THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 268, No. 12, Issue of April 25, pp. 9023-9034, 1993
Printed in U.S.A.

Three-dimensional Models of Four Mouse Mast Cell Chymases

IDENTIFICATION OF PROTEOGLYCAN BINDING REGIONS AND PROTEASE-SPECIFIC ANTIGENIC EPITOPES*

(Received for publication, October 22, 1992)

Andrej Šali‡§, Ryoji Matsumoto¶, H. Patrick McNeil¶‖, Martin Karplus‡, and Richard L. Stevens¶****

From the ‡Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138 and the ¶Department of Medicine, Harvard Medical School and Department of Rheumatology and Immunology, Brigham and Women's Hospital, Boston, Massachusetts 02115

Mouse mast cell protease (mMCP) 1, mMCP-2, mMCP-4, and mMCP-5 are serine proteases which are predicted to have chymotryptic specificity (chymases). They are bound to negatively charged heparin or chondroitin sulfate proteoglycans and are stored in secretory granules. Three-dimensional (3D) models of these four proteases were constructed with a comparative molecular modeling technique based on satisfaction of spatial constraints. The models were used to predict immunogenic epitopes and surface regions that are likely to interact with proteoglycans. Nine potential antigenic segments in the four chymases were identified on the basis of solvent accessibility, protrusion, flexibility, and sequence variability. These segments are suitable epitopes for preparation of protease-specific antipeptide immunoglobulin. Two regions with net charges ranging from +6 to +10 at neutral pH were found on the surfaces of mMCP-4 and mMCP-5. The two regions are located far from the substrate binding cleft at diametrically opposite ends of the folded proteases. A strong positive electrostatic potential surrounds the two regions. Thus, they are good candidates for binding sites that interact with heparin proteoglycan in the granules of serosal mast cells. In contrast, mMCP-1 and mMCP-2, which are present in granules of mucosal mast cells that contain chondroitin sulfate, lack one of these regions and have a lower charge density in the other. The differences between the 3D models provide a structural basis for the selective localization of specific chymases within mouse mast cells that contain different proteoglycans.

Mast cells contain many serine proteases in their secretory granules. These effector cells of the immune response are the major source of neutral proteases in connective tissues. The physiologic functions of mast cell granule proteases have not been determined, although they have been implicated in the metabolism of cytokines and hormones (1–6), extracellular matrix proteins (7–11), metalloproteases (12), and plasma proteins (13–15).

At least seven different 26–32-kDa mast cell serine proteases, designated mouse mast cell protease (mMCP)[1] 1 to mMCP-7, have been identified in the granules of mouse mast cells (16–24). Based on the homology arguments and the amino acid sequences deduced from their cDNAs and genes, mMCP-6 and mMCP-7 (tryptases) are predicted to have substrate specificity for a positively charged residue at the amino-terminal side of the scissile bond, whereas mMCP-1 to mMCP-5 (chymases) are predicted to have specificity for a large hydrophobic residue at the corresponding substrate position. mMCP-1 (16) and mMCP-2 (20) are preferentially expressed in mucosal mast cells, a subclass of mast cells that increases in the intestines of helminth-infected BALB/c mice. In contrast, mMCP-4 (21), mMCP-5 (22), and mMCP-6 (18) are preferentially expressed in the mouse mast cells that reside in the serosal cavity of BALB/c mice. mMCP-7 is not synthesized in serosal or mucosal mast cells, but it is transiently expressed in *in vitro* differentiated mast cells derived from bone marrow (19). Transformed (17) and nontransformed (25, 26) mouse mast cells have been obtained *in vitro* that express mixed protease phenotypes, raising the possibility that the protease phenotypes of mouse mast cells might be more heterogeneous *in vivo* than recognized previously.

Despite being derived from distinct genes, the mast cell chymases are similar to each other and possess identical regions of up to 15 amino acid residues in length. Thus, to identify immunohistochemically which chymases are expressed in a particular tissue-localized mast cell, the antiprotease immunoglobulins need to be specific to regions on the protein surface that differ among the proteases. By means of an antipeptide approach, rabbit anti-mMCP-$5_{146-162}$ immunoglobulin was generated against a synthetic peptide that corresponds to a variable region within this protease (27).

Many of the 26–32-kDa mouse, rat, and human mast cell serine proteases are exocytosed from the effector cell in fully active form as $>10^7$ Da macromolecular complexes bound to serglycin proteoglycans (3, 28–30). Serglycin proteoglycans, which are highly acidic, consist of a serine/glycine-rich pep-

* This work was supported in part by Grants AI-23483, AI-31599, HL-36110, GM-30804, and RR-05950 from the National Institutes of Health, by a grant from the Hyde and Watson Foundation and by a grant from The Jane Coffin Childs Memorial Fund for Medical Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Fellow of The Jane Coffin Childs Memorial Fund for Medical Research.

‖ Recipient of a Heald Fellowship from the Arthritis Foundation of Australia.

** To whom correspondence should be addressed: Dept. of Chemistry, Harvard University, 12 Oxford St, Cambridge, MA 02138. Tel.: 617-495-4018; Fax: 617-496-3204 (or R. Stevens at Harvard Medical School, Seeley G. Mudd Bldg., Rm. 617, 250 Longwood Ave., Boston, MA 02115. Tel.: 617-432-1512; Fax: 617-432-0979.)

[1] The abbreviations used are: mMCP, mouse mast cell protease; chymases, mMCPs with chymotryptic specificity; rMCP, rat mast cell protease; 3D, three-dimensional; tryptases, mMCPs with tryptic specificity.

tide core (31, 32) whose serine residues are covalently linked to different types of glycosaminoglycan chains (33–37). When the positively charged serine proteases and carboxypeptidases of mast cells are bound to proteoglycans, autolysis is minimized. In at least one instance, interaction with proteoglycan has been found to influence the substrate specificity of the protease (3). It is possible that binding of a protease by a proteoglycan increases the retention period of a protease in the inflammation site and shields it against inactivation by circulating protease inhibitors.

Little is known about how mast cell granule proteases interact with serglycin proteoglycans. Because mast cell chymases are positively charged at neutral pH and because they do not have the Trp-Ser-*X*-Trp heparin-binding motif (38), it has been presumed that they are electrostatically bound to the negatively charged glycosaminoglycans of serglycin proteoglycans. Commercially prepared porcine heparin glycosaminoglycan binds to numerous proteins that have a consensus amino acid sequence of either *X*-B-B-*X*-B-*X* or *X*-B-B-B-*X*-*X*-B-*X*, where *X* and B are noncharged and basic amino acids, respectively (39). These two patterns are a result of the periodicity of helices and strands, and of the requirement for basic residues to be exposed so that they can interact with a negative ligand (39). However, many other heparin-binding proteins exist that do not have either of these patterns (39). In fact, neither pattern is present in mMCP-4 (21), mMCP-5 (22), or mast cell carboxypeptidase A (40), even though all three proteases are stored in the secretory granules of mouse serosal mast cells (17) in complex with heparin proteoglycan (34). Although the predominant proteoglycan in mouse mucosal mast cells has not been identified, rat mucosal mast cells contain chondroitin sulfate di-B/E proteoglycan rather than heparin proteoglycan (41). Thus, it is likely that mMCP-1 and mMCP-2 are preferentially complexed to highly sulfated chondroitin proteoglycans.

To understand how mMCPs interact with proteoglycans and antibodies, it is important to know their tertiary structure. Although none of the mouse chymases has been purified and therefore no experimentally determined structures of mouse chymases are available, comparative molecular modeling can be used to predict their structures. It has been shown that the three-dimensional (3D) structure of a protein can be calculated if its amino acid sequence is sufficiently similar to that of a protein with known tertiary structure (42–46). This modeling technique is particularly useful when only low to medium resolution results are required, such as prediction of exposed regions that may interact with antibodies (47) and models of interaction based on electrostatic complementarity (48). Comparative modeling based on the crystallographic structures of homologous proteases has been applied to obtain 3D models of rat mast cell protease (rMCP) I (49), rat mast cell carboxypeptidase A (50), and mast cell tryptases (24). These 3D models were used to identify putative heparin binding sites by focusing on the regions that contain many positively charged amino acid residues.

In this study, we have used a method of comparative modeling that is based on the satisfaction of spatial constraints (51) to predict the 3D structures of four mouse mast cell chymases. The modeling was based on the 3D structure of rMCP-II determined by x-ray crystallography (49) and on its sequence identity of 55–75% to the mouse chymases. Although the overall structure of the mouse chymases is shown to be similar to that of rMCP-II, a detailed examination of the models provided additional information about the interaction between the mMCPs and proteoglycans. Two regions containing a large number of Lys and Arg residues were identified on the faces of mMCP-4 and mMCP-5 away from the substrate-binding cleft of each protease. These two regions are likely to interact with the negatively charged heparin of the granular matrix, leaving the active site of the enzyme exposed for substrate hydrolysis. Variable regions were also identified on the surface of mMCP-1, mMCP-2, mMCP-4, and mMCP-5 that would be suitable epitopes for preparation of protease-specific antipeptide immunoglobulins.

MATERIALS AND METHODS

Comparative Modeling of Mouse Mast Cell Chymases—The high-resolution 3D structures of nine serine proteases (Table I) have been determined by x-ray crystallography and deposited in the Brookhaven Protein Data Bank (52); they form the data base that was used to predict the 3D structures of mMCP-1, mMCP-2, mMCP-4, and mMCP-5. Because of the size of the family, the number of insertions and deletions, and the poor degree of similarity between the nine known proteases, their structures were superimposed with the use of COMPARER (53, 54), a computer program that relies on a number of features of protein 3D structure to obtain the best match between their amino acid sequences. These features included residue type identity, residue type properties, side chain accessibility, main chain accessibility, hydrogen bonds, position of C_α atoms, Φ dihedral angle, Ψ dihedral angle, and local main chain direction (53). The multiple structural alignment of the nine known proteases and the sequences of mMCP-1 to mMCP-7 were then compared in the second step to obtain the final alignment of the whole serine protease family. The second sequence comparison between mMCP-1, mMCP-2, mMCP-4, mMCP-5 and the structurally known rMCP-II was done by hand because of their 55–75% sequence identities. This sequence similarity allowed an unambiguous alignment between the mouse and rat chymases and thus also an unambiguous alignment between the mouse chymases and the other eight known structures.

A 3D model is usually most accurate if only the proteins most similar to the sequence being modeled are used (55). To find which of the nine serine proteases with known 3D structure are most similar to each mMCP, the table of percentage sequence identities for all pairs of the proteins was calculated from the multiple alignment. This matrix was used with the KITSCH computer program (56) to calculate a tree that expresses the relationships among the sequences

TABLE I

Sources of structural and sequence data used in the comparative modeling of the mouse mast cell chymases

The structures were obtained from the fall 1991 release of the Brookhaven Protein Data Bank (52). When more than one molecule is present in a file, the first molecule is used. The deduced amino acid sequences of the mMCPs were obtained from the GenBank data base (86), except for mMCP-1 and mMCP-7.

Protease	Brookhaven code	Resolution	Ref.
		Å	
A. Serine proteases with 3D structures determined by x-ray crystallography			
Rat tonin	1TON	1.8	87
Porcine kallikrein	2PKA	2.0	88
Bovine trypsin	2PTN	1.5	89
Bovine chymotrypsin	4CHA	1.7	90
Porcine elastase	3EST	1.6	91
Rat trypsin	1TRM	2.3	92
Human neutrophil elastase	1HNE	1.8	93
rMCP-II	3RP2	1.9	94
S. griseus trypsin	1SGT	1.7	95

Name	GenBank code	Ref.
B. Mouse mast cell proteases		
mMCP-1		23
mMCP-2	J05177	20
mMCP-4	M55616, M55617, M57401	21
mMCP-5	M73759, M73760	22
mMCP-6	M57625	18
mMCP-7	L00653, L00654	19

of serine proteases, similar to the trees used to deduce the evolution of protein families. In this tree, the differences between two groups of sequences are approximated by a vertical distance from the top of the tree to the highest node from which the two groups of sequences branch off.

Starting with the alignment of mMCPs with the selected proteins of known structure, a method of comparative modeling by satisfaction of spatial constraints (51) was used to predict the 3D structure of each mMCP without further manual intervention or subjective decisions. This method is implemented in the MODELLER program. The spatial constraints needed to calculate the structure of the unknown protein are determined and then these constraints are maximally satisfied to obtain the 3D model. The spatial constraints are derived by transferring the spatial features from the structures of known proteins to the sequence of the unknown protein. For example, if there is a conserved hydrogen bond at an equivalent position in the 3D structures of all known serine proteases, it is assumed that each mMCP has this hydrogen bond in its structure. Such a hypothesis represents a distance constraint on the 3D structure of the unknown protein. The more strong constraints are available, the more precise is the overall 3D model of the unknown protein.

As an example of the constraints, we consider the C_α–C_α distances. The C_α–C_α distance constraints between all pairs of C_α atoms are derived in the following way. A survey of many pairs of aligned protein structures shows that a conditional probability density function for a C_α–C_α distance, given an equivalent C_α–C_α distance in a related protein determined from the sequence alignment, is a Gaussian function,

$$p(d'/d) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{d'-d}{\sigma}\right)^2\right] \qquad \text{(Eq. 1)}$$

where d' and d are the two distances and σ is standard deviation of the distribution. When d is known, this relationship can be used to constrain d'. The most likely value of d' is equal to d. Since the standard deviation is only about 1 Å, this constraint is strong. The precise value of the standard deviation is calculated from a polynomial that depends on the similarity between the two proteins, on the solvent exposure of the 2 residues spanning the distance, and on the proximity of the nearest gaps in the alignment of the two proteins.

The side chain dihedral angles were modeled using the rotamer library (57) that also takes into account side chain conformations and types of equivalent residues in related structures (51). This homologue-dependent rotamer library consists of entries for each existing combination of residue and dihedral angle types. The entries were derived from the alignments of related structures by tabulating relative frequencies of up to three possible side chain dihedral angle classes for each existing combination of equivalent residue and dihedral angle types in a related protein. For example, the probabilities of conformations −, *t*, and + for a Ser χ_1 angle are 0.6, 0.2, and 0.2, respectively, if there is a Cys residue in conformation − at an equivalent position in a related structure. Once these weights are determined, the probability density function for the side chain dihedral angle can be modeled by a weighted sum of three Gaussian distributions with means and standard deviations corresponding to the observed distribution of this dihedral angle in the known protein structures.

Similar analyses of alignments of related structures gave the probability density functions for several other spatial features, including main chain dihedral angles Φ and Ψ, and distances between all pairs of main chain N and O atoms (51). The numbers of constraints for each of the 16 features used to model mMCP-5 are listed in Table II. The initial structure for the sequence to be modeled is a polypeptide chain, consisting of all non-hydrogen atoms, with random main chain and side chain dihedral angles. The final model is obtained by iteratively changing the structure to satisfy optimally all the individual constraints combined in the molecular probability density function. The optimization procedure employed in MODELLER is the variable target function method (58). This stage of comparative modeling is technically similar to the refinement of protein 3D structures from the distance and dihedral angle constraints obtained from multidimensional nuclear magnetic resonance spectroscopy. Five slightly different models were calculated for each mMCP by using different initial conformations; the root mean square difference for superposition of these models was generally less than 0.2 Å. The structure with the highest value of the molecular probability density function was selected as the representative model; the deviations from the most likely values of the constraints for the mMCP-5 model are given in Table II.

TABLE II

Constraints used to model mMCP-5

Type	Number[a]	Root mean square[b]
Bond lengths	1,822	0.005 Å
Bond angles	2,465	1.96°
Dihedral angles[c]	994	3.14°
Van der Waals contacts[d]	559	0.02 Å
C_α–C_α distances	11,294	0.05 Å
Main chain N–O distances	3,280	0.14 Å
Main chain Φ dihedral angles	225	21.5°
Main chain Ψ dihedral angles	225	21.3°
Side chain χ_1 dihedral angles	181	9.1°
Side chain χ_2 dihedral angles	135	9.6°
Side chain χ_3 dihedral angles	54	14.5°
Side chain χ_4 dihedral angles	29	12.2°
Disulfide bridge bonds	3	0.003 Å
Disulfide bridge angles	6	2.28°
Disulfide bridge dihedral angles	3	15.7°
cis-Peptides[e]	1	4.0°

[a] Number of constraints of a given type that were used to model mMCP-5.

[b] Deviation between the actual values and the most likely values.

[c] These dihedral angles constrain the planarity of peptide bonds and rings as well as chirality of the chiral carbon atoms.

[d] All pairs of atoms that are not constrained by any of the bond, or bond angle terms, are constrained by the minimal contact distance. The number of pairs that violate this constraint in the final model is listed.

[e] The $205–206_{(224-225)}$ peptide bond in rMCP-II is in the *cis* conformation. Since the mMCP-5 sequence also has a Pro residue at an equivalent position, it was constrained to a *cis* conformation.

Electrostatic Potential of Mouse Mast Cell Chymases—Electrostatic terms in the potential energy often give rise to specific interactions in complexes (*e.g.* that between a Lys and a sulfate at contact distance). However, for trying to understand or to predict the nature of a complex between two macromolecules it is often useful to look at the global electrostatic potential of the two ligands involved. If the structure of only one ligand is known it is particularly helpful to examine its electrostatic potential for possible binding sites of the other ligand. This is true in the present case where the interaction between a positive (the protein) and a negative (the glycosaminoglycan) polyion is considered and the detailed structure of the latter is not available.

The electrostatic potential of each mMCP was calculated with UHBD 2.2 (59), a computer program that uses the finite difference method to solve the linearized Poisson-Boltzmann equation. Solvent and protein are treated as two dielectric continuums with different dielectric constants. Calculation of the electrostatic potential of a protein can be sensitive to details of the protein model, nature of the solvent, and the parameters in the electrostatic model. Therefore, the potential surfaces were calculated under a number of different conditions to determine which features of the potential were conserved and thus most likely to be correct. The ionic strength of the aqueous solvent was varied between 1 and 100 mM and temperature was set to 300 K. The relative dielectric constant used for the protein was in the range from 2 to 10. The grid size for the calculation varied between 1 and 2 Å and the box size between 100 and 150 Å, approximately three times the diameter of the molecule. Three types of atomic models were used: a MODELLER model consisting of all heavy atoms, but no hydrogens; a MODELLER model with hydrogen atoms added by CHARMM (60); and an average MODELLER model obtained by averaging the positions of side chain heavy atoms. The basis for using the average structure to calculate the average electrostatic potential, as opposed to averaging the potentials from individual molecules, is provided by the empirical observation that the electrostatic potential of an average structure from a dynamics trajectory is close to the average of potentials from individual coordinate sets of that trajectory (61). The side chain conformations were calculated by generating all possible side chain rotamers and then averaging them with the weights from the side chain probability density functions used to derive the model (51). Because the exocytosed serosal mast cell chymases remain bound to heparin proteogly-

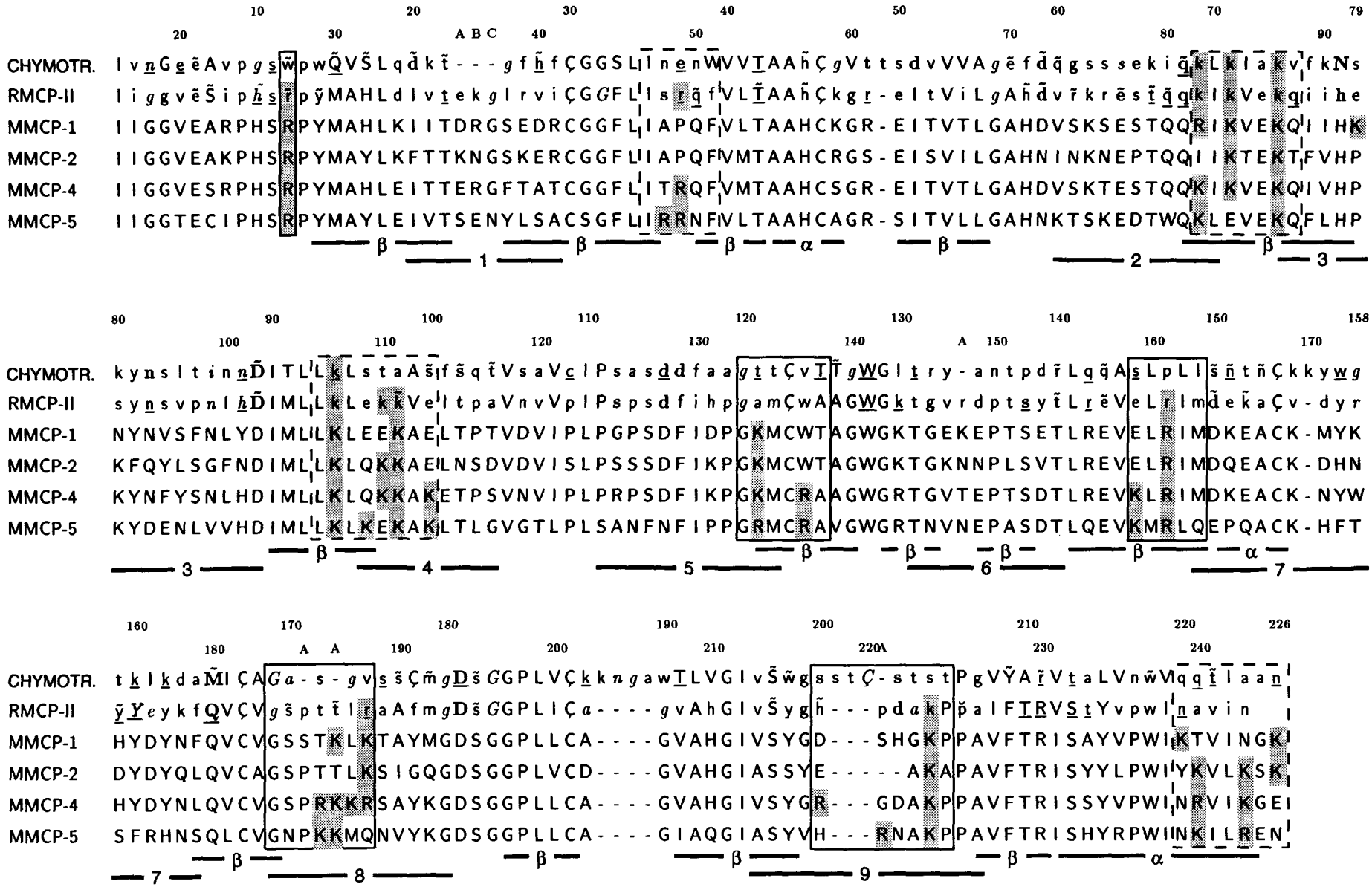

FIG. 1. **Alignment of the amino acid sequences of bovine chymotrypsin (*CHYMOTR.*), rMCP-II, mMCP-1, mMCP-2, mMCP-4, and mMCP-5.** The formatting convention of the program JOY (96) applies to the sequences with known 3D structures. *Uppercase letters*, solvent-inaccessible amino acid residues; *lowercase letters*, solvent-accessible amino acid residues; *bold type*, hydrogen bonds to main chain amide; *underline*, hydrogen bonds to main chain carbonyl; *~dash*, side chain-side chain hydrogen bonds; *italic*, positive main chain dihedral angle Φ. The numbers in the top line refer to the actual position of amino acids in rMCP-II, mMCP-1, mMCP-4, and mMCP-5. The numbers in the second line correspond to the residues in chymotrypsin. Gaps in the chymotrypsin sequence are indicated by the letters *A*, *B*, and *C*. The second *line* from the *bottom* contains the secondary structure assignment for rMCP-II, and the *bottom line* indicates the epitope segments predicted in Fig. 4. The *boxes* enclose the residues in the positively charged region 1 (*continuous lines*) and the positively charged region 2 (*dashed lines*). The Lys and Arg residues in these two regions are *highlighted*.

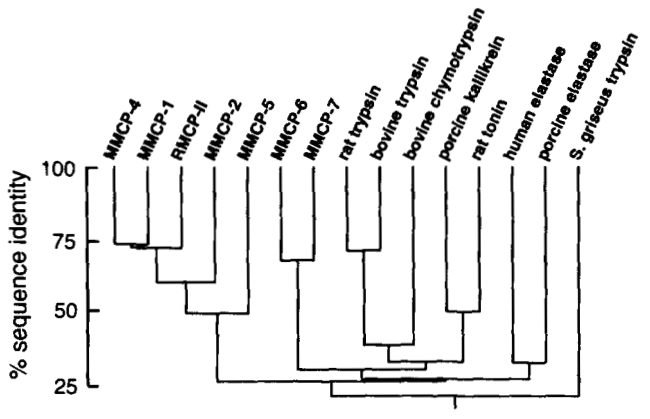

FIG. 2. **Clustering of serine proteases.** The scale on the *y* axis is the percentage sequence identity. The tree shows that mMCP-1 and mMCP-4 are most similar to rMCP-II. mMCP-2 and mMCP-5 also belong to the same group of proteins, but they are less similar to rMCP-II than mMCP-1 and mMCP-4. The mouse tryptases, mMCP-6 and mMCP-7, form a small subgroup which clusters in a separate group with all the remaining proteases, including trypsin and chymotrypsin. An exception is yeast trypsin from *Streptomyces griseus*, which is an outlier in the family that otherwise contains only mammalian serine proteases.

can in an extracellular environment where the pH is approximately 7, the net charge of −1 was assigned to each Asp and Glu residue and a net charge of +1 to each Lys and Arg residue. The His residues, which usually have pK_a of ≈6.5, were considered neutral as were all other amino acid residues. The partial atomic charges were taken from the CHARMM 22 force field.[2] When hydrogen atoms were omitted from the protein model, the charges of the remaining heavy atoms were corrected by adding the charges of hydrogen atoms covalently bound to them. In the analysis that follows, we consider only qualitative features of the electrostatic potential. These were preserved in the different models that were compared.

Location of Potential Protease-specific Antigenic Sites in Mouse Mast Cell Chymases—To find continuous epitopes in the models that would elicit antipeptide immunoglobulins specific for an individual mMCP, four features of protein structure and sequence were optimized (62–64) in addition to the factors that influence cellular uptake, processing, and presentation of the antigen. First, such a stretch of residues should be on the surface of the folded protein (65). The PSA computer program[3] that applies the method of Richmond and Richards (66) was used to calculate solvent-accessible areas for each residue. These areas were normalized to a range between 0 and 100% as described (53) and then used to highlight the exposed segments of a molecule. Second, a good epitope should not only be on the surface but also protrude out of it (62). A protrusion index, calculated by the program ELLIPSOID, was used to quantitate how far out of the surface a segment of residues protrudes. Third, antipeptide immunoglobulins bind more strongly to mobile parts of the chain than to more rigid parts (67, 68). Thus, the average main chain (including N, C_α, C, and O atoms) isotropic temperature factors of rMCP-II were examined to identify the more mobile segments in each mMCP. Because rMCP-II molecules A and B in the crystallographic unit cell (49) have virtually the same mobility, only molecule A was used. The three short segments of the main chain whose conformation could not be determined (residues 83–85$_{(96-98)}$, 155–157$_{(169-170B)}$, and 224$_{(243)}$),[4] presumably due to high mobility, were assigned isotropic temperature factors of 40 Å^2. Additionally, no correction for intermolecular contacts in the crystal had to be made, since the two loops around residues 28$_{(40)}$ and 136$_{(150)}$ that form intermolecular contacts have high temperature factors. Fourth, to obtain an antibody capable of distinguishing between individual mMCPs, the immunogen must have an amino acid sequence that is specific to the protein of interest. Moreover, a high mutation rate resulting in specific sequences is also correlated with antigenicity (63). The average difference for all pairs of residues at each position in the alignment of mMCPs was plotted to highlight these variable segments. The difference between two residue types (53) is defined as being proportional to the difference in residue size, hydrophobicity, refractivity index, and secondary structure propensities. These properties were used because they are statistically the most conserved combination of residue type features in evolution (69) and thus the most reliable indicator of the differences among the protein structures. Before plotting accessibility, protrusion index, mobility, and variability for each residue position in a serine protease, the values were smoothed by the running average method; a value at position i is an average of values from positions $i - 2$ to $i + 2$. The four features are highly correlated and have about the same overall prediction success, but they do not always give the same prediction (62, 63). Thus, it is useful to employ them in combination and to check if there are significant deviations in the results for any potential epitope.

RESULTS

Comparative Modeling of Mouse Mast Cell Chymases—The amino acid sequence alignment of mMCP-1, mMCP-2, mMCP-4, mMCP-5, and rMCP-II with pancreatic chymotrypsin is shown in Fig. 1. The grouping of the amino acid sequences of all mMCPs and nine serine proteases of known 3D structure is shown in Fig. 2. mMCP-1, mMCP-2, mMCP-4, and mMCP-5 are more similar to rMCP-II than to any other serine protease with known 3D structure. No insertions or deletions are needed to align the amino acid sequences of these mast cell chymases with rMCP-II, except for a single 2-residue deletion in mMCP-2 at positions 201–202$_{(220A-221)}$ (Fig. 1). Consequently, the structure of rMCP-II determined by x-ray crystallography (49) was chosen as the template protein for comparative modeling of the four mouse mast cell chymases. The amino acid sequences of mMCP-6 and mMCP-7 were significantly different from other mMCPs and from rMCP-II (Fig. 2). Although the models of the 3D structure of mMCP-6 and mMCP-7 were not constructed in this investigation, these two tryptases were recently modeled by Johnson and Barton (24). mMCP-3 was not modeled because only the NH_2-terminal amino acid sequence of this serosal mast cell protease has been determined.

When the 3D models of mMCP-1, mMCP-2, mMCP-4, and mMCP-5 were calculated, their backbones were found to be virtually indistinguishable from the backbone of rMCP-II (Fig. 3). The root mean square differences for superposition of the C_α atoms are generally less than 0.2 Å which is similar to the root mean square deviation between the independently determined rMCP-II molecules A and B (49). The only large main chain differences among the mMCPs occur for the last 2 residues at the COOH terminus (Fig. 3). These differences are probably a consequence of the lack of conformational constraints used in modeling which resulted from the absence of equivalent residues in rMCP-II (Fig. 1). The only significant differences among the four models and rMCP-II are the types and orientations of the side chains. The fact that violations of the input constraints are also small (see Table II for the constraints used to model mMCP-5) indicates that the amino acid sequences of the four mouse mast cell chymases are consistent with the 3D structure of rMCP-II.

A strong correlation between the sequence similarity of two proteins and the root mean square deviation of their backbone structures has been described (70, 71). Because the error in

[2] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, and M. Karplus, manuscript in preparation.

[3] The computer programs PSA and ELLIPSOID were written by A. Šali and are available upon request. The MODELLER computer program was written by A. Šali and T. Blundell, manuscript in preparation. For more information about MODELLER, contact A. Šali.

[4] Two numbering systems are used to indicate the position of an amino acid residue in a sequence. The first number corresponds to the position of the residue in the mature protease. The subsequent subscript number in parentheses refers to the equivalent residue in chymotrypsin (Fig. 1).

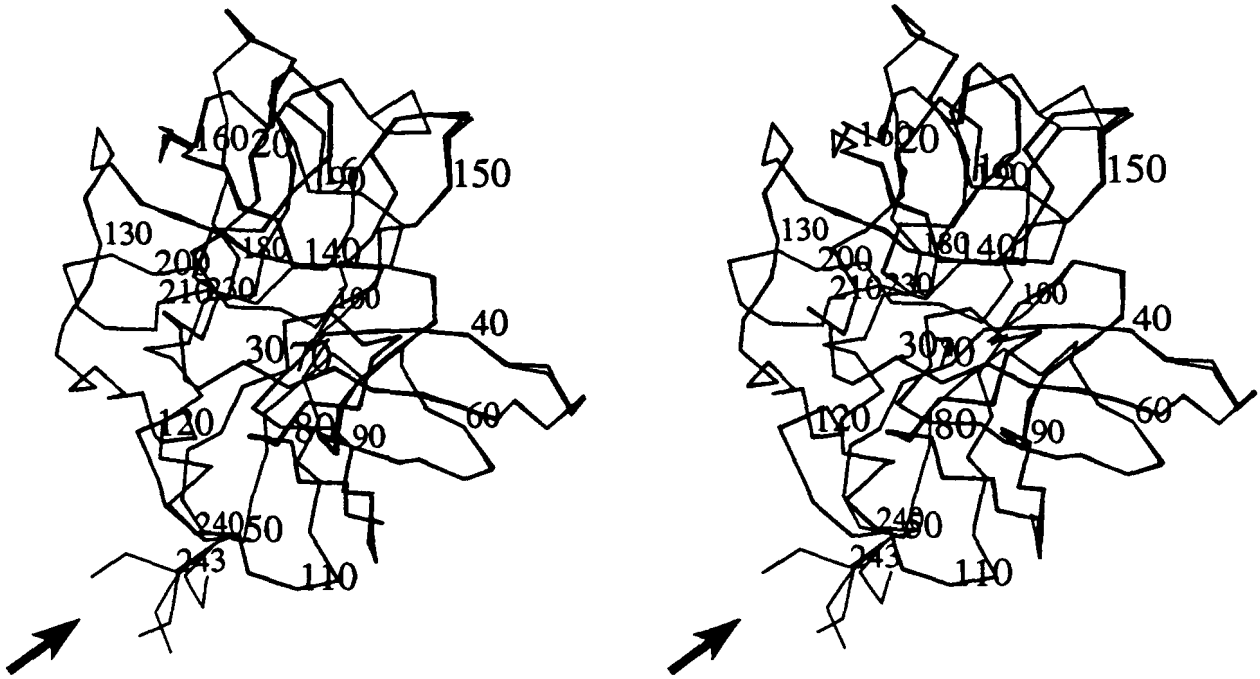

FIG. 3. **A stereo C_α plot comparing the 3D models of mMCP-1, mMCP-2, mMCP-4, and mMCP-5 with the 3D structure of rMCP-II.** The C_α atoms of the four mMCP models are superposed on the crystallographic structure of rMCP-II. The *arrow* indicates the largest difference in the main chain conformation of rMCP-II and the mMCP models. Chymotrypsin numbering is used. All black and white plots of protein structures were created by the program MOLSCRIPT (97).

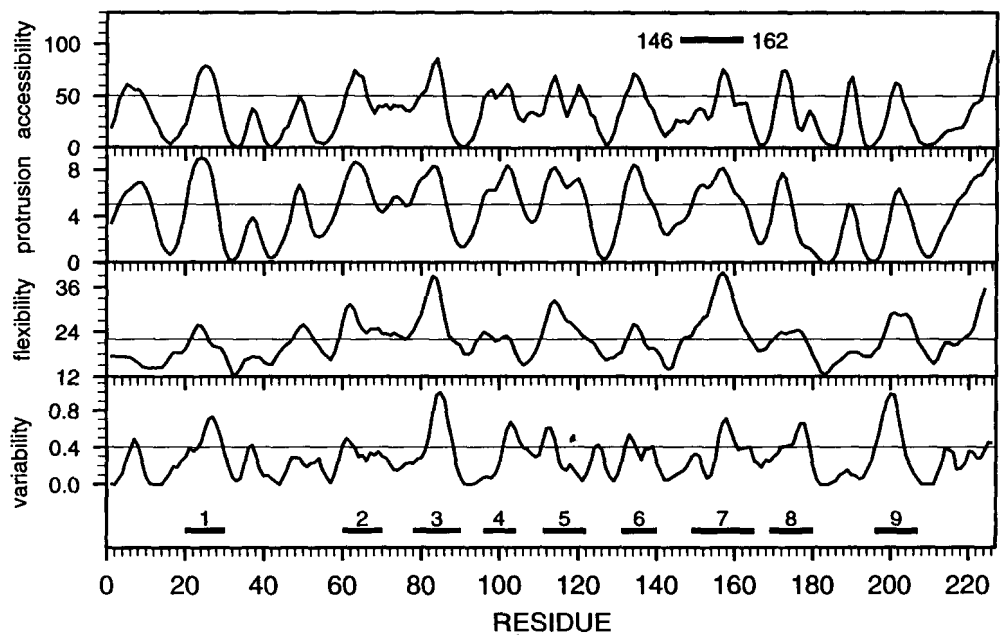

FIG. 4. **Identification of sequence segments that can be used to obtain protease-specific antipeptide immunoglobulins.** Depicted are accessibility (the smoothed fractional side chain accessibility of the mMCP-5 model, measured in percentage points from 0 to 100), protrusion (the smoothed protrusion index for the mMCP-5 model, measured in relative units from 1 to 10), flexibility (the smoothed main chain isotropic temperature factor of rMCP-II, measured in $Å^2$), and variability (the smoothed variability of mMCP-1, mMCP-2, mMCP-4, and mMCP-5, in relative units from 0 to 1). The position of peptide 146–162 in mMCP-5 (27) is indicated at the *top* of the figure. The *thin horizontal lines* (at arbitrary height) help to locate minima and maxima in the four curves. Antigenic segments are identified as the regions where at least three of the four features have a pronounced optimum; the boundaries of antigenic segments are approximate, but they are all between 8 and 15 residues long, which is an optimal size for an antigenic peptide (68). The nine predicted epitope segments are indicated at the *bottom* of the figure. Epitope 1 is a protruding tip of a β-hairpin; epitopes 2, 4, and 5 are extended segments; epitopes 3 and 9 are protruding reverse turns; epitope 6 is a β-hairpin; epitope 7 is a protruding tip of the α-turn-β motif; and epitope 8 is a reverse turn. The residue numbers correspond to those in rMCP-II, mMCP-1, mMCP-4, and mMCP-5, rather than to those in chymotrypsin.

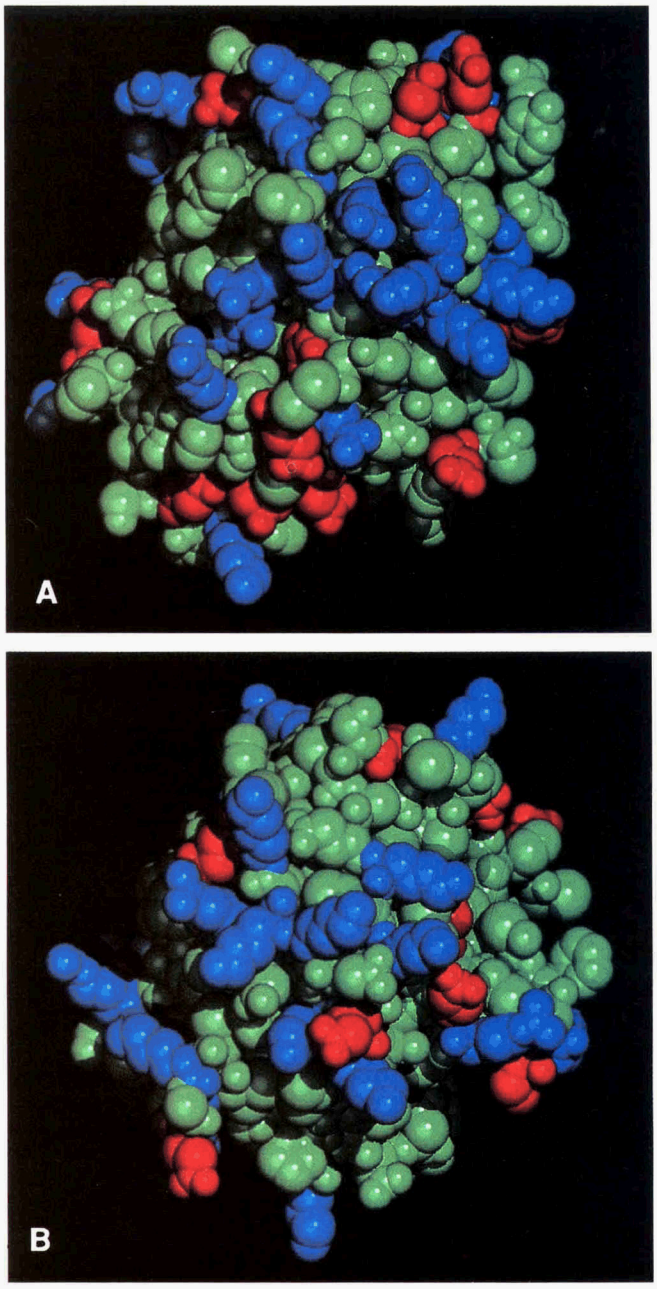

FIG. 5. **Top view of the two positive regions in mMCP-4.** All heavy atoms are shown as *spheres*. The Lys and Arg residues are colored in *blue*, the Asp and Glu residues are *red*, and the rest of the residues are *green*. The two orientations of the molecule are obtained from that in Fig. 3 by a rotation of ±90° around a horizontal axis. All color plots of protein structures were created by QUANTA (Molecular Simulations Inc., Waltham, MA). *A*, region 1; *B*, region 2. Both regions are seen as *blue strips* running approximately horizontally across the center of the plot.

the model is usually similar to the difference between the template protein and the actual structure of the unknown protein, this relationship indicates that the error in the models of the four mouse mast cell chymases is approximately 0.6 Å for the buried and 1.3 Å for the exposed backbone atoms. Approximately 80, 80, 70, and 75% of the side chains are expected to have dihedral angles χ_1, χ_2, χ_3, and χ_4, respectively, in the correct optima.[5] Thus, there are some uncertainties in the positions of positive charges at the end of long Lys and Arg side chains on the protein surface. However, these have a small effect on the global features of the electrostatic potential considered below.

Location of Potential Protease-specific Antigenic Sites in Mouse Mast Cell Chymases—To maximize the likelihood that a linear peptide will elicit an antibody that recognizes one and only one protease, the stretch of residues in the native protein should have maximal solvent exposure, protrusion out of surface, conformational mobility, and residue type variability among the mouse mast cell chymases. A plot of these features for mMCP-5 is shown in Fig. 4. If accessibility and protrusion of mMCP-5 are replaced by those of rMCP-II, mMCP-1, mMCP-2, or mMCP-4, essentially the same results are obtained (data not shown). The nine segments that are predicted to be the best for obtaining protease-specific antipeptide immunoglobulins are shown in Figs. 1 and 4. The two most favored segments, according to the four criteria used, correspond to residues $74–89_{(87–102)}$ (segment 3) and $196–207_{(212–224)}$ (segment 9). Both segments are protruding reverse turns.

Location of Proteoglycan Binding Regions in mMCP-4 and mMCP-5—Based on their 3D models, mMCP-4 and mMCP-5 have two distinct regions on their surfaces that contain considerably more positive than negative charges (Figs. 1, 5, 6, and Table III). These regions are located at opposite ends of the molecule. The regions are in separate domains and are equidistant from the active site at the interface between the two domains of the protein. The two positively charged regions are convex strips approximately 20 Å long and 10 Å wide. The chain segments comprising these regions are not contiguous. Region 1 consists of $Arg\text{-}12_{(27)}$, two antiparallel strands ($120–125_{(133–138)}$, $145–149_{(159–163)}$), and tips of two loops ($169–175_{(184–188)}$, $200–205_{(217–224)}$). Region 2 consists of a turn ($35–39_{(47–51)}$), two anti-parallel strands ($69–75_{(82–88)}$, $93–100_{(106–113)}$), and two turns of the COOH-terminal α-helix ($220–226_{(239–245)}$). In mMCP-4, each of these segments has 2 positively charged residues, except for segment $93–100_{(106–113)}$ which has 4 of them. Both regions exploit the amphiphilic periodicity of surface β-strands and extended segments to have every other residue in a sequence charged and on the surface. For example, there is a run of four positively charged residues at $94_{(107)}$, $96_{(109)}$, $98_{(111)}$, and $100_{(113)}$ in region 2 of mMCP-5. Assembling the positive regions from secondary structure elements, which are rich in main chain-main chain hydrogen bonds, contributes to the stability of these regions with a large number of positive charges close to each other. The destabilizing effect of repulsion between like charges is also reduced by the strong screening effect of water, which diminishes charge-charge interactions on protein surfaces (72, 73).

Region 1 has a net charge of +10 in mMCP-4 and +9 in mMCP-5 (Table III); region 2 has a smaller but still large net charge of +8 in mMCP-4 and +6 in mMCP-5 (Table III). Region 2 has a smaller net charge than region 1, because there are more negatively charged residues in it, not because of a smaller number of positively charged residues. Other serine proteases listed in Table I have a smaller number of positive charges in both region 1 and region 2 (Table III). Whether or not these regions bind heparin is likely to be determined by the overall electrostatic potential which is a sum of the contributions from positive and negative charges.

Electrostatic Potential of Mouse Mast Cell Chymases—A stereoplot of electrostatic potential contours around mMCP-4 is shown in Fig. 7*A*. The contour maps of the electrostatic potential for mMCP-1, mMCP-2, mMCP-4, mMCP-5, rMCP-II, and chymotrypsin are compared in Fig. 8. For each mole-

[5] A. Šali and T. L. Blundell, unpublished findings.

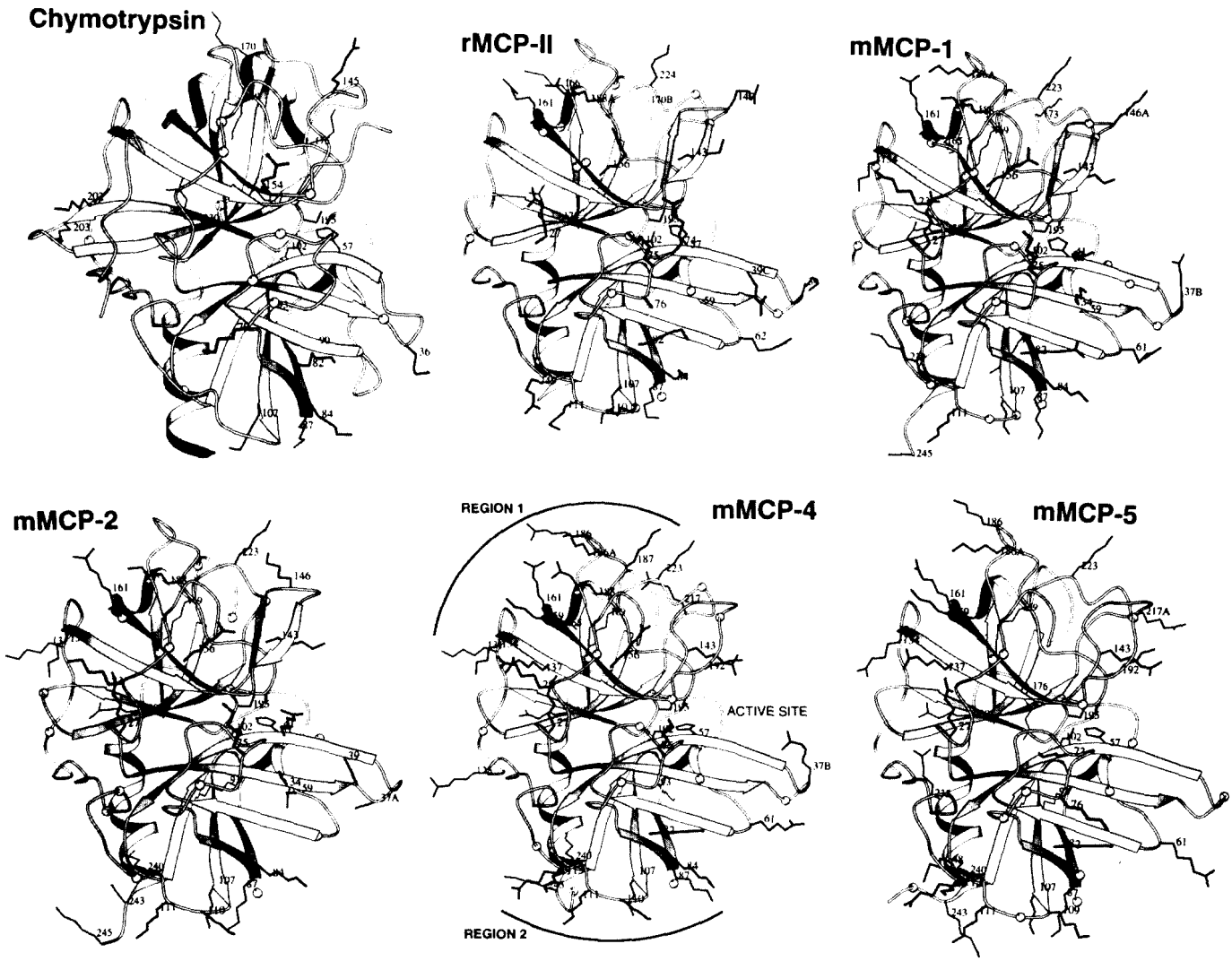

FIG. 6. **3D structures of chymotrypsin, rMCP-II, and models of the mouse mast cell chymases.** Charge distribution is compared. The two areas where positive charges predominate, regions 1 and 2, and the putative active site are identified in mMCP-4. Secondary structure definitions were calculated by the program SSTRUC written by D. Smith. Arg and Lys side chains are drawn in full and labeled. Asp and Glu residues are represented by their enlarged C_α atoms colored in *gray*. The active site residues are also drawn. The two domains correspond to the upper and lower halves of the molecule. Residue numbers correspond to those in chymotrypsin. The orientation of the molecules is the same as in Fig. 3.

cule, the main features of the electrostatic potential do not depend on the exact atomic model and parameters used in the calculation. For mMCP-4 and mMCP-5, almost all of the envelope about 10 Å from the surface has a positive potential. However, mMCP-4 and mMCP-5 have two large bulges in their contour maps that are due to the large number of positively charged residues in regions 1 and 2. The potential remains as high as 0.3 kcal/electron mol at an ionic strength of 100 mM even 20 Å away from the protein surface. Both the location and shape of regions 1 and 2 are the same in mMCP-4 and mMCP-5. The pronounced positive electrostatic potential above regions 1 and 2 is eliminated if 3 residues in the center of region 1 and 2 residues in the center of region 2 are mutated to glutamic acid residues (Fig. 7*B*).

Region 1 in mMCP-1, mMCP-2, and rMCP-II does not have a pronounced positive electrostatic potential, in contrast to region 1 of mMCP-4 and mMCP-5 (Fig. 8). This is due to a smaller number of positive residues (Table III) and to an almost even spatial distribution of the positive and negative residues on the surface of these proteases (Fig. 6). In contrast, region 2 is present in all four mast cell chymases and in rMCP-II (Fig. 8). However, whereas the location of the positive field over region 2 in mMCP-1, mMCP-2, and rMCP-II is almost identical, this potential covers a smaller area than in mMCP-4 and mMCP-5. This difference is due to a smaller number of positive residues and to a different spatial distribution of positive and negative residues. For example, 2 positive residues ($96_{(109)}$, $100_{(113)}$) and 1 neutral residue ($97_{(110)}$) that are on the periphery of region 2 in mMCP-4 are negative residues in mMCP-1. As a result, the area covered by a strong positive potential above region 2 is smaller in mMCP-1.

Although regions 1 and 2 are on the opposite sides of the molecule, a strip of weak positive potential covers the surface away from the active site, and connects the two regions (Figs. 7*A* and 8). As with regions 1 and 2, this connecting strip is more pronounced in mMCP-4 and mMCP-5 (Fig. 8) but is also present in mMCP-1, mMCP-2, and rMCP-II. In mMCP-4, the connecting strip includes residues $128_{(131)}$, $112_{(125)}$, and $211_{(230)}$.

None of the other serine proteases listed in Table I has a

TABLE III

Basic and acidic residues in each serine protease (the total number and those in regions 1 and 2)

Basic (+) residues at pH 7 consist of Lys and Arg; acidic (−) residues consist of Asp and Glu; His is not included. The 15 serine proteases are arranged in increasing order of their overall net charge.

Serine protease	Whole molecule			Region 1			Region 2		
	+	−	Net	+	−	Net	+	−	Net
Porcine kallikrein	13	30	−17	2	2	0	3	4	−1
mMCP-7	15	25	−10	1	2	−1	2	0	2
Rat trypsin	13	19	−6	1	4	−3	3	2	+1
mMCP-6	18	23	−5	4	2	+2	2	3	−1
Rat tonin	21	24	−3	3	3	0	6	4	+2
S. griseus trypsin	16	16	0	1	2	−1	5	1	+4
Bovine chymotrypsin	17	14	+3	0	0	0	4	1	+3
mMCP-1	28	25	+3	6	2	+4	8	4	+4
rMCP-II	25	21	+4	4	2	+2	7	3	+4
Porcine elastase	15	11	+4	4	1	+3	3	0	+3
Bovine trypsin	16	10	+6	4	2	+2	4	0	+4
mMCP-2	27	20	+7	5	2	+3	8	2	+6
Human neutrophil elastase	19	9	+10	3	0	+3	1	1	0
mMCP-5	29	17	+12	9	0	+9	10	4	+6
mMCP-4	34	19	+15	11	1	+10	10	2	+8

positive potential in regions 1 and 2 as pronounced as that of the four mouse mast cell chymases and rMCP-II.

The four mouse chymases and rMCP-II have a region with negative electrostatic potential located close to the active site (Figs. 7*A* and 8). However, this region does not overlap with any of the substrate binding sites previously identified in complexes between serine proteases and their protein inhibitors (74). In mMCP-5, this negative region includes residues $6_{(21)}$, $64_{(77)}$, $65_{(78)}$, $135_{(149)}$, $139_{(153)}$, and $143_{(157)}$.

DISCUSSION

Mammalian serine proteases are enzymes consisting of two domains and approximately 230 residues, with the active site located in the cleft at the interface between the two domains. Each of the two domains consists of a distorted six-stranded β-barrel with a buried structurally conserved core and of one or two helices (75). The main structural differences between the members of this family are in the length and conformation of the exposed loop segments connecting the conserved strands and helices. Except for the most similar pair, the sequence identities of protein pairs range from 26 to 42%. Generally, about 85% of the C_α atoms can be superimposed with a root mean square deviation from 0.8 to 1.1 Å.

In this study, comparative molecular modeling was used to determine the 3D structure of mMCP-1, mMCP-2, mMCP-4, and mMCP-5. Because these four mMCPs were found to be more similar to rMCP-II than to any other serine protease with known 3D structure (Figs. 1 and 2), their 3D models were all based on the 3D structure of rMCP-II determined by x-ray crystallography (49). As depicted in Fig. 3, the backbones of these four models are virtually indistinguishable from the backbone of rMCP-II; the root mean square deviation is less than 0.2 Å. Like in rMCP-II (49), the amino acid insertions, deletions, and the loss of a disulfide bond caused conformational changes in the mouse mast cell chymases relative to chymotrypsin which created new binding sites for interaction with the P_3 and P_1' residues of a substrate, thereby restricting their specificities compared with that of chymotrypsin. In particular, the mast cell chymases are expected to have a preference for a large hydrophobic residue at P_3 and for a hydrophobic residue at P_1', in addition to the chymotryptic specificity for a large hydrophobic residue at P_1 (49). Unlike the pancreatic chymotrypsin gene which resides on chromosome 8, the mast cell chymase genes are all clustered on chromosome 14.[6] Apparently, this family of genes developed when a primordial gene encoding a protease with more restricted specificity than chymotrypsin underwent duplication and divergence.

From the 3D models of mMCP-1, mMCP-2, mMCP-4, and mMCP-5, nine peptide segments were identified that could be tested to obtain protease-specific antipeptide immunoglobulins. The degrees of accessibility, protrusion, flexibility, and variability of the peptide sequences were the parameters used for identification. The two most favored regions generally correspond to residues $74–89_{(87–102)}$ (segment 3) and $196–207_{(212–224)}$ (segment 9). However, segment 1 in mMCP-2 and segment 3 in mMCP-1 contain *N*-glycosylation sites and therefore may not be good epitopes for those two proteases. McNeil *et al.* (27) obtained an antibody against mMCP-5 using as an immunogen a synthetic peptide that corresponds to residues $146–162_{(160–177)}$ (segment 7) of the mature protease. Additionally, segment 2 was used to prepare an antibody that reacted with mMCP-2 but not with any of the other known mMCPs (76).

Many proteins with exposed basic residues bind to acidic proteoglycans (77). Although several 3D structures of glycosaminoglycans (78) and of their protein ligands are available, no 3D structure of a proteoglycan-protein complex has been determined experimentally. Nevertheless, structural aspects of this interaction have been studied theoretically and indirectly by experiment (39, 79–83). It appears in some cases (*e.g.* in antithrombin-heparin complex) that the binding involves specific interactions (80), whereas in others (*e.g.* in the complex of heparin cofactor II with heparin or dermatan sulfate) more general charge effects are involved (79).

An aim of this study was to use the 3D models of the four mast cell chymases to identify positively charged regions on their surfaces that could enable them to interact with negatively charged serglycin proteoglycans. There are two such potential heparin binding regions in mMCP-4 and mMCP-5. Their location away from the active site explains why mMCP-4 and mMCP-5 are still active when in complex with proteoglycans. Additionally, the active site region has a net negative charge that is likely to repel the negatively charged heparin glycosaminoglycan. In contrast to mMCP-4 and mMCP-5, mMCP-1 and mMCP-2 have a weaker region 2 and only a few positively charged residues in region 1. Of the rat chymases, rMCP-I has both positive regions (49) and therefore resembles mMCP-4 and mMCP-5; rMCP-II is similar to mMCP-1 and mMCP-2. The two positive regions in these mast cell chymases are different from the single proposed heparin binding region in thrombin (83) as well as from that in mast cell tryptases (24). Therefore, the absence of a larger number of basic residues in regions 1 and 2 in the serine proteases listed in Table III does not necessarily imply that they do not bind proteoglycans, but only that proteoglycans probably do not bind to regions 1 and 2.

The importance of the large positive electrostatic potential in regions 1 and 2 for binding to glycosaminoglycans can be tested by site-directed mutagenesis. The relatively large differences in the total charge and distribution of positive residues among the chymases that have a pronounced positive potential (Table III, Fig. 1) indicate that a single point mutation is not likely to remove the binding capacity of regions 1 and 2, unless specific ion pair interactions are involved. A

[6] Gurish, M. F., Nadeau, J. H., Johnson, K. R., McNeil, H. P., Grattan, K. M., Austen, K. F., and Stevens, R. L. (1993) *J. Biol. Chem.* **268**, in press.

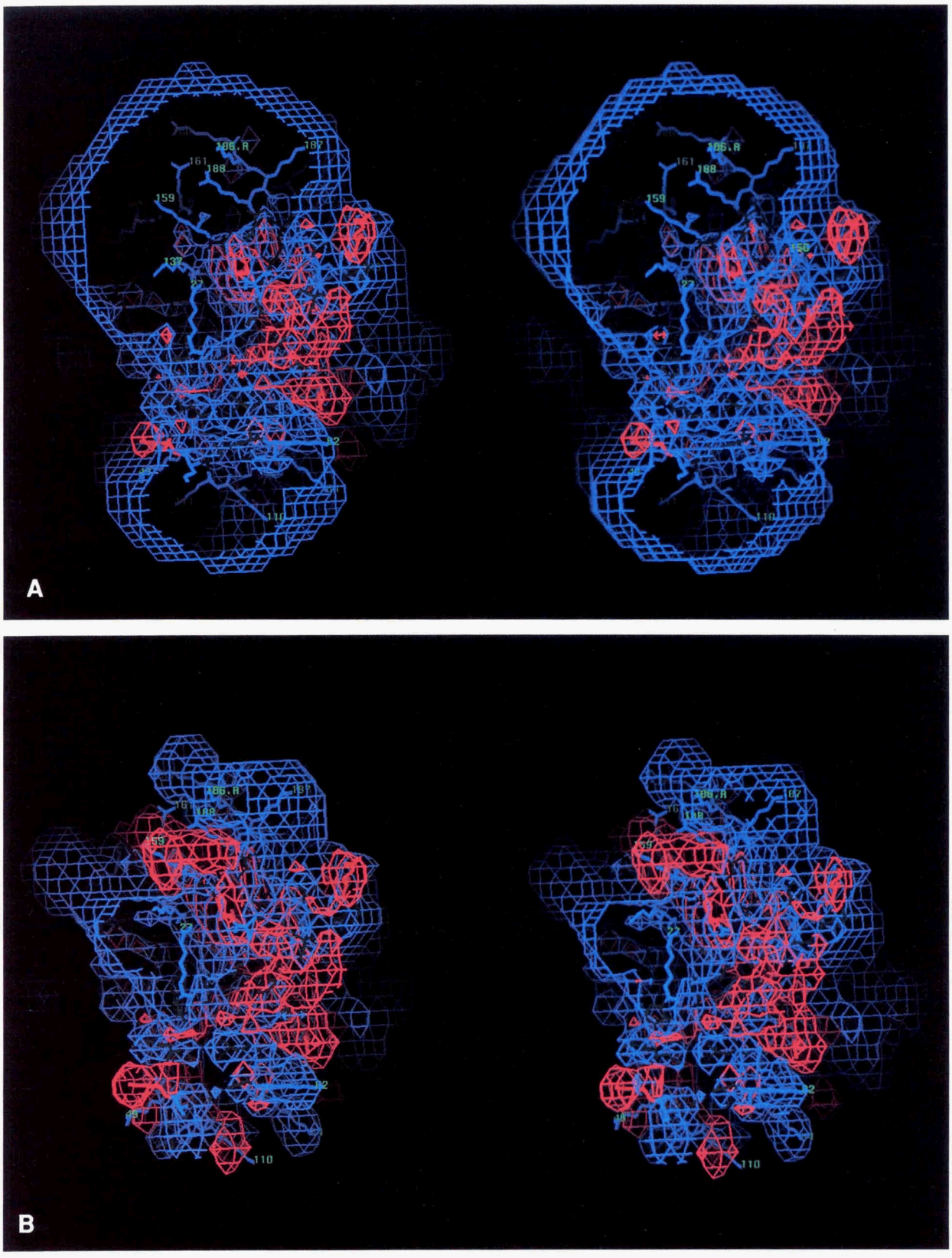

FIG. 7. **Stereoplot of the electrostatic potential around native and altered mMCP-4.** The Lys and Arg side chains are colored in *blue*, the Asp and Glu side chains are *red*, and the rest of the protein is *green*. The orientation of the molecule is similar to that in Fig. 3. The electrostatic potential is contoured at 0.6 (*light blue*) and −0.6 kcal/electron mol (*light red*). *A*, native protein. *B*, mMCP-4 with diminished electrostatic potential. The latter model is obtained by replacing the positively charged amino acid residues $145_{(159)}$, $147_{(161)}$, $173_{(188)}$ (region 1), $37_{(49)}$, and $97_{(110)}$ (region 2) with glutamic acid residues.

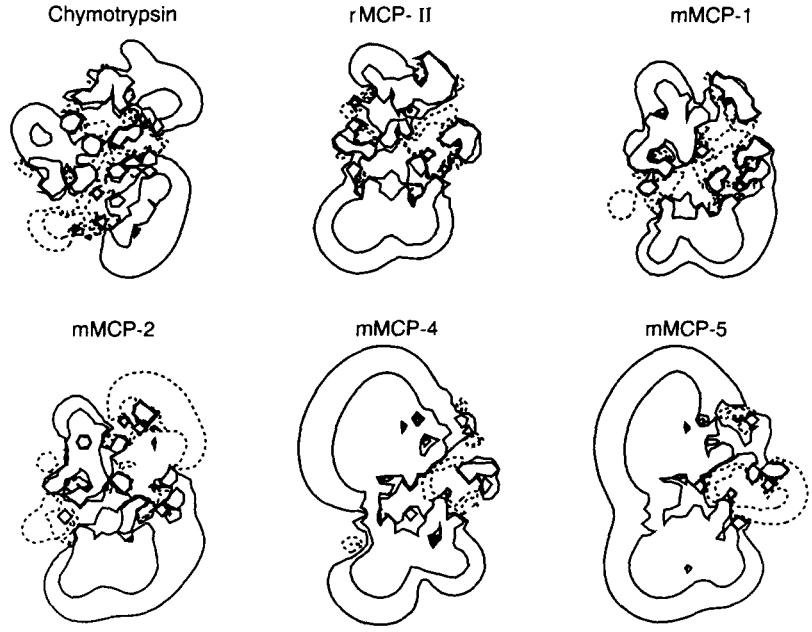

FIG. 8. **Comparison of electrostatic potentials of serine proteases.** Contour plots for electrostatic potential are compared for the molecules plotted in Fig. 6. The molecules are in the same orientation as in Fig. 3. The electrostatic potential is contoured at levels −0.9 and −0.3 (*dashed lines*) and 0.3 and 0.9 kcal/electron mol (*continuous lines*).

model of mMCP-4 in which the two strongly positive regions disappeared was constructed by changing 3 positively charged residues in the center of region 1 and 2 residues in region 2 to glutamic acids (Fig. 7*B*). In this altered molecule, the net charge of both regions is decreased to the value typical for the proteases that do not bind heparin (Table III).

Even though mMCP-4 has an amino acid sequence most similar to that of mMCP-1 (Figs. 1 and 2), only the proteases packaged in serosal mast cell granules in complex with heparin proteoglycans (mMCP-4, mMCP-5, and rMCP-I) have both regions 1 and 2 with high positive charge density. In contrast, proteases packaged in the chondroitin sulfate-rich granules of mucosal mast cells (mMCP-1, mMCP-2, and rMCP-II) have a significant positive potential only around region 2 and even that is weaker than the corresponding potential in the proteases of serosal mast cells. This localization of chymases can be interpreted as follows. The minimal positive charge density on a protease that is required for the binding of heparin is larger than that required for the binding of chondroitin sulfate because of the higher negative charge density on heparin. Evolution without the selective pressure for regions with high charge density tends to remove such regions because these regions do not have random composition. The proteases are therefore expected to have only as much positive charge as is required for binding their ligands. Consequently, it is possible that chymases with the lower concentration of positive charges (mMCP-1, mMCP-2, and rMCP-II) evolved to bind the weaker electrolyte (chondroitin sulfate), and chymases with the higher concentration of charges (mMCP-4, mMCP-5, and rMCP-I) evolved to bind the stronger electrolyte (heparin). Thus, the 3D models suggest a structural explanation for the selective localization of specific chymases within mouse and rat mast cells that contain different proteoglycans.

For the tryptase-heparin interaction inside the acidic granule, it has been suggested that histidine residues play an important role (24). This assumption was used to explain dissociation of tryptases from heparin when the pH is raised from 5.5 to 7.0 during exocytosis and degranulation. The absence of histidines in positive regions 1 and 2 of chymases is consistent with the fact that those proteases remain bound to heparin even after exocytosis.

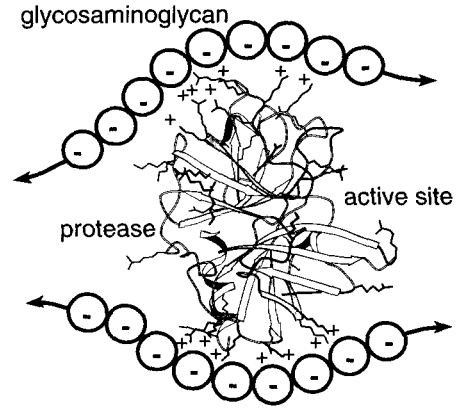

FIG. 9. **Schematic model of the interaction of heparin with mMCP-4.** A protease molecule interacting with two glycosaminoglycan chains is shown. It is also possible that the two regions are covered by one segment of a single glycosaminoglycan chain embracing the whole top, left, and bottom part of the molecule; this alternative would be facilitated by a strip of weak positive electrostatic potential connecting regions 1 and 2. A similar mode of binding to heparin is expected for mMCP-5 and rMCP-I.

The crystallographic analyses of several types of polysaccharide chains (78) show that they exist in left-handed helical conformation with four, six, eight, or sixteen monosaccharides per turn. In the case of chondroitin sulfate with eight monosaccharides per turn (code 1C4S in the Brookhaven Protein Data Bank), the dimension of an arc formed by a part of one turn of the polysaccharide helix is about 20 Å, corresponding to four to five monosaccharides. The size of the arc matches the length of the specific antithrombin binding pentasaccharide (81). Its size and curvature are also complementary to the positively charged regions 1 and 2 in mMCP-4 and mMCP-5. Additionally, there is a charge complementarity between approximately seven negative charges on the heparin pentasaccharide and six to ten positive charges on the two regions in mMCP-4 and mMCP-5. The large number of charged residues involved is in accord with significant cooperativity and high affinity binding (84, 85); the application of Manning's condensation model suggests that a minimum of five charges is required for this effect.

A schematic model of the interaction of mMCP-4, mMCP-5, and rMCP-I with heparin is shown in Fig. 9. The model for interaction is not sufficiently detailed to distinguish between a specific electrostatic interaction that requires a certain oligosaccharide sequence and an interaction that relies on charge density without many steric restrictions. Nevertheless, the model can serve as the basis for informed site-directed mutagenesis experiments that should provide more information on the binding between mast cell chymases and heparin.

REFERENCES

1. Wintroub, B. U., Schechter, N. B., Lazarus, G. S., Kaempfer, C. E., and Schwartz, L. B. (1984) *J. Invest. Dermatol.* **83,** 336–339
2. Cromlish, J. A., Seidah, N. G., Marcinkiewicz, M., Hamelin, J., Johnson, D. A., and Chrétien, M. (1987) *J. Biol. Chem.* **262,** 1363–1373
3. Le Trong, H., Neurath, H., and Woodbury, R. G. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 364–367
4. Franconi, G. M., Graf, P. D., Lazarus, S. C., Nadel, J. A., and Caughey, G. H. (1989) *J. Pharmacol. Exp. Ther.* **248,** 947–951
5. Urata, H., Kinoshita, A., Misono, K. S., Bumpus, F. M., and Husain, A. (1990) *J. Biol. Chem.* **265,** 22348–22357
6. Mizutani, H., Schechter, N., Lazarus, G., Black, R. A., and Kupper, T. S. (1991) *J. Exp. Med.* **174,** 821–825
7. Seppä, H., Väänänen, K., and Korhonen, K. (1979) *Acta Histochem.* **64,** 64–70
8. Vartio, T., Seppä, H., and Vaheri, A. (1981) *J. Biol. Chem.* **256,** 471–477
9. DuBuske, L., Austen, K. F., Czop, J., and Stevens, R. L. (1984) *J. Immunol.* **133,** 1535–1541
10. Briggaman, R. A., Schechter, N. M., Fraki, J., and Lazarus, G. S. (1984) *J. Exp. Med.* **160,** 1027–1042
11. Stevens, R. L., Somerville, L. L., Sewell, D., Swafford, J. R., Caulfield, J. P., Levi-Schaffer, F., Hubbard, J. R., and Dayton, E. T. (1992) *Arthritis Rheum.* **35,** 325–335
12. Gruber, B. L., Schwartz, L. B., Ramamurthy, N. S., Irani, A. M., and Marchese, M. J. (1988) *J. Immunol.* **140,** 3936–3942
13. Maier, M., Spragg, J., and Schwartz, L. B. (1983) *J. Immunol.* **130,** 2352–2356
14. Schwartz, L. B., Bradford, T. R., Littman, B. H., and Wintroub, B. U. (1985) *J. Immunol.* **135,** 2762–2767
15. Kokkonen, J. O., and Kovanen, P. T. (1989) *J. Biol. Chem.* **264,** 10749–10755
16. Le Trong, H., Newlands, G. F. J., Miller, H. R. P., Charbonneau, H., Neurath, H., and Woodbury, R. G. (1989) *Biochemistry* **28,** 391–395
17. Reynolds, D. S., Stevens, R. L., Lane, W. S., Carr, M. H., Austen, K. F., and Serafin, W. E. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 3230–3234
18. Reynolds, D. S., Gurley, D. S., Austen, K. F., and Serafin, W. E. (1991) *J. Biol. Chem.* **266,** 3847–3853
19. McNeil, H. P., Reynolds, D. S., Schiller, V., Ghildyal, N., Gurley, D. S., Austen, K. F., and Stevens, R. L. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 11174–11178
20. Serafin, W. E., Reynolds, D. S., Rogelj, S., Lane, W. S., Conder, G. A., Johnson, S. S., Austen, K. F., and Stevens, R. L. (1990) *J. Biol. Chem.* **265,** 423–429
21. Serafin, W. E., Sullivan, T. P., Conder, G. A., Ebrahimi, A., Marcham, P., Johnson, S. S., Austen, K. F., and Reynolds, D. S. (1991) *J. Biol. Chem.* **266,** 1934–1941
22. McNeil, H. P., Austen, K. F., Somerville, L. L., Gurish, M. F., and Stevens, R. L. (1991) *J. Biol. Chem.* **266,** 20316–20322
23. Huang, R., Blom, T., and Hellman, L. (1991) *Eur. J. Immunol.* **21,** 1611–1621
24. Johnson, D. A., and Barton, G. J. (1992) *Protein Sci.* **1,** 370–377
25. Ghildyal, N., McNeil, H. P., Gurish, M. F., Austen, K. F., and Stevens, R. L. (1992) *J. Biol. Chem.* **267,** 8473–8477
26. Ghildyal, N., McNeil, H. P., Stechschulte, S., Austen, K. F., Silberstein, D., Gurish, M. F., Somerville, L. L., and Stevens, R. L. (1992) *J. Immunol.* **149,** 2123–2129
27. McNeil, H. P., Frenkel, D. P., Austen, K. F., Friend, D. S., and Stevens, R. L. (1992) *J. Immunol.* **149,** 2466–2472
28. Schwartz, L. B., Riedel, C., Caulfield, J. P., Wasserman, S. I., and Austen, K. F. (1981) *J. Immunol.* **126,** 2071–2078
29. Serafin, W. E., Katz, H. R., Austen, K. F., and Stevens, R. L. (1986) *J. Biol. Chem.* **261,** 15017–15021
30. Goldstein, R. A., Luthey-Schulten, Z. A., and Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 4918–4922
31. Bourdon, M. A., Oldberg, A., Pierschbacher, M., and Ruoslahti, E. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1321–1325
32. Avraham, S., Austen, K. F., Nicodemus, C. F., Gartner, M. C., and Stevens, R. L. (1989) *J. Biol. Chem.* **264,** 16719–16726
33. Yurt, R. W., Leid, R. W., Jr., Austen, K. F., and Silbert, J. E. (1977) *J. Biol. Chem.* **252,** 518–521
34. Razin, E., Stevens, R. L., Akiyama, F., Schmid, K., and Austen, K. F. (1982) *J. Biol. Chem.* **257,** 7229–7236
35. Seldin, D. C., Austen, K. F., and Stevens, R. L. (1985) *J. Biol. Chem.* **260,** 11131–11139
36. Forsberg, L. S., Lazarus, S. C., Seno, N., DeVinney, R., Caughey, G. H., and Gold, W. M. (1988) *Biochim. Biophys. Acta* **416,** 416–428
37. Davidson, S., Gilead, L., Amira, M., Ginsburg, H., and Razin, E. (1990) *J. Biol. Chem.* **265,** 12324–12330
38. Guo, N., Krutzsch, H. C., Nègre, E., Zabrenetzky, V. S., and Roberts, D. D. (1992) *J. Biol. Chem.* **267,** 19349–19355
39. Cardin, A. D., and Weintraub, H. J. R. (1989) *Arteriosclerosis* **9,** 21–32
40. Reynolds, D. S., Stevens, R. L., Gurley, D. S., Lane, W. S., Austen, K. F., and Serafin, W. E. (1989) *J. Biol. Chem.* **264,** 20094–20099
41. Stevens, R. L., Lee, T. D. G., Seldin, D. C., Austen, K. F., Befus, A. D., and Bienenstock, J. (1986) *J. Immunol.* **137,** 291–295
42. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. (1987) *Nature* **326,** 347–352
43. Summers, N. L., and Karplus, M. K. (1989) *J. Mol. Biol.* **210,** 785–811
44. Šali, A., Overington, J. P., Johnson, M. S., and Blundell, T. L. (1990) *Trends Biochem. Sci.* **15,** 235–240
45. Greer, J. (1990) *Proteins* **7,** 317–334
46. Swindells, M. B., and Thornton, J. M. (1991) *Curr. Opin. Struct. Biol.* **1,** 219–223
47. de la Paz, P., Sutton, B. J., Darsley, M. J., and Rees, A. R. (1986) *EMBO J.* **5,** 415–425
48. Salemme, F. R. (1976) *J. Mol. Biol.* **102,** 562–572
49. Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W., and Neurath, H. (1988) *Biochemistry* **27,** 8097–8105
50. Cole, K. R., Kumar, S., Le Trong, H., Woodbury, R. G., Walsh, K. A., and Neurath, H. (1991) *Biochemistry* **30,** 648–655
51. Šali, A. (1991) *Modeling the Three-dimensional Structure of Proteins from Their Sequence of Amino Acid Residues.* Ph.D. Thesis, University of London, London
52. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542
53. Šali, A., and Blundell, T. L. (1990) *J. Mol. Biol.* **212,** 403–428
54. Zhu, Z.-Y., Šali, A., and Blundell, T. L. (1992) *Protein Eng.* **5,** 43–51
55. Sutcliffe, M. J., Haneef, I., Carney, D., and Blundell, T. L. (1987) *Protein Eng.* **1,** 377–384
56. Felsenstein, J. (1985) *Evolution* **39,** 783–791
57. Ponder, J. W., and Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791
58. Braun, W., and Gō, N. (1985) *J. Mol. Biol.* **186,** 611–626
59. Davis, M. E., Madura, J. D., Luty, B. A., and McCammon, J. A. (1990) *Comp. Phys. Comm.* **62,** 187–197
60. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. K. (1983) *J. Comp. Chem.* **4,** 187–217
61. Wendoloski, J. J., and Matthew, J. B. (1989) *Proteins* **5,** 313–321
62. Thornton, J. M., Edwards, M. S., Taylor, W. R., and Barlow, D. J. (1986) *EMBO J.* **5,** 409–413
63. Frömmel, C. (1988) *J. Theor. Biol.* **132,** 171–177
64. Novotny, J., Handschumacher, M., and Bruccoleri, R. E. (1987) *Immunol. Today* **8,** 26–31
65. Davies, D. R., Padlan, E. A., and Sheriff, S. (1990) *Annu. Rev. Biochem.* **59,** 439–473
66. Richmond, T. J., and Richards, F. M. (1978) *J. Mol. Biol.* **119,** 537–555
67. Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A., and van Regenmortel, M. H. V. (1984) *Nature* **311,** 123–126
68. Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A., and Hendrickson, W. A. (1984) *Nature* **312,** 127–134
69. Argos, P. (1987) *J. Mol. Biol.* **193,** 385–396
70. Hubbard, T. J. P., and Blundell, T. L. (1987) *Protein Eng.* **1,** 159–171
71. Chothia, C., and Lesk, A. M. (1986) *EMBO J.* **5,** 823–826
72. Šali, D., Bycroft, M., and Fersht, A. R. (1991) *J. Mol. Biol.* **220,** 779–788
73. Dao-Pin, S., Soderlind, E., Baase, W. A., Wozniak, J. A., Sauer, U., and Matthews, B. W. (1991) *J. Mol. Biol.* **221,** 873–887
74. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., and Huber, R. (1983) *Acta Crystallogr. Sect. B Struct. Sci.* **39,** 480–490
75. McLachlan, A. D. (1979) *J. Mol. Biol.* **128,** 49–79
76. Gurish, M. F., Stevens, R. L., Ghildyal, N., McNeil, H. P., Stechschulte, S., and Nicodemus, C. F. (1992) *FASEB J.* **6,** A1723
77. Kjellén, L., and Lindahl, U. (1991) *Annu. Rev. Biochem.* **60,** 443–475
78. Chakrabarti, B., and Park, J. W. (1980) *CRC Crit. Rev. Biochem.* **8,** 225–313
79. Whinna, H. C., Blinder, M. A., Szewczyk, M., Tollefsen, D. M., and Church, F. C. (1991) *J. Biol. Chem.* **266,** 8129–8135
80. Huber, R., and Carrell, R. W. (1989) *Biochemistry* **28,** 8951–8966
81. Lindahl, U., Thunberg, L., Bäckström, G., Riesenfeld, J., Nordling, K., and Bjork, I. (1984) *J. Biol. Chem.* **259,** 12368–12376
82. Church, F. C., Pratt, C. W., Noyes, C. M., Kalayanamit, T., Sherrill, G. B., Tobin, R. B., and Meade, J. B. (1989) *J. Biol. Chem.* **264,** 18419–18425
83. Bode, W., Turk, D., and Karshikov, A. (1992) *Protein Sci.* **1,** 426–471
84. Heuck, C. C., Schiele, U., Horn, D., Fronda, D., and Ritz, E. (1985) *J. Biol. Chem.* **260,** 4598–4603
85. Horn, D. (1980) in *Polymeric Amines and Ammonium Salt* (Goethals, E. J., ed) pp. 333–355, Pergamon Press, New York
86. Burks, C., Fickett, J. W., Goad, W. B., Kanehisha, M., Lewitter, F. I., Rindone, W. P., Swindell, C. D., Tung, C. S., and Bilofsky, H. S. (1985) *Comput. Appl. Biosci.* **1,** 225–233
87. Fujinaga, M., and James, M. N. G. (1987) *J. Mol. Biol.* **195,** 373–396
88. Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., and Bartunik, H. (1983) *J. Mol. Biol.* **164,** 237–282
89. Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W., and Huber, R. (1982) *Acta Crystallogr. Sect. B Struct. Sci.* **38,** 1462–1472
90. Tsukada, H., and Blow, D. M. (1985) *J. Mol. Biol.* **184,** 703–711
91. Meyer, E., Cole, G., Radakrishnan, R., and Epp, O. (1988) *Acta Crystallogr. Sect. B Struct. Sci.* **44,** 26–38
92. Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N.-H., Hamlin, R., Rutter, W. J., and Craik, C. S. (1987) *Science* **237,** 905–909
93. Navia, M. A., McKeever, B. M., Springer, J. P., Lin, T.-Y., Williams, H. R., Fluder, E. M., Dorn, C. P., and Hoogsteen, K. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7–11
94. Reynolds, R. A., Remington, J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L., and Matthews, B. W. (1985) *Acta Crystallogr. Sect. B Struct. Sci.* **41,** 139–147
95. Read, R. J., and James, M. N. G. (1988) *J. Mol. Biol.* **200,** 523–551
96. Overington, J., Johnson, M. S., Šali, A., and Blundell, T. L. (1990) *Proc. R. Soc. Lond. B Biol. Sci.* **241,** 132–145
97. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 946–950