Commentary

Can small cyclic peptides have the activity and specificity of proteolytic enzymes?

B. W. Matthews*†, C. S. Craik‡, and H. Neurath§

*Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, OR 97403; ‡Departments of Pharmaceutical Chemistry and Biochemistry/Biophysics, University of California, San Francisco, CA 94143; and §Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195

"Catalysis by an enzyme requires the specific binding of one or more substrate molecules to a catalytic site and a chemical interaction with this site, which may directly utilize the binding forces to decrease the free energy of activation of the catalyzed reaction" (1). In this preamble to his book on the subject, Jencks (1) aptly summarizes the widely accepted view of enzymes as catalysts. It is understood that the function of an enzyme is to provide not only catalytic groups but also a substrate binding site relative to which the catalytic groups must be correctly positioned. As proposed by Haldane and Pauling (2), part of the energy of substrate binding is also understood to favor the formation of the transition state in catalysis.

The need to satisfy these multiple requirements is one of the factors that determines the minimum size of an enzyme. The smallest globular proteins [e.g., crambin (3)] have 40–50 amino acids, the smallest enzymes [e.g., cytochrome c_{551} (4)] have 80 or so amino acids, and the smallest structurally characterized serine proteases [e.g., *Streptomyces griseus* protease A (SGPA) (5)] have about 180 residues.

The serine proteases, which are the subject of this commentary, are especially well studied (6). Starting with α-chymotrypsin (7), the three-dimensional structures of at least 20 such proteases have been determined. These include representatives of the mammalian pancreatic and the microbial serine proteases (typified by α-chymotrypsin, trypsin, and SGPA), the bacterial serine proteases such as subtilisin (8), and the serine carboxypeptidases (9). Although these families of serine proteases have different three-dimensional structures and apparently unrelated amino acid sequences, their respective active sites all have elements in common. In particular, they all share the so-called "catalytic triad" (10) in which a serine, a histidine, and an aspartic acid are brought together in a characteristic three-dimensional juxtaposition. They all have a similar "oxyanion hole" (11) that is thought to help align the substrate and stabilize the transition state during catalysis. In the case of

α-chymotrypsin and trypsin, and to a greater or lesser degree in other cases, they have a well-developed pocket that favors the binding of certain amino acid side chains and helps determine the specificity of cleavage. Thus, the serine proteases exemplify the general attributes that are anticipated for enzyme catalysts (1, 12).

The primary binding site of trypsin is a cavity that is approximately 10 Å deep and 4 Å × 6 Å in cross section with an aspartic acid located at its base. When a peptide substrate binds to the enzyme the binding pocket not only provides specificity by preferentially binding an arginine or lysine side chain but also helps align the scissile peptide bond relative to the catalytic triad and the oxyanion hole. It is this synergistic effect that contributes to rate enhancement of nine orders of magnitude for the serine proteases when the mechanisms of enzyme- and non-enzyme-catalyzed reactions are compared (13). Parts of the enzyme that do not directly interact with the substrate can, nevertheless, be critical in helping to align those parts that do. It is this direct and indirect participation of many amino acids that contributes to the size of enzymes. The sensitivity of this "chemical machine" to small perturbations can be seen in the loss of activity that results from small changes in the structure of the substrate or the enzyme. More than 99% of the rate enhancement can be lost by changing the substrate at a site distant from the atoms involved in bond making or bond breaking (1, 14). Similarly, conservative amino acid substitutions that introduce minor changes in the binding pocket of trypsin can decrease catalytic turnover rates by up to four orders of magnitude while having a less than 10-fold effect on substrate binding (15).

In contrast to the well-defined shapes of typical globular proteins, shorter polypeptides (e.g., 1–40 amino acids) usually have little if any detectable structure in aqueous solution (16). There are some peptides that display partial helical character (17, 19) or the tendency to form a hairpin bend (18) but these are the exception rather than the rule. Also some "leucine-zipper" peptides are known to

form structures based on side-by-side stacking of α-helices (19). Certain peptides can also adopt well-defined structures when bound to metal ions, surfaces, or receptors (19). Except for such special cases, however, peptides are expected to be either weakly structured or structureless in aqueous solutions at room temperature.

Against this background it was, therefore, very surprising to read the report of Atassi and Manshouri (20) in which two 29-residue cyclic peptides, ChPepz and TrPepz (Fig. 1), were claimed, respectively, to have both activity and specificity compared to that of chymotrypsin and trypsin. Such a result was without precedent.

The peptides described by Atassi and Manshouri (20) were designed to incorporate amino acids thought to be important in the catalytic activity of trypsin and chymotrypsin. Based, for example, on inspection of the three-dimensional structure of α-chymotrypsin, a sequence was designed that linked together individual amino acids, or short amino acid sequences, that were located in the active site of the enzyme. To allow the appropriate spatial separation between these elements, glycine spacers were also incorporated (Fig. 1). To retain topological correspondence, it was also necessary to reverse the sequence of the heptapeptide Ser^{195}–Ser^{189} in the peptide relative to that in α-chymotrypsin (Fig. 1). The overall rationale, therefore, was to try to reconstruct the active site of an enzyme by the simplistic device of linking together those residues thought to be important for catalysis. Except for the cyclization of the peptide, no attempt was made to incorporate some sort of structural framework or scaffold whereby the constituent amino acids would be held in a defined three-dimensional arrangement resembling that seen for the intact enzyme. Indeed, the incorporation of glycines as "spacers" would be expected to have the opposite effect. Overall, 38% of ChPepz consists of glycine. The same is true of TrPepz (Fig. 1). Glycine is more

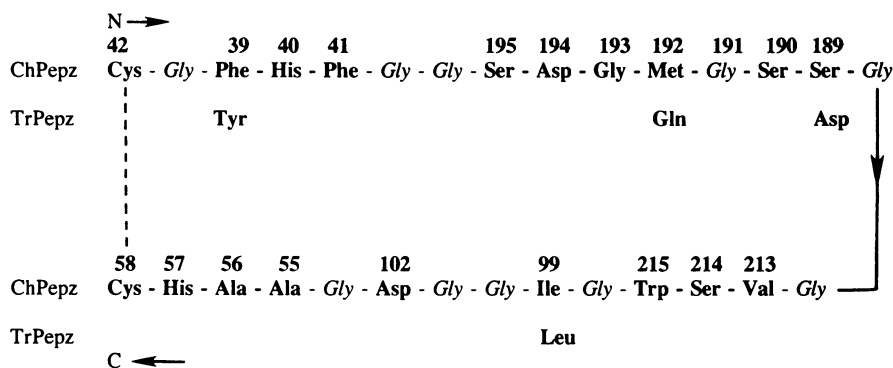---

†To whom reprint requests should be addressed.

N →
| | 42 | | 39 | 40 | 41 | | | 195 | 194 | 193 | 192 | 191 | 190 | 189 |
ChPepz Cys - *Gly* - Phe - His - Phe - *Gly* - *Gly* - Ser - Asp - Gly - Met - *Gly* - Ser - Ser - *Gly*

TrPepz   Tyr   Gln   Asp

| | 58 | 57 | 56 | 55 | | 102 | | | 99 | | 215 | 214 | 213 |
ChPepz Cys - His - Ala - Ala - *Gly* - Asp - *Gly* - *Gly* - Ile - *Gly* - Trp - Ser - Val - *Gly*

TrPepz   Leu

C ◄—

FIG. 1. Amino acid sequences of ChPepz and TrPepz [after Atassi and Manshouri (20)]. The letters N and C denote, respectively, the amino and carboxyl terminus. The complete sequence of ChPepz is shown, with the four amino acid substitutions in TrPepz given below. Amino acids that correspond to segments of α-chymotrypsin (21) [or trypsin (22)] are shown in boldface type with the α-chymotrypsin sequence numbering (21) given above. In α-chymotrypsin and trypsin, residue 191 is a cysteine and forms a disulfide bridge with Cys[220]. In ChPepz and TrPepz, Cys[191] is replaced by a glycine. Cys[58] in α-chymotrypsin and trypsin forms a disulfide bridge to Cys[42].

flexible than any of the other naturally occurring amino acids and, therefore, has a higher entropy cost to maintain in an ordered structure (23, 24). In addition, glycine lacks a side chain and, therefore, can contribute only weakly to the hydrophobic stabilization that is the principal factor in stabilizing the three-dimensional structures of proteins (25, 26). For these reasons, the high proportion of glycine would tend to ensure that both ChPepz and TrPepz would remain unstructured, even when cyclized (see below). Therefore, there was no reason to expect that ChPepz or TrPepz would adopt a defined structure resembling the active site of chymotrypsin or trypsin, let alone have comparable catalytic activity.

It was also hard to understand why cyclization of ChPepz or TrPepz would affect its catalytic properties. According to Atassi and Manshouri (20), the cyclic form of the peptides was fully active whereas the linear molecules had no activity whatsoever. A peptide of 29 amino acids has many degrees of freedom. If we make the simplifying assumption that the backbone of each amino acid can adopt three conformations (which might be categorized as "helical," "extended," and "other"), then by ignoring excluded volume, the peptide could, in principle, adopt $3^{29} = 7 \times 10^{13}$ conformations. It can be estimated (27, 28) that the formation of a disulfide loop encompassing 29 amino acids will reduce the number of backbone conformations available to the loop by a factor on the order of 1000. (This estimate is necessarily crude, but the reducing factor is clearly small relative to the very large number of conformations available.) Thus the backbone conformation of a relatively large loop, as is the case here, is only very weakly constrained by the incorporation of a disulfide linkage between its ends. It is therefore very unlikely that the closure of the disulfide bridge in ChPepz or TrPepz

will cause such a peptide to adopt an ordered three-dimensional fold.

The same argument can be made in terms of free energy. Based on theoretical (29) and experimental (30) estimates of the conformational entropy of a peptide chain as 4.1–4.2 cal per mol per deg per residue (1 cal = 4.184 J), the overall folding energy required to maintain a protein in a defined three-dimensional structure is about 1.2 kcal per mol per residue. For ChPepz or TrPepz, this corresponds to an overall free energy of folding of about 35 kcal/mol. The formation of a disulfide bridge encompassing *n* amino acids has been estimated (31) to reduce the conformational entropy of the backbone chain by

$$\Delta S_{conf} = -2.1 - (3/2)R \ln n.$$

For *n* = 29, this corresponds to a free energy at room temperature of 3.6 kcal/mol. Thus, the disulfide bridge is expected to provide only 10% or so of the overall energy required to fold a peptide such as ChPepz. Cyclization of ChPepz or TrPepz would, therefore, be expected to contribute relatively weakly to a well-defined three-dimensional structure and to any concomitant catalytic properties.

It might be argued that the binding of ChPepz or TrPepz to a substrate could organize the peptide into a defined catalytically active conformation. The binding energy between trypsin and bovine pancreatic trypsin inhibitor (BPTI), one of the tightest protein–protein complexes known, is about 19 kcal/mol (32, 33). A major reason for this high affinity, however, is that the enzyme and the inhibitor have preformed three-dimensional structures that are complementary and change little on association (33). A model for the association of a ligand with an unstructured receptor is provided by the binding of BPTI to trypsinogen. In contrast to trypsin, the 39-residue "activation do-

main" (33) in trypsinogen is flexible and unstructured. From the difference in binding energy of BPTI to trypsin as compared to trypsinogen, Huber and Bode (33) estimated the free energy necessary to order the activation domain in trypsinogen as 9 kcal/mol. Huber and Bode (33) also pointed out that the binding affinity for good substrates of trypsin is around 6 kcal/mol and that such substrates would, therefore, be essentially incapable of binding to trypsinogen (i.e., in a catalytically productive mode). For the same reason, there is little reason to think that the binding of such a substrate would be capable of organizing ChPepz or TrPepz into a folded three-dimensional structure or that such a structure would be catalytically active.

Another surprising claim of Atassi and Manshouri (20) was that the catalytic activity of ChPepz could be changed from chymotrypsin-like to trypsin-like by the simple device of substituting four amino acids (Fig. 1). The four substitutions made by Atassi and Manshouri (20) are a limited subset. According to Blow (34), a total of 10-amino acid substitutions plus a deletion differentiate the substrate binding region of chymotrypsin (21) relative to trypsin (22). Therefore, it would not be expected that the four substitutions included by Atassi and Manshouri (20) would fully account for the difference in specificity of the two enzymes. In addition, a series of mutagenesis experiments has shown that it is not a simple matter to cleanly change the specificity of trypsin to chymotrypsin even with multiple amino acid substitutions (35). If such experiments are difficult in the context of the folded native enzyme, why should they work so well in the peptides ChPepz and TrPepz?

In the present issue of the *Proceedings*, two attempts to reproduce the results of Atassi and Manshouri (20) are described (36, 37). In both cases the results are negative. Neither Wells *et al.* (36) nor Corey and Phillips (37) are able to substantiate the claim (20) that TrPepz mimics trypsin. The experimental data of Wells *et al.* (36) and Corey and Phillips (37) are well documented, and the fact that their results are in agreement gives additional credence to their respective findings. The NMR spectrum of cyclized TrPepz (36) also shows no evidence of a stable conformation, consistent with the general expectation outlined above, that the peptide would remain unstructured in solution. Thus, these two reports (36, 37) provide strong evidence against the claim of Atassi and Manshouri (20) that TrPepz mimics trypsin.

The ultimate test of our understanding of the various factors involved in enzyme catalysis is the design and synthesis of molecules that can duplicate the mode of action of their natural macromolecular

counterparts (38). The use of protein engineering to dissect and redesign preexisting enzymes has provided trenchant insights regarding the nature of enzymes. However, the *de novo* synthesis of an enzyme with prescribed function from first principles has remained elusive. This has been particularly true for peptidases. Extensive efforts have been focused on designing an organic model for each of the steps in chymotrypsin catalysis. Chymotrypsin has a substrate binding pocket that prefers large hydrophobic amino acids like phenylalanine. By using cyclodextrin to mimic the binding cavity and serve as a scaffold for functional groups, a molecule was created that cleaved ester substrates with efficiency comparable to the natural enzyme (39). Although chymotrypsin does display esterase activity, its physiological function is to hydrolyze peptide bonds. When peptide substrates were tested with the artificial enzyme, no significant amidase activity was observed.

It has been more than 60 years since the crystallization of urease by Sumner (2) and of pepsin, trypsin, and chymotrypsin by Northrop and Kunitz (2). Subsequent to these pioneering discoveries and the characterization of these and many other proteins, there has been much speculation regarding the "bigness" of enzymes (40–43). Clearly geometric and energetic as well as chemical factors are implicated in their effectiveness as catalysts. As yet there has been no compelling evidence that small polypeptides can approach the larger naturally occurring proteolytic enzymes in terms of activity and specificity.

1. Jencks, W. P. (1969) *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York).
2. Fersht, A. (1977) *Enzyme Structure and Mechanism* (Freeman, New York).
3. Hendrickson, W. A. & Teeter, M. M. (1981) *Nature (London)* 290, 109–113.
4. Almassy, R. J. & Dickerson, R. E. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2674–2678.
5. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C.-A. (1980) *J. Mol. Biol.* 144, 43–88.
6. Neurath, H. (1984) *Science* 224, 350–357.
7. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967) *Nature (London)* 214, 652–656.
8. Wright, C. S., Alden, R. A. & Kraut, J. (1969) *Nature (London)* 221, 235–242.
9. Liao, D.-I. & Remington, S. J. (1990) *J. Biol. Chem.* 265, 6528–6531.
10. Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969) *Nature (London)* 221, 337–340.
11. Robertus, J. D., Kraut, J., Alden, R. & Birktoft, J. J. (1972) *Biochemistry* 11, 4293–4303.
12. Kraut, J. (1977) *Annu. Rev. Biochem.* 46, 331–358.
13. Kahne, D. & Still, W. C. (1988) *J. Am. Chem. Soc.* 110, 7529–7534.
14. White, H. & Jencks, W. P. (1976) *J. Biol. Chem.* 251, 1688–1699.
15. Evnin, L., Vasquez, J. & Craik, C. S. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6659–6663.
16. Epand, R. M. & Scheraga, H. A. (1968) *Biochemistry* 7, 2864–2872.
17. Brown, J. E. & Klee, W. A. (1971) *Biochemistry* 10, 470–476.
18. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E. & Lerner, R. A. (1985) *Nature (London)* 318, 480–483.
19. DeGrado, W. F., Wasserman, Z. R. & Lear, J. D. (1989) *Science* 243, 622–628.
20. Atassi, M. Z. & Manshouri, T. (1993) *Proc. Natl. Acad. Sci. USA* 90, 8282–8286.
21. Hartley, B. S. (1964) *Nature (London)* 201, 1284–1287.
22. Walsh, K. A. & Neurath, H. (1964) *Proc. Natl. Acad. Sci. USA* 52, 884–889.
23. Nemethy, G., Leach, S. J. & Scheraga, H. A. (1966) *J. Phys. Chem.* 70, 998–1004.
24. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6663–6667.
25. Bernal, J. D. (1939) *Nature (London)* 143, 663–667.
26. Kauzmann, W. (1959) *Adv. Protein Chem.* 14, 1–64.
27. Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim.* 29, 230–259.
28. Kauzmann, W. (1959) in *Sulfur in Proteins*, eds. Benesch, R., Boyer, P. D., Klotz, I. M., Szent-Györgii, A. G. & Schwartz, D. R. (Academic, New York), p. 70.
29. Kauzmann, W. (1954) in *The Mechanism of Enzyme Action*, eds. McElroy, W. D. & Glass, B. (Johns Hopkins Press, Baltimore).
30. Privalov, P. L. (1979) *Adv. Protein Chem.* 33, 167–241.
31. Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. (1988) *J. Biol. Chem.* 263, 11820–11825.
32. Vincent, J. P. & Lazdunski, M. (1972) *Biochemistry* 11, 2967–2977.
33. Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* 11, 114–122.
34. Blow, D. M. (1974) in *Bayer-Symposium V: Proteinase Inhibitors*, eds. Fritz, H., Tschesche, H., Greene, L. J. & Truscheit, E. (Springer, Berlin), pp. 473–483.
35. Hedstrom, L., Szilagyi, L. & Rutter, W. J. (1992) *Science* 255, 1249–1253.
36. Wells, J. A., Fairbrother, W. J., Otlewski, J., Laskowski, M., Jr. & Burnier, J. (1994) *Proc. Natl. Acad. Sci. USA* 91, 4110–4114.
37. Corey, D. R. & Phillips, M. A. (1994) *Proc. Natl. Acad. Sci. USA* 91, 4106–4109.
38. Breslow, R. (1972) *Chem. Soc. Rev.* 1, 553–580.
39. Mallick, I., D'Souza, V. T., Yamaguchi, M., Lee, J., Chalabi, P., Godwood, R. C. & Bender, M. L. (1984) *J. Am. Chem. Soc.* 106, 7252–7254.
40. Rittenberg, D. (1956) in *Essays in Biochemistry*, ed. Graff, S. (Wiley, New York), pp. 232–240.
41. Monod, J., Wyman, J. & Changeux, J.-P. (1965) *J. Mol. Biol.* 12, 88–118.
42. Srere, P. A. (1984) *Trends Biochem. Sci.* 9, 387–390.
43. Goodsell, D. S. & Olson, A. J. (1993) *Trends Biochem. Sci.* 18, 65–68.