PROTEINS: Structure, Function, and Genetics 4:190-204 (1988)

Comparative Molecular Model Building of Two Serine Proteinases From Cytotoxic T Lymphocytes

Michael E.P. Murphy,[1,2] John Moult,[1,2] R. Chris Bleackley,[2] Howard Gershenfeld,[3] Irving L. Weissman,[3] and Michael N.G. James[1,2]

[1]*Medical Research Council of Canada, Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G2H7;* [2]*Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7;* [3]*Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305*

ABSTRACT **Two genes that are expressed when precursor cytotoxic T lymphocytes are transformed to T killer cells have been cloned and sequenced. The derived amino acid sequences, coding for cytotoxic cell protease 1 (CCP1) and Hannuka factor (HF) are highly homologous to members of the serine proteinase family. Comparative molecular model building using the known three-dimensional structures and the derived amino acid sequences of the lymphocyte enzymes has provided useful structural information, especially in predicting the conformations of the substrate binding sites. In applying this modelling procedure, we used the X-ray structures of four serine proteinases to provide a structurally based sequence alignment: α-chymotrypsin (CHT), bovine trypsin (BT), *Streptomyces griseus* trypsin (SGT), and rat mast cell protease 2 (RMCP2). The root mean square differences in α-carbon atom positions among these four structures when compared in a pairwise fashion range from 0.79 to 0.97 Å for structurally equivalent residues. The sequences of the two lymphocyte enzymes were then aligned to these proteinases using chemical criteria and the superimposed X-ray structures as guides. The alignment showed that the sequence of CCP1 was most similar to RMCP2, whereas HF has regions of homology with both RMCP2 and BT. With RMCP2 as a template for CCP1 and the two enzymes RMCP2 and BT as templates for HF, the molecular models were constructed. Intramolecular steric clashes that resulted from the replacement of amino acid side chains of the templates by the aligned residues of CCP1 and HF were relieved by adjustment of the side chain conformational angles in an interactive computer graphics device. This process was followed by energy minimization of the enzyme model to optimize the stereochemical geometry and to relieve any remaining unacceptably close nonbonded contacts. The resulting model of CCP1 has an arginine residue at position 226 in the specificity pocket, thereby predicting a substrate preference for P_1 aspartate or glutamate residues. The model also predicts favorable binding for a small hydrophobic residue at the P_2 position of the substrate. The primary specificity pocket of HF resembles that of BT and therefore predicts a lysine or arginine preference for the P_1 residue. The arginine at position 99 in the model of HF suggests a preference for aspartate or glutamate side chains in the P_2 position of the substrate. Both CCP1 and HF have a free cysteine in the segment of polypeptide 88 to 93. Models of the dimeric form of these enzymes can be constructed by forming disulfide bridges between the suitably oriented monomers, i.e., Cys88–Cys88′ for CCP1 and Cys93–Cys93′ for HF.**

Key words: computer graphics, energy minimization, proteolytic enzymes

INTRODUCTION

Cytotoxic T lymphocytes (CTL) bind to cells that express foreign molecules and induce them to lyse. [1,2] This class of lymphocyte is thus one of the major effectors of cell-mediated immune responses that include the destruction of virally transformed cells, organ transplants, and possibly tumor cells. It is therefore of great interest to understand the mechanism of CTL-mediated lysis. However, despite extensive efforts, the nature and role of the molecules involved remains largely unclear.

Our approach to the study of this mechanism is based on the hypothesis that the proteins involved in the killing machinery are uniquely expressed in CTL. Recently, the cloning of a number of CTL-specific genes was described.[3–6] These genes are of particular interest, as their expression clearly parallels the induction of cytotoxicity. The derived amino acid sequences suggest molecules with high sequence homology to members of the serine proteinase family.[4,5] Two of the enzymes coded by the gene sequences have been named *cytotoxic cell protease* 1 (CCP1) and *Hannuka factor* (HF). Using antipeptide antibodies, we have demonstrated that one of the proteinases (CCP1) is localized within cytoplasmic granules of CTL.[7] It therefore appears likely that at least one of these enzymes plays a key role in the

Received May 20, 1988; revision accepted August 4, 1988.

Address reprint requests to Dr. Michael N.G. James, Department of Biochemistrty, University of Alberta, Edmonton, Alberta, Canada T6G 2H7.

John Moult is now at the Center for Advanced Research in Biotechnology, 9600 Gudesky Drive, Rockville, MD 20850.

cytolytic mechanism[8] either directly or in some ancilliary capacity as a component of a lytic proteinase cascade.

Several serine proteinases from cytoplasmic granules of CTLs have been purified recently.[9,10] Amino terminal sequence analyses of these proteins reveal that some of them have sequences identical to the derived sequences of CCP1 and HF. These proteinases have been called *granzymes*. A comparison of our predicted sequence with the determined protein sequences of the granzymes also indicates that CCP1 and HF are synthesized with a hydrophobic leader sequence and may exist in a proenzyme form by virtue of the presence of two extra amino acids at the amino terminus.[4]

Structural models of proteinases from CTLs will be useful in the design of experiments to define better their role in the mechanism of cell lysis. In particular, predicting the substrate specificity of the enzymes from the models could aid in the discovery of the physiologically important substrate. Alternatively, the design and synthesis of inhibitors of these proteinases could be of benefit in the suppression of the immune response in organ transplantation.

The technique of comparative molecular modelling[11] has come a long way since it was first used to explain the substrate specificity of bovine trypsin (BT) from the α-chymotrypsin (CHT) structure as template.[12] The early modelling of α-lytic protease based on porcine elastase[13] was shown to be seriously in error[14] primarily because of the very low sequence homology (<18% identity) between these two enzymes. Large deletions and insertions are necessary to align the bacterial and pancreatic serine proteinase sequences correctly.[15,16] On the other hand, when the sequence homology between two proteins is much higher, it is possible to make structural deductions that have reasonable fidelity.[17] A structural model of *Streptomyces griseus* trypsin (SGT) based on the structure of BT was better than 80% faithfully reproduced. It differed seriously from the X-ray structure in regions where the amino acid sequence alignment was completely ambiguous, i.e., nonhomologous regions. There were several less serious errors in regions where BT has disulfide bridges that are absent in SGT. Final comparison of the two molecular structures (SGT and BT) shows that of the 190 structurally equivalent residues, 70 (38.8%) are identical in both enzymes. This relatively high sequence identity results in the similar folding of the polypeptide chains of the two enzymes.[17]

The initial sequence alignment of CCP1 with RMCP2 indicated a very high homology (~48% identity) that portends very similar tertiary structures. Recently, a sequence alignment of RMCP1, RMCP2, human neutrophil cathepsin G (CATG), CCP1, and human lymphocyte protease (HLP) suggested that the serine proteinases produced in granulocytes or lymphocytes represent a distinct subclass of serine proteinase with common structural features.[18,19] This subclass is characterized by the tertiary structure of RMCP2[20] that is missing the disulfide bridge Cys191–Cys220 present in the proteinases of the mammalian pancreas and plasma. It is also found in those bacterial serine proteinases of the same family.[15,16] The granulocyte/lymphocyte family has only the three disulfide bridges that are homologous to Cys42–Cys58, Cys136–Cys201, and Cys168–Cys182 of the chymotrypsin family.

Comparative molecular modelling relies on a large data base of refined crystal structures of the member proteins. The serine proteinase family is one of the best represented families in terms of available high resolution single crystal structure analyses. In addition to the data base with α-chymotrypsin,[21,22] elastase,[23] and trypsin,[24,25] several other serine proteinase structures have been determined recently: SGT,[17,26] kallikrein,[27] CHT complexed to OMTKY3, the third domain of turkey ovomucoid,[28] and RMCP2.[20] These structures were used as templates for the structurally based sequence alignment of mouse CCP1 and human HF and to provide the framework on which to construct the enzyme models.

MATERIALS AND METHODS

Sequence Alignment

The first step in the construction of an accurate three-dimensional model of an unknown protein is to deduce correctly the sequence alignment with the template protein(s). Alignment of sequences of proteins that have a high degree of identity is straightforward and relatively foolproof.[29] However, there is no guarantee that the homologous protein of unknown structure will in fact adopt the conformation of the template structure in all regions.[17] The strength of the comparative molecular modelling methodology lies in there being a large data base of known structures available to select structures that will be close to that of the unknown. Since the initial work of Greer[11] on serine proteinases, this data base has been increased substantially, with several new crystal structures having been determined.

The sequence alignment of proteins of known three-dimensional structure is best accomplished by superimposing the α-carbon atom coordinates of the proteins and noting those residues that are structurally equivalent.[17] The four serine proteinases that we selected to construct our templates are CHT, SGT, BT and RMCP2. The coordinates of CHT are from the molecular complex of CHT with OMTKY3[28]; SGT coordinates are those of Read et al.[17] and Read and James[26]; the coordinates of BT and RMCP2 were obtained from the Brookhaven Protein Data Bank.[30]

The structural alignment of the proteins was done following the methodology developed by Read et al.[17] We used the criterion that will superimpose the α-carbon atoms of each enzyme by a least-squares procedure provided they agree in position to within 1.9 Å. In this way the root mean square (rms) devia-

tions of the structurally equivalent α-carbon atoms between pairs of proteins were derived (Table I).

To align the sequences of CCP1 and HF to these four proteinases we used a program based on the method of McLachlan.[31] Any insertions or deletions that were required to maximize the homology of the proteins to the template molecules were kept out of the structurally conserved regions by paying strict attention to the structural alignment described above. CCP1 is highly homologous to RMCP2, and the alignment was straightforward. The sequence alignment of HF required more attention to detail.

Coordinate Generation of the CCP1 and HF Models

The computer program MUTATE (R. J. Read, unpublished) was used to replace the amino acid side chains of RMCP2 by the side chains of CCP1 using the sequence alignment shown in Figure 1. The template for constructing HF was a combination of the proteins BT and RMCP2 (Table II). When the sequences of template and model were identical, the atomic coordinates of the template were used directly for the model. In the case of an amino acid change or substitution, the atomic coordinates were replaced by those for the new side chain. In selecting the new side chain conformation, the torsional angles were adjusted so that it follows the path of the template side chain as far as possible; otherwise, the program will select from a dictionary those preferred side chain conformational angles.[32–35]

Following the initial generation of the model, the coordinates were viewed on the interactive graphics. The CCP1 model was optimized by removing unacceptably close nonbonded intramolecular contacts using the MMS-X interactive graphics running the program M3 written by C. Broughton.[36] Similar adjustments were made for the HF model using the program MMS (written by S. Dempsey, University of California at San Diego) on a Silicon Graphics IRIS 3030. The resulting two sets of visually optimized atomic coordinates were further analyzed by computing all intramolecular contacts (program INTRA written by A. Sielecki, University of Alberta). The output of this program was used to make further adjustments to the model in the graphical devices.

The above procedure does not consider the additional problems resulting from insertions or deletions as a result of the sequence alignments. For example, CCP1 requires a deletion relative to RMCP2 in the region 35–41; on the other hand, it has two extra residues over RMCP2 in the disulfide-bridged loop 168–182, and it has an additional Ser-Ser dipeptide at the C terminus (Fig. 1).

To build the loop with the deletion Ile35–Ile41, the following strategy was employed. Building was done in the interactive graphics devices. First, the several X-ray structures (Fig. 1) were examined in the region of the deletion to determine those residues closest to the deletion that are structurally conserved. In this case residues 35 and 41 of the four enzymes are conserved. Then the enzyme structure with the loop closest in length to the segment of chain that is being built is selected (in this case RMCP2). The intervening peptide was generated, i.e., spanning the residues from Ile35 to Ile41. One end of the peptide (Ile35) was overlapped with the structurally conserved residue of

TABLE I. Results of α-Carbon Atom Superpositions of Crystal Structures and the Degree of Sequence Identity in Structurally Equivalent Regions*

	CHT (230)	SGT (223)	BT (223)	RMCP (224)
CHT	—	0.97 (182)	0.82 (199)	0.80 (188)
SGT	66 (36)	—	0.88 (185)	0.86 (172)
BT	96 (48)	71 (38)	—	0.79 (195)
RMCP	71 (38)	58 (34)	63 (32)	—

*The upper triangular portion of this matrix contains the root mean square (rms) differences (Å) in α-carbon atom positions between the enzymes compared in a pairwise fashion. The numbers in parentheses are the numbers of structurally equivalent residues that have α-carbon atoms that agree to within 1.9Å. The lower triangular part of this matrix contains the number of residues that are identical in the structurally equivalent regions of each pair of proteins. The number following in parentheses is the percentage of the structurally equivalent residues that are identical.

TABLE II. Modelling Strategy and Models for Loop Construction

CCP1 Model	Template	HF Model	Template
16–245	RMCP2	16–87	RMCP2
		88–123	BT
		124–167	RMCP2
		168–246	BT
CCP1 Loops	**Template**	**HF loops**	**Templates**
35–41	RMCP2 with deletion at Gly 37C	35–41	BT, CHT
		58–64	CHT
168–182	SGT	168–178	Insertion of four residues
		217–220	Insertion of one residue
		231–237	Insertion of one residue

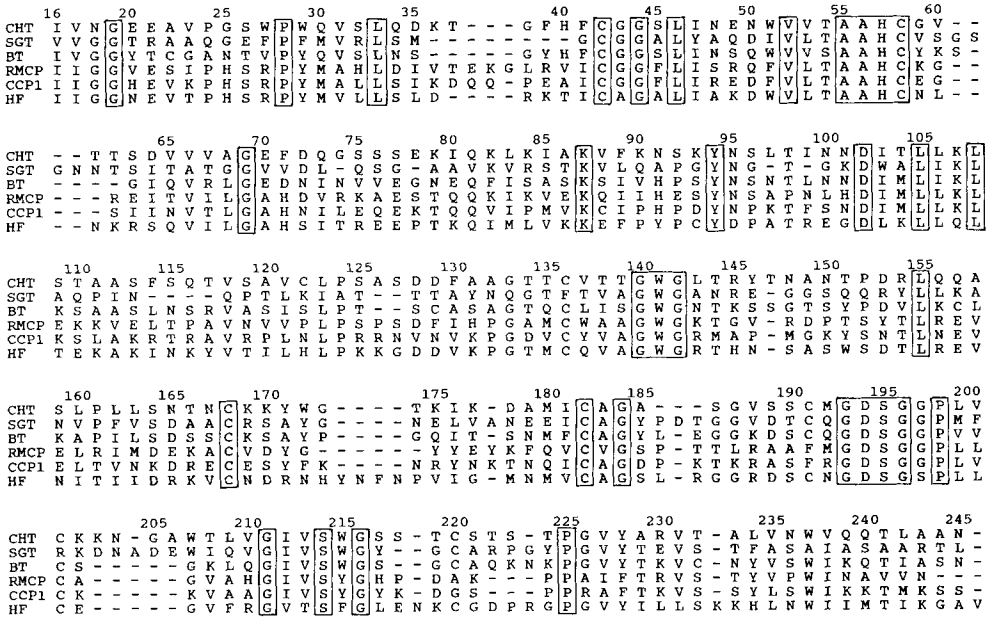

Fig. 1. Alignment of the amino acid sequences of CCP1 [4] and human HF (HF, Gershenfeld and Weissman, unpublished data) with the sequences of α-chymotrypsin (CHT), *Streptomyces griseus* trypsin (SGT), bovine trypsin (BT), and rat mast cell protease 2 (RMCP). The sequences of CHT, SGT, BT, and RMCP were all taken from the Brookhaven Protein Data Bank.[30] The alignment shown here was based on the three-dimensional structural alignment of CHT, SGT, BT, and RMCP as described in the text. The numbering is that of CHT. Only those residues that are conserved in all six sequences as aligned are enclosed in boxes.

the model (Ile35 in RMCP2). The ϕ,ψ angles of the residues were then adjusted to form the loop so that the other terminus was overlapped with the other structurally conserved residue (Ile41 of RMCP2). Particular attention was paid to the overall conformation of the loop in RMCP2. The resulting ϕ,ψ angles of the residues in this newly constructed peptide were compared with a data base of observed, allowed torsional angles to ensure that only energetically accessible ϕ,ψ values were used in building the loop. Alternative methods of loop building have been described recently.[35,37,38] These were not attempted in the present study.

The structurally conserved segments of polypeptide that were used in constructing the models of CCP1 and HF are given in Table II. For CCP1 the Ser-Ser dipeptide at 244–245 was added with an α-helical conformation, and in the case of HF the additional Val246 at the C terminus was added also with an α-helical conformation.

Energy Minimization and Structure Analysis

To ensure that all unfavorable nonbonded contacts were relieved and to idealize the peptide bond geometry in regions where the polypeptide chain had to be annealed because of the model building, energy minimizations of each model (CCP1 and HF) were done using the GROMOS package.[39,40] In addition, energy minimization of the X-ray structural coordinates of RMCP2 was done to provide a standard on which to judge the results of the calculations on the hypothetical models. Conjugate gradient energy minimization (1,000 steps), including in the model polar hydrogen atoms and using the potential energy function with the parameter set 37D (8 Å cut-off), was performed. These parameters have electrostatically neutral groups on all potentially charged residues to mitigate the absence of solvent. The experimental X-ray coordinates and the model built coordinates of CCP1 and HF were minimized directly with no constraints on the atomic positions. In addition, an alternative protocol for CCP1 and HF was followed. In this second process, all of the atoms except for those of residues in the loops in which it was necessary to adjust the main chain were restrained to their original positions. After 100 steps of this restrained minimization, the restraints were relieved on all atoms and the minimization was continued on to the 1,000 steps as previously. The coordinates for CCP1 and HF for the energy minimization done by the restrained atom protocol are the ones we used in the following discussion. These coordinates have been deposited with the Brookhaven Protein Data Bank. [30]

Several analyses of the atomic coordinates of CCP1 and HF models were carried out to evaluate as well as possible the results of the model building procedure. Solvent accessible areas of RMCP2, CCP1, HF, and the other three enzymes listed in Figure 1 were

evaluated by the method of Lee and Richards[41] using a program written by J. Moult with a probe size of 1.4 Å. Implicit polar hydrogens were not included in these calculations. The relative distribution of the solvent accessible surface areas of polar and nonpolar residues were compared for these several coordinate sets.

RESULTS AND DISCUSSION

Sequence Alignment

In regions of a protein where the sequences have been correctly aligned, the chances that the comparative modelling will yield a structure close to the true structure are relatively high. When the alignment is wrong the model is guaranteed to be wrong.[11,17] Although sequence alignments of CCP1 and HF to the several serine proteinases have already been published,[4,5,18] a new alignment based on the three-dimensional structures was necessary for model building (Fig. 1). The popular Dayhoff algorithm [42,43] used in the previous studies places more emphasis on maximizing sequence identity than does our procedure.[17] That method can result in the placement of gaps or insertions in structurally conserved regions of the molecule. For example, in the alignment of the sequences of several proteinases from granulocytes and leukocytes to that of chymotrypsinogen A (CHTA), a gap of four residues was positioned after Ala206.[18] The structurally correct position for this deletion based on our alignment occurs after Lys202. Residues Lys203 to Ala206 form a β-hairpin turn in CHT and are the ones most likely to be deleted in CCP1 and HF, as is observed experimentally for BT and RMCP2.

The data base that we used for sequence alignment and template selection is presented in Figure 1. Table I gives the extent of structural equivalence among the four enzymes of known three-dimensional structure. The rms agreement among the α-carbon atom coordinates of the mammalian enzymes (CHT, BT, and RMCP2) is marginally better than the structural agreement of SGT with the same three enzymes. Better than 82% of the residues of the mammalian enzymes form the structurally equivalent regions (α-carbon atoms agree to within 1.9 Å) with rms differences of 0.79 to 0.82 Å. It is of interest to note that although RMCP2 and BT have one of the highest percentages of structural equivalence, the sequence identity between this pair is one of the lowest in the samples (32% of the structurally equivalent residues). As expected, bovine trypsin and bovine chymotrypsin have the most extensive sequence identity in the sample (48%).

The alignment of CCP1 and HF to the sequences of the four enzymes in Figure 1 shows extensive overall identity of CCP1 with RMCP2 (48%). Somewhat less homology of HF to the other enzymes (38 and 33% with RMCP2 and BT, respectively) is observed. All of these enzymes clearly belong to the same class.[4,5] However, the disulfide bridge content and distribution have suggested that the different enzymes will make better templates upon which to construct CCP1 and HF (Table II).

The very high sequence homology between CCP1 and RMCP2, the identical disulfide bridge pattern, and the fact that the disulfide bridge Cys191 to Cys220, characteristic of many of the mammalian serine proteinases, is missing in both clearly shows that RMCP2 will be an excellent template for CCP1. On the other hand, we have chosen two of the proteinases of known structure for the construction of HF. As in RMCP2, the HF sequence does not have the disulfide between residues 22 and 157, but, as in BT, there is a disulfide between Cys191 and Cys220. Therefore, RMCP2 and BT make the templates upon which HF was constructed (Table II).

Insertions and deletions present the greatest problems in the modelling project.[11,17,35] Figure 1 shows that CCP1 differs in size from RMCP2 at three places: a deletion at Gly37C, an insertion of two residues at 173 and 174, and a longer C terminus (Ser244–Ser245). The model of the deletion at Gly37C is less likely to represent the true structure than the insertion at 173, since the structure of SGT[26] could be used as a guide for the insertion.[11] The differences of HF from the templates used for its construction are much more extensive. The major one is the tetrapeptide insertion at 173A–173D in which there is no satisfactory model for its construction. This segment was constructed in the interactive graphics, but the process clearly would have benefitted by the application of the loop building algorithm of J. Moult.[35]

Clearly, the segments (loops) listed in Table II for CCP1 and HF are the regions of these structural models that are likely to differ from the actual structures. Of course there is at present no way to predict which segments of CCP1 and HF that are highly homologous to the template are likely to differ as well, but there will undoubtedly be such regions.[17]

Model Evaluation

The data in Table I and analyses of other groups of homologous structures[44] give estimates of how much our models of CCP1 and HF might deviate from the true structures. In regions of the molecular core (the structurally equivalent regions in Table I), coordinate differences of ~0.8 to 1.1 Å can be expected. The divergence of core structure, main chain atoms that deviate by less than 3 Å as a function of sequence identity has been well correlated by comparing related X-ray structures.[44] Based on that work, the core of the CCP1 model coordinates would deviate from the true structure by 1.1 Å, i.e., 48% sequence identity to RMCP2. The calculated deviation for HF would range from 1.2 to 1.4 Å based on the sequence identity to BT, RMCP2, and the RMCP2/BT sequence mix. The coordinate difference of the noncore parts (the external loops) is expected to depart more from the true structure because of the lower level of homology

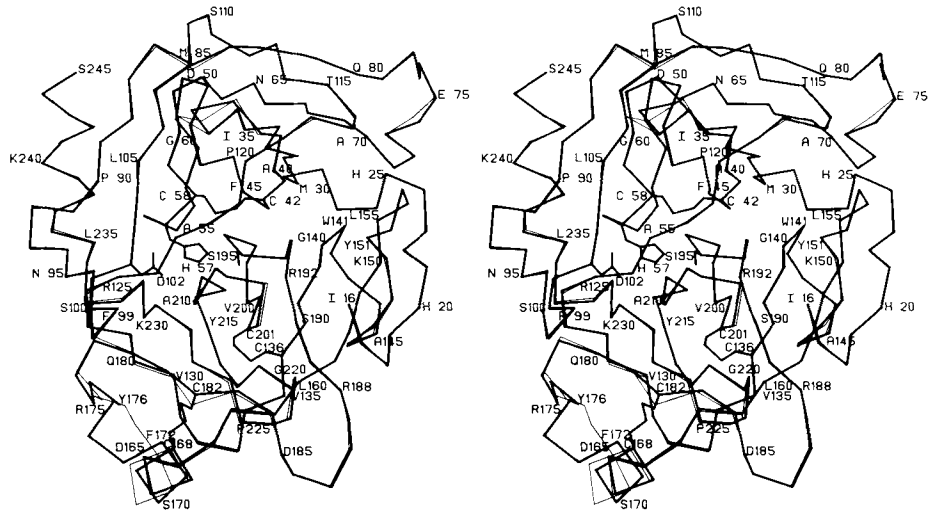

Fig. 2. Stereorepresentation of the polypeptide chain fold (α-carbon atoms only) for the model of CCP1 (thick lines) and the X-ray crystal structure of RMCP2 (thin lines). The amino acid sequence and residue numbers are for CCP1 (see Fig. 1). The residues of the catalytic triad (His57, Asp102, and Ser195) and the three disulfide bridges of CCP1 are included with thick lines.

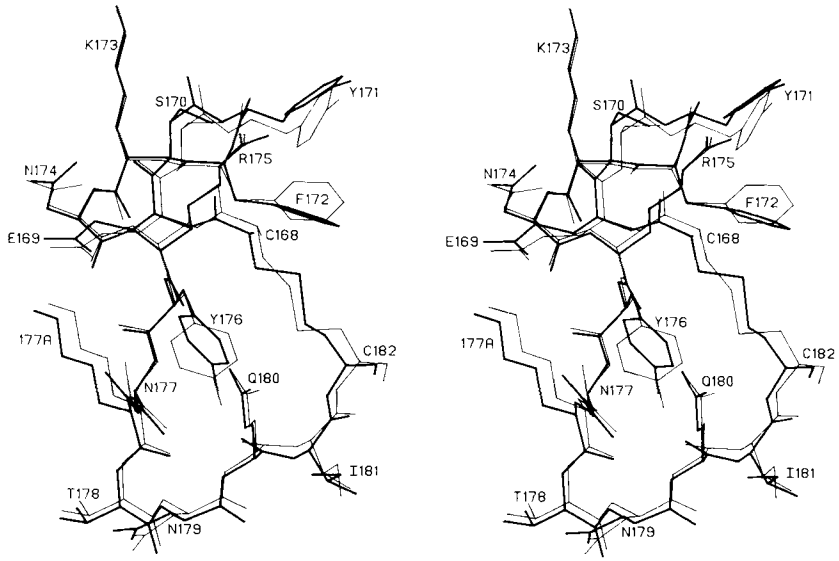

Fig. 3. Stereographic representation of the disulfide-bridged loop Cys168–Cys182. Relative to RMCP2, CCP1 has two additional residues at Lys173 and Asn174. The model for this loop was SGT, as it has the same number of residues. CCP1 after energy minimization is in thick lines; before energy minimization, thin lines. Major adjustments are in the orientations of the aromatic rings and the disulfide bridge (see sequence in Fig. 1).

and the presence of insertions and deletions. By these criteria, the CCP1 and HF segments with greatest deviation from the actual structure would be the disulfide-bridged loop from Cys168 to Cys182, where the two leukocyte-derived enzymes have large sequence insertions and relatively low sequence identity with the templates used for construction.

Energy minimization was carried out to improve the stereochemistry of the models. Initially, the models were energy minimized directly after graphical manipulation. However, it was observed that some steric clashes were relieved at the expense of adjusting the molecular core, which is expected to be highly structurally conserved because of the extensive sequence identity. Constraints on the model at all atomic positions except those segments in which large energies were observed allowed minimization to comparable global energies at the same time conserving the homologous core structure. This energy minimization strategy resulted in the greatest displacement from initial structure for the constructed loops.

Energy parameters, given in Table III, provide relative energies per atom for the constrained and unconstrained minimization of the two models and

RMCP2. The HF model was constructed by starting with a combination of RMCP2 and BT structures that created some van der Waals clashes in the molecular core that were not easily removed with the interactive graphics. These clashes resulted in the higher starting energy of the HF model than that of CCP1. The total energy and the electrostatic energy found for both of the models is less than that of RMCP2 on a per atom basis. These values cannot be used as an estimate of the structural reliability, since there are different numbers of charged residues in each case: 54 for CCP1 and 51 for HF, whereas RMCP2 contains only 44. The total energy minimum found for RMCP2 is simliar in magnitude to the energy minima calculated by Novotony et al.[45] for hemerythrin (−5.7 kJ/mol atom) and mouse κ-chain variable domain (−3.8 kJ/mol atom). It should be noted that these comparisons are not useful for determining the reliability of the models.

The results of the energy minimization on the two model structures are presented in Figures 2 and 3. The superposition of α-carbon atoms of RMCP2 and CCP1 after energy minimization (Fig. 2) shows that the new side chains of CCP1 are accommodated in the RMCP2 chain fold. The single residue deletion (Gly37C) and two-residue insertion in CCP1 relative to RMCP2 (Lys173, Asn174) are clearly visible from the relatively large changes in main chain conformation. The extension of the C-terminal helix by the two serine residues is a trivial difference provided that these residues adopt an α-helical conformation. Smaller changes in the protein core and periphery are a result of the energy minimization. The rms shift in the core regions is ~0.4 Å. In Figure 3 the modelled loop of CCP1 between Cys168 and Cys182 is shown before and after energy minimization. The model before the minimization is essentially that of SGT with the superposed sequence changes (Fig. 1).

TABLE III. Results of Energy Minimization Calculations

	Unconstrained minimization*			Constrained minimization	
Parameters	CCP1	HF	RMCP2	CCP1	HF
Total starting	25.86	53.67	−1.829	25.86	53.67
Total final	−6.175	−5.617	−5.505	−6.167	−5.418
Electrostatic	−4.076	−3.839	−3.285	−4.067	−3.639
rms shift (Å)	0.43	0.62	0.37	0.42	0.60
Number of atoms†	2,203	2,237	2,100	2,203	2,237

*Potential energy values are in Kjoules and are quoted per atom.
†This number includes the polar hydrogen atoms added to the O and N atoms in each protein.

TABLE IV. Solvent Accessible Surface Areas*

A. Energy Minimized Structures	Solvent accessibility (Å^2) per atom		
Atom type	CCP1	HF	RMCP2
Polar side chain atoms	19.472 (185)	19.260 (184)	16.307 (160)
Nonpolar side chain atoms	8.588 (678)	8.298 (685)	8.233 (669)
Ratio (nonpolar/polar) surface area/surface area	1.62	1.60	2.11
Main chain atoms	2.138 (909)	2.403 (937)	2.283 (897)
All atoms	6.416 (1772)	6.357 (1806)	5.889 (1726)

B. X-ray Crystallographic Structures				
Atom type	RMCP2	BT	SGT	CHY
Polar side chain atoms	16.73 (160)	16.20 (156)	16.71 (155)	16.36 (159)
Nonpolar side chain atoms	7.99 (669)	6.92 (580)	7.03 (572)	7.30 (613)
Ratio (nonpolar/polar) surface area/surface area	1.98	1.92	1.55	1.72
Main chain atoms	2.36 (897)	2.84 (893)	2.66 (893)	2.92 (919)
All atoms	5.87 (1726)	5.58 (1629)	5.55 (1620)	5.77 (1691)

*The numbers in parentheses are the numbers of nonhydrogen atoms in each protein in each category.

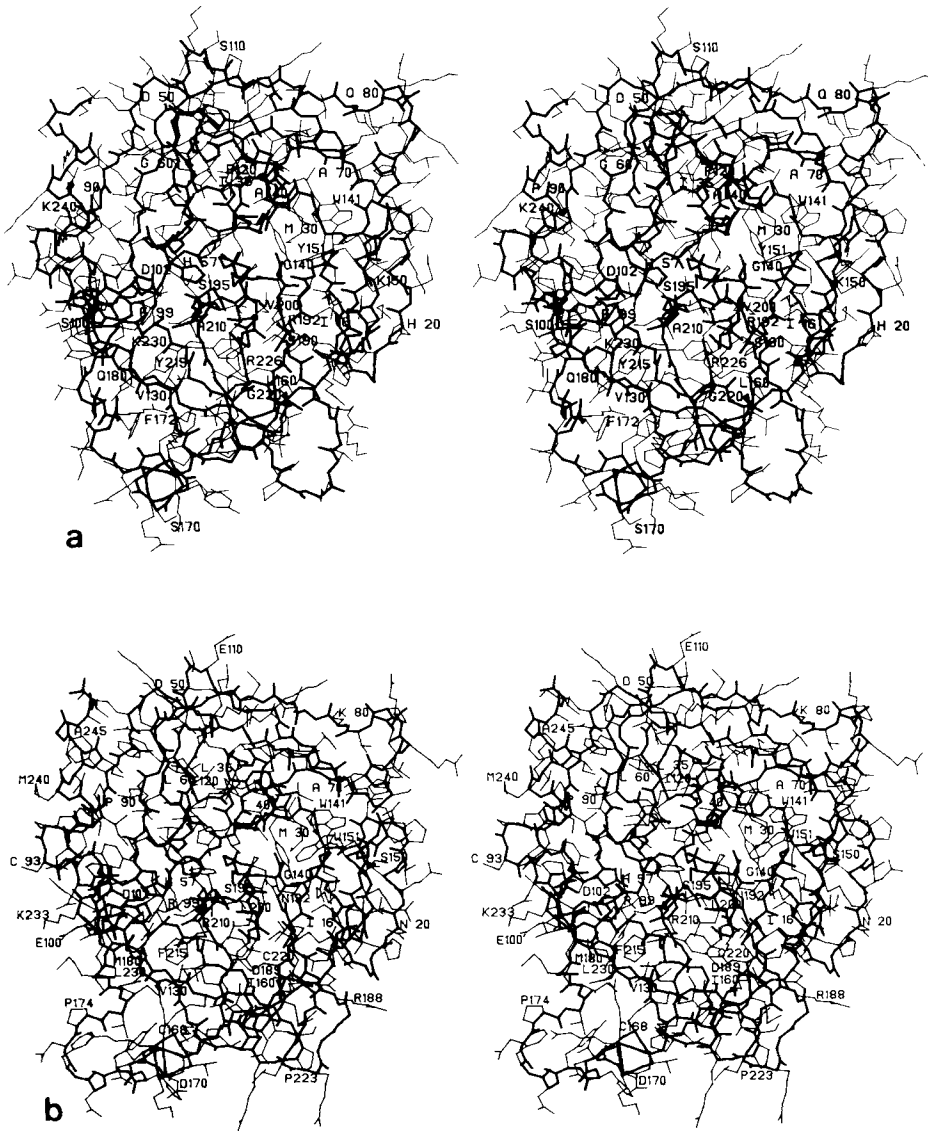

Fig. 4. **a:** Full atom representation of the molecular model for CCP1. Every tenth residue is labeled with the one letter code and sequence number. The view is approximately that of CCP1 in Figure 2. Thick lines highlight the polypeptide chain; thin lines represent the amino acid side chains. The side chains of His57, Asp102, Ser195, and Arg226 are also in thick lines. **b:** Final stereographic representation of the model of HF. Thick lines and thin lines have the same meaning as in a. The side chain of Asp189 is also represented by a thick line. Regions of this molecular model that required extensive manual manipulation and are thus less reliable are: Leu35 to Thr40, the four-residue insertion Tyr173A to Asn173D, and the insertion Asn218A.

A large change occurs in the conformation of the disulfide bridge. The procedure also rotates χ^2 of aromatic residues in the loop away from the most commonly observed value of 90°. This is most clearly seen for Tyr176 (Fig. 3). The plane of the aromatic ring has rotated by approximately 60° to accommodate the changed conformation of the Cys168–Cys182 bridge and the change of Ala177A to a lysine residue. The side chain of Lys177A is predicted to make a stabilizing hydrogen-bonded ion pair with Glu169 (Fig. 3).

The observation that the modelled loop regions moved the most during the energy minimization procedure would indicate that these regions of the model are the ones likely to deviate significantly from the true structure. However, the low energy parameters and the small changes in the core structure of the enzymes, which includes much of the active site region, suggest that the remainder of the models are likely to represent accurately the structures of CCP1 and HF.

A comparison of correctly and incorrectly folded proteins has also shown that correctly folded proteins have smaller areas of solvent-exposed nonpolar surfaces.[45] The solvent accessibilities for the two models and RMCP2 after energy minimization as well as those for some X-ray determined structures are given

TABLE V. Possible Ion-Pair Interactions in CCP1 and HF*

CCP1	HF
Ile16N†. . . .Asp194†	Ile16N†. . . .Asp194†
His20†. . . .Glu157	His25†. . . .Glu77†
His25†. . . .Glu77†	Arg27†. . . .Glu157
Arg27†. . . .Glu157	Arg38†. . . .Asp36
Lys36†. . . .Glu39	His57†. . . .Asp102†
His57†. . . .Asp102†	Arg63†. . . .Asp36
His71†. . . .Glu21	His71†. . . .Glu21
His91†. . . .Asp93	Lys86†. . . .Val2460
Lys107†. . . .Asp50†	Lys104†. . . .Glu88
Arg114†. . . .Glu49†	Lys111†. . . .Asp50†
Lys177A†. . . .Asp165	Arg156†. . . .Glu21
Lys177A†. . . .Glu169	Arg210†. . . .Asp128
Lys185C†. . . .Asp185	Arg210†. . . .Glu206
Arg188†. . . .Glu159	Lys232A†. . . .Asp129
Lys202†. . . .Asp134†	Lys233†. . . .Glu100

*Favorable electrostatic interactions exist in the present models, with the distances between the charged groups ranging from 2.5 to 4.0 Å.
†These residues are either buried or partially buried in the current models of CCP1 and HF.

in Table IV. The polar side chain atoms of the two models are considerably more exposed than in any of the X-ray structures, whereas the nonpolar atoms are approximately equally exposed. As a result, the total surface exposure of the models is higher than found for some of the crystal structures. The ratio of nonpolar to polar accessibility is lower than for some of the X-ray structures. The energy minimization procedure does not include solvent. In spite of this, some long polar side chains protrude directly out from the surface of the molecule (see Fig. 4) instead of lying compactly on the surface and reducing the total surface area.

Molecular Description of CCP1 and HF

Figure 4 contains a stereoview of the complete models of CCP1 (Fig. 4a) and HF (Fig. 4b) in approximately the same orientation as the C^{α} representation of CCP1 in Figure 2. Both models are slightly larger than that of RMCP2, with calculated molecular weights of 25 and 27 kD for CCP1 and HF, respectively, compared with 24 kD for RMCP2. Clearly, the models are similar to RMCP2, and, like all serine proteinases, the active site is located at the junction of two domains that form the surface cleft of the substrate binding site.

The structures of RMCP1 and RMCP2 are thought to have sites of lysyl clustering, whereas the arginine residues are spread over the entire molecules.[18] No such clustering of lysyl residues is observed in the models of HF or CCP1; however, the active site face of CCP1 contains a strip of basic residues on the bottom half of the molecule (in the orientation shown in Fig. 4a) and another strip of acidic residues on the top. This suggests specific electrostatic recognition sites on the surfaces of the natural substrates of these enzymes.

The internal hydrophobic surfaces between domains are highly homologous between RMCP2 and CCP1. The junction between the two β barrels has virtually the same contact surface residues in both enzymes. Also, the packing of the C-terminal α helix against the β sheet in the N-terminal domain is similar in both enzymes. The only major difference was a leucine for valine substitution at position 236 (Figs. 1, 4a).

Another indicator of the basic correctness of these molecular models is the environment of the charged residues. Indeed, the majority of the charged groups are on the molecular surfaces, with the side chains freely exposed to solvent. In addition, many of the side chains already form, or could be made to, form ion-pair interactions. These potential charge–charge interactions are listed in Table V. Some of the charged groups are buried or only partially exposed to solvent on the surfaces. In all cases there is a neighboring compensating charge for each of the buried residues in the model. The most unusual buried charge is Lys104 in HF. In most serine proteinases this is a hydrophobic residue (Fig. 1). However, in human HF the lysine-positive charge can be compensated by the carboxylate of Glu88 on a neighboring β strand. This glutamate is also particular to the human sequence (Fig. 1) and seems to be a compensatory difference specifically to accommodate the lysine at position 104. In CCP1, only six of the possible ion-pair interactions are conserved in the template RMCP2 structure. The other nine in CCP1 (Table V) are not possible in RMCP2 because of sequence differences in either one or both partners.

In Figure 5 the active sites of the X-ray structure of RMCP2 and the constrained energy minimized CCP1 model are given in stereo for comparison. In Figure 5b it can be seen that the S_1 pocket of the CCP1 model is partially filled by an Arg side chain resulting from the sequence difference at position Ala226 in RMCP2. This difference suggests that CCP1 would prefer an aspartate residue in the P_1 position of a substrate instead of a large hydrophobic P_1 residue as in RMCP2.[46,47] The high sequence homology between CCP1 and RMCP2 in this structurally conserved region around position 226 suggests that alternative sequence alignments or local refolding of the main chain are very unlikely. Therefore, the position for Arg226 seems to be the only one possible.

To explore other specific binding interactions between CCP1 and possible substrates, we have positioned the reactive site residues of the turkey ovomucoid third domain (OMTKY3), P_4 Ala to P_3' Arg, in the active site of CCP1. The structure of OMTKY3 in complex with SGPB and with CHT was used.[28,48] The inhibitor orientation was achieved by superimposing the active site residues of each enzyme with the active site residues of CCP1. The result is shown in Figure 5c. The sequence of residues P_4 to P_3' in OMTKY3 is Ala15-Cys-Thr-Leu-Glu-Tyr-Arg21. These have been changed to suitable residues that

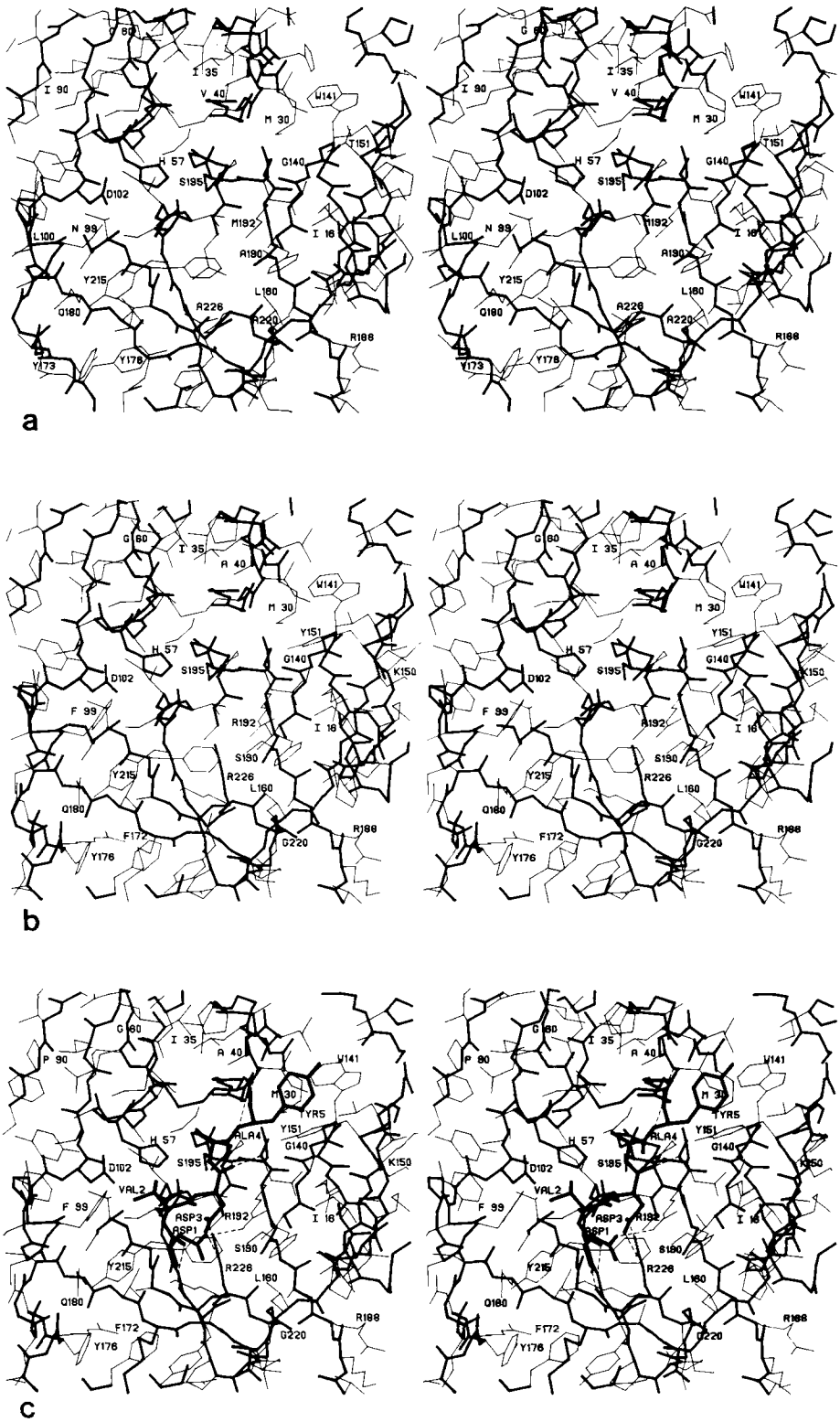

Fig. 5. **a:** A close-up stereoview of the active site of RMCP2. The atoms of the main chain are connected by thick lines representing the bonds, whereas side chain atoms are connected by thin lines. The active site residues His57, Asp102, Ser195, and Ala226 are represented by thick lines for emphasis. The residue numbering is that given in Figure 1. **b:** Close-up stereoview of the active site of CCP1 from the same vantage point as that of RMCP2 in a. The thick and thin lines have the same meaning as in a. The replacement Arg226 for Ala226 is evident in the central S_1 binding pocket of CCP1. **c:** Close-up stereoview of the active site of CCP1 with a possible substrate shown bound. The residues from P_3 to P_2' are represented by the thick lines. The presence of basic residues Arg175, Arg192, and Lys218 make it possible that P_4 and P_3 acidic residues would be favored in good substrates. The relatively large aromatic residue in S_2 (Phe99) limits the size of P_2 residues in substrates. An acidic aspartate residue is shown in the P_1 position of the possible substrate. There seems to be no particular preference for P_1' and P_2' residues, so they are shown as Ala and Tyr residues, respectively.

could reflect the proposed specificity of CCP1. It can be seen that hydrogen-bonding interactions that are well characterized between protein inhibitors of the serine proteinase and the enzymes[49] are entirely possible with this proposed binding mode to CCP1.

The S_2 site in CCP1 is altered from that in RMCP2 by a replacement of Asn99 by Phe99. This change could restrict the size of residues in the P_2 position of a substrate to small, preferably hydrophobic, side chains (Ala,Val). The side chains of residues in the P_3 position of a substrate of the serine proteinases points away from the enzyme surface. Therefore, almost any side chain would be suitable in this position. The presence of Arg192 and Lys218 could influence the enzyme specificity additionally toward a P_3 glutamate. To accommodate a residue at P_4, the conformation of the side chain of Arg175 would have to be changed.

In analyzing the possible interactions at the S_n' binding region of CCP1, it should be remembered that the conformation of the loop from Ile35 to Ile41 was modified from that in RMCP2. Therefore, these possible contacts should be viewed as tentative. There are two hydrophobic regions, one that includes Ile35, Ala40 and the disulfide bridge Cys42–Cys58; the other involves residues Leu32, Ile41, Ile73, Trp141, and Tyr171. The latter patch suggests possible binding sites for P_2', P_3' and P_4' hydrophobic residues. A similar but more extensive hydrophobic patch has been noted on the surface of tonin,[50] a kallikrein-like serine proteinase from rat submaxillary glands.

Two other serine proteinases contain unusual residues at position 226. In cathepsin G (CATG) there is a glutamate at 226, implying lysine specificity at the P_1 site. However, the P_1 specificity of CATG has been shown to be for large aromatic groups such as phenylalanine.[46,47] On the other hand, synthetic peptide substrates with a Phe at P_1 are about 15 times less reactive to CATG compared with CHT.[46] As in CCP1, human lymphocyte proteinase (HLP) also has an Arg at position 226.[18] There is another sequence change in HLP, that of the aromatic residue normally at position 228 found in most serine proteinases to a cysteine probably further altering the S_1 binding pocket.

Stereoplots of the active sites of the HF model (after constrained energy minimization) and the X-ray structure of BT are presented in Figure 6. The S_1 specificity pocket of HF has an insertion relative to BT of two residues after Ser217, which extends the entrance to the pocket. This proposed enzyme structure includes the residue Asp189 found in BT, suggesting Lys or Arg specificity. The S_1' site of HF derived from RMCP2 is similar to that of both CCP1 and RMCP2 with a hydrophobic patch. However, the loop between residues 35 and 41 contains a deletion of four residues relative to RMCP2. The S_2 site of HF is also derived from BT. It has an Arg at position 99 instead of a Leu as in BT, suggesting a possible P_2 specificity for an aspartate.

Several peptides have been synthesized and tested as substrates for granzymes B, G, and H by Masson and Tschopp[9] without success. The novel specificity predicted by the CCP1 model may be an aid in the logical design of new test substrates. The residues making up the S_1 pocket of HF are all similar in sequence to that of BT, including Asp189, and as expected the S_1 specificity of HF has been shown to be in favor of Arg or Lys side chains.[9,10]

The derived amino acid sequences of both CCP1[4] and HF[5] indicate the presence of free cysteine residues on the external β strand (Cys88 and Cys 93, respectively) of the N-terminal domain. Studies of the purified HF enzyme [10] indicate that homodimers of HF may form. We propose that dimerization occurs in the cysteine region of the two proteinases through formation of intermolecular disulfide bridges and extension of the β-sheet interactions in the case of CCP1.

Modelling dimer formation with HF is relatively easy. Cys93 is in position 3 of a type I 3_{10}-turn (Tyr91-Pro92-Cys93-Tyr94). The –SH group is fully exposed to solvent (Fig. 7a). The dimer was made by a 180° rotation of one HF molecule through an axis close to the Sγ atom. The intermolecular –S–S– distance was then adjusted to 2.1 Å and the C–S–S–C torsion angle adjusted to ~90°. Intermolecular contacts are not extensive in the HF dimer.

Although there is as yet no reported experimental evidence for CCP1 dimerization, we have also explored this possibility. To construct the CCP1 dimer, we used a strategy similar to that used for HF. Initially, the torsional angle, χ^1, of Cys88 was set to −60°. Then one copy of CCP1 was rotated by 180° relative to the other copy of CCP1 about an axis that ran approximately perpendicular to the strand from Val86 to Pro90 in the β sheet of the N-terminal domain. The two protein molecules were then translationally adjusted so that the intermolecular disulfide –S–S– distance was ~ 2.1 Å. This operation resulted in an intermolecular antiparallel β-sheet interaction for the strand Val86 to Pro90. Hydrogen bonding of the main chain at Cys88 and Cys88′ was thus possible in addition to the disulfide bridge formation. The presence of Pro90 in each strand prevents an extensive β-sheet interaction for the two strands. The second CCP1 molecule was then rotated by approximately 10° about an axis running through the β-sheet so as to achieve the commonly observed twist.

With these global movements in docking the two proteins, we observed an exceptional self-complementarity of the surfaces (Fig. 7b,c). There were only a few bad intermolecular nonbonded contacts that could be relieved by relatively small readjustments of individual side chains. The too-close contact of the side chains of Lys87 and Ile89′ could be obviated by adjusting Lys87 so that it made an intermolecular ion pair with the C-terminal carboxylate of Ser245. How-

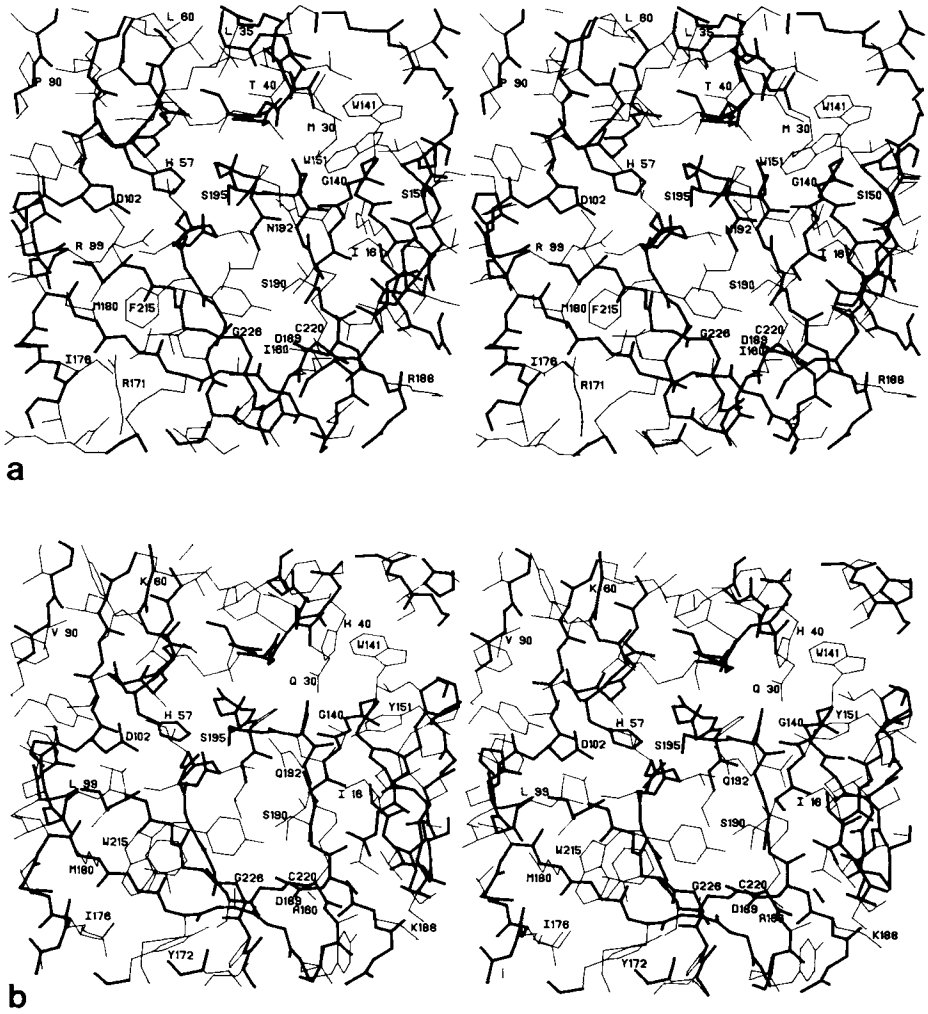

Fig. 6. **a:** Close-up stereoview of the active site of human HF. Thick and thin lines have the same designation as described in Figure 5a. The residue numbering is that for HF in Figure 1. The presence of Asp 189, homologous to the trypsin-like enzymes, suggests that HF has a specificity for P_1 lysine and arginine residues. There is a two-residue insertion, Glu218–Asn218A, in the active site binding loop from Phe215 to Cys220 relative to BT (Fig. 1). **b:** Similar orientation for a close-up view of the active site of BT to that shown for HF in a. As in the other figures, His57, Asp102, Ser195, and selected other residues have been highlighted by thick lines.

ever, to achieve this conformation, the side chain of Lys107 had to be adjusted to make a close intramolecular ion pair interaction with Asp50.

Another region required side chain adjustment to remove close contacts with twofold related Glu59 and Glu59′. These side chains and the carboxyl groups of Asp37 and Asp37′ are all closely interacting. Hydrogen bonding to these carboxyl side chains could come from Gln37A. This local concentration of negatively charged groups may not be unfavorable if they share hydrogen bonds in carboxyl-carboxylate interactions.[51]

In addition to the favorable electrostatic interaction Lys87 to Ser245′ COO^- at the dimer interface, there are two more possible ion pairs that are a consequence of this model for dimer formation in CCP1. These are Lys109 to Asp93′ and Lys97 to Glu39′. Since it is a symmetric dimer, there are six ion-pair interactions that stabilize this contact. The total surface area buried in forming this dimer is 650 $Å^2$ per monomer, a value comparable with the surface areas buried at subunit contacts in multimeric proteins.[52] The formation of this symmetric homodimer could have an influence on the substrate specificity of CCP1. The conformation of the extended β hairpin from Lys36 to Ala40 brings the side chain of Gln37A close to a possible binding site of P_1' or P_3' residues. In this way a more stringent substrate specificity for CCP1 may result from this proposed dimer formation. Alternatively, it may restrict the substrate specificity or binding capacity.

CONCLUSIONS

Molecular models of two hypothetic serine proteinases from cytotoxic T killer cells have been con-

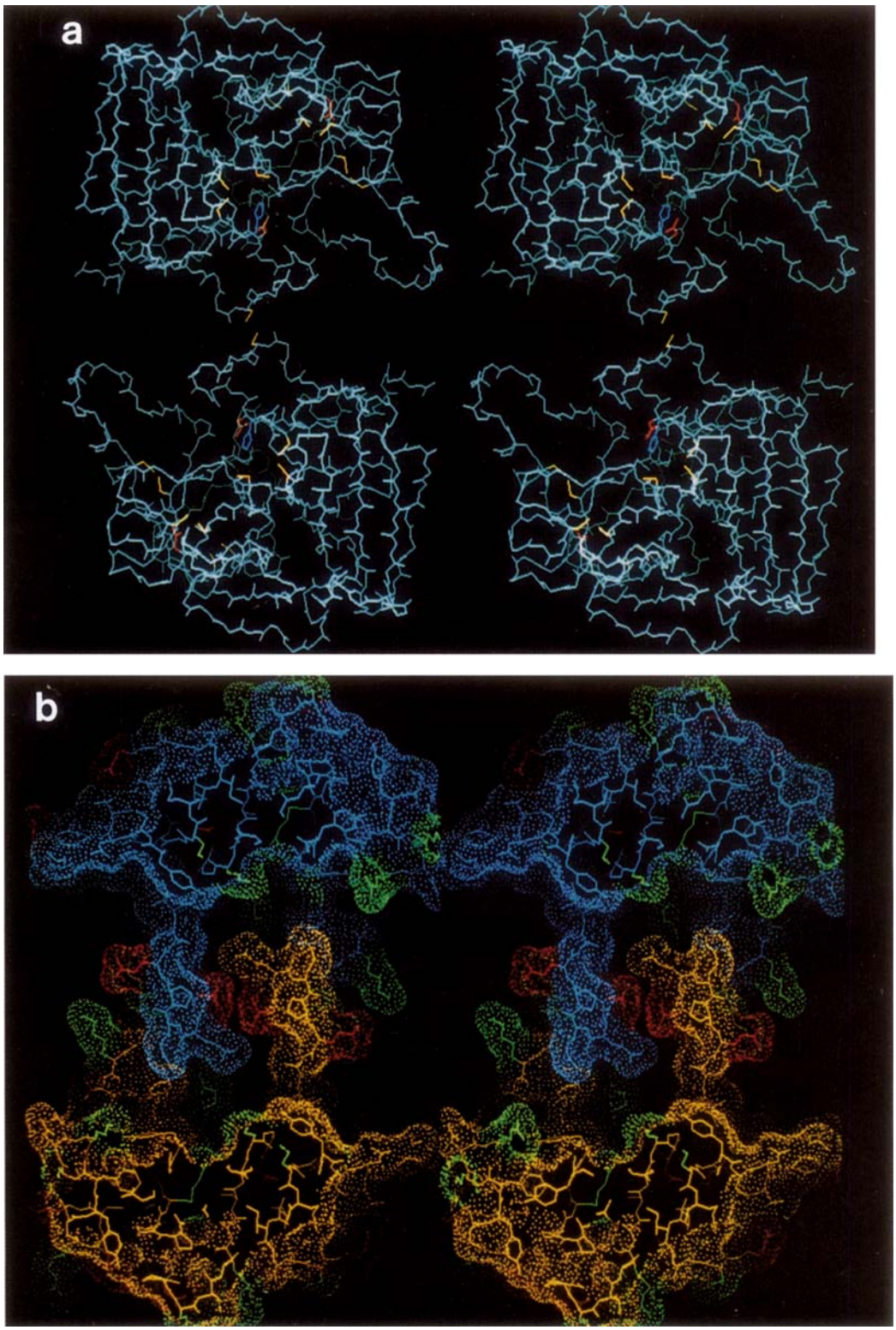

Fig. 7. **a:** A stereoview of the proposed dimer of HF. The intermolecular disulfide bridge Cys 93-Cys93′, can be seen (in yellow). The sulfur atoms (also yellow) of the several intramolecular disulfide bridges in HF are also not connected in this rendering. For each molecule, only the main chain atoms are represented, with the addition of the residues of the catalytic triad in color (His57 [blue], Asp102 [red], Ser195 [orange], and Asp189 [red]). **b:** A stereorepresentation of the self-complementarity of the model of CCP1 in the region of the proposed intermolecular contact that results from the disulfide bridge formation, Cys88–Cys88′. The twofold dimer axis is approximately perpendicular to the page, and the two extended β loops Lys36 to Ala40 are closely related by it and are seen near the center. The active site binding region is seen as a deep cleft on each monomer coming out of the page toward the viewer The molecular surfaces are represented by dots. The color coding in this and in c is as follows: The upper monomer is yellow, the lower is blue, acidic groups on CCP1 are red, and basic groups are green. **c:** A stereoview of a 15-Å-thick slab through the dimer interface of CCP1. The main chain in the region of Cys88–Cys88′ is seen in the central part. The strand extends from Pro84 to Pro92. One of the proposed stabilizing ion pair interactions is seen in the region of Pro92 (Asp93 [red] to Lys109′ [green]). This section of the dimer lies directly below that shown in b.

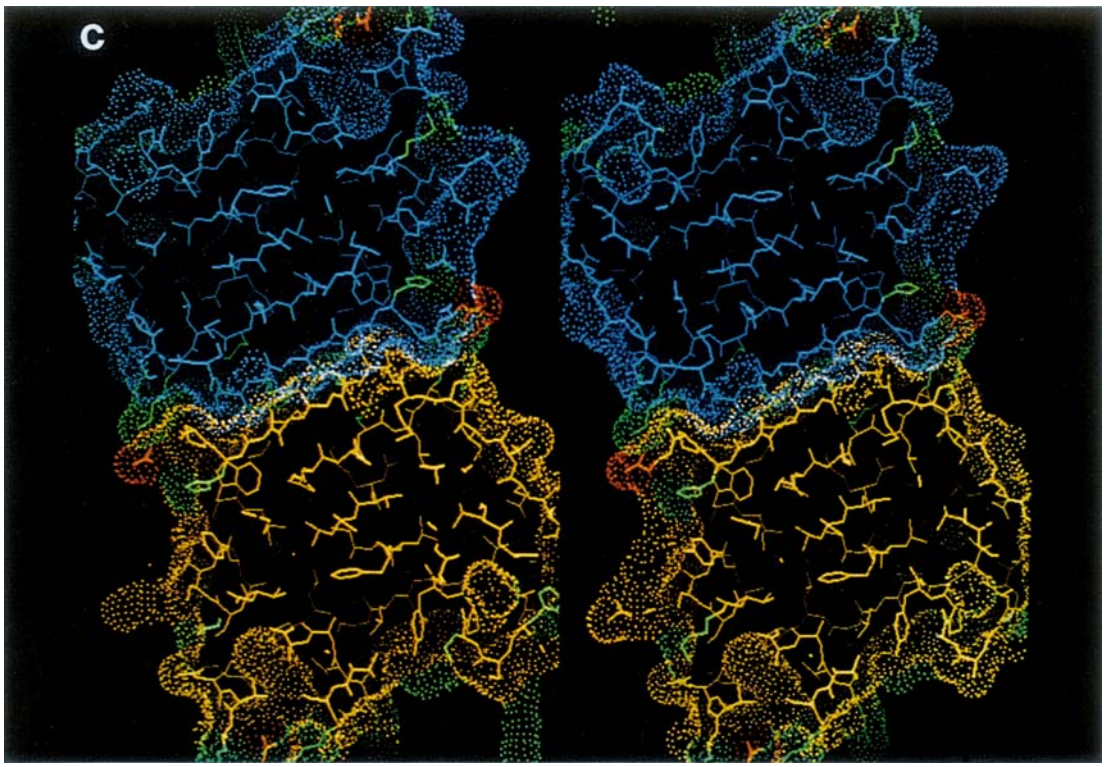

Figure 7 continued

structed from the cDNA sequences of CCP1 and human HF. These molecular models predict a tryptic specificity for HF and an unusual P_1 aspartic acid specificity for CCP1 on the basis of the conformations in the region of the major specificity pockets. Such data should be valuable in the search for the natural substrates of these serine proteinases.

ACKNOWLEDGMENTS

We thank Wilfred van Gunsteren for the use of GROMOS for energy minimization. Mae Wylie has done an excellent typing of the original manuscript. We thank Anthony Hawrylechko for his help in preparing the figures. This research was supported by the Medical Research Council of Canada, the Alberta Heritage Foundation for Medical Research (a scholarship to R.C.B. and a summer studentship to M.E.P.M.), and the NIH.

REFERENCES

1. Berke, G. Cytotoxic T-lymphocytes: How do they function? Immunol. Rev. 72:5–38, 1983.
2. Nabholtz, M., MacDonald, H.R. Cytolytic T-lymphocytes. Ann. Rev. Immunol. 1:273–306, 1983.
3. Lobe, C.G., Havele, C., Bleackley, R.C. Cloning of two genes which are specifically expressed in activated cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. U.S.A. 83:1448–1452, 1986.
4. Lobe, C.G., Finlay, B.B., Paranchych, W., Paetkau, V.H., Bleackley, R.C. Two cytotoxic T lymphocyte-specific genes encode unique serine proteases. Science 232:858–861, 1986.
5. Gershenfeld, H.K., Weissman, I.L. Cloning of cDNA for a cell-specific serine protease from a cytotoxic T lymphocyte. Science 232:854–858, 1986.
6. Brunet, J.F., Dosseto, M., Denizot, F., Mattei, M.G., Clark, W.R., Haqqi, T.M., Ferrier, P., Nabholz, M., Schmitt-Verhulst, A.M., Luciani, M.F., Golstein, P. The inducible cytotoxic T-lymphocyte-associated gene transcript CTLA-1 sequence and gene localization to mouse chromosome 14. Nature 322:268–271, 1986.
7. Redmond, M.J., Letellier, M., Parker, J.M.R., Lobe, C.G., Havele, C., Paetkau, V.H., Bleackley, R.C. A serine protease (CCP1) is sequestered in the cytoplasmic granules of cytotoxic T lymphocytes. J. Immunol. 139:3184–3189, 1987.
8. Tschopp, J., Nabholz, M. The role of cytoplasmic granule components in cytolytic lymphocyte-mediated cytolysis. Ann. Inst. Pasteur (Immunol.) 138:290–295, 1987.
9. Masson, D., Tschopp, J. A family of serine esterases in lytic granules of cytolytic T lymphocytes. Cell 49:679–685, 1987.
10. Pasternak, M.S., Verret, C.R., Liu, M.A., Eisen, H.N. Serine esterase in cytolytic T lymphocytes. Nature 322:740–743, 1986.
11. Greer, J. Comparative model-building of the mammalian serine proteases. J. Mol. Biol. 153:1027–1042, 1981.
12. Hartley, B.S. Homologies in serine proteinases. Phil. Trans. R. Soc. Lond. [Biol.] 257:77–87, 1970.
13. McLachlan, A.D., Shotton, D.M. Structural similarities between α-lytic protease of Myxobacter 495 and elastase. Nature 229:202–205, 1971.
14. Delbaere, L.T.J., Brayer, G.D., James, M.N.G. Comparison of the predicted model of α-lytic protease with the X-ray structure. Nature 279: 165–168, 1979.
15. James, M.N.G., Delbaere, L.T.J., Brayer, G.D. Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. Can. J. Biochem. 56:396–402, 1978.
16. Fujinaga, M., Delbaere, L.T.J., Brayer, G.D., James, M.N.G. Refined structure of α-lytic protease at 1.7 Å resolution: Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. 183:479–502, 1985.
17. Read, R.J., Brayer, G.D., Jurášek, L., James, M.N.G. A critical evaluation of comparative model-building of *Streptomyces griseus* trypsin. Biochemistry 23:6570–6575, 1984.
18. LeTrong, H., Parmelee, D.C., Walsh, K.A., Neurath, H., Woodbury, R.G. Amino acid sequence of rat mast cell pro-

tease 1 (Chymase). Biochemistry 26:6988–6994, 1987.
19. Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., Travis, J. Molecular cloning of human cathepsin G: Structural similarity to mast cell and cytotoxic T lymphocyte proteinases. Biochemistry 26:2289–2293, 1987.
20. Reynolds, R.A., Remington, S.J., Weaver, L.H., Fisher, R.G., Anderson, W.F., Ammon, H.L., Matthews, B.W. Structure of a serine protease from rat mast cells determined from twinned crystals by isomorphous and molecular replacement. Acta Cryst. B41:139–147, 1985.
21. Tsukada, H., Blow, D.M. Structure of α-chymotrypsin refined at 1.68 Å resolution. J. Mol. Biol. 184:703–711, 1985.
22. Blevins, R.A., Tulinsky, A. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. J. Biol. Chem. 260:4264–4275, 1985.
23. Sawyer, L., Shotton, D.M., Campbell, J.W., Wendel, P.L., Muirhead, H., Watson, H.C., Diamond, R., Ladner, R.C. The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution. Comparisons with the structure of α-chymotrypsin. J. Mol. Biol. 118:137–208, 1978.
24. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Diesenhofer, J., Steigemann, W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol. 89:73–101, 1974.
25. Chambers, J.L., Stroud, R.M. The accuracy of refined protein structures: Comparison of two independently refined models of bovine trypsin. Acta Cryst. B35:1861–1874, 1979.
26. Read, R.J., James, M.N.G. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. J. Mol. Biol. 200:523–551, 1988.
27. Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., Bartunik, H. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. J. Mol. Biol. 164:237–282, 1983.
28. Fujinaga, M., Sielecki, A.R., Read, R.J., Ardelt, W., Laskowski, M. Jr., James, M.N.G. The crystal and molecular structures of the complex of turkey ovomucoid inhibitor third domain with α-chymotrypsin at 1.8 Å resolution. J. Mol. Biol. 195:397–418, 1987.
29. Strynadka, N.C.J., James, M.N.G. Two trifluoperazine binding sites on calmodulin predicted from comparative molecular modelling with troponin-C. Proteins Struct. Function Genet. 3:1–17, 1988.
30. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasmui, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
31. MacLachlan, A.D. Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome C_{551}. J. Mol. Biol. 61:409–424, 1971.
32. Bhat, T.N., Sasisekharan, V., Vijayan, M. An analysis of side chain conformations in proteins. Int. J. Peptide Protein Res. 13: 170–184, 1978.
33. Janin, J., Wodak, S., Levitt, M., Maigret, B. Conformation of amino-acid side chains in proteins. J. Mol. Biol. 125:357–386, 1978.
34. James, M.N.G., Sielecki, A.R. Structure and refinement of penicillopepsin at 1.8Å resolution. J. Mol. Biol. 163:299–361, 1983.
35. Moult, J., James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins Struct. Function Genet. 1:146–163, 1986.
36. Sielecki, A.R., James, M.N.G., Broughton, C.G. Fourier and graphical methods in the refinement of macromolecules. In: Sayre, D. (ed.): "Computational Crystallography." Oxford: Clarendon Press, 409–419, 1982.
37. Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., Levinthal, C. Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of MCP603 from many randomly generated loop conformations. Proteins Struct. Function Genet. 1:342–362, 1986.
38. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–822, 1986.
39. van Gunsteren, W.F., Berendsen, H.J.C. (distributors). Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.
40. Herzberg, O., Moult, J., James, M.N.G. A model for the Ca^{2+} induced conformational transition of troponin C: A trigger for muscle contraction. J. Biol. Chem. 261:2638–2644, 1986.
41. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
42. Dayhoff, M.O. (ed.). "Atlas of Protein Structure." Washington, DC: National Biomedical Research Foundation, Vol. 5, 1979:1.
43. Dayhoff, M.O., Barker, W.C., Hunt, L.T. Establishing homologies in protein sequences. Methods Enzymol. 91:524–545, 1983.
44. Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823–826, 1986.
45. Novotny, J., Bruccoleri, R., Karplus, M. An analysis of incorrectly folded protein models. Implications for structure predictions. J. Mol. Biol. 177:787–818, 1984.
46. Powers, J.C., Tanaka, T., Harper, J.W., Minematsu, Y., Barker, L., Lincoln, D., Crumely, K.V., Fraki, J.E., Schechter, N.M., Lazarus, G.G., Nakajima, K., Nakashino, K., Neurath, H., Woodbury, R.G. Mammalian chymotrypsin-like enzymes. Comparative reactivities of rat mast cell proteases, human and dog skin chymases, and human cathepsin G with peptide 4-nitroanilide substrates and with peptide chloromethyl ketone and sulfonyl fluoride inhibitors. Biochemistry 24:2048–2058, 1985.
47. Harper, J.W., Cook, R.R., Roberts, J.C., McLaughlin, B.J., Powers, J.C. Active site mapping of the serine proteases human leukocyte elastase, cathepsin G, porcine pancreatic elastase, rat mast cell proteases I and II, bovine chymotrypsin A_α, and *Staphylococcus aureus* protease V-8 using tripeptide thiobenzyl ester substrates. Biochemistry 23:2995–3002, 1984.
48. Read, R.J., Fujinaga, M., Sielecki, A.R., James, M.N.G. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. Biochemistry 22:4420–4433, 1983.
49. Read, R.J., James, M.N.G. Introduction to the protein inhibitors: X-ray crystallography. In: Barrett, A.J., Salvesen, G. (eds.): "Proteinase Inhibitors." Amsterdam: Elsevier, 1986:301–336.
50. Fujinaga, M., James, M.N.G. Rat submaxillary gland serine protease, tonin: Structure solution and refinement at 1.8 Å resolution. J. Mol. Biol. 195:373–396, 1987.
51. Sawyer, L., James, M.N.G. Carboxyl-carboxylate interactions in proteins. Nature 295:79–80, 1982.
52. Miller, S., Lesk, A.M., Janin, J., Chothia, C. The accessible surface area and stability of oligomeric proteins. Nature 328:834–836, 1987.