Complexation of Two Proteic Insect Inhibitors to the Active Site of Chymotrypsin Suggests Decoupled Roles for Binding and Selectivity*

Received for publication, June 20, 2001
Published, JBC Papers in Press, August 8, 2001, DOI 10.1074/jbc.M105707200

The crystal structures of two homologous inhibitors (PMP-C and PMP-D2v) from the insect Locusta migratoria have been determined in complex with bovine α-chymotrypsin at 2.1- and 3.0-Å resolution, respectively. PMP-C is a potent bovine α-chymotrypsin inhibitor whereas native PMP-D2 is a weak inhibitor of bovine trypsin. One unique mutation at the P1 position converts PMP-D2 into a potent bovine α-chymotrypsin inhibitor. The two peptides have a similar overall conformation, which consists of a triple-stranded antiparallel β-sheet connected by three disulfide bridges, thus defining a novel family of serine protease inhibitors. They have in common the protease interaction site, which is composed of the classical protease binding loop (position P5 to P’4, corresponding to residues 26–34) and of an internal segment (residues 15–18), held together by two disulfide bridges. Structural divergences between the two inhibitors result in an additional interaction site between PMP-D2v (position P10 to P6, residues 21–25) and the residues 172–175 of α-chymotrypsin. This unusual interaction may be responsible for species selectivity. A careful comparison of data on bound and free inhibitors (from this study and previous NMR studies, respectively) suggests that complexation to the protease stabilizes the flexible binding loop (from P5 to P’4).

Small canonical serine protease inhibitors are widely distributed among living organisms. They have been classified by Bode and Huber (1) into 16 structural families. One of the novel families that has emerged since then is the grasshopper family (2). The first members were characterized from the brain (pars intercerebralis) and the hemolymph of the insect Locusta migratoria (3–5). Furthermore, similar peptides (named SGPI) were isolated from Schistocerca gregaria (6); they share 40–80% homology (including the six conserved cysteines) with the Locusta peptides. More recently, the same sequence motif was also identified in Pacifastin, a 155-kDa protein from the crayfish Pacifastacus leniusculus and composed of two domains with different activities. One of these domains contains nine repeats similar to the locust peptides and has a protease inhibitory activity (7).

We have carried out extensive investigations on two locust peptides, PMP-C and PMP-D2 (pars intercerebralis major peptide). They consist of 36 and 35 residues, respectively, have three disulfide bridges, and are 40% identical in sequence. The three-dimensional structures of PMP-D2 and PMP-C were determined by 1H NMR (8, 9). These peptides display a new fold with an unusual disulfide bond pattern. PMP-C is a very potent bovine α-chymotrypsin inhibitor (Kᵢ 0.13 nM) and a weak human leukocyte elastase inhibitor (Kᵢ 180 nM) and is devoid of activity toward porcine trypsin. The nature of its P1 residue (nomenclature according to Schechter and Berger (10), Leu32, is in accordance with the literature and its inhibitory properties. Indeed, chymotrypsin inhibitors have bulky and aromatic residues such as Phe, Leu, or Met as their P1 residue whereas trypsins require basic amino acids (Lys or Arg). In PMP-D2, the corresponding residue is Arg29. However, PMP-D2 has no effect on porcine trypsin and is a weak inhibitor of bovine trypsin and chymotrypsin (Kᵢ 100 and 1,500 nM, respectively), whereas it can be converted to a potent bovine α-chymotrypsin inhibitor by a single mutation R29L at P1 (Kᵢ 0.8 nM). In addition, a second point mutation at P’1 (K30M) converts PMP-D2 into a fairly potent human leukocyte elastase inhibitor (Kᵢ 2 nM) (5). Very recently, we showed that PMP-D2 is a strong inhibitor of trypsins isolated from L. migratoria. The inhibitory properties of PMP-D2 are of great interest, because they highlight selectivity not observed yet for PMP-C.

To provide the first description of the recognition mechanism in the grasshopper family, we solved the crystal structure of PMP-C in complex with bovine α-chymotrypsin. We also determined the structure of the R29L/K30M PMP-D2 variant (named PMP-D2v) bound to chymotrypsin, because the weak affinity of native PMP-D2 prevented any complex crystallization with available proteases.

EXPERIMENTAL PROCEDURES

Biological Material—The peptides were synthesized on solid phase using Fmoc strategy and were refolded by air oxidation of the six cysteines as described previously (5). Bovine α-chymotrypsin was purchased from Sigma (product code C3142).

Crystallization and X-ray Diffraction Studies—The peptides, as well as the enzyme, were solubilized in a buffer containing 50 mM Tris, pH 8.0 (buffer A). 

1. C. Kellenberger, G. Ferrat, and H. Darbon, manuscript in preparation.
7.5, and mixed together in a molar ratio of 1:3 (enzyme:inhibitor). The crystallization trials were performed after 1 h of incubation. Single crystals of PMP-C peptide complexed to bovine α-chymotrypsin were obtained at 20 °C using the hanging drop vapor diffusion method. Protein drops were equilibrated against a 1-ml reservoir solution containing 100 mM sodium acetate, 29% polyethylene glycol 400, and 100 mM CdCl₂ at pH 5.0. The drops consisted of 2 ml of protein solution at 7.5, and mixed together in a molar ratio of 1:3 (enzyme:inhibitor). The final model has good geometry, with an Rfree of 22.8% for an Rfactor of 18.9%. The three PMP-C molecules are complete, respectively, between residues 2 and 35, 4 and 34, and 5 and 35. The Ramachandran plot shows that 86.9% of the non-glycine residues fall within the most favorable regions. The remaining 13.1% lies within the additionally allowed regions. The statistics of the refinement are given in Table I.

### RESULTS AND DISCUSSION

#### General Description of the Structures

The crystals of the PMP-C-α-chymotrypsin complex belong to the space group P6₅ and diffract up to 2.1 Å resolution. After phase determination by molecular replacement and phase improvement by density modification, refinement yielded an Rfactor and an Rfree of 18.9 and 22.8%, respectively (Table I).

PMP-C folds into a brick-shaped structure of overall dimensions 30 × 15 × 10 Å³. It is formed of one β-sheet of three antiparallel β-strands (β₁, residues 8–11; β₂, residues 16–19; β₃, residues 26–29) connected by two loops (λ₁, residues 12–15; λ₂, residues 20–25) and stabilized by three disulfide bridges, Cys³–Cys¹⁹, Cys¹⁷–Cys²⁸, and Cys¹⁴–Cys¹³ (Fig. 1, A and B).

The asymmetric unit contains three copies of the complex displaying nearly identical conformations. Pairwise comparisons of the Cu atoms of the three proteases yield root mean square deviations (r.m.s.d.)² values of 0.184, 0.176, and 0.207 Å. Larger r.m.s.d. values are obtained for the three inhibitor molecules (0.350, 0.366, and 0.358 Å). This is because of a poor definition of the electron density map of the N termini of PMP-C molecules. However, as the Cu r.m.s.d. are in the range

| Data collection | α-Chymotrypsin/PMP-C | α-Chymotrypsin/PMP-D2v |
|-----------------|---------------------|-----------------------|
| Resolution limit (Å) | 20–2.1 | 15–3.0 |
| Completion (%) | 98.2 (89.6) | 99.8 (99.9) |
| Redundancy | 3.9 (2.1) | 7.3 (7.5) |
| R_free | 7.9 (37.1) | 15.1 (42.3) |
| Refinement | | |
| Number of reflections (test set) | 46359 (2311) | 8699 (437) |
| Number of protein atoms | 5966 | 2000 |
| Number of water molecules | 369 | 56 |
| Final R_Final/R_Free (% | 18.9/22.8 | 16.4/19.3 |
| a-α-Chymotrypsin | 27.3 | 21.5 |
| Inhibitor | 63.4 | 67.8 |
| Waters | 37.0 | 35.7 |
| r.m.s.d. deviations | | |
| Bond (Å) | 0.006 | 0.006 |
| Angles (°) | 1.31 | 1.32 |
| Improper/dihedral angles (°) | 0.7526.1 | 0.74/25.7 |

² The abbreviation used is: r.m.s.d., root mean square deviation.
of the estimated coordinate errors, it can be considered that the three complexes are identical. The main chain of the α-chymotrypsin component can be traced from Cys1 to Gly12 (activation peptide), from Ile16 to Tyr146, and from Asn150 to Asn245. The main chain of PMP-C shows no ordered electron density at both the amino (Glu1) and the carboxyl termini (Asn36) (Fig. 1A). The temperature factors of the α-chymotrypsins are in average of 27.3 Å², whereas those of the inhibitors are higher, with an average value of 63.4 Å². Nevertheless, the residues 14–18 and 25–33 in contact with chymotrypsin are better defined and display average B-factor values of 49.4 and 32.1 Å². Fig. 1B shows the overall fold of PMP-C colored according to the temperature factors.

The molecular replacement method was also used to solve the structure of PMP-D2v-α-chymotrypsin complex. Crystals belong to the space group P6_522 and diffract up to 3.0 Å resolution. The refinement yields an R-factor of 16.4% and an Rfree of 19.5%. The model of the complex accounts for 242 residues for the protease (1–12 and 16–245) and for 33 residues for the inhibitor. The N-terminal (Glu1) and C-terminal (Ala36) of the inhibitor residues are not modeled because of lack of electron density. The refined average B-factor are 21.5 Å² for all protease atoms and 67.8 Å² for the inhibitor. PMP-D2v exhibits the same overall structure as PMP-C with one β-sheet composed of three antiparallel β-strands (β1, residues 8–11; β2, residues 16–19; and β3, residues 25–28) connected by two loops (λ1, residues 12–15; λ2, residues 20–24) and stabilized by three disulfide bridges (Cys4–Cys19, Cys17–Cys27', and Cys14–Cys22).

**Binding Loop of PMP-C and Its Interactions with α-Chymotrypsin**—PMP-C buries 875 Å² of its solvent-accessible surface upon binding to the protease. As illustrated in Fig. 2A, the interaction site of PMP-C is composed of two regions, the binding loop (residues 26–34, 596 Å²) and an internal segment (residues 15–18, 153 Å²). Leu40 (P1) binds to the S1 pocket of the protease, and its side chain conformation is located in the deepest energy minimum (angular values of $\chi_1 = -57^\circ$, $\chi_2 = 173^\circ$) and superimposes well with the P1 side chain of inhibitors such as ascaris inhibitor of chymotrypsin/elastase ($\chi_1 = -71^\circ$, $\chi_2 = 164^\circ$; see Ref. 16) and OMTKY3 ($\chi_1 = -48^\circ$, $\chi_2 = 157^\circ$; see Ref. 17). The P3-P5 segment (residues 26–28) forms an antiparallel β-sheet with chymotrypsin residues 218–216. Ala22 (P2) interacts with Phe41 through a hydrogen bond, and Pro34 (P4) makes a stacking interaction with Phe39's side chain. The intermolecular hydrogen bonds are listed in Table II, top. The binding loop of PMP-C is maintained in a rigid conformation by intramolecular hydrogen bonds to an internal segment (residues 14–20), as shown in Table II, bottom. In particular, a network of four hydrogen bonds (denoted by * in Table II, bottom), involving Asn15 and Thr29's side chains, highly stabilizes P1 and P2 and maintains the local conformation of P1. Reviewing the grasshopper family inhibitors, we found out that Asn15 and Thr29 are the only conserved residues, apart from...
the six cysteines (Fig. 1C). Although unrelated to grasshopper inhibitors, Ecotin (18, 19) displays a segment (residues 47–56) similar to that of PMP-C (residues 11–20), which is also connected to the binding site by a disulfide bridge (Cys^{17}–Cys^{87}). However, the second disulfide bridge present in PMP-C (Cys^{17}–Cys^{87}) is missing, and the cysteines are replaced by His^{53} and Ser^{82}, respectively. This local conformation is maintained by a similar hydrogen bond network, as described previously for PMP-C, where Asn^{31} and Thr^{33} side chains play the same role as Asn^{15} and Thr^{29} (Fig. 3A).

Comparison of PMP-C and PMP-D2v—From P3 (Cys^{27}) to P2 (Gly^{31}), the intermolecular interactions between PMP-D2v and α-chymotrypsin are are similar to those observed for PMP-C. The P1 residue of PMP-D2v (Leu^{29}) is exactly in the same conformation as that of PMP-C. However, some differences are observed between the two inhibitors. The nature of the P’4 residue in PMP-D2v (Gln^{33} instead of Pro^{34}) suppresses the stacking with Phe^{39}. In addition, the hydrogen bond between Ser^{218} and Ala^{26} (P5) in PMP-C is not present in PMP-D2v. A water-accessible surface area of 946 Å^{2} is buried upon chymotrypsin binding to PMP-D2v (Fig. 2A). Beside the primary interaction site, composed of the binding loop (residues 25–33, 584 Å^{2}) and the internal segment (residues 15–18, 108 Å^{2}), PMP-D2v displays a secondary site (residues 20–24, 175 Å^{2}). This region interacts with residues 172–175 of α-chymotrypsin through van der Waals contacts (Thr^{20}–Trp^{172}, Pro^{21}–
TABLE II

| Atom in ω-chymotrypsin | Atom in PMP-C at P5-P3 | Distance (Å) |
|------------------------|------------------------|--------------|
| Phe41 O                | Ala32 N                | 3.0          |
| Gly139 N               | Leu140 O               | 2.7          |
| Ser135 OG              | Leu140 N               | 2.9          |
| Gly216 N               | Cys217 O               | 3.0          |
| Gly216 O               | Cys217 N               | 2.9          |
| Ser218 N               | Ala219 O               | 3.1          |

**Intramolecular hydrogen bonds that maintain the P5-P4 segment of PMP-C in a rigid conformation**

| Atom at P5-P3 segment | Atom at residues 14 to 20 | Distance (Å) |
|-----------------------|---------------------------|--------------|
| Ser114 O              | Gly120 N                  | 2.6          |
| Ala127 N              | Arg128 O                  | 3.1          |
| Ala137 O              | Arg138 N                  | 3.0          |
| Thr139 N              | Thr140 N                  | 3.2          |
| *Thr139 OG1           | Asn140 OD1                | 2.8          |
| *Thr139 OG1           | Thr140 N                  | 3.2          |
| *Lys139 N             | Asn140 OD1                | 2.7          |
| *Lys139 O             | Asn140 ND2                | 2.9          |

**Fig. 3.** A, close view of the hydrogen bond network between the binding loop and the internal segment for PMP-C (in atom-type color; see legend of Fig. 1A) and Erotin (in dark blue), with the four residues in PMP-C labeled. B, superimposition of the crystal structure (in yellow) and the 36 models from the NMR study (in blue) of PMP-C. The r.m.s.d. from superimposition are smaller than 1 Å for regions 3–11 and 16–28 and 1.8 and 2.4 Å for regions 12–15 and 29–35, respectively.

Gly173 and Thr174, Thr22–Gly178, Val24–Trp172). The structures of PMP-C and PMP-D2v were superimposed on each other (Fig. 2B), yielding an overall r.m.s.d. value of 0.7 Å for 17 residues (residues 8–17 and 27–33, according to PMP-C numbering). A large structural divergence, with deviations between Cα atoms higher than 5 Å, is observed for residues 4–7 and 18–27 (λ2 loop). Accordingly, PMP-C and PMP-D2v can be dissected into two structural subdomains as illustrated in Fig. 2B. On one hand, the so-called “subdomain I” is formed by residues 8–17 and 28–34, and these two regions are bound together by the disulfide bridge Cys14–Cys33 (numbering of PMP-C). The subdomains I of the two inhibitors carry the primary binding site and are therefore structurally similar. On the other hand, residues 4–7 and 18–27, which are linked together by the disulfide bridge Cys4–Cys19, constitute subdomains II. The structural divergences that characterize subdomains II may be caused by differences in λ2 loop, which is one residue shorter and possesses a Pro at position 21 in PMP-D2v. As a result, subdomain II of PMP-D2v is shifted closer to chymotrypsin and may be responsible of additional contacts (either favorable interaction or steric hindrance) and could therefore be considered as an element for discriminating between protease targets. PMP-C may be regarded as more permissive toward various enzymes, as also observed with S. gregaria inhibitors (20).

Comparison of Complexed (X-ray) and Free (NMR) Inhibitors—The x-ray structure of bound PMP-C was compared with the free form (36 structures; Protein Data Bank code 1PMC), solved in solution by 1H NMR (9). The β-sheets of the two structures superimpose well, with an r.m.s.d. smaller than 1 Å whereas two regions display larger deviations, λ1 loop (residues 12–15) and C terminus (residues 29–35), with r.m.s.d. values of 1.8 and 2.4 Å, respectively (see Fig. 3B). The NMR structures display an average r.m.s.d. of 0.33 (±0.14) Å for the backbone atoms of the well defined segments 3–11, 14–19, and 24–30 (calculation performed with 164 nuclear Overhauser effect-derived distances using the program X-PLOR) and below 1 Å for the regions 12–15 and 29–33. As the r.m.s.d. values from NMR study are smaller than the deviations resulting from NMR/x-ray structure superimposition, these latter may be considered as conformationally different in the regions 12–15 and 29–33. However, it should be pointed out that the conformation of region 29–33 was built from only four medium and long range nuclear Overhauser effects. Therefore, we believe that the scarcity of NMR data would rather account for more flexibility in this region. Our assumption is supported by the examination of PMP-D2. Although the superimposition of the free (20 structures) and bound structures yields r.m.s.d. values similar to that of PMP-C (within 1 Å for the core and 1.3 and 2.1 Å for the loop λ1 and the C terminus, respectively), free PMP-D2v is globally less structurally defined than PMP-C, with NMR r.m.s.d. values in the range of 1–2 Å for loop λ1 and of 2–6 Å for C terminus region (residues 29–35). This demonstrates a clear tendency to flexibility in these regions.

A scenario of events can therefore be put forward; Leu30 (P1) penetrates deeply into the active site pocket S1 of the protease, followed by its neighbors Thr29 (P2) and residues 31–33, resulting in a rather rigid conformation of this stretch. The Cys4–Cys19 bridge, together with the hydrogen bond network involving Asn15 and Thr29, stabilizes residues 12–15. This cascade of events is supported by a detailed comparison of the main chain conformations of the P5–P3 residues in bound and free PMP-C, in relation with the canonical values reported by Bode and Huber (1). The dihedral angles of the binding loop in our study are in accordance with the canonical values whereas some given by the NMR study are not, in particular for the P1 and P′1 residues (Table III). We therefore infer that the disordered C terminus of free PMP-C achieves conformational stabilization upon chymotrypsin binding.

Such a hypothesis is supported by the example of Elafin, a
potent elastase inhibitor. Its structure was studied in complexed (21) and free (22) forms. The binding loop of the free inhibitor shows a high degree of flexibility, most of the φ and ψ angles being out of the range defined for a canonical conformation. In the crystal structure, the φ and ψ angles for P3 to P′3 residues fall in the defined range.

| P3 | P2 | P1 | P′1 | P′2 | P′3 |
|----|----|----|-----|-----|-----|
| φ  | ψ  | φ  | ψ  | φ  | ψ  |
| −131 | 161 | −72 | 163 | −112 | 42 |
| −127 | −157 | −141 | 163 | −110 | 70/167 |
| −140/−120 | 140/170 | −100/−60 | 139/180 | −120/−95 | 9/50 |

| φ  | ψ  | φ  | ψ  | φ  | ψ  |
|----|----|----|----|----|----|
| −80/165 | 182 | −84 | 151 |
| −100/−60 | 139/180 | −140/−99 | 70/120 |
| −140/−99 | 70/120 |

a Minor populations in NMR studies.
b Major populations in NMR studies.

The peculiar structure of insect proteases and the structural requirements for inhibition.

The preeminent feature of the recently sequenced insect genome of Drosophila melanogaster was the high number of putative protease open reading frames (over 300 identified). Such potentially dangerous enzymes need to be handled carefully by insects along their development pathway, a task that might be devoted to the short peptides inhibitors described in this study.

REFERENCES
1. Bode, W., and Huber, R. (1992) Eur. J. Biochem. 204, 433–451
2. Laskowski, M., and Qasim, M. A. (2000) Biochim. Biophys. Acta 1477, 324–337
3. Nakamura, N., Hietter, H., Van Dorselaer, A., and Luu, B. (1992) Eur. J. Biochem. 204, 147–153
4. Boigeon, R. A., Mattras, H., Brechel, M., Paroutaud, P., and Coletti-Proviero, M. A. (1992) Biochim. Biophys. Acta 1099, 790–793
5. Kellenberger, C., Boudier, C., Bermudez, I., Bie, J. G., Luu, B., and Hietter, H. (1995) J. Biol. Chem. 270, 25514–25519
6. Hamadaoui, A., Wataleb, S., Devreese, B., Choui, S. J., Vanden Broeck, J., Van Beemlen, J., De Looft, A., and Schoofs, L. (1998) FEBS Lett. 422, 74–78
7. Liang, Z., Sotrup-Jensen, L., Ascan, A., Hall, M., and Soderhall, R. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 6682–6687
8. Mer, G., Kellenberger, K., Koehl, P., Stote, R., Sorokin, E., Van Dorselaer, A., Luu, B., Hietter, H., and Lefevre, J. F. (1994) Biochemistry 33, 15397–15407
9. Mer, G., Hietter, H., Kellenberger, C., Renatus, M., Luu, B., and Lefevre, J. F. (1989) J. Mol. Biol. 215, 158–171
10. Sreebier, L., and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162
11. Ovinnikov, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
12. Navaza, J. (1994) Acta Crystallogr. Sect. A 50, 157–163
13. Velioux, F. M., Hunt, J. F., Roy, S., and Read, R. J. (1993) J. Appl. Crystallogr. 26, 347–351
14. Brunger, A. T. (1996) X-PLOR, Version 3.843, Yale University Press, New Haven, CT
15. Roussel, A., and Cambillau, C. (1991) in Silicon Graphics Geometry Parameters Directory (Graphic, S., ed) pp 86, Silicon Graphics, Mountain View, CA
16. Huang, K., Strychnak, N. C., Bernard, V. D., Peanasky, R. J., and James, M. N. (1994) Structure 2, 679–689
17. Fujimaru, M., Sielecki, A. R., Sarbu, W., Laskowski, M., and James, M. N. (1987) J. Mol. Biol. 195, 397–418
18. McGrath, M. E., Erpel, T., Bystroff, C., and Fletterick, R. J. (1994) EMBO J. 13, 1502–1507
19. Gillmor, S. A., Takeuchi, T., Yang, S. Q., Craik, C. S., and Fletterick, R. J. (2000) J. Mol. Biol. 299, 993–1005
20. Malik, Z., Amir, S., Pal, G., Buzas, Z., Varallyay, E., Antal, J., Szilagyi, Z., Vekey, K., Ashoth, B., Patthy, A., and Graf, L. (1999) Biochem. Biophys. Acta 1434, 143–150
21. Tsunemi, M., Matsuura, Y., Sakakibara, S., and Katsube, Y. (1996) Biochemistry 35, 11570–11576
22. Francart, C., Duachez, M., Aïx, A. J., and Lippens, G. (1997) J. Mol. Biol. 268, 666–677
23. Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A., and Kossiakoff, A. A. (1997) Protein Sci. 6, 1806–1824
24. Bode, W., and Huber, R. (2000) Biochim. Biophys. Acta 1477, 241–252
25. Egu, M. (2000) Comp. Biochem. Physiol. A 127, 449–456
26. Boto, M., Meyer, E., Nguyen, M., Swanson, H. S., Koonen, M. J., Russell, D. H., and Meyer, E. F. (2000) J. Mol. Biol. 298, 895–901
27. Lam, W., Coast, G. M., and Rayne, C. R. (1999) Insect Biochem. Mol. Biol. 29, 653–660
28. Lam, W., Coast, G. M., and Rayne, R. C. (2000) Insect Biochem. Mol. Biol. 30, 85–94
29. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950