserved backbone–backbone H-bonds in the enzyme–inhibitor complexes.

Results: We synthesized an array of backbone-engineered ester analogs of eglin c using native chemical ligation to yield five inhibitor proteins each containing a single backbone ester bond from P3 to P2′ (i.e. –CONH– to –COO–). The structure at the ligation site (P6–P5) is essentially unaltered as shown by a high-resolution analysis of the subtilisin-BPN′–eglin c complex. The free-energy changes ($\Delta\Delta G^{NH\rightarrow O}$) associated with the binding of ester analogs at P3, P1 and P2′ with bovine α-chymotrypsin, subtilisin Carlsberg and porcine pancreatic elastase range from 0–4.5 kcal/mol. Most markedly, the NH→O substitution at P2 not only stabilizes the inhibitor but also enhances binding to the enzymes by as much as 500-fold.

Conclusions: Backbone H-bond contributions are context dependent in the enzyme–eglin c complexes. The interplay of rigidity and adaptability of the binding loop of eglin c seems to play a prominent role in defining the binding action.

Addresses: [1]Gryphon Sciences, 250 East Grand Avenue, Suite 90, South San Francisco, CA 94080, USA. [2]Department of Protein Engineering, Genentech, Inc, 1 DNA Way, South San Francisco, CA 94080, USA.

*Present address: Department of Biochemistry and Molecular Biology, the University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA.

Correspondence: Wuyuan Lu
E-mail: wuyuanlu@midway.uchicago.edu

Key words: eglin c, hydrogen bonding, native chemical ligation, protein backbone engineering, solid-phase peptide synthesis

Received: **19 February 1999**
Revisions requested: **17 March 1999**
Revisions received: **25 March 1999**
Accepted: **7 April 1999**

Published: **10 June 1999**

Chemistry & Biology July 1999, **6**:419–427
http://biomednet.com/elecref/1074552100600419

 ISSN 1074-5521

Introduction

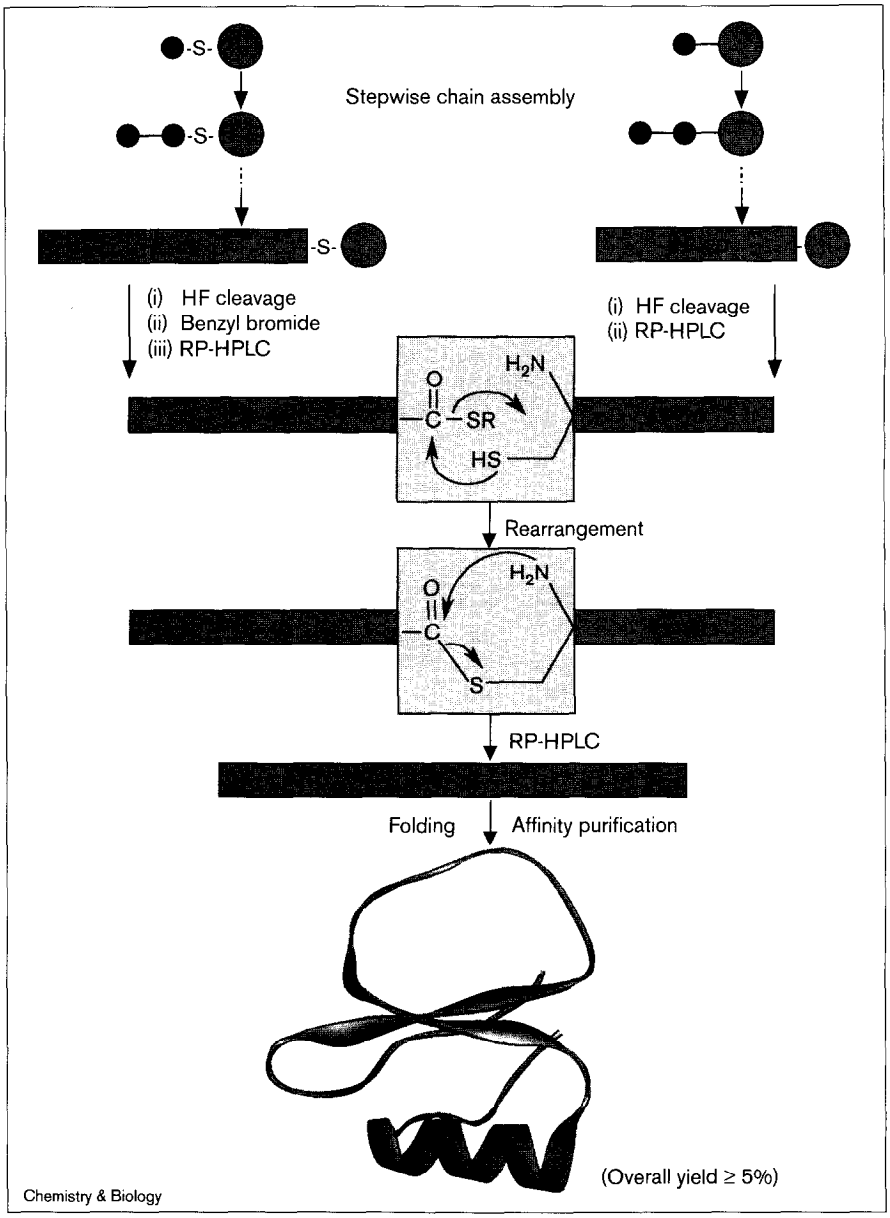
The energetic significance of H-bonding in protein folding, stability and recognition remains a debated issue in protein chemistry [1–4]. Despite significant progress in the past decade, our understanding of the energetic parameters of H-bonding is still limited because of our very limited ability to engineer H-bonds to systematically evaluate the individual effects such as electrostatics, entropy and solvation [5–6].

An especially important class of H-bonds in proteins involve backbone–backbone interactions between the amide NH group (donor) and the carbonyl C=O group (acceptor) [7–9]. Backbone–backbone H-bonding is thought to be critically important in stabilizing α helices and β sheets in a folded protein [7]. Furthermore, backbone H-bonds are commonly encountered at the interface of many protein complexes, suggesting they have a prominent role in protein recognition [5]. By their nature, backbone H-bonds are less affected by entropic factors than their sidechain counterparts. Consequently, studying this major group of H-bonds will focus our understanding on the solvation and electrostatic properties of H-bonding. Studies on backbone–backbone H-bonding have so far been severely hindered, however, by the lack of the experimental means to directly probe its role.

Chemical synthesis [10–11] and *in vitro* protein translation [12] are probably the only two viable techniques available for protein backbone engineering. Recently, the Schultz laboratory [13] used unnatural amino acid mutagenesis to replace three backbone peptide bonds with an ester bond in an α helix of T4 lysozyme. It was found that the loss of terminal H-bonds in the α helix destabilized the protein by ~0.8 kcal/mol, whereas breakage of the H-bond in the middle destabilized the protein by 1.7 kcal/mol.

Previously we reported the total synthesis, via native chemical ligation [14], of a backbone-engineered analog of turkey ovomucoid third domain (OMTKY3, a protein inhibitor of serine proteinases), in which the amide peptide bond between Thr17 and Leu18 was replaced by an ester bond, that is, –CONH– to –COO– [15]. This chemical mutation eliminated a highly conserved backbone H-bond donated by the NH group of Leu18 to the carbonyl C=O groups in the enzyme. Consequently, the binding energy of the inhibitor to six different serine proteinases decreased by an average of 1.5 kcal/mol, which was attributed to the loss of the intermolecular

Figure 1



Strategy for the synthesis of 'wild-type' (8–70)S41C eglin c by native chemical ligation.

backbone–backbone H-bond in the enzyme–inhibitor complex. Recently, the ester-containing OMTKY3 complexed with *Streptomyces griseus* proteinase B has been crystallized, and the resultant structure did not show any appreciable difference, both locally and globally, to the structure of wild-type OMTKY3 complexed with the same enzyme (K.S. Bateman and M.N.G James, personal communication). To further our understanding of the biophysical roles of H-bonding, we have used the native chemical ligation technique to extend protein backbone engineering and the synthesis of an array of ester-substituting backbone analogs of the protein eglin c, and examined the effect on binding to serine proteinases.

Eglin c of the leech *Hirudo medicinalis* is a potent protein inhibitor of many serine proteinases in both the trypsin and subtilisin families of enzymes. Crystal structures of eglin c complexed with bovine α-chymotrypsin (CHYM) [16] and three different subtilisins [17–20] have shown that eglin c, like many other standard-mechanism protein proteinase inhibitors [21], adopts a canonical binding-loop conformation upon complexation with its cognate enzymes [22]. At the interface of the eglin c complexes, there exist a number of backbone H-bonds, which have also been observed in most other serine proteinase–protein inhibitor complexes. The energetic significance of these H-bonds to binding and catalysis has not been determined, however.

Table 1

Association equilibrium constants for recombinant eglin c and synthetic (8–70)S41C eglin c at pH 7.0 with three serine proteinases.

	K_a (M^{-1})	
Enzyme	Recombinant	Synthetic
CHYM	1.9×10^{11}	2.0×10^{11}
SCAR	9.1×10^{11}	6.8×10^{11}
PPE	6.7×10^{7}	6.9×10^{7}

To investigate the energetic contributions of intermolecular H-bonding in these enzyme–protein inhibitor complexes, we have used the native chemical ligation technique to prepare five backbone-engineered analogs of eglin c, in which the amide peptide bond from P3 to P2′ [23] was replaced, each in turn, by an ester bond via the incorporation of 'leucic acid' (i.e. –CONH– → –COO–). As a control, all corresponding reference compounds with 'leucine' substitution were also synthesized, including 'wild-type' eglin c. In this paper we report on syntheses and thermodynamic studies of these synthetic proteins interacting with three different serine proteinases: CHYM and porcine pancreatic elastase (PPE) of the trypsin family and subtilisin Carlsberg (SCAR) of the subtilisin family. In addition, the crystal structure of the chemically synthesized 'wild-type' eglin c in complex with subtilisin BPN′ (SBPN) at 2.0 Å resolution will also be discussed.

Results and discussion

Synthesis of 'wild-type' (8–70)S41C eglin c

Naturally occurring eglin c consists of 70 amino acid residues with no cysteines. X-ray crystallographic studies of eglin c complexed with CHYM and subtilisins showed that the first seven residues at the amino terminus of the inhibitor were not visible in the electron-density map [17–19]. It was reported that truncation at the amino terminus of eglin c by six residues [24] and by seven residues [25] gave rise to a functionally identical inhibitor to the full-length eglin c. For these reasons, we chose to synthesize eglin c with the first seven residues deleted. In addition, our 'wild-type' eglin c construct contained a Ser→Cys mutation at position 41 (P5). This mutation is necessary to introduce a ligation site, -Gly40–Cys41- (Figure 1). As the sidechain of the natural P5 residue (serine) makes no direct contact with the enzyme as was judged by the known crystal structures, (8–70)S41C eglin c was expected to be functionally identical to full-length eglin c. To support this, our synthetic 'wild-type' (8–70)S41C eglin c was carefully compared with full-length recombinant eglin c with regard to the binding to three different serine proteinases, CHYM, SCAR and PPE. As shown in Table 1, association equilibrium constants (K_a) for synthetic (8–70)S41C and for recombinant eglin c are identical within experimental error, indicating that the two different forms of eglin c are functionally indistinguishable.

Backbone geometry at Cys41

Figure 2 shows the electron density from a $2F_O–F_C$ omit map in the region of Cys41 where residues 40–42 were left out of the phasing model. The map clearly shows that this region is well defined. It also shows density extending off the S^γ of Cys41, suggesting that the cysteine had been oxidized. Because there was no indication from the mass spectrometry analysis that the cysteine had undergone oxidation, this probably occurred during crystallization or was X-ray induced during data collection.

Residues 40–42 from this analysis and the 1SBN structure are superimposed in Figure 3. The root mean square deviation (rmsd) of the mainchain atoms (N, C_α and C) for these residues is 0.15 Å, which indicates that the stereochemistry of these residues is identical within the error limits of the data. There is a small change in the puckering of Pro42, but it does not alter the conformation of the mainchain. These data show that native chemical ligation characterized by the introduction of a cysteine does not affect the conformation of the protein in any appreciable way.

Figure 2

Electron-density map in the region of Cys41 of synthetic (8–70)S41C eglin c complexed with SBPN.

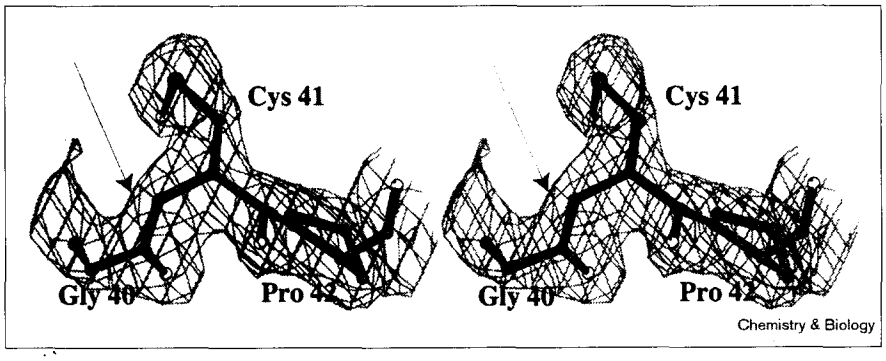
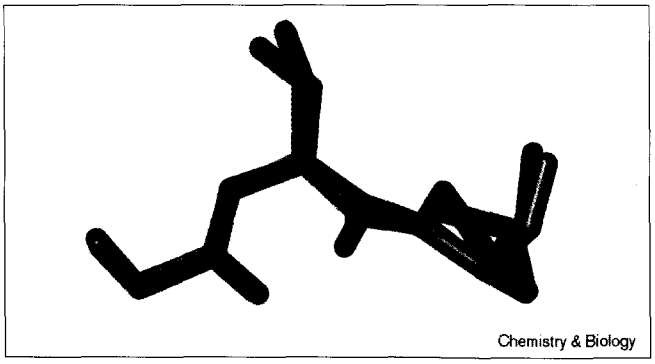


Figure 3



Superposition of residues 40–42: Gly–Cys–Pro (orange) from synthetic eglin c and Gly–Ser–Pro (green) from native sequences (PDB code 1sbn, see [19]). The rmsd of the mainchain atoms (N, C_α, C) for these residues is 0.15 Å.

Effects of NH→O substitutions

Several factors may contribute to the apparent free-energy change $\Delta\Delta G^{amide\rightarrow ester}$ associated with the NH→O substitution.

$$[E>C=O\ldots H_2O]+[I>N-H\ldots H_2O]= [E>C=O\ldots H-N<I]+2H_2O \qquad \Delta G_1 \quad (1)$$

$$[E>C=O\ldots H_2O]+[I>O\ldots H_2O]= [E>C=O\quad O<I]+2H_2O \qquad \Delta G_2 \quad (2)$$

In equations (1) and (2) (E, enzyme; I, inhibitor), ΔG_1 and ΔG_2 are the apparent free-energy changes associated with the binding to the enzyme of reference analog and ester-containing analog, respectively. $\Delta\Delta G^{amide\rightarrow ester}$ can be therefore defined as:

$$\Delta\Delta G^{amide\rightarrow ester} = RT\ln(K_a{}^{amide}/K_a{}^{ester}) = \Delta G_2 - G_1 = \Delta G^{O\ldots O} + (\Delta G^{amide}{}_{solvation} - \Delta G^{ester}{}_{solvation}) - \Delta G^{H\text{-}bond} \quad (3)$$

As illustrated in the equations, $\Delta\Delta G^{amide\rightarrow ester}$ can potentially be affected by: loss of the backbone H-bond, $\Delta G^{H\text{-}bond}$; the differential dehydration energy of inhibitors in the free form, $\Delta G^{amide}{}_{solvation} - \Delta G^{ester}{}_{solvation}$; and the electrostatic and van der Waals interaction between the two oxygen atoms in the H-bond-depleted complex, $\Delta G^{O\cdots O}$. We have suggested that although it is difficult to dissect $(\Delta G^{O\cdots O} + \Delta G^{amide}{}_{solvation} - \Delta G^{ester}{}_{solvation})$, its overall contribution to $\Delta\Delta G^{amide\rightarrow ester}$ is insignificant [15]. In other words, the experimentally measured $\Delta\Delta G^{amide\rightarrow ester}$ value is more or less a reflection of the energetic consequence of loss of a backbone H-bond. Listed in Table 2 are K_a values and $\Delta\Delta G^{amide\rightarrow ester}$ values for four pairs of eglin c analogs interacting with CHYM, SCAR and PPE at pH 7.0.

Table 2

Association equilibrium constants for synthetic eglin c analogs interacting with three serine proteinases measured at pH 7.0.

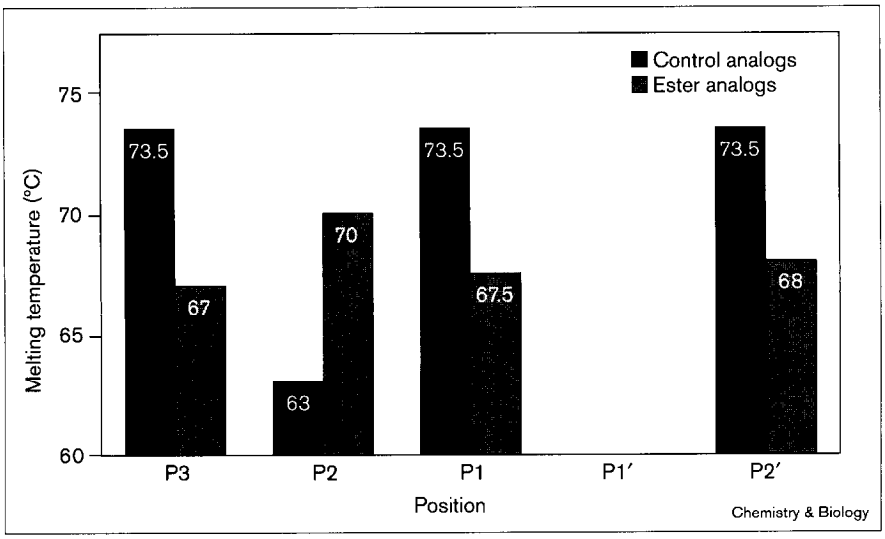
	CHYM	SCAR	PPE
P3			
V43L (K_a; M^{-1})	8.0×10^9	1.7×10^9	9.0×10^7
V43(O)L (K_a; M^{-1})	1.0×10^9	2.1×10^9	5.9×10^5
K_a ratio	10	1	150
ΔΔG (kcal/mol)	1.2	0	2.9
P2			
T44L (K_a; M^{-1})	3.5×10^7	2.9×10^7	1.6×10^6
T44(O)L (K_a; M^{-1})	1.7×10^{10}	1.0×10^8	3.4×10^6
K_a ratio	0.002	0.3	0.5
ΔΔG (kcal/mol)	−3.6	−0.7	−0.4
P1			
Wild type (K_a; M^{-1})	2.0×10^{11}	6.8×10^{11}	6.9×10^7
L45(O)L (K_a; M^{-1})	3.8×10^8	2.8×10^8	1.9×10^6
K_a ratio	500	2500	40
ΔΔG (kcal/mol)	3.7	4.5	2.1
P2′			
Wild type (K_a; M^{-1})	2.0×10^{11}	6.8×10^{11}	6.9×10^7
L47(O)L (K_a; M^{-1})	6.8×10^9	3.2×10^{10}	5.8×10^6
K_a ratio	30	20	10
ΔΔG (kcal/mol)	2.0	1.8	1.4

It should be pointed out that one factor that can potentially complicate the energetic assessment of $\Delta\Delta G^{amide\rightarrow ester}$ is protein destabilization associated with the amide→ester substitution. Incorporation of an ester bond into the backbone of a polypeptide is generally expected to destabilize the protein owing to increased conformational entropy in the denatured state (extra rotational freedom). Although destabilization of the denatured state itself may not be relevant to the binding of the native protein, compromised rigidity of the binding loop in the native form could significantly affect the outcome. For instance, as the binding loop of the inhibitor tends to rigidify upon complexation with the enzyme [26], elevated flexibility in the same region may weaken the binding to the enzyme as a result of an extra entropy loss. On the other hand, added flexibility in the backbone may lead to a lower energy barrier for conformational adjustment upon complexation, therefore resulting in a better adaptation in subsite interactions in the enzyme–inhibitor complex, and consequently a tighter binding.

Thermal denaturation studies indicate that the NH→O substitutions at positions P1, P3 and P2′ destabilized these three backbone-engineered ester analogs uniformly by ~6°C with respect to their reference analogs, which is a marginal decrease in protein stability, whereas the amide→ester substitution at the P2 position enhanced the stability of the protein by 7°C (Figure 4).

Figure 4

Melting temperatures determined by Nano-DSC at pH 5.4 for four pairs of eglin c analogs at positions P3, P2, P1 and P2′.



The P1 position

The amide NH group of the P1 residue (Leu45 in eglin c) H-bonds to the carbonyl C=O group of Ser214 and Ser195O$^{\gamma}$ in chymotrypsin (Ser125 and Ser221 in subtilisins, respectively; Figure 5). This H-bonding pattern has been observed in most inhibitor–enzyme complexes studied [22,26]. In eglin c, the P1 NH→O substitution weakened the binding to CHYM and SCAR by 500- and 2400-fold, that is, 3.7 and 4.5 kcal/mol, respectively. These results were surprising in light of our previous work with OMTKY3, in which the same P1 NH→O substitution caused a decrease in binding to either enzyme by only tenfold (1.3 kcal/mol) [15].

It has recently been reported that eglin c and OMTKY3, with superimposable binding loop conformations in complexes, are two well-behaved interframe-additive inhibitors with respect to CHYM and SCAR [27]. In OMTKY3 and most other small-protein proteinase inhibitors, however, the binding loop is mainly stabilized by flanking disulfide bridges between the loop and the scaffold of the inhibitor. In the case of eglin c, electrostatic/H-bonding interactions are found to be the only forces that support the binding loop. Not surprisingly, although the Leu→leucic acid replacement at the P1 position did not destabilize OMTKY3 [15], it did destabilize eglin c by 6°C at pH 5.4. It is not clear why large effects have been observed at P1 for CHYM and SCAR, but not for PPE, where the decrease in binding to PPE upon the P1 NH→O substitution is only 40-fold, or 2.1 kcal/mol.

The P2 position

The amide NH group of the P2 threonine in eglin c is H-bonded to Gly100O in SCAR. This H-bond is absent in the complex with CHYM (Figure 5). It was therefore anticipated that the P2 NH→O substitution would

Figure 5

Close-up view of the intermolecular backbone H-bonding at the interface of **(a)** CHYM–eglin c [16] and **(b)** SBPN–(8–70)S41C eglin c. The residues of eglin c (P3 to P2′, in Schechter and Berger notation) and of the enzyme are depicted in gold and green, respectively. The mainchain N atoms are blue (cyan for the sidechain N atom) and the carbonyl O atoms are red (pink for the sidechain O atom). Only the atoms involved in H-bonding are displayed in ball-rendering mode.

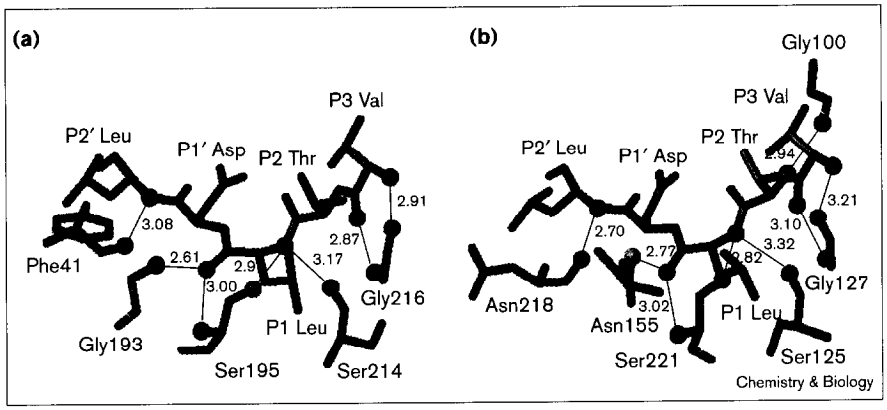
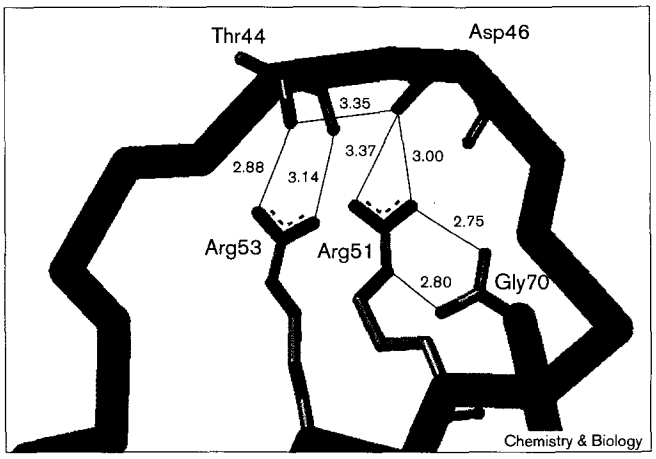


Figure 6



Intramolecular H-bonding networks supporting the binding loop as seen in synthetic (8–70)S41C eglin c complexed with SBPN.

weaken the binding to SCAR, but not to CHYM. Surprisingly, although the NH→O substitution improved the binding to SCAR by a factor of 3, it enhanced the binding to CHYM by 500-fold. Furthermore, the amide-to-ester replacement at the P2 position unexpectedly stabilized eglin c by as much as 7°C at pH 5.4.

The fact that the P2 ester analog (8–70)T44[O]L is more stable than (8–70)T44L suggests that the ester bond incorporation may have resulted in an enthalpically more stable native conformer, enough to compensate for the entropical destabilization in the denatured state. Apparently, a significant local conformational adjustment in the binding loop had to occur upon ester incorporation. It is plausible that the observed enhancement in binding resulted from a better binding adaptation in the complex as a consequence of the conformational adjustment in the binding loop of the inhibitor. A definitive answer requires structural determination of both the free inhibitor and inhibitor–enzyme complex.

Notably, a much smaller gain in binding observed for SCAR (–0.7 kcal/mol) can at least partially be attributed to the loss of the H-bond between Thr44N and Gly100, which is not present in the complex with chymotrypsin. Nonetheless, it is not apparent why the NH→O substitution at the P2 position enhanced the binding to PPE by only a factor of 2.

The significant decrease in binding and stability for the control analog (8–70)T44L eglin c deserves some comments. In wild-type eglin c, the sidechain of Thr44 is H-bonded to both Arg53 and Asp46 (Figure 6). Replacement of the P2 threonine by leucine will undoubtedly weaken the internal rigidity of the reactive-site peptide bond. It is known that a rigid reactive-site peptide bond will effectively elevate the activation energy barrier to the transition state, deterring a productive nucleophilic attack on the peptide bond by the enzyme. Heinz *et al.* [28] reported that although an Arg53→Lys mutation weakened the binding to SCAR, replacement of Thr44 by proline in eglin c converted the inhibitor to a substrate. Our data have shown that the P2 Thr→Leu substitution not only destabilized the protein (by more than 10°C), but also significantly weakened the binding to CHYM and SCAR by four orders of magnitude. It is worth pointing out that PPE is much more tolerant towards a Thr44→Leu mutation at the P2 position, which can probably be attributed to a substantially different subsite-binding pocket in the enzyme [29].

The P3 position

The amide NH group at the P3 position is H-bonded to Gly216O in CHYM and Gly127O in SCAR. The P3 NH→O substitution weakened the binding to CHYM by tenfold, but did not affect the binding to SCAR. In contrast, it weakened the binding to PPE by 150-fold. These discriminating effects seen at the P3 position are difficult to understand and will be further investigated.

The P1′ position

We were unable to purify the two P1′ analogs (8–70)D46L and (8–70)D46[O]L using methyl-chymotrypsin affinity chromatography. The P1′ aspartate in eglin c is a structurally important residue participating in extensive electrostatic/H-bonding interactions involving Arg48, Arg51 and Gly70 (Figure 6). These interactions are believed to be critical in maintaining the rigidity of the enzyme-binding loop of the inhibitor. Heinz *et al.* [28] demonstrated that mutation of P1′ Asp or Arg51 resulted in conversion of the inhibitor into a substrate with trypsin and SCAR. A similar result has also been observed in our laboratory with respect to Gly70. Using total chemical synthesis, we have shown that replacement of the carboxylate group ($-COO^-$) of Gly70 by a neutral amide terminus ($-CONH_2$) destabilized the inhibitor by as much as 20°C at pH 5.4, and caused a weak binding and much accelerated hydrolysis at the reactive site of the inhibitor by both SCAR and CHYM (W-Y. Lu, M. Starovasnik, J. Dwyer, A.K., S.B.H.K. and W.L., unpublished observations). Incorporation of either leucine or leucic acid at the P1′ position apparently compromised the electrostatic interactions helping to rigidify the binding loop of the inhibitor, therefore undermining the binding of mutant eglin c to the affinity column. We expect that the direct incorporation of Asp-derived α-hydroxyl carboxylic acid at P1′ of eglin c will help eliminate this problem.

The P2′ position

The amide NH group of the P2′ residue is H-bonded to Phe41O in CHYM and Asn218O in SCAR. Removal of

this H-bond caused a decrease in binding to CHYM, SCAR and PPE by 30-fold, 20-fold and tenfold, respectively. Our results are in good agreement with that reported by Groeger *et al.* [30] in which the P2′ NH→O substitution in R17G-BPTI decreased the binding of BPTI to trypsin by 25-fold (1.9 kcal/mol).

Significance

Recent advances in the total chemical synthesis of proteins and development of chemical ligation techniques have presented us with an opportunity to explore protein structure and function relationships in a highly systematic fashion. In the work reported here, high-resolution X-ray crystallography showed that chemical synthesis of eglin c, a protein inhibitor of many serine proteases, by native chemical ligation did not alter the protein structure. Chemical protein engineering using noncoded amino acids and unnatural backbone structures will broaden our understanding of the molecular basis of how proteins function. This paper represents a major example of the growing efforts in using chemistry to investigate the energetic significance of H-bonding in protein recognition, in this case using backbone–backbone interactions.

H-bonding contributions are believed to be context-dependent. Some of the apparent free-energy changes, ΔΔG, associated with the amide→ester replacements in the binding loop of eglin c are, however, clearly beyond the interpretation of loss of backbone H-bonds. Our data suggest that interplay of rigidity and adaptability of the binding loop in eglin c plays an important role in dictating its binding action. Of particular interest is the NH→O substitution at the P2 position that leads to an unexpected enhancement in both binding and stability. This observation clearly deserves further analysis using nuclear magnetic resonance and X-ray crystallography.

Materials and methods

Boc-L-amino acids and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate (HBTU) were purchased from Novabiochem; Boc-Gly-OCH_2-Pam-resin and *N,N*-diisopropylethylamine (DIEA) were obtained from Applied Biosystems. Aminomethyl-resin was prepared by Michael Carrasco according to published procedures [31]. L-Leucic acid (S-2-hydroxyisocaproic acid), benzyl bromide, benzyl mercaptan and thiophenol were purchased from Aldrich Chemical Co CHYM and PPE were purchased from Worthington Biochemical Co; SCAR was obtained from Sigma Chemical Co Chromogenic substrates used in binding assays were purchased from Bachem Bioscience Inc. All other commonly used chemicals were obtained in analytical grade or better, as indicated in previous publications.

Methyl-chymotrypsin was prepared according to Ryan and Feeney [32], and was coupled to CNBr-activated Sepharose 4B (Pharmacia) according to the coupling procedures provided by the manufacturer. Analytical reversed phase (RP)-HPLC was performed on an Hewlett Packard Series 1050 equipped with a Vydac C-18 column (5 μm, 4.6 × 150 mm); Preparative RP-HPLC was carried out on a Waters Delta Prep 4000 system using a Vydac C-18 column (15–20 μm, 50 x 250 mm); Solvent A for HPLC was water containing 0.1% TFA; Solvent B was 90% ACN containing 0.09% TFA. Mass spectrometry analysis was carried out on a PE Sciex API-III quadrupole electrospray ionization (ESI) mass spectrometer. Thermal denaturation was performed on a differential scanning calorimeter (Nano-DSC from Calorimetry Sciences) at pH 5.4.

The inhibitor–enzyme association equilibrium constants (Ka) were measured on a Hewlett Packard HP8450A spectrophotometer by a modified version of the Green and Work method [33,34]. The experiment was carried out at room temperature in 0.1 M Bis-Tris buffer containing 0.02 M $CaCl_2$, 0.005% Triton X-100 and 1 mM β-mercaptoethanol, pH 7.0. For most enzymes, the dynamic range of K_a values is from 10^3 to 10^{12} with an accuracy of ± 20%.

Synthesis of 'wild-type' (8–70)S41C eglin c

The synthetic strategy for (8–70)S41C eglin c is outlined in Figure 1. In brief, syntheses of two peptide segments, (8–40)αCOSH and (41–70)S41C, were carried out manually in a stepwise fashion on Boc-Gly-(thioester-linker)-aminomethyl-resin [35] and Boc-Gly-OCH_2-Pam-resin, respectively, using the published *in situ* neutralization/HBTU activation protocol for Boc chemistry [36]. After deprotection and cleavage in HF, crude thioacid peptide (8–40)αCOSH was allowed to react with benzyl bromide to yield thioester peptide (8–40)αCOSBzl , which was then purified by preparative C-18 RP-HPLC. The crude carboxy-terminal peptide (41–70)S41C was also purified by preparative C-18 RP-HPLC prior to ligation.

The ligation of (8–40)αCOSBzl and (41-70)S41C was carried out at pH 7.5 in the presence of 6 M GuHCl, 1% benzyl mercaptan and 3% thiophenol, as described previously [37,38] (Figure 7). The reaction was allowed to proceed to completion in 24 h, and the resultant ligation product was purified on preparative C-18 RP-HPLC. The purified, synthetic (8–70)S41C eglin c was characterized by ESI–MS (observed mass 7342.5 ± 1.1 Da, calculated average isotope mass 7342.4 Da).

Protein folding was carried out by dissolving (8–70)S41C eglin c at 1.2 mg/ml in 6 M GuHCl, 6 mM EDTA, which was then diluted with 50 mM MES, 1 mM DTT, pH 6.0 buffer and water (final protein concentration ~0.2 mg/ml). After 3 h, the folding solution was loaded onto a methyl-chymotrypsin-Sepharose column. The active, folded protein was eluted by lowering the pH to 2.0. The overall yield was 70 mg of pure, active protein out of a typical 0.2 mmol scale synthesis, i.e. 5%.

Syntheses of analogs of (8–70)S41C eglin c

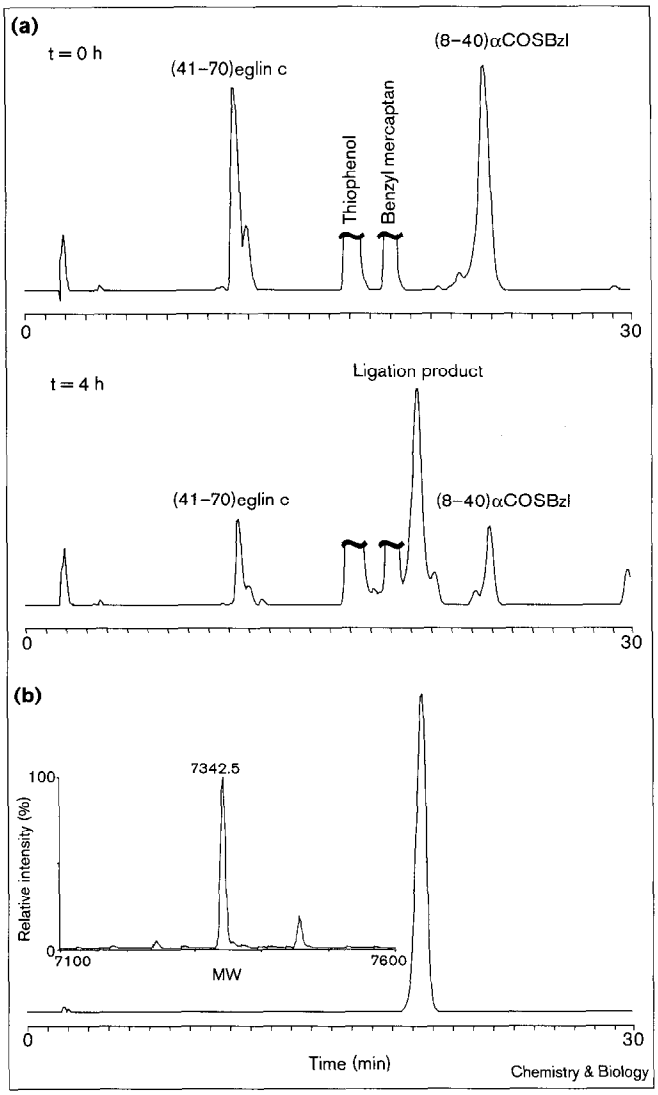
Parallel syntheses of five backbone-engineered ester analogs and three reference proteins (excluding wild type) were performed essentially as for the 'wild-type' molecule. The target sequences are illustrated in Figure 8. Principles for the incorporation of an ester bond into peptides have been described previously [15]. After loading of leucic acid, triple couplings (2 h each) were used for the subsequent esterification reaction. Small amounts of unreacted peptides were acetylated by suspending resins in 5 ml of DMF/DCM (1:1) containing 240 μl acetic anhydride, 120 μl *N*-ethylmorpholine, and 50 mg 4-(dimethylamino)-pyridine for 60 min. After deprotection and cleavage in HF, crude ester-containing and Leu-substituted peptides (41–70)S41C were directly ligated with (8–40)αCOSBzl without further purification.

Conditions under which ligation, purification and folding were conducted were identical to those used in the synthesis of 'wild-type' (8–70)S41C eglin c. Prior to refolding, all the analogs were analyzed by ESI–MS and observed molecular weights were within experimental error of calculated average isotope masses without exception (data not shown).

Crystallization and data collection

The SBPN–eglin complex was crystallized using the procedure of Heinz *et al.* [19]. These conditions produced small prisms of trigonal cross-section having the space group $P3_221$. The cell constants are a = b = 86 Å, c = 106 Å giving a Matthews constant of $V_m = 3.0$ and a

Figure 7



Native chemical ligation of (8-40)αCOSBzl and (41-70)S41C eglin c. **(a)** Ligation reaction. **(b)** Ligated eglin c after purification. The ligation reaction was monitored by analytical C-18 RP-HPLC (25–50% B over 24 min) and ESI-MS (inset in (b); for details see the Materials and methods section).

Figure 8

Ester analogs

P4	P3′
...Pro-	**[O]Leu**-Thr-Leu-Asp-Leu-Arg...
...Pro-Val-	**[O]Leu**-Leu-Asp-Leu-Arg...
...Pro-Val-Thr-	**[O]Leu**-Asp-Leu-Arg...
...Pro-Val-Thr-Leu-	**[O]Leu**-Leu-Arg...
...Pro-Val-Thr-Leu-Asp-	**[O]Leu**-Arg...

Control analogs

P4	P3′
...Pro-	**Leu**-Thr-Leu-Asp-Leu-Arg...
...Pro-Val-	**Leu**-Leu-Asp-Leu-Arg...
...Pro-Val-Thr-	**Leu**-Asp-Leu-Arg...*
...Pro-Val-Thr-Leu-	**Leu**-Leu-Arg...
...Pro-Val-Thr-Leu-Asp-	**Leu**-Arg...*

Chemistry & Biology

Amino acid sequences of the binding loop of eglin c where leucic acid and leucine were inserted. It is worth noting that our 'wild-type' eglin c (labeled with *), although functionally identical to the naturally occurring counterpart, differs on two counts. First, our version is amino-terminally truncated. Second, it contains a Ser→Cys mutation at position 41.

solvent content of about 60%. These crystals had slightly different cell constants, however, than those previously reported by Heinz *et al.*: a = b = 85 Å, c = 90 Å [19]. Interestingly, the space groups are also different $P3_221$ versus $P3_121$; the rather simple packing adjustment along the c axis therefore alters the internal symmetry of the crystal.

Data were collected from one crystal that was exposed to a mother liquor containing 15% glycerol. The treated crystal was then scooped into a small nylon loop and flash frozen by manually plunging the loop into liquid nitrogen. The crystal was mounted on a MAR image plate detector at a crystal to detector distance of 25 cm. The X-ray source was a Rigaku RU2000 set at 120 mA and 45 kV. The crystal was maintained at 100 K during data collection using an Oxford Cryo-stream system. In all, 101,149 observations were collected between 20.0 Å to 1.98 Å resolution. The data were reduced by MARXDS and scaled using MARSCALE [39] resulting in an R_{merge} of 7% (29% between 2.1 to 1.98 Å). These data reduced to 28,638 unique reflections having 96% completeness over the total range. Approximately 5% of the data were removed to be used in calculation of the free R value (R_f) during refinement.

Structure solution and refinement

The structure was solved using the rotation and translation search routines in Amore [40]. The model used was that from the Heinz *et al.* analysis (PDB 1SBN) [41]. The procedure gave a distinctive rotation solution with a peak height of four times the next unrelated solution. The rotation solution was then subjected to a translation search in both possible space groups, $P3_121$ and $P3_221$, which cannot be differentiated based on the symmetry of the diffracted intensities alone. It is only through this translation search that the space group ambiguity can be sorted out. In space group $P3_121$ the best solution had an R-value of 0.53; the best solution for space group $P3_221$ was R = 0.41. This comparison clearly indicated the latter space group to be the correct one.

The structure was refined using XPLOR [42]. The initial model included residues 1–275 for SBPN and 8–70 for eglin. All data from 20 Å to 2.0 Å were used. The data were corrected for anisotropic scattering (correction a = b = −1.9 Å^2, c = 3.7 Å^2). The final model also included 307 water molecules and one calcium ion. This model gave an R value of 0.169 and an R_f of 0.213. A full description of the experimental details and structure analysis will be reported elsewhere (A.K. and M.R., unpublished observations).

Acknowledgements

We are indebted to Michael N.G. James and Kathy Bateman for sharing with us unpublished data, and to Guy Salvesen for providing us with recombinant eglin c. We also thank Michael Carrasco for preparing the aminomethyl-resin and Edwin Madison for useful discussion. This research was supported by funds from NIH GM48870 and GM48897 (S.B.H.K.).

References

1. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.
2. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155.
3. Honig, B. & Yang, A.S. (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27-58.

4. Myers, J.K. & Pace, C.N. (1996). Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71**, 2033-2039.
5. Fersht, A.R. (1987). The hydrogen bond in molecular recognition. *Trends Biochem Sci* **12**, 301-304.
6. Connelly, P.R., Thomson, J.A., *et al.*, (1994). Enthalpy of hydrogen bond formation in a protein-ligand binding reaction. *Proc. Natl. Acad. Sci. USA* **91**, 1964-1968.
7. Pauling, L., Corey, R.B. & Branson, H.R. (1951). The structure of protein: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA* **37**, 205-211.
8. Baker, E.N. & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97-179.
9. Stickle, D.F., Presta, L.G., Dill, K.A. & Rose, G.D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159.
10. Kent, S.B. (1988). Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **57**, 957-989.
11. Muir, T.W. & Kent, S.B. (1993). The chemical synthesis of proteins. *Curr. Opin. Biotechnol.* **4**, 420-427.
12. Mendel, D., Cornish, V.W. & Schultz, P.G. (1995). Site-directed mutagenesis with an expanded genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 435-462.
13. Koh, J.T., Cornish, V.W. & Schultz, P.G. (1997). An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36**, 11314-11322.
14. Dawson, P.E., Muir, T.W., Clark-Lewis, I. & Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779.
15. Lu, W., Qasim, M.A., Laskowski, M., Jr. & Kent, S.B.H. (1997). Probing intermolecular main chain hydrogen bonding in serine proteinase-protein inhibitor complexes: chemical synthesis of backbone-engineered turkey ovomucoid third domain. *Biochemistry* **36**, 673-679.
16. Frigerio, F., *et al.*, & Bolognesi, M. (1992). Crystal and molecular structure of the bovine alpha-chymotrypsin-eglin c complex at 2.0 Å resolution. *J. Mol. Biol.* **225**, 107-123.
17. Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166**, 673-692.
18. McPhalen, C.A. & James, M.N. (1988). Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27**, 6582-6598.
19. Heinz, D.W., Priestle, J.P., Rahuel, J., Wilson, K.S. & Grutter, M.G. (1991). Refined crystal structures of subtilisin novo in complex with wild-type and two mutant eglins. Comparison with other serine proteinase inhibitor complexes. *J. Mol. Biol.* **217**, 353-371.
20. Gros, P., Teplyakov, A.V. & Hol, W.G. (1992). Effects of eglin-c binding to thermitase: three-dimensional structure comparison of native thermitase and thermitase eglin-c complexes. *Proteins* **12**, 63-74.
21. Laskowski, M., Jr. & Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593-626.
22. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433-451.
23. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.
24. Dodt, J., Seemuller, U. & Fritz, H. (1987). Influence of chain shortening on the inhibitor properties of hirudin and eglin c. *Biol. Chem. Hoppe. Seyler.* **368**, 1447-1453.
25. Okada, Y., Tsuboi, S., Tsuda, Y., Nagamatsu, Y. & Yamamoto, J. (1990). Synthesis of a trihexacontapeptide corresponding to the sequence 8-70 of eglin c and studies on the relationship between the structure and the inhibitory activity against human leukocyte elastase, cathepsin G and alpha-chymotrypsin. *FEBS Lett.* **272**, 113-116.
26. Read, R.J. & James, M.N.G. (1986). Introduction to the Protein Inhibitors. In *Proteinase Inhibitors.* (Barrett, A.J. & Salvesen, G., eds) pp. 301-336, Elsevier, Amsterdam.
27. Qasim, M.A., Ganz, P.J., Saunders, C.W., Bateman, K.S., James, M.N. & Laskowski, M., Jr. (1997). Interscaffolding additivity. Association of P1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases. *Biochemistry* **36**, 1598-1607.
28. Heinz, D.W., Hyberts, S.G., Peng, J.W., Priestle, J.P., Wagner, G. & Grutter, M.G. (1992). Changing the inhibitory specificity and function of the proteinase inhibitor eglin c by site-directed mutagenesis: functional and structural investigation. *Biochemistry* **31**, 8755-8766.
29. Bode, W., Meyer E., Jr. & Powers, J. C. (1989). Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity and mechanism-based inhibitors. *Biochemistry* **28**, 1951-1963.
30. Groeger, C., Wenzel, H.R. & Tschesche, H. (1994). BPTI backbone variants and implications for inhibitory activity. *Int. J. Pept. Protein. Res.* **44**, 166-172.
31. Mitchell, A.R., Kent, S.B.H., Engelhard, M. & Merrifield, R.B. (1978). A new synthetic route to tert-butyloxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin, an improved support for solid-phase peptide synthesis. *J. Org. Chem.* **43**, 2845-2852.
32. Ryan, D.S. & Feeney, R.E. (1975). The interaction of inhibitors of proteolytic enzymes with 3-methylhistidine-57-chymotrypsin. *J. Biol. Chem.* **250**, 843-847.
33. Green, N.M. & Work, E. (1953). Pancreatic trypsin inhibitor: 2. reaction with trypsin. *Biochem. J.* **54**, 347-352.
34. Empie, M.W. & Laskowski, M., Jr. (1982). Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: effect on inhibitor interactions with serine proteinases. *Biochemistry* **21**, 2274-2284.
35. Canne, L.E., Walker, S.M. & Kent, S.B.H. (1995). A general method for the synthesis of thioester resin linkers for use in the solid phase synthesis of peptide-alpha-thioacids. *Tetrahedron Lett.* **36**, 1217-1220.
36. Schnölzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. (1992). In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int. J. Pept. Prot. Res.* **40**, 180-93.
37. Dawson, P.E., Churchill, M.J., Ghadiri, M.R. & Kent, S.B.H. (1997). Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* **119**, 4325-4329.
38. Lu, W., Qasim, M.A. & Kent, S.B.H. (1996). Comparative total syntheses of turkey ovomucoid third domain by both stepwise solid phase peptide synthesis and native chemical ligation. *J. Am. Chem. Soc.* **118**, 8518-8523.
39. Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 1916-1934.
40. Navaza, J. (1994). AMORE - An automated package for molecular replacement. *Acta. Crystallogr. D* **50**, 157-163.
41. Bernstein, F.C., *et al.*,. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
42. Brünger, A.T. (1992) *X-PLOR: A System for X-ray Crystallography and NMR.* Yale University Press, New Haven CT, USA.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.