The Structure of Rat Mast Cell Protease II at 1.9-Å Resolution†

Stephen J. Remington,*,‡ Richard G. Woodbury,§ Ross A. Reynolds,‡,∥ Brian W. Matthews,‡ and Hans Neurath§

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, and Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received March 21, 1988; Revised Manuscript Received June 21, 1988

ABSTRACT: The structure of rat mast cell protease II (RMCP II), a serine protease with chymotrypsin-like primary specificity, has been determined to a nominal resolution of 1.9 Å by single isomorphous replacement, molecular replacement, and restrained crystallographic refinement to a final *R*-factor of 0.191. There are two independent molecules of RMCP II in the asymmetric unit of the crystal. The rms deviation from ideal bond lengths is 0.016 Å and from ideal bond angles is 2.7°. The overall structure of RMCP II is extremely similar to that of chymotrypsin, but the largest differences between the two structures are clustered around the active-site region in a manner which suggests that the unusual substrate specificity of RMCP II is due to these changes. Unlike chymotrypsin, RMCP II has a deep cleft around the active site. An insertion of three residues between residues 35 and 41 of chymotrypsin, combined with concerted changes in sequence and a deletion near residue 61, allows residues 35–41 of RMCP II to adopt a conformation not seen in any other serine protease. Additionally, the loss of the disulfide bridge between residues 191 and 220 of chymotrypsin leads to the formation of an additional substrate binding pocket that we propose to interact with the P_3 side chain of bound substrate. RMCP II is a member of a homologous subclass of serine proteases that are expressed by mast cells, neutrophils, lymphocytes, and cytotoxic T-cells. Thus, the structure of RMCP II forms a basis for an explanation of the unusual properties of other members of this class.

Mast cells contain high levels of proteolytic enzymes that are segregated in secretory granules. They include chymotrypsin- (Woodbury et al., 1978a,b) or trypsin-like (Kido et al., 1985) serine proteases of distinct specificity toward low molecular weight substrates, as well as a carboxypeptidase A like enzyme (Everitt & Neurath, 1980). Although several of these proteolytic enzymes are stored within the granules in active form (Woodbury & Neurath, 1978; Everitt & Neurath, 1980; Woodbury et al., 1981), their activities may be regulated by physical adsorption onto proteoglycans, e.g., heparin (Shick et al., 1984), which excludes most protein substrates larger than 15–20 kDa (Le Trong et al., 1987a,b). Moreover, the intracellular pH of the granules is reported to be 5–6 (De Young et al., 1987), which is signficantly lower than that (pH 7–8) required for maximal activity of the proteases (Yoshida et al., 1980).

Other granulocytes, such as neutrophils (Salvesen et al., 1987), and lymphocytes, such as cytotoxic T-cells (Lobe et al., 1986; Schmid & Weissmann, 1987), contain proteases that are closely related to two chymotrypsin-like enzymes, i.e., rat mast cell proteases I and II (RMCP I and II).[1] Collectively, these proteases exhibit a homologous relationship to each other (approximately 50% amino acid sequence identity). Significantly, all of the chymotrypsin-like proteases lack a disulfide bond that in other serine proteases is located near the primary and secondary substrate binding sites (Cys 191–Cys 220 in chymotrypsin) (Le Trong et al., 1987a,b; Woodbury et al., 1978c). For these reasons, it has been proposed that the chymotrypsin-like enzymes of granulocytes and lymphocytes comprise a subclass of serine proteases that has diverged in the course of evolution from the pancreatic and hepatic proteases (Woodbury & Neurath, 1981; Woodbury et al., 1987).

The substrate specificities of RMCP I, isolated from mast cells of the peritoneal cavity, and RMCP II, isolated from mucosal mast cells, have been extensively studied by using peptide-*p*-nitroanilide (Yoshida et al., 1980; Powers et al., 1985) and protein substrates (Kobayashi et al., 1978; Kobayashi & Katunuma, 1978; Le Trong et al., 1987a,b). Overall, the specificities of these mast cell proteases are similar, but the rate of hydrolysis of a given substrate by RMCP I is usually significantly higher than that by RMCP II (Yoshida et al., 1980; Powers et al., 1985). Unlike pancreatic chymotrypsin, which efficiently hydrolyzes substrates containing a hydrophobic residue only at the P_1 position (Segal, 1972), the mast cell proteases are most active toward substrates containing hydrophobic residues also at the P_2 and P_3 positions (Yoshida et al., 1980). Additionally, RMCP I exhibits a preference for hydrolyzing peptide bonds formed by hydrophobic residues at P_1 and P_1' (Le Trong et al., 1987a,b).

In order to understand how the unusual substrate specificities of these mast cell proteases relate to their structures, we have determined and previously reported the three-dimensional structure of RMCP II (Reynolds et al., 1985). A refinement of the X-ray structure is presented herein, together with a detailed comparison of the structure of RMCP II to those of chymotrypsin and elastase. The results indicate that there are significant differences in the structure of RMCP II compared to those of the pancreatic enzymes and that these differences are consistent with the substrate specificity differences observed for these proteases.

Prior to the structural analysis of RMCP II, an attempt was made to predict the structure of this molecule [called group-specific protease or GSP in older literature (Katunuma et al.,

† This work was supported in part by grants from the National Science Foundation (DMB-8517785 and DMB-8611084) and the National Institutes of Health (HL-36114).

* Address correspondence to this author.

‡ University of Oregon.

§ University of Washington.

∥ Present address: Department of Physics, Colby College, Waterville, ME 04901.

[1] Abbreviations: RMCP I, rat mast cell protease I; RMCP II, rat mast cell protease II.

0006-2960/88/0427-8097$01.50/0

1975)] from the amino acid sequence (Greer, 1981). A reliable method to estimate three-dimensional structure from sequence is needed, and it is important to evaluate the success of such attempts. In the proposed method, the assumption is made that nonhomologous loops of identical length would have similar conformations in homologous proteins. We show that this assumption failed for a loop forming part of the active-site cavity and discuss reasons for the failure.

Materials and Methods

RMCP II crystallizes from 23% saturated ammonium sulfate as described previously (Anderson et al., 1978) in space group $P3_1$ with cell dimensions $a = b = 78.2$ Å and $c = 96.8$ Å and two molecules per asymmetric unit related by a noncrystallographic 2-fold axis nearly parallel to the x axis. The crystals are merohedrally twinned about the local 2-fold axis with the result that each observed intensity (hkl) consists of the superposition of two components, hkl and $kh\bar{l}$.

Native and heavy atom derivative data collection, detwinning to 1.9-Å resolution and the location of two chymotrypsin-like molecules in the unit cell using real space rotation and translation in a single isomorphous replacement (SIR) electron density map have already been discussed (Reynolds et al., 1985).

Crystallographic refinement proceeded first with reciprocal-space, rigid-body refinement of the position and orientation of the two molecules using the CORELS program (Sussman et al., 1977), followed by energy-restrained least-squares refinement using the EREF program (Jack & Levitt, 1978). Electron density maps were calculated with coefficients ($2F_o - F_c$) phased in the early stages by combining the SIR and calculated phases (Hendrickson & Lattmann, 1978) and in later stages by Sim-weighted calculated phases (Sim, 1959). The maps were averaged about the local 2-fold axis by using the PROTEIN program (Steigemann, 1974). The position of this local axis was determined by the algorithm of Kabsch (1978) in which the two molecules in the asymmetric unit were superimposed. The maps were inspected and model alterations carried out by using the FRODO program (Jones, 1978) on an MMSX graphics system (Molnar et al., 1976) and an Evans and Sutherland PS300. Structural comparisons were performed according to the technique of Rossmann and Argos (1976) with the program OVRLAP by W. S. Bennett. Residue numbering is based on that of chymotrypsin with tentative locations of insertions marked by appending a letter to the residue number.

For structural comparisons, highly refined coordinates of γ-chymotrypsin at pH 7.0 (R-factor 17.3% at 1.6 Å) were used (Cohen et al., 1981; M. Dixon and B. W. Matthews, manuscript in preparation), as were those for elastase (Sawyer et al., 1978), subsequently refined by others (R-factor 18.4% at 1.65 Å) (Meyer et al., 1986; coordinates kindly provided by Edgar Meyer).

Results

Crystallographic Refinement. The starting model consisted of a 2-fold dimer of chymotrypsin (Birktoft & Blow, 1972) with coordinates obtained from the Protein Data Bank (Bernstein et al., 1972), oriented according to the rotation/translation function solution described by Reynolds et al. (1985). The models were truncated to 72% of the atoms expected to have similar conformation in RMCP II (Greer, 1981). Residues 35–37 and 74–77, which were poorly defined in the original chymotrypsin structure, were also removed. As discussed later, we note that there is no relation between the dimerization of α-chymotrypsin discussed by Birktoff and Blow and that reported here; the orientation of the subunits in the two dimers is quite different. This model had an R-factor of 50.9% for data from 6.0 to 3.5 Å. The rigid body refinement proceeded in two stages. Rotations of 1–2° and shifts of a few tenths angstroms, followed by geometry idealization, lowered the R-factor to 49.3% for these data.

Five cycles of model building in averaged maps, each followed by EREF refinement, resulted in a nearly complete model with an R-factor of 19.1% for the 28782 reflections recorded between 5.0- and 1.9-Å resolution. After each cycle of model building into an averaged map, the dimer was regenerated by using the local symmetry operator derived from the previous model. The final electron density map is ambiguous for the location of a number of atoms. Some of these (typically at the end of side chains) were included in the model with zero weight (occupancy) or were deleted if a smaller side chain was felt to be an adequate description of the electron density. Three sections of the main chain have weak or no interpretable electron density associated with them. These are residues 96, 97, 98, 169, 170, 170A, 170B, and C-terminal residue 243, probably representing flexible regions of the chain. Although they have been included in the model with zero weight, the conformations of these residues are strictly hypothetical. Four side chains are represented by smaller amino acids in the model; these are Arg 76 (Ala), Val 97 (Ala), Arg 170B (Gly), and Ile 242 (Val).

In the fifth cycle, the electron density was not averaged, and a careful inspection was undertaken to see if there were detectable differences between the two molecules in the asymmetric unit. No large differences were detected, but minor changes were made to some side-chain positions. In particular, density did not appear in any of the regions where the main chain appeared to be disordered. Therefore, the lack of density in these regions in the averaged maps was not an artifact of the averaging process. In the sixth cycle, the handedness of several chiral centers, which had become inverted during the EREF refinement, was corrected.

In the final stages, water molecules were added in locations that showed peaks in $2F_o - F_c$ maps as well as $F_o - F_c$ maps and were within about 3.0 Å of hydrogen-bond acceptors or donors. In total, 123 water molecules are associated with the dimer. Forty-four pairs of these solvent molecules closely obey the local 2-fold axis and show an rms deviation of 0.6 Å from exact 2-fold symmetry. There is evidence for additional water molecules in the final map, but these have not yet been taken into account. The refined coordinates have been deposited with the Protein Data Bank.

Accuracy of the Model. The stereochemistry of the final model is excellent, with rms deviations from ideal bond lengths of 0.016 Å and from ideal bond angles of 2.7°. After the final round of crystallographic refinement, superposition of 1620 atoms with nonzero weight of one molecule onto the other resulted in an rms deviation of position of 0.39 Å for all atoms.

The largest deviations from one molecule to the next occur in side-chain positions that may be different due to crystal contacts. After removing from consideration 62 atoms with positional deviations larger than twice the rms value, we found that the remaining 1558 atoms superimpose to an accuracy to 0.26-Å rms. This provides a conservative estimate of the coordinate error in this investigation, since there may be distortions induced by crystal packing forces. If we superimpose the 24 main- and side-chain atoms comprising the catalytic triad Asp 102, His 57, and Ser 195 of the two molecules, these atoms superimpose to 0.13-Å rms, providing a lower bound for the coordinate error.

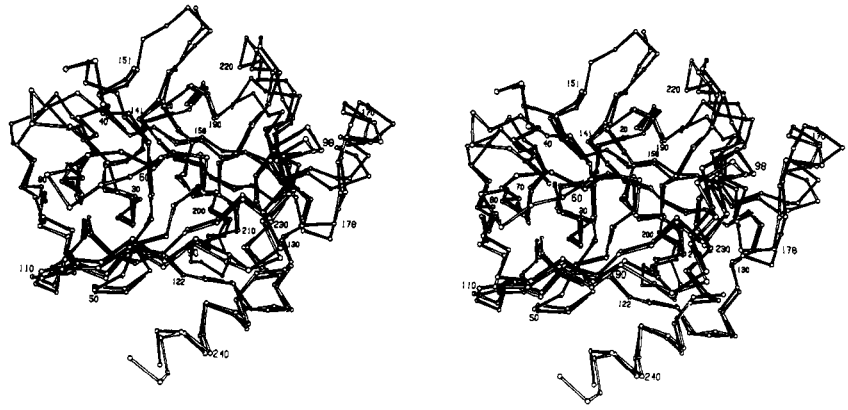

FIGURE 1: α-Carbon drawing of chymotrypsin (open bonds) superimposed on RMCP II (filled bonds). The residue numbering is that of chymotrypsin.

```
          16      20        25        30        35              40        45
CHT        I V N G E E A V P G S W P W Q V S L Q D k t g f       h F C G G S
RMCP II    I I G G V E S I P H S R P Y M A H K D I v t e k g l r v I C G G F
PRED       I I G G V E S i p h S R P Y M A H L D I v t e k g l R V I C G G F

                50        55        60        65        70        75
CHT        L I N E N W V V T A A H C G V T t S D V V V A G E F D Q G S
RMCP II    L I S R Q F V L T A A H C K G R   E I T V I L G A H D V R K
PRED       L I S R Q F V L T A A H C k g r   E I T V I L G A H d v r k

                  80        85        90        95       100       105
CHT        S s e   k I Q K L K I A K V F K N S K Y N S L T I N N D I T L
RMCP II    R e s   t Q Q K I K V E K Q I I H E S Y N . . . N L H D I M L
PRED         r e s t q Q K I K V E K Q I I H E S Y N s v p n l h D I M L

                110       115       120       125       130       135
CHT        L K L S T A A S F S Q T V S A V C L P S A S D D F A A G T T
RMCP II    L K L E K K V E L T P A V N V V P L P S P S D F I H P G A M
PRED       L K L E K K V E L T p A V N V V P L P s p s d f i h p g A M

                140       145       150       155       160       165
CHT        C V T T G W G L T R Y . . . . T P D R I Q Q A S L P L L S N
RMCP II    C W A A G W G K T G V r d p t S Y T L R E V E L R I N D E K
PRED       C W A A G W G K T G v r d p t S Y T L R E V E L R I M D E k

                170       175       180       185       190       195
CHT        T N C K k y w g t K I K   D A M I C A G     A S G   V S S C M G D S
RMCP II    A C . . . .       Y Y E y K F Q V C V G s   P T T 1 R A A F M G D S
PRED       a c v d     y r y y e y   k f Q V C V G s p t t l   R A A F M G D S

                200       205       210       215       220       225
CHT        G G P L V C K k n g a W T L V G I V S W G S S T C S t s   T P
RMCP II    G G P L L C A         G V A H G I V S Y G H P D A K     P P
PRED       G G P L L C A         G V A H G I V S Y G h p       d a k p P

                230       235       240       245
CHT        G V Y A R V T A L V N W V Q Q T L A a n
RMCP II    A I F T R V S T Y V P T I N A V I .
PRED       A I F T R V S T Y V P T I N A V I N
```

FIGURE 2: Alignment of amino acid sequences of γ-chymotrypsin and RMCP II based on the correspondence of their three-dimensional structures as assessed by the method of Rossmann and Argos. Upper case letters indicate residues that are structurally equivalent, lower case letters indicate residues that are not structurally equivalent, and periods indicate flexible loops where the electron density is too weak to permit an unambiguous interpretation. Deletions are indicated by blanks. PRED is the prediction by Greer (1981). The residue numbering is that of chymotrypsin.

Backbone Conformation and Structural Comparisons. As expected, the overall fold of the molecules is typical of those serine proteases that show sequence similarity with chymotrypsin. The structure consists of two six-stranded β barrels with the active site located in the cleft between the barrels. There are, however, significant differences between RMCP II and the other serine proteases of known structure, which are likely to account for the observed difference in specificity.

If we compare chymotrypsin with RMCP II by the method of Rossmann and Argos (1976) using values of E1 and E2 of 3.0 Å, 206 α-carbons are determined to be structurally equivalent and to superimpose to an rms distance of 1.21 Å. More conservative values for E1 and E2 of 1.0 Å yield 162 equivalent residues that superimpose to 0.66-Å rms. The first superposition is illustrated in Figure 1 and summarized in Figure 2. Of the 206 structurally equivalenced amino acids, 65 are identical, leading to a sequence identity of 32%. Alternatively, 65 of 224 residues total in RMCP II yield 29% sequence identity.

From Figure 1, it is clear that there are several insertions and deletions in loops surrounding the active site. Of particular significance is a three-residue insertion somewhere between residues 34 and 41 (chymotrypsin numbering) and a deletion of one residue somewhere between residues 59 and 65. Due

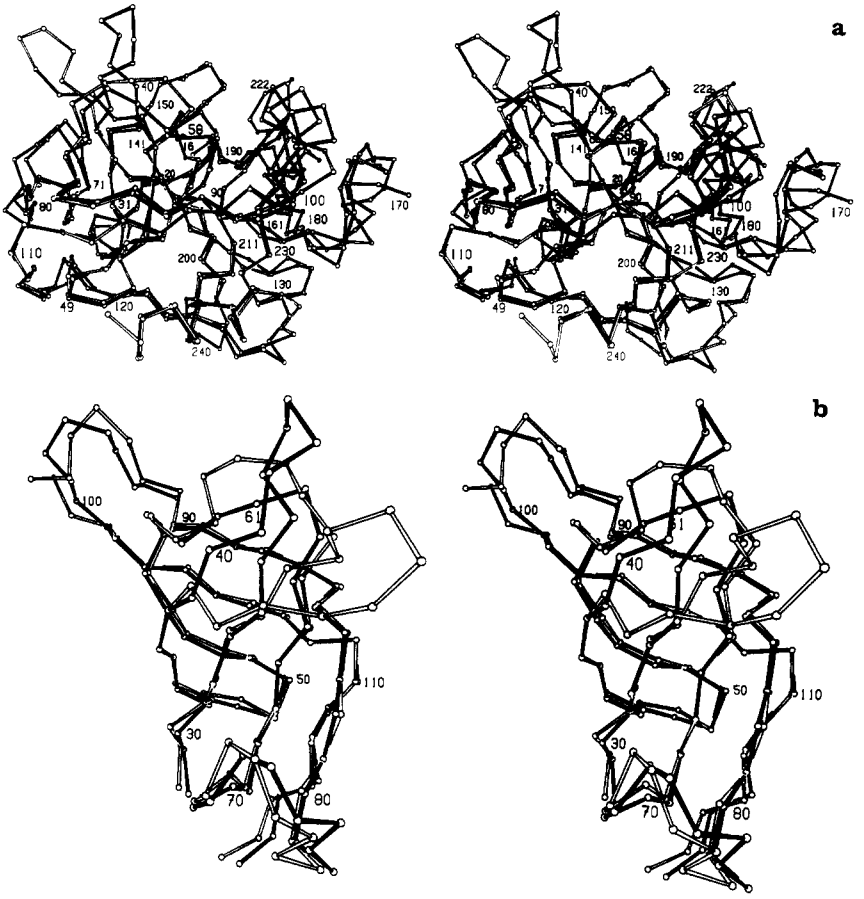

FIGURE 3: (a) α-Carbon drawing of elastase (open bonds) superimposed on RMCP II (filled bonds). The residue numbering is that of elastase. (b) α-Carbon drawing of the N-terminal β barrel of elastase (open bonds) superimposed on RMCP II (filled bonds). The residue numbering is that of RMCP II.

to these changes, the loops consisting of residues 34–41 and 59–65 are in substantially different conformation than the corresponding residues of chymotrypsin and tower over the active-site region, resulting in a deep cleft in RMCP II that is not present in chymotrypsin. There are also significant rearrangements of residues 191–222 (chymotrypsin numbering) due to the loss of the disulfide 191–220. This will be discussed in more detail later.

Similar results are obtained when one superimposes elastase on RMCP II. With values of E1 and E2 of 1.0 Å, 152 residues are equivalent with an rms deviation of 0.68 Å (data not shown). The α-carbon trace of the two backbones and of the N-terminal β barrel for this superposition is shown in parts a and b of Figure 3. Although elastase also has a three-residue insertion between residues corresponding to 34–41 of chymotrypsin and is thus identical in length with that of RMCP II, these loops have substantially different conformations. Main-chains atoms at the tip of these loops are separated by over 12 Å after superposition of structurally equivalent residues. Surprisingly, this loop has a more regular conformation in RMCP II than either that of chymotrypsin or elastase in that the barrel structure is more symmetrical (see Figure 3b).

Noncrystallographic Symmetry. The transformation relating the two molecules in the asymmetric unit can be decomposed into a rotation about an axis combined with a translation along that axis. In polar coordinates [angular system defined by Rossmann and Blow (1962)], the rotation is given by ϕ = 88.6°, ψ = 2.3°, and κ = 179.6° with a net translation of 0.1 Å along the axis. The transformation is therefore very accurately represented by a pure 2-fold rotation axis about 3° from parallel to the x axis. The two molecules in the asymmetric unit interact through two surface loops (residues 39, 39B, and 39C of one molecule with residues 147, 148, 149, and 150 from the second molecule; see Figure 4), leaving a large solvent-filled cavity between them. One of the local 2-fold-related sets of interactions is shown in Figure 5 and is summarized in Table I.

Table I: Hydrogen Bonds between the Two Molecules in the Asymmetric Unit[a]

molecule 1	molecule 2	distance (Å)	HNO angle[b] (deg)
Arg 62 NH1	Arg 174 O	2.7	37
Lys 39 O	Lys 143 NZ	2.7	33
Arg 39C NH1	Thr 150 O	2.8	65
Arg 39C NH2	Thr 150 O	3.3	71
Arg 39C NH1	Thr 150 OG1	3.1	71
Sol 313 O	Thr 150 N	3.0	17
Gly 39A O	Sol 313 O	2.7	*c*

[a] The actual number of hydrogen bonds is twice the above by application of the local 2-fold operator. [b] Obtained from calculated hydrogen atom coordinates. Ideally this angle, a measure of the linearity of the bond, should be about 10°. [c] Arbitrary.

There are five intermolecular hydrogen bonds as well as two mediated through a water molecule, plus a van der Waals interaction between Pro 149 of one molecule and Val 36 of the other. Several of the hydrogen bonds are clearly not ideal in length or orientation, and all involve side-chain–main-chain interactions (except for Sol 313), so we expect this dimerization to be an artifact of crystallization not anticipated in dilute solution. Although residues 35–41 of RMCP II participate in this intermolecular contact, inspection of the structure

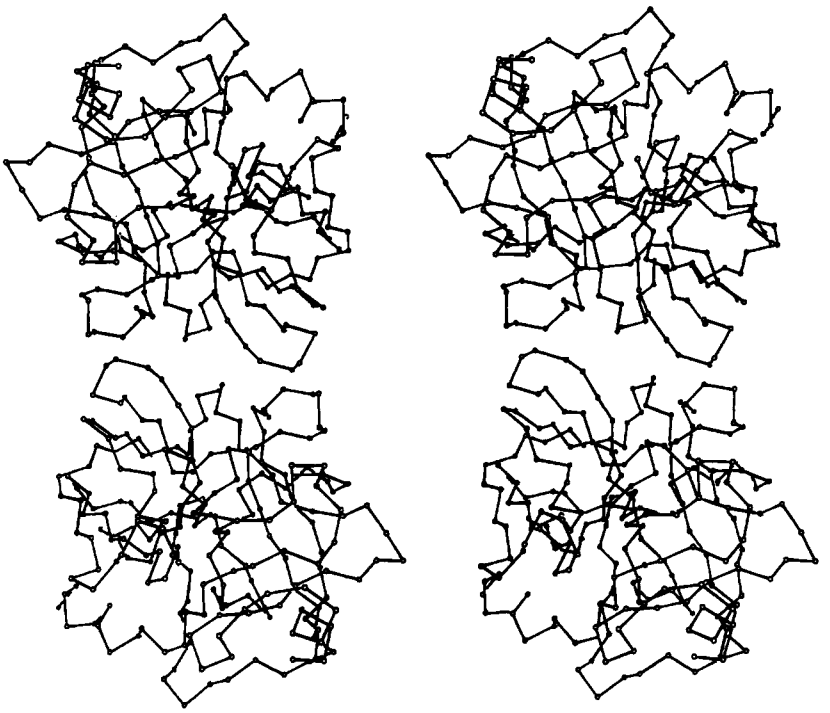

FIGURE 4: α-Carbon drawing of the dimer of RMCP II that forms the asymmetric unit in the crystal.

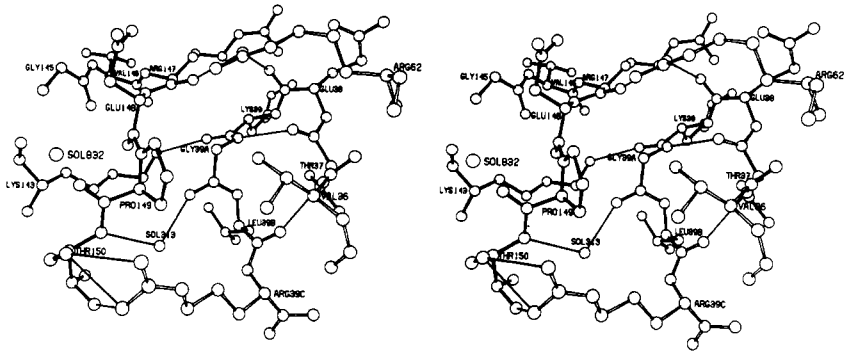

FIGURE 5: Detail of one of the local 2-fold related sets of contacts forming the dimer interface. Molecule 1 is shown with open bonds; molecule 2 is shown with filled bonds. The thin bonds represent possible hydrogen bonds.

suggests that the conformation of this loop is dominated by its participation in the regular hydrogen-bond network and the hydrophobic core of the N-terminal β barrel. Note that the active sites of the two molecules are not occluded upon dimerization, as in the case for the dimer formed by α-chymotrypsin upon crystallization [for an extensive discussion, see Blevins and Tulinsky (1985)].

The Catalytic Triad. Upon isolation of the 24 main- and side-chain atoms of the catalytic triad His 57, Asp 102, and Ser 195 of either molecule of RMCP II, these are found by the Kabsch algorithm to superimpose on the same atoms of γ-chymotrypsin with an rms error of 0.20 Å. This is probably within the coordinate error of both structure determinations combined, so we concluded that the conformation of these residues is identical in the two serine proteases.

Conformation and Substrate Specificity. The most striking difference between chymotrypsin, elastase, and RMCP II is the pronounced cleft formed in part by residues 35–41 and 60–65 of RMCP II. Since this structure could be involved in substrate binding and may be important for specificity, we have undertaken a detailed examination of this structure in RMCP II, elastase, and chymotrypsin. The results indicate that a number of concerted amino acid substitutions have occurred that can account for the movement and support of this major surface structure in RMCP II. A comparison of elastase and RMCP II was made since there are approximately the same number of residues in these structures. The superposition of this region using the transformation discussed previously, superimposing 152 α-carbons to 0.68 Å, is shown in Figure 6.

In elastase, residues 35–40 and 64–65A form three strands of β sheet supported on the outside by the stacking of large side chains (Trp 38 and Arg 65A) and on the active-site side by Phe 65 and Tyr 35. In RMCP II, all of these residues are replaced by smaller side chains (note, for example, Phe 65 and Tyr 35 in elastase are structurally equivalent to Ile 64 and Ile 35 in RMCP II). In addition, two residues, apparently residues 60 and 61 in elastase, are deleted in RMCP II. This smaller loop is positioned in RMCP II by main-chain–side-chain hydrogen bonds to Gln 88. Although residues 59–65 and 35–40 still form three strands of β sheet in RMCP II, the net effect is to allow residues 35–40 to rotate inward toward the active site, supported by hydrophobic interactions on the active-site side of this loop. Later the possibility is raised that this structure may interact with substrate in the P_1' position and possibly other positions as well.

The backbones of RMCP II and chymotrypsin are nearly identical in the vicinity of the primary specificity pocket (the

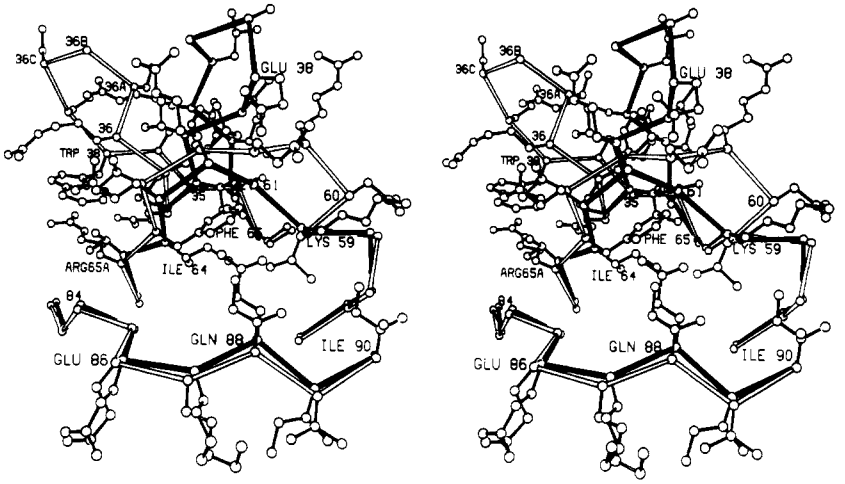

FIGURE 6: Detail showing concerted structural changes leading to a major conformational difference between elastase (open bonds) and RMCP II (filled bonds). The residue numbering is that of both molecules.

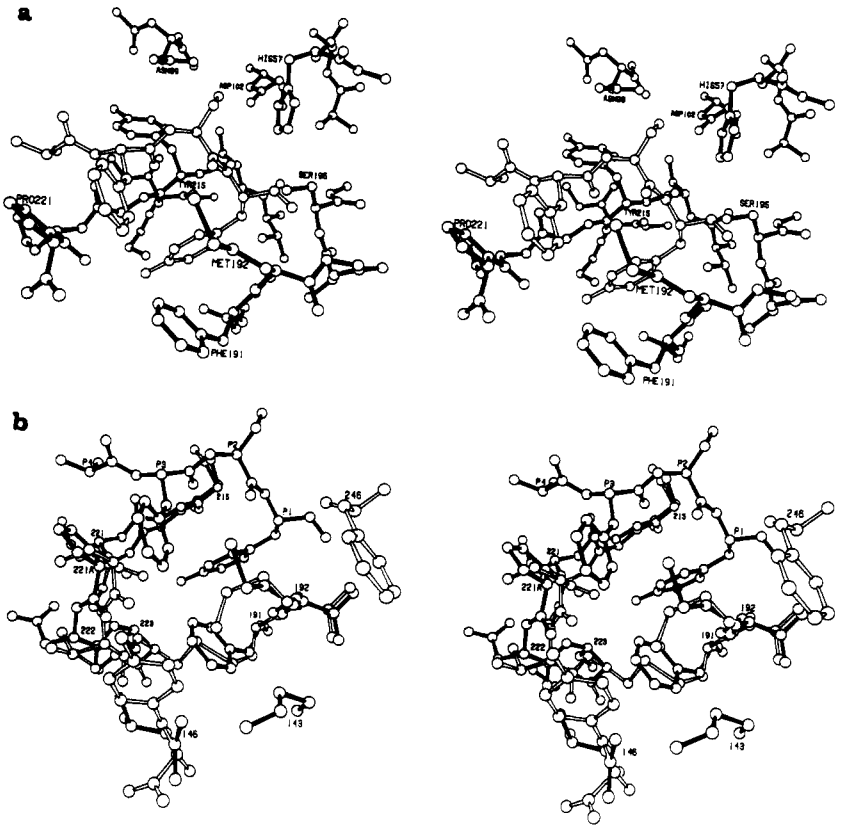

FIGURE 7: (a) Hypothetical model of the N-terminal section of the substrate Ala-Phe-Ser-Tyr- (open bonds) binding to RMCP II for the substrate showing the phenylalanyl side chain in the proposed P_3 binding site. (b) As in (a) but with the corresponding atoms of chymotrypsin overlaid (open bonds).

"tosyl hole") and the active-site seryl residue Ser 195. This region includes the disulfide bond (Cys 191–Cys 220 in chymotrypsin) found in all serine proteases except RMCP II, RMCP I, and related granulocyte and lymphocyte enzymes (Woodbury et al., 1978c; Lobe et al., 1986; Schmid & Weissmann, 1987; Le Trong et al., 1987a,b). The absence of this disulfide bond and the substitution in RMCP II of a phenylalanyl residue for Cys 191 force a β turn (residues labeled 221–223) and the side chain of Met 192 outward by about 1.5 Å compared to the analogous residues 217–220 and Met 192 of chymotrypsin. In RMCP II, residues Pro 221A and Met 192 form a small hydrophobic pocket not found in chymotrypsin.

To investigate whether this pocket could account for the specificity for hydrophobic residues at P_3 (Yoshida et al., 1980), a model substrate was constructed. We took advantage of the fact that the native γ-chymotrypsin structure apparently contains the tetrapeptide Pro-Gly-Ala-Tyr (M. Dixon and B. W. Matthews, manuscript in preparation) bound in the active site of the protease. The tetrapeptide was transformed onto the RMCP II active site by the transformation that superimposes 162 α-carbons of chymotrypsin onto RMCP II with an rms error of 0.66 Å. Inspection of the model revealed that such a peptide could interact with RMCP II in the manner typical of peptide inhibitors of the serine proteases. These inhibitors form a short section of extended antiparallel β sheet with residues 214, 215, and 216 of the binding site (Segal et al., 1971). The lengths and orientations of these putative

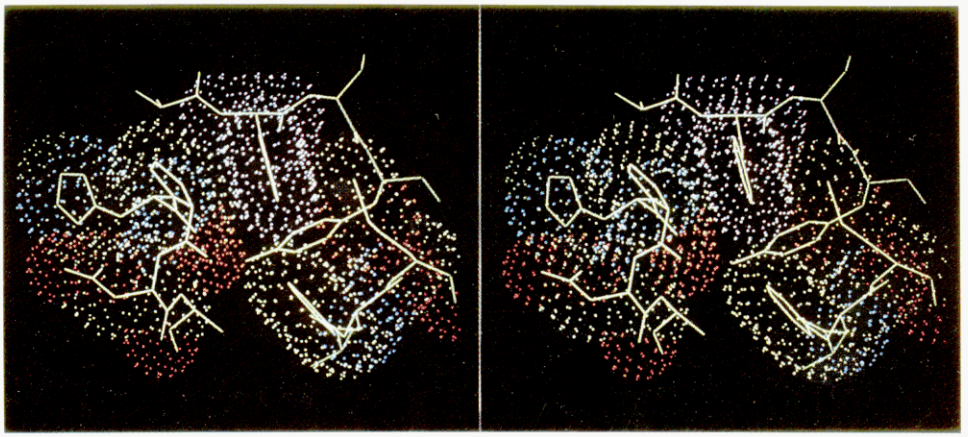

FIGURE 8: van der Waals dot surface representation of the proposed substrate in the P_3 binding site (purple). Other colors represent carbon (green), oxygen (red-brown), and nitrogen (blue).

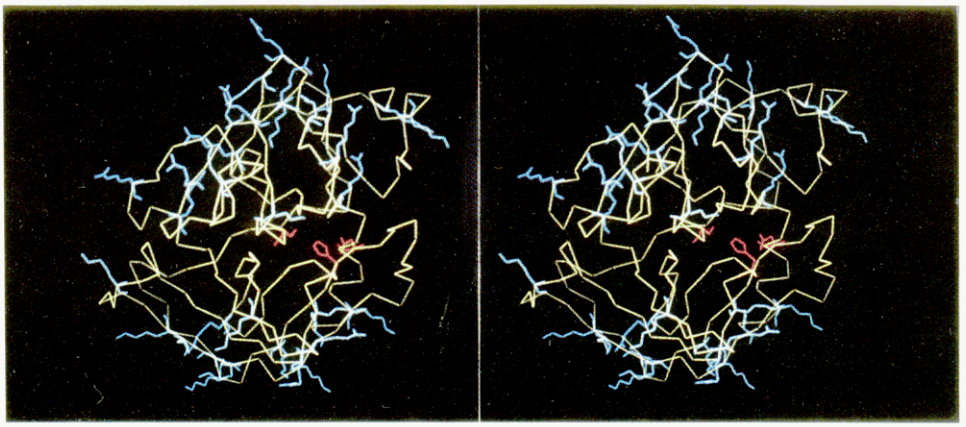

FIGURE 9: Backbone drawing (yellow) of the hypothetical model of RMCP I showing the localization of positively charged lysyl and arginyl residues (blue) to two patches on the molecular surface. The atoms of the charge relay system are shown in red.

β-sheet hydrogen bonds in our model are satisfactory. Substitution of the phenylalanine side chain for glycine revealed that the side chain of the phenylalanine can be completely accommodated by the hydrophobic pocket formed by Pro 221A, Met 192, and Phe 191 with nearly perfect van der Waals contacts and acceptable stereochemistry but that little, if any, interaction with a phenylalanine at P_3 is possible in this region of chymotrypsin. This hypothetical model is diagrammed in Figures 7 and 8.

The position and orientation of the modeled peptide substrate indicate that residues at the P_1', P_2', and P_3' of an extended substrate could interact with residues in the loop region formed by residues 35–42 in RMCP II. The difference in structure in this region relative to that of chymotrypsin may therefore play a major role in determining substrate specificity.

Modeling of RMCP I. The amino acid sequence of RMCP I, obtained from peritoneal mast cells, has been determined, and a comparison of this structure to that of RMCP II indicated that the amino acid sequences were 73% identical (Le Trong et al., 1987a,b). Where there are differences in the sequences of the two proteases, the residues of RMCP I were substituted for those of RMCP II by using FRODO in order to obtain a putative three-dimensional model of RMCP I. The results indicate that only six amino acid substitutions have occurred in the interior of the protein and all of these are conservative. The other changes all occur on the protein surface. No change in the backbone conformation was necessary to accommodate the sequence changes. The model obtained is consistent with a prediction that the major changes in the three-dimensional structure of RMCP II relative to chymotrypsin and elastase should also be observed for RMCP I and other similarly related granulocyte and lymphocyte proteases. A major difference between the two proteases is that RMCP I has 10 more lysyl residues than RMCP II, and these were suggested to be important for the very strong binding of RMCP I to the heparin contained in the mast cell granules (Woodbury & Neurath, 1978). Our hypothetical model of RMCP I (Figure 9) suggests that these residues and the arginyl residues are clustered in two patches at either end of the molecule, relatively far from the active site.

Discussion

Although the overall folding of RMCP II is similar to that of other members of the mammalian serine protease family whose three-dimensional structures have been determined, some areas of the surface structure of RMCP II are significantly different from the corresponding regions in both chymotrypsin and elastase. The major structural changes in RMCP II all occur very near the active site of the protease and appear to be related to the differences in substrate specificities observed between RMCP II and chymotrypsin (Yoshida et al., 1980). Similar observations have recently been made in a comparison of rat submaxillary gland tonin and kallikrein (Fujinga & James, 1987).

The substrate specificities of RMCP I and RMCP II are similar to each other but quite different from that of chymotrypsin (except at the P_1 position). For example, the nature of the residue at the P_3 position of the substrate has little effect on the hydrolysis by chymotrypsin, whereas it significantly affects the activities of RMCP I and RMCP II. Each mast

cell protease hydrolyzes Phe-Leu-Phe-*p*-nitroanilide much more rapidly than Gly-Leu-Phe-*p*-nitroanilide. Kinetic analyses indicate that for each protease the K_m of a substrate with Gly at P_3 is approximately 100-fold higher than that for a substrate with Phe at that position. For each enzyme, k_{cat} is similar for the two substrates. Additionally, it was observed that RMCP I and RMCP II, but not chymotrypsin, readily hydrolyze substrates with a prolyl residue at P_3 (Yoshida et al., 1980). These observations are consistent with the hypothesis that the structures of the mast cell proteases allow extended interaction with substrate, thus providing greater specificity than that possible with chymotrypsin (Woodbury & Neurath, 1981).

On the basis of structural (Woodbury et al., 1978c) and substrate specificity studies (Yoshida et al., 1980; Le Trong et al., 1987a,b), it was proposed that RMCP I and RMCP II have extended, possibly hydrophobic, binding sites not present in the pancreatic serine proteases (Woodbury et al., 1987). These extended sites were presumed to result, in part, from the deletion of the disulfide bond present in other mammalian serine proteases (Cys 191–Cys 220 in chymotrypsin). Analysis of the three-dimensional structure of RMCP II has now provided evidence that this protease, and probably RMCP I as well, has a hydrophobic binding pocket that can interact with the side chain of the P_3 residue of the substrate and which apparently accounts for the unusual substrate specificity of the mast cell enzymes.

The significance of the major difference between the structure of RMCP II and chymotrypsin, near His 40 in chymotrypsin, is not yet clear. This segment of the polypeptide backbone of RMCP II differs from the corresponding region in γ-chymotrypsin by as much as 15–20 Å. In RMCP II, this structural change results in a prominent protrusion extending from the protein surface, causing the formation of a deep cleft near the active site (Figures 1 and 3a). Model building suggests that, in the RMCP II structure, this cleft provides additional binding sites that could interact with the P_1' and other residues of the substrate on the carboxy-terminal side of the scissle peptide bond. Since the structures of chymotrypsin and elastase lack this cleft, there is little possibility for similar interactions between these proteases and substrates.

It is highly desirable that a reliable method be developed to predict tertiary structures from sequence data. While we are still far from this goal, an approach suggested by Greer (1981) seems applicable to members of a family of similar structures. Members of unknown structure are modeled, as far as possible, on those of known structure. In particular, the assumption is made that corresponding loops of the same length will adopt similar conformations in different molecules, even if their amino acid sequences are different. In the present comparison, the prediction was made that residues 35–41 of RMCP II would adopt a conformation similar to that of the corresponding residues of elastase. As was noted, the assumption failed in the case of RMCP II. Furthermore, the incorrect prediction suggested a structure that would have had very different substrate specificity relative to the structure actually adopted by the protein.

Inspection of Figure 2 also reveals that insertions and deletions do not necessarily lead to a markedly different conformation. For example, variable regions 3, 8, and 9, corresponding to residues 59–62, 166–179, and 185–187 of chymotrypsin, all have differing numbers of residues in RMCP II as compared to chymotrypsin, yet their structures are very similar. In the absence of the three-dimensional structure, one might be tempted to conclude that these loops in chymotrypsin would not be a good model for the corresponding loops in RMCP II, but this is not the case.

While the proposed approach seems reasonable, it is not clear how one would take into account the possibility of concerted changes similar to those described herein, which ultimately may be critical in identifying residues responsible for the biological role of the molecule of interest. The structure of RMCP II, on the other hand, does support the notion that "structurally conserved regions" in the serine proteases (as defined by Greer) are indeed conserved, as is shown by the extreme similarity of the backbone structure in spite of low sequence identity.

Due to the overall high sequence identity (73%) between RMCP II and RMCP I, a structural prediction based on homology will be on somewhat safer ground. Our analysis indicates that the amino acid sequence of RMCP I will allow the protein to fold in a manner very similar to that of RMCP II, and we predict that RMCP I also has a prominent cleft at its active site. Specificity studies using different peptide and protein substrates indicate that RMCP I, but not chymotrypsin, preferentially hydrolyzes peptides between adjacent hydrophobic residues (Le Trong et al., 1987a,b). This suggests that RMCP I may have a binding site for the P_1' residue and that proteolytic activity is influenced by this interaction. Although the three-dimensional structure of RMCP II, and that predicted for RMCP I, may account for the unusual substrate specificity observed for RMCP I, the structure of RMCP I needs to be determined, and the substrate specificity studies must be extended to RMCP II before this hypothesis can be confirmed.

In conclusion, the three-dimensional structure of RMCP II is different from those of other known mammalian serine proteases. The high degree of correspondence at the amino acid sequence level between the mast cell proteases and similar chymotrypsin-like enzymes of granulocyte and lymphocyte origin suggests that these proteases all have similar three-dimensional structures and that the lack of the 191–220 disulfide bond contributes to enhanced substrate specificity. The results of this study provide further evidence that the mast cell proteases are representative of a family of cellular proteases that have evolved as a subclass of the serine proteases and that they may be functionally distinct from the pancreatic and hepatic proteolytic enzyme families (Woodbury & Neurath, 1981).

Registry No. RMCP II, 82599-74-4.

References

Anderson, W. F., Matthews, B. W., & Woodbury, R. G. (1978) *Biochemistry 17*, 819.

Bernstein, F. C., Koetzel, T. F., Williams, G. J. G., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535.

Birktoft, J. J., & Blow, D. M. (1972) *J. Mol. Biol. 68*, 187.

Blevins, Z. A., & Tulinsky, A. (1985) *J. Biol. Chem. 260*, 4264.

Cohen, G. H., Silverton, E. W., & Davies, D. R. (1981) *J. Mol. Biol. 148*, 449.

De Young, M. B., Nemeth, E. F., & Scarpa, A. (1987) *Arch. Biochem. Biophys. 254*, 222.

Everitt, M. T., & Neurath, H. (1980) *FEBS Lett. 110*, 292.

Fujinaga, M., & James, M. N. G. (1987) *J. Mol. Biol. 195*, 373.

Greer, J. (1981) *J. Mol. Biol. 153*, 1027.

Hendrickson, W. A., & Lattmann, E. E. (1970) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B26*, 136.

Jack, A., & Levitt, M. (1978) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A34*, 931.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268.
Kabsch, W. (1978) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A34*, 827.
Katunuma, N., Komiyama, E., Kobayashi, K., Banne, Y., Suzuki, K., Chichibu, K., Hamaguchi, Y., & Katsunum, T. (1975) *Eur. J. Biochem. 52*, 37.
Kido, H., Fukusen, N., & Katunuma, N. (1985) *Arch. Biochem. Biophys. 239*, 436.
Kobayashi, K., & Katunuma, N. (1978) *J. Biochem. (Tokyo) 84*, 65.
Kobayashi, F., Sanada, Y., & Katunuma, N. (1978) *J. Biochem. (Tokyo) 84*, 477.
Le Trong, H., Neurath, H., & Woodbury, R. G. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 364.
Le Trong, H., Parmelee, D. D., Walsh, K. A., Neurath, H., & Woodbury, R. G. (1987b) *Biochemistry* 26, 6988.
Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H., & Bleackley, R. C. (1986) *Science (Washington, D.C.) 232*, 858.
Meyer, E. F., Jr., Radhakrishnan, G. M. C., & Presta, L. G. (1986) *J. Mol. Biol. 189*, 533.
Molnar, C. E., Barry, C. D., & Rosenberger, F. U. (1976) Technical Memorandum No. 229, Computer systems Laboratory, Washington University, St. Louis, MO.
Powers, J. C., Tanaka, T., Harper, J. W., Minematus, Y., Barker, L., Lincoln, D., Crumley, K. V., Fraki, J. E., Schecter, N. M., Lazarus, G. G., Nakajima, K., Nakashino, K., Neurath, H., & Woodbury, R. G. (1985) *Biochemistry 24*, 2048.
Reynolds, R. A., Remington, S. J., Weaver, L. H., Fischer, R. G., Anderson, W. F., Ammon, H. L., & Matthews, B. W. (1985) *Acta Crystallogr., Sect. B: Struct. Sci. B41*, 139.
Rossmann, M. G., & Blow, D. M. (1962) *Acta Crystallogr. 15*, 24.
Rossmann, M. G., & Argos, P. (1976) *J. Mol. Biol. 105*, 75.
Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., & Travis, J. (1987) *Biochemistry 26*, 2289.
Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H., Watson, H. C., Diamond, R., & Ladner, R. C. (1978) *J. Mol. Biol. 118*, 137.
Schick, B., Austen, K. F., & Schwartz, L. B. (1984) *J. Immunol. 132*, 2571.
Schmid, J., & Weissman, C. (1987) *J. Immunol. 139*, 250.
Segal, D. M. (1972) *Biochemistry 11*, 349.
Segal, D. M., Cohen, G. H., Davies, D. R., Powers, J. C., & Wilcox, P. E. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 85.
Sim, G. A. (1959) *Acta Crystallogr. 12*, 813.
Steigmann, W. (1974) Doctoral Thesis, Technical University, Munich.
Sussmann, J. L., Holbrook, S. R., Church, G. M., & Kim, S. (1977) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A33*, 800.
Woodbury, R. G., & Neurath, H. (1978) *Biochemistry 17*, 4289.
Woodbury, R. G., & Neurath, H. (1981) in *Metabolic Interconversion of Enzymes* (Holzer, H., Ed.) p 145, Springer-Verlag, Berlin.
Woodbury, R. G., Gruzenski, G. M., Lagunoff, D., & Neurath, H. (1978a) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2785.
Woodbury, R. G., Everitt, M. T., Sanada, Y., Katunuma, N., Lagunoff, D., & Neurath, H. (1978b) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5311.
Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K., & Neurath, H. (1978c) *Biochemistry 17*, 811.
Woodbury, R. G., Everitt, M. T., & Neurath, H. (1981) *Methods Enzymol. 80*, 588.
Woodbury, R. G., Le Trong, H., & Neurath, H. (1987) *Acta Histochem. Cytochem. 20*, 261.
Yoshida, N., Everitt, M. T., Neurath, H., Woodbury, R. G., & Powers J. C. (1980) *Biochemistry 19*, 799.