JOURNAL OF VIROLOGY, Mar. 1997, p. 2436–2448 Vol. 71, No. 3
0022-538X/97/$04.00+0

The Refined Crystal Structure of the 3C Gene Product from Hepatitis A Virus: Specific Proteinase Activity and RNA Recognition

ERNST M. BERGMANN,[1,2] STEVEN C. MOSIMANN,[1,2] MAIA M. CHERNAIA,[1,2] BRUCE A. MALCOLM,[3] AND MICHAEL N. G. JAMES[1,2]*

Department of Biochemistry,[1] Medical Research Council of Canada Group in Protein Structure & Function,[2] and Department of Medical Microbiology and Immunology,[3] University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Received 5 August 1996/Accepted 11 November 1996

The virally encoded 3C proteinases of picornaviruses process the polyprotein produced by the translation of polycistronic viral mRNA. The X-ray crystallographic structure of a catalytically active mutant of the hepatitis A virus (HAV) 3C proteinase (C24S) has been determined. Crystals of this mutant of HAV 3C are triclinic with unit cell dimensions $a = 53.6$ Å, $b = 53.5$ Å, $c = 53.2$ Å, $\alpha = 99.1°$, $\beta = 129.0°$, and $\gamma = 103.3°$. There are two molecules of HAV 3C in the unit cell of this crystal form. The structure has been refined to an R factor of 0.211 ($R_{free} = 0.265$) at 2.0-Å resolution. Both molecules fold into the characteristic two-domain structure of the chymotrypsin-like serine proteinases. The active-site and substrate-binding regions are located in a surface groove between the two β-barrel domains. The catalytic Cys 172 S^{γ} and His 44 $N^{\epsilon 2}$ are separated by 3.9 Å; the oxyanion hole adopts the same conformation as that seen in the serine proteinases. The side chain of Asp 84, the residue expected to form the third member of the catalytic triad, is pointed away from the side chain of His 44 and is locked in an ion pair interaction with the ε-amino group of Lys 202. A water molecule is hydrogen bonded to His 44 $N^{\delta 1}$. The side-chain phenolic hydroxyl group of Tyr 143 is close to this water and to His 44 $N^{\delta 1}$ and may be negatively charged. The glutamine specificity for P_1 residues of substrate cleavage sites is attributed to the presence of a highly conserved His 191 in the S_1 pocket. A very unusual environment of two water molecules and a buried glutamate contribute to the imidazole tautomer believed to be important in the P_1 specificity. HAV 3C proteinase has the conserved RNA recognition sequence KFRDI located in the interdomain connection loop on the side of the molecule diametrically opposite the proteolytic site. This segment of polypeptide is located between the N- and C-terminal helices, and its conformation results in the formation of a well-defined surface with a strongly charged electrostatic potential. Presumably, this surface of HAV 3C participates in the recognition of the 5′ and 3′ nontranslated regions of the RNA genome during viral replication.

Once released into the cytoplasm of a host cell, the single-stranded genome of a picornavirus acts directly as an mRNA and is translated into a single large polyprotein. Proteolytic cleavages of the polyprotein are then necessary to generate the individual structural and functional proteins of the replicating virus (14, 34). There are no suitable host-derived proteolytic activities in the cellular cytoplasm to effect the precise cleavages of the picornaviral polyprotein. The 3C element of the polyprotein does, however, constitute such a proteolytic activity in all picornaviruses (50). 3C is a relatively slow proteinase compared to extracellular mammalian proteinases such as trypsin and chymotrypsin. This is in keeping with the requirement that a virus could not afford to unleash a very active proteinase on the cellular, cytosolic proteins which are required for its own replication. The 3C proteinase is also quite specific for the appropriate cleavage sites within the viral polyprotein (34).

In hepatitis A virus (HAV), the 3C gene product appears to be the only proteolytic activity present. It processes all cleavage sites of the viral polyprotein (18, 23, 30, 41, 61, 62, 64). Because the polyprotein processing requires 10 catalytic cleavages at specific sites, effective inhibition of the 3C proteinase activity prevents viral maturation (25, 46). Therefore, the 3C proteinase activity has long been identified as an obvious target for structure-based drug design of antiviral compounds (34).

As is common for the gene products of small RNA viruses, 3C has functions in addition to its proteolytic activity. A sequence motif, KFRDI, conserved among the 3C proteinases of all picornaviruses (20), has been implicated in the recognition of the 5′ and 3′ termini of the RNA genome (21, 24). There is strong evidence that in some picornavirus systems, the recognition of the 5′ and 3′ nontranslated region (NTR) of the viral RNA by protein complexes, which include 3C, is important for the initiation and regulation of RNA replication (2, 24). No investigations of the RNA binding activity of 3C have been carried out in the HAV system; however, this region of the 3C sequence is the most highly conserved in all picornaviruses.

Site-directed mutagenesis studies identified the picornavirus 3C proteinase as a cysteine proteinase (27). Predictions of the polypeptide fold did, however, suggest structural similarity to serine proteinases of the chymotrypsin family (4, 19).

HAV 3C proteinase cleaves the polyprotein between a glutamine residue and a small residue such as alanine, serine, or glycine (54). Residues from at least P_4 to $P_{2'}$ (in the nomenclature of Schechter and Berger [60]) contribute to the cleavage specificity, but the most important determinant is the glutane residue in the P_1 position of a substrate (29). HAV 3C can utilize asparagine or glutamate residues in this position as a

* Corresponding author. Phone: (403) 492-4550. Fax: (403) 492-0886. E-mail: michael.james@ualberta.ca.

substrate, but it does so very poorly, in the context of either a peptide substrate (38) or the polyprotein (41). Thus, the 3C proteinase is able to distinguish a glutamine residue electrostatically from glutamate and sterically from asparagine. Gorbalenya et al. (20) predicted that the major determinant of substrate specificity among the 3C proteinases would be a conserved His 191, located in the S_1-binding pocket of the enzyme.

The first crystal structure obtained for a 3C proteinase was that of an inactive double mutant of the enzyme from HAV (1). This protein (C24S, C172A) had both cysteines in the sequence mutated to facilitate crystallization. It had an overall fold and arrangement of the active site which were very similar to those of the serine proteinases of the chymotrypsin family but which was lacking the postulated third member of the catalytic triad. In addition, the conformation of the electrophilic oxyanion hole in the active site was distorted from that found in chymotrypsin-like serine proteinase structures. Allaire et al. attributed this distortion to the alanine substitution of the catalytic cysteine nucleophile (Cys 172).

Shortly after the crystal structure of HAV 3C was reported, the structure of the 3C proteinase from human rhinovirus type 14 (HRV14) was also published (42). The 3C proteinase from rhinovirus belongs to another subclass of the viral chymotrypsin-like proteinases. It is smaller than the HAV enzyme (182 versus 219 residues). The overall fold and active-site structure of the HRV14 3C proteinase, once again, show similarities to the chymotrypsin family of serine proteinases, as was predicted. A third member of the postulated catalytic triad, glutamate 71, was found interacting with the general acid-base catalyst His 40 in the crystal structure of the HRV14 3C proteinase. The conformation of Glu 71 in the crystal structure of Matthews et al. (42) is unusual. Glu 71 of HRV 14 3C interacts with the $N^{\delta 1}$ of the general acid-base catalyst His 40 through one of the anti-orbitals of the carboxylate group. Whether a functional catalytic triad is required in these cysteine proteinases remains unclear.

We now present the refined crystal structure of an active, single mutant of the 3C proteinase from HAV [HAV 3C (C24S)] (Fig. 1). This mutant retains the catalytic nucleophile Cys 172 in the active site, and its proteolytic activity, at least in vitro, is indistinguishable from that of the wild-type enzyme (38).

MATERIALS AND METHODS

The C24S mutant of HAV 3C was purified from an overexpression system as described previously (40). The HAV 3C protein from the bacterial expression system consists of residues 1 to 217. The last two residues at the carboxy terminus are missing in the expressed HAV 3C. Crystals were grown by the hanging-drop vapor diffusion method. Hanging drops were prepared from 2 μl of the protein (10 mg/ml) in Tris buffer and 10 μl of the reservoir solution consisting of 20% polyethylene glycol (PEG)-monoethyl ether 2000, 1.0 M LiCl, 3 mM β-mercaptoethanol, and 0.1 M Tris-HCl (pH 8.5). Crystals grew slowly over a period of 4 to 5 weeks to a maximum size of about 0.2 by 0.3 by 0.8 mm. Macroseeding was routinely used to improve the size of the crystals.

Intensity data were collected from a single crystal of HAV 3C over a period of 2 days on an SDMS twin-area detector system (67) with graphite-monochromated Cu K_α radiation generated by a Rigaku RTP 200 X-ray generator (40 kV and 150 mA). Data were scaled and analyzed with the area detector software and programs of the CCP4 suite of crystallographic software (13).

The structure was solved by the method of molecular replacement (59) with the software package AMoRe (47). The search model was a partially refined structure of the inactive double mutant (C24S, C172A) of HAV 3C of Allaire et al. (1). Crystallographic refinement was carried out with the program X-Plor version 3.1. (7) and interspersed with cycles of manual inspection and rebuilding. Crystallographic refinement consisted of cycles of simulated annealing (9) followed by conjugate gradient minimization of atomic coordinates and isotropic temperature factors.

All rebuilding and visualization was performed with the program O (31, 32) running on a Silicon Graphics Indigo^2 Extreme workstation. The progress of the refinement was monitored with the aid of the free *R*-factor (8), the residual real-space correlation coefficient (6), refined isomorphous *B*-factors, and the fit into the features of electron density maps calculated from coefficients $2|F_O| - |F_C|, 4|F_O| - 3|F_C|$, and $|F_O| - |F_C|$, where $|F_O|$ and $|F_C|$ are the observed and calculated structure factor amplitudes, respectively. The first two rounds of refinement used all observed reflections with 20.0- to 2.3-Å resolution and $|F_O| > 2.0\sigma(|F_O|)$, and later rounds used all observed reflections with 20.0- to 2.0-Å resolution and $|F_O| > 2.0\sigma(|F_O|)$.

At the same time, we were given the opportunity to try an experimental refinement procedure implemented within X-Plor (52). The new program allows the use of a likelihood function as the target function in crystallographic refinement. Throughout the rest of the crystallographic refinement, we performed a parallel refinement with the crystallographic residual as well as the likelihood target (data not shown). After initial tests which convinced us of the superiority of the likelihood target function, we used only the model from this refinement for manual rebuilding but continued the parallel refinement throughout. Three observations convinced us of the superiority of the likelihood target: (i) the refinement converged to a minimum with fewer cycles, (ii) the discrepancy between the free *R*-factor and the conventional *R*-factor was significantly diminished, and (iii) even the conventional *R*-factor faithfully recorded attempts to overfit the structure.

The resulting structure was analyzed with the programs Procheck (35) and Promotif (26). The stereochemical parameter set of Engh and Huber (15) was used throughout the refinement and in all visualization and analysis programs.

RESULTS AND DISCUSSION

Crystallization and data collection. Crystals of HAV 3C are triclinic, space group P1 with two molecules in the unit cell. X-ray diffraction data were collected from a single crystal at room temperature. Crystal parameters and the statistics of the diffraction data are presented in Table 1. Figure 2 shows a detailed analysis of the intensity and completeness of the data. The crystals diffract to a maximum resolution of 2.0 Å, but the intensities become weak at resolutions higher than 2.3 Å. Only 20% of reflections in the highest-resolution shell were deemed measurable at a level of 2.0σ on the structure factor amplitudes.

Structure solution and refinement. The structure was solved by molecular replacement with the inactive double mutant (C24S, C172A) described by Allaire et al. (1) as the search model. The two top peaks of the rotation search results yielded a clear solution. The center of mass of molecule A, corresponding to the top peak in the rotation function, was chosen and fixed as the origin of the triclinic unit cell. After an initial rigid-body refinement, the model was extensively rebuilt. At this point, it was possible to place the side chains that were missing in the model of Allaire et al. (1). Several rounds of crystallographic refinement and manual rebuilding, with inclusion of solvent and a single chloride ion in the last cycles, yielded the statistics of the final refined model shown in Table 2. This model consists of residues 1 to 216 in molecule A, 1 to 214 in molecule B, 1 chloride ion, and 107 water molecules. One and three residues at the carboxy terminus of molecules A and B, respectively, are disordered in our structure.

Quality of the refined structure. Figure 3 shows the refined temperature factors of the main chain of the two molecules and the residual real-space correlation coefficient (6). Representative electron density is shown in Fig. 4. The structure is well defined throughout, with the exception of a flexible or disordered region consisting of residues 147 to 151 in both molecules. This region is poorly defined in the electron density maps, and the refined temperature factors are very high (Fig. 3). Because it is also the region of the structure with the most solvent-accessible surface (data not shown), it is assumed to be flexible. Three residues are flagged as outliers in the Ramachandran plot of the refined structure: Asp A36, Asp B36, and Asp A149. Residue 36 constitutes the $i + 1$ residue of a type II reverse turn ($\varepsilon\alpha_R$ in the Ramachandran nomenclature of Wilmot and Thornton [65]). This position of a type II re-

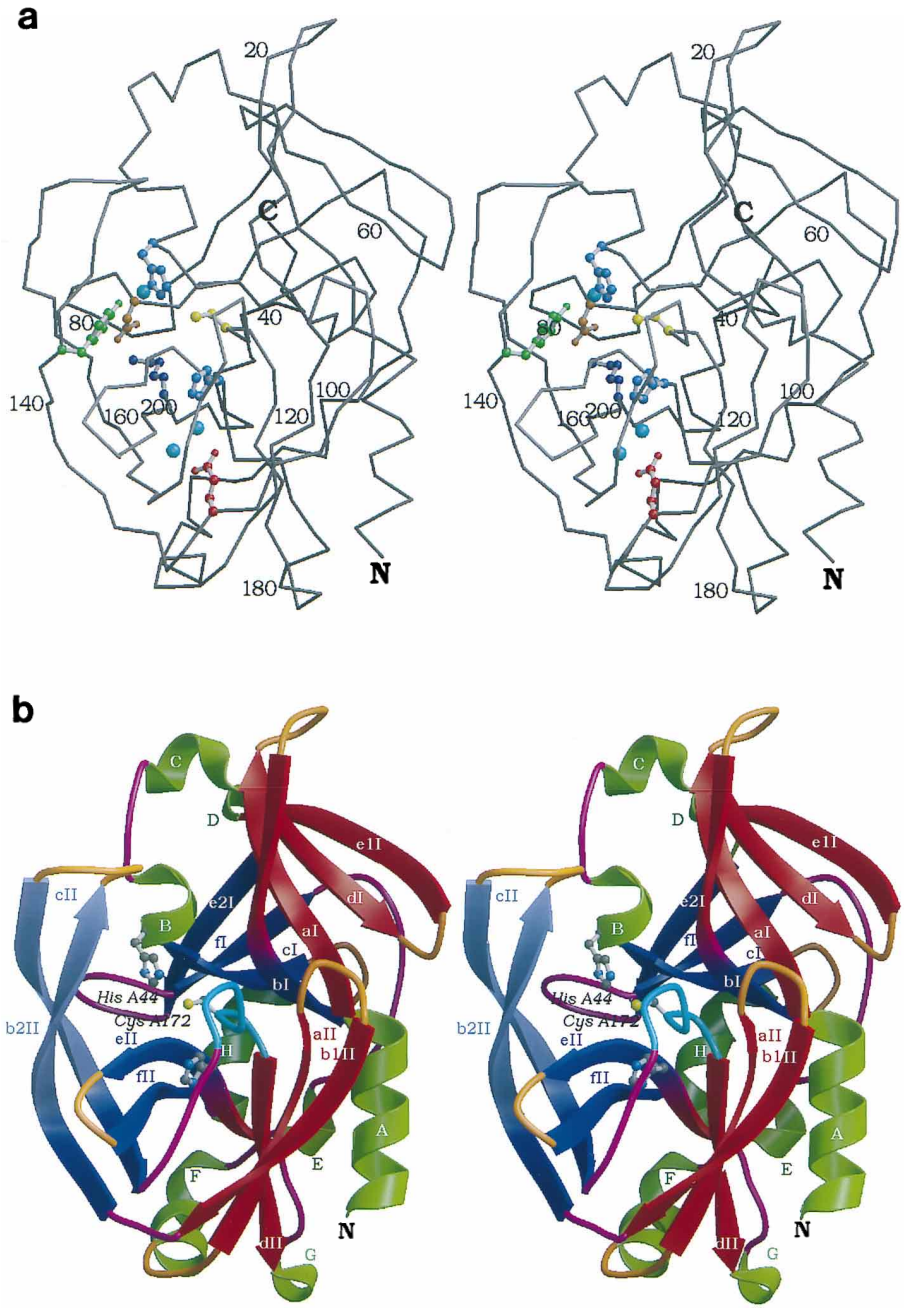

FIG. 1. (a) Stereo view of the C^{α} coordinates of molecule A of HAV 3C proteinase. Every 20th residue is labeled. The amino and carboxy termini are labeled N and C, respectively. The nucleophile Cys A172 is included in a ball-and-stick representation and colored yellow. Similarly represented are the general acid-base catalyst His A44 and the specificity determinant in the S_1 pocket of the enzyme His A191 in light blue, Asp A84 and Glu A132 in red, Tyr A143 in green, and Lys A202 in dark blue. One water molecule, hydrogen bonded to the imidazole side chain of His A44, and two water molecules, trapped between His A191 and Glu A132, are represented by spheres colored in cyan blue. The vantage point is into the proteolytic active site. This figure, as well as panel b and Fig. 10a, was prepared with Robert Esnouf's version of Molscript (33) and Raster-3D (3, 43). (b) Stereo view of the secondary-structure features of molecule A of HAV 3C proteinase from the same viewpoint as in panel a. The elements of the secondary structure were determined with the program Promotif (26). The residue ranges for helices, β-sheet structure, and reverse turns are listed in Table 3. Helices (α and 3_{10}) are represented by green ribbons, the β-sheets of the two β-barrel domains are represented by red and blue arrows, and the isolated antiparallel β-ribbon (Lys A140 to Asp A158) is indicated by lilac arrows. Reverse turns are colored yellow, and the random-coil structure is in magenta. The segment of backbone corresponding to the oxyanion hole in the active site is light blue. The side chains of His A44, Cys A172, and His A191 are included in a ball-and-stick representation.

verse turn usually requires a glycine, but Asp 36 is well defined in the electron density (Fig. 4a) and also forms part of the proposed RNA recognition site (see below). Presumably, the charge of the side chain of Asp 36 is so important for the RNA recognition site that the unfavorable main-chain conformation is tolerated by the structure. Asp A149 belongs to the same region (residues 147 to 151) that is indicated to be flexible by high-temperature factors. It is also poorly defined in the electron density maps as documented by the low real-space correlation coefficients (Fig. 3). Therefore, the conformation of

TABLE 1. Crystal parameters and statistics of the diffraction data

Parameter	Value[a]
Space group	P1 (2 molecules/unit cell)
Unit cell	a = 53.6 Å, b = 53.5 Å, c = 53.2 Å, α = 99.1°, β = 129.0°, γ = 103.3°
Resolution range	20.0–2.0 Å (2.15–2.0 Å)
Observed reflections	64,410 (5,802)
Unique reflections	25,299 (3,841)
Completeness	86.6% (67.2%)
Avg intensity	9.1σ (1.4σ)
R_{merge}[b]	6.73% (27.13%)

[a] Numbers in parentheses are for the highest-resolution shell, 2.15 to 2.0 Å.
[b] $R = \Sigma_{hkl}(\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma_i I_i)$.

TABLE 2. Refinement statistics

Parameter	Value
No. of reflections used	20,059 ($\mid F_O \mid > 2.0\sigma$ and 20.0- to 2.0-Å resolution)
No. of nonhydrogen atoms	3,430 (107 water molecules and 1 chloride ion)
No. of degrees of freedom	13,173
R-factor[a]	21.1%
Free R-factor[b]	26.5%
Root mean square deviations from ideality	
Bonds	0.011 Å
Bond angles	1.79°
Improper dihedral angles	1.64°
No. of Φ and Ψ angle outliers[c]	3 (of 366 non-Pro and non-Gly residues) (Asp A36, Asp B36, and Asp A149)

[a] $R = \Sigma(\|F_O\| - \|F_C\|)/\Sigma \mid F_O \mid$ for the reflections used in refinement.
[b] For 1,958 reflections not used in the refinement (10%).
[c] In the "disallowed" and "generous" regions defined by Morris et al. (45).

residues 147 to 151, in both molecules of the refined crystal structure of HAV 3C, is not very dependable.

Differences between molecules A and B. The two molecules in the triclinic unit cell of the HAV 3C crystal structure are remarkably different. Refined main-chain temperature factors are shown in Fig. 3, and the differences in atomic coordinates of the two superimposed molecules are shown in Fig. 5. Early in the refinement process, we noticed difference electron density in the active site. The difference electron density in the active site of molecule B could be interpreted as a cysteic acid, and eventually Cys B172 was refined as such. Whether the oxidation state of Cys B172 is truly $-SO_3^-$ or an ensemble of conformations of $-SO_2^-$ or even a mixture of the above cannot be decided with the available diffraction data (Fig. 4b). As a result of the oxidation of molecule B, the side chain of Tyr B143 is rotated about 120° out of the active site and Lys B147 closes over the active site to form a salt bridge with the cysteic acid B172. Residual difference density in the region of the active site of molecule A could not be interpreted or refined successfully as either oxidation or a covalent modification (e.g., a dithiothreitol adduct) of Cys 172. The difference density has therefore been modeled and refined as ordered solvent. All parts of the structure that are significantly different in conformation and refined temperature factors, except Cys 172 but including Tyr 143 and Lys 147, form crystal-packing contacts in one or the other of the two molecules in the unit cell. Tyr B143 packs against the Tyr A65 at the amino terminus of β-eI in another molecule in an adjacent unit cell. Therefore, we suggest that crystal growth selected reduced molecules for position A in the unit cell and oxidized molecules for B. We consider molecule A to resemble the enzymatically active, reduced form of the enzyme.

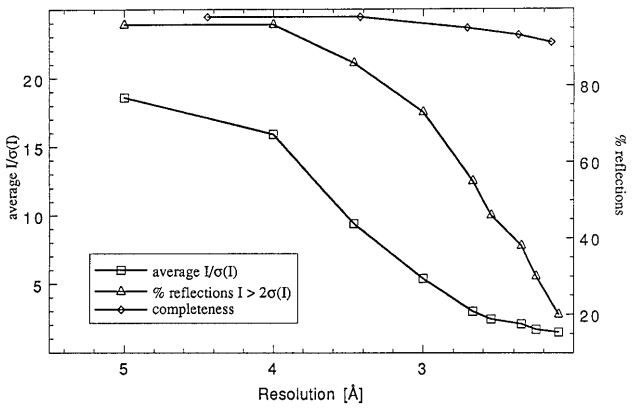

FIG. 2. Plot of the completeness of the data (right y axis) and the intensity distribution, $I/\sigma(I)$ (left y axis), as a function of the resolution for the X-ray diffraction data measured from the triclinic crystals of HAV 3C.

Residues with the largest coordinate differences between the two molecules are clustered around Phe 82. The loop preceding β-strand β-fI from Gly A80 to Gln A83 forms another crystal contact between unit cells and has relatively low temperature factors ($B_{iso} < 20$ Å^2). The corresponding residues in molecule B have high temperature factors ($B_{iso} > 30$ Å^2) and are accessible to solvent and presumably flexible. The largest differences between molecules A and B in coordinates and refined temperature factors (Fig. 3 and 5) can all be rationalized with crystal-packing contacts. The conformation of those residues that display different conformations in the two molecules A and B (Fig. 5b) and are involved in crystal packing contacts may simply be a result of crystal-packing forces.

Domain structure. The overall domain structure is formed by two antiparallel β-barrel domains, similar to all chymotrypsin-like proteinases (53). They are best described as a barrel formed by the stacking together of two orthogonal, four-stranded β-sheets (11). The four-stranded β-sheets of the two β-barrels in HAV 3C are colored red and blue in Fig. 1b. The edge strands of the two β-sheets (β-strands b and e) participate in both sheets, and as a result two corners of the double-sheeted β-barrel are closed. In the amino-terminal domain (I), strand β-bI continues uninterrupted from one sheet to the other whereas strand β-eI is interrupted by the single turn of helix D. In the carboxy-terminal domain (II), however, strand β-eII is continuous from one β-sheet to the second. The secondary-structure features of molecule A of HAV 3C are summarized in Table 3. Characteristic features of HAV 3C are the helices at the amino and carboxy termini of the molecule and a long, antiparallel β-ribbon extending out from one of the β-sheets of domain II (colored lilac in Fig. 1b).

β-strands β-b2II and β-cII continue out of one of the β-sheets of the C-terminal domain as a long, antiparallel β-ribbon (58) which wraps around the active site (Fig. 1). An antiparallel, bent β-bulge (26) introduces an almost 90° kink into this ribbon at residues 139 and 159. The residues of β-strand β-b2II up to residue 138 and from residue 160 on, in strand β-cII, form part of one of the β-sheets of domain II (colored blue in Fig. 1b). The tip of this ribbon (residues 147 to 151) is the most flexible and solvent-exposed part of the structure (Fig. 3). These residues are also the most poorly defined part of the crystal structure. This long, antiparallel β-ribbon, form-

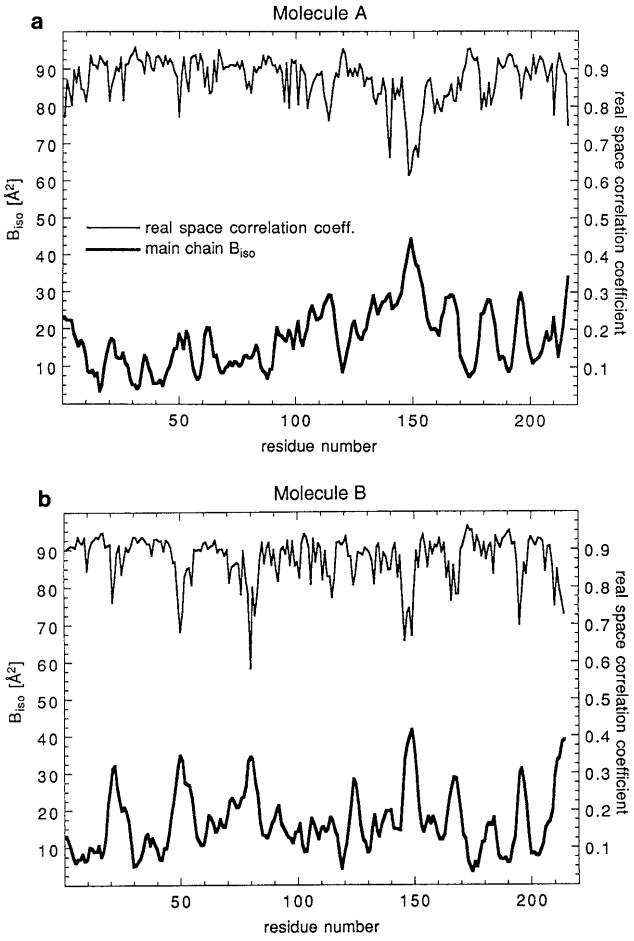

FIG. 3. Real-space correlation coefficient (6) for each residue (top curves) and averaged, isotropic temperature factor for the main-chain atoms of each residue (bottom curve) plotted against the residue number for molecules A (a) and B (b) in the unit cell of the refined structure of HAV 3C.

ing part of the active site, is a unique feature of the HAV 3C. It is present, but shorter, in the structure of HRV14 3C (42) and some bacterial proteinases (16, 49, 56). The corresponding loop in the mammalian chymotrypsin-like proteinases assumes a completely different conformation (17).

The active site of HAV 3C is found between the two β-barrel domains and the antiparallel β-ribbon extending from the carboxy-terminal domain. As a result, it is narrow and partly occluded (Fig. 1). In molecule B, the side chain of Lys B147 reaches across the active site to form a salt bridge with the cysteic acid B172 (Fig. 4b). The active site of molecule B is effectively inaccessible, due to the oxidation of Cys B172.

Two helices form the amino and carboxy termini of the structure. The amino-terminal helix packs against the carboxy-terminal domain, and the carboxy-terminal helix packs against the amino-terminal domain, thereby acting like two latches in connecting the domains (Fig. 1). The amino and carboxy termini of 3C are far apart, an unusual feature in protein structures (58). The amino-terminal helix is an α-helix; all other helical secondary structures in 3C are closer to the conformation of a 3_{10}-helix (Table 3).

Proteolytic active site. The active site of the catalytically competent HAV 3C resembles that of the chymotrypsin-like serine proteinases, as had been predicted by Allaire et al. (1). The disposition of the nucleophilic Cys A172, the general acid-base catalyst His A44, and the electrophilic oxyanion hole (Pro A169, Gly A170, Met A171, and Cys A172) is similar to that of Ser 195, His 57, and the oxyanion hole of chymotrypsin (Met 192 to Ser 195), but the active site is larger. The distance between the sulfur atom of the nucleophilic Cys A172 and the general acid-base His A44 $N^{\varepsilon 2}$ is 3.9 Å (Fig. 6 and 7). Also present is a β-strand (β-eII), part of the β-sheets of domain II, which forms the bottom of the active-site pocket (Fig. 1b and 7). This β-strand is in a position to form an antiparallel, β-sheet-type hydrogen bond interaction with a bound substrate. This is also a common feature of the chymotrypsin-like serine proteinases and their inhibitor complexes (5, 57).

There is no equivalent to the third member of the postulated catalytic triad in HAV 3C. A water molecule (Wat V6) occupies the position of the carboxylate of a catalytic triad residue of the serine proteinases in the HAV 3C structure (Fig. 6). The side chain hydroxyl of Tyr A143 is 3.45 Å from this water and 3.1 Å from the $N^{\varepsilon 1}$ of His A44. The geometry of neither interaction qualifies as a hydrogen bond. Water molecule Wat V6 is hydrogen bonded to another water molecule (Wat V29), the backbone carbonyl of Asp A84, and the imidazole of His A44 (Fig. 6). It is in the plane of the His A44 imidazole and forms a strong hydrogen bond to the His A44 $N^{\varepsilon 1}$ (2.7 Å).

The side chain of Asp A84 points away from His A44, even though it is topologically in an equivalent position to the third members of the catalytic triad of the chymotrypsin-like serine proteinases (Asp 102), as had been predicted by Gorbalenya et al. (20). Its carboxylate is locked in an ionic interaction with the ε-amino group of Lys A202 (Fig. 1a and 6). In computer-modeling studies (results not shown), we disrupted this interaction and tried to move the side chain of Asp A84 into the

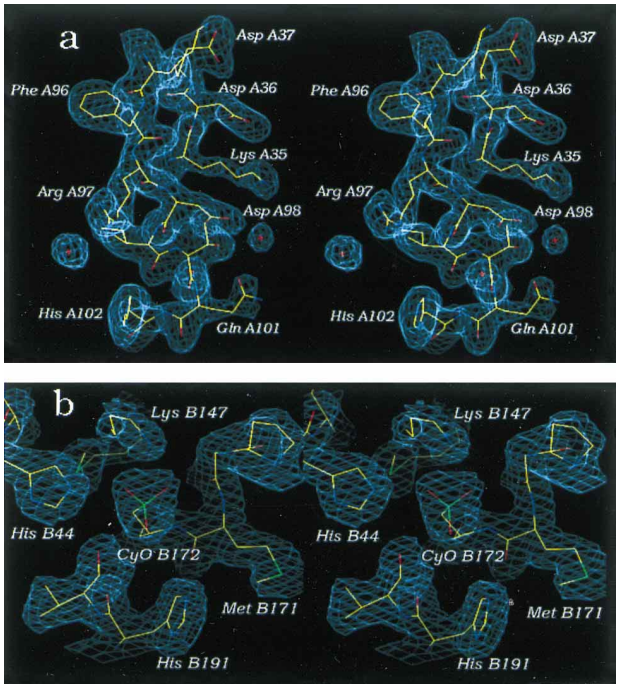

FIG. 4. Sample electron density map of representative regions of the refined structure of HAV 3C, calculated with coefficients $2|F_O| - |F_C|$ and contoured at a level of 1.0σ above the mean of the electron density. (a) Part of the interdomain connection from residues Phe A96 to His A102, including the residues of the reverse turn that connects β-strands β-bI and β-cI (Lys A35 to Asp A37). (b) Oxidized cysteic acid Cys B172 and surrounding residues in the active site of molecule B.

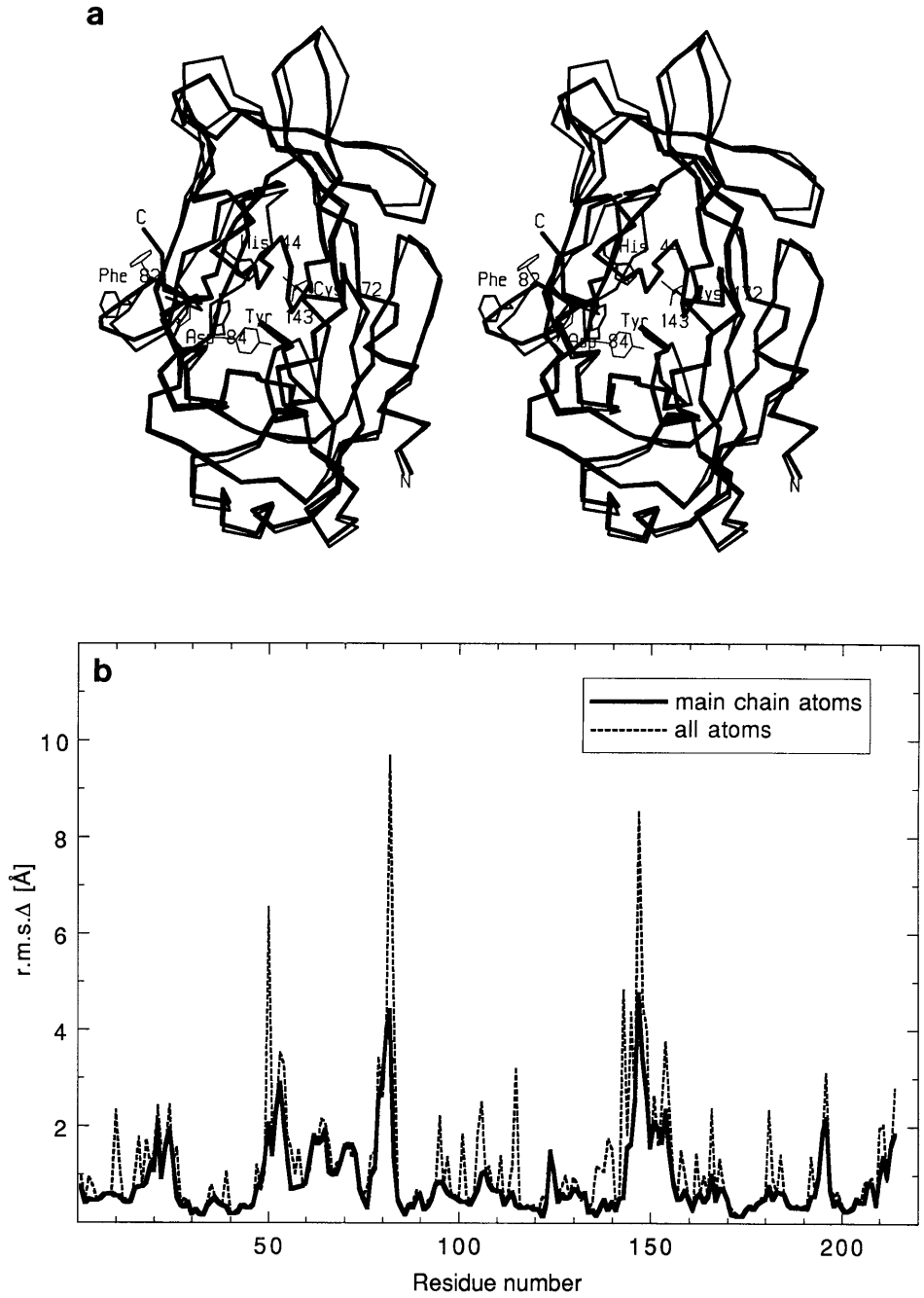

FIG. 5. Superposition of the two molecules in the unit cell of the triclinic crystals of HAV 3C. The molecules were superimposed by using the lsq-option of program O (31, 32). The overall root mean square differences in atomic coordinates for the superimposed molecules are 1.1 Å for the main-chain atoms and 1.86 Å for all atoms. (a) C^{α} traces of the two superimposed molecules; molecule A is represented by thicker lines. Also included are the side chains of residues His 44, Phe 82, Asp 84, Tyr 143, and Cys 172, with the side chains of molecule A drawn with thicker lines. (b) Root mean square differences between the atomic coordinates of the superimposed molecules A and B of the refined structure of HAV 3C for main-chain atoms and all atoms for each residue.

position of the carboxylate of a putative catalytic triad. We found that it is impossible to move the Asp A84 side chain close enough to the imidazole of His A44 without dramatically changing main-chain conformations in the amino-terminal β-barrel domain.

Because the side chain of Tyr A143 does not form hydrogen bonds, we suggest that it might be deprotonated and negatively charged. Tyr A143 is almost perpendicular to the plane of the imidazole and 3.1 Å away from the $N^{\varepsilon 1}$ of His A44 (Fig. 6). Unlike Wat V6, it is therefore not in a position for a direct proton transfer to or from His A44.

It is generally assumed that the native state of the active site of papain-like cysteine proteinases is a thiolate-imidazolium ion pair formed by the nucleophilic cysteine and the histidine general acid-base catalyst (reference 63 and references therein). Whether the active site of 3C-like proteinases also contains a thiolate-imidazolium ion pair has not been investigated. We speculate that the function of Tyr A143 in HAV 3C is to

TABLE 3. Secondary-structure elements[a] in HAV 3C (C24S) molecule A

Structure and residues	Sequence	Domain or type	Label
Helix			
A2–A12		α	A
A43–A45		3_{10}	B
A52–A54		3_{10}	C
A70–A72		3_{10}	D
A99–A102		3_{10}	E
A106–A111		3_{10}	F
A181–A183		3_{10}	G
A206–A213		3_{10}	H
β-Sheet			
A13–A19		I	aI
A25–A35		I	bI
A38–A42		I	cI
A56–A61		I	dI
A64–A69		I	e1I
A74–A77		I	e2I
A85–A89		I	fI
A117–A123		II	aII
A126–A132		II	b1II
A136–A146		II	b2II
A152–A164		II	cII
A175–A178		II	dII
A187–A194		II	eII
A198–A203		II	fII
Reverse turns			
A20–A23	EKNG	II	
A35–A38	KDDW	II′	
A46–A49	YKFE	I	
A49–A52	EKDY	VIII	
A61–A64	RGGT	I′	
A90–A93	VPTI	I	
A112–A115	ALNR	II	
A123–A126	VNGT	I′	
A147–A150	KNDG	IV	
A168–A171	LPGM	II	
A171–A174	MCGG	II	
A195–A198	GNSI	IV	

[a] Assignment according to Hutchinson and Thornton (26).

stabilize electrostatically a positive charge on the imidazole of His A44 and possibly the delocalization of this charge, involving Wat V6. We suggest that the 3C proteinases do not rely on a catalytic triad, because the HAV 3C proteinase lacks the third member of the catalytic triad and the carboxylate of Glu 71 in the HRV14 3C proteinase is in a conformation that makes proton transfer from the imidazole of the general acid-base catalyst (His 40) improbable (42). Instead, the electrostatic environment in the active site of 3C proteinases is designed to stabilize a developing positive charge on the catalytic histidine, at least during the rate-limiting transition state.

Substrate binding. The HAV 3C-specific cleavage sites are formed by at least six residues surrounding the scissile bond (P_4 to P_2') in the sequence of the viral polyprotein (29). Assuming that a peptide substrate would bind to the active site of HAV 3C in a fashion similar to the substrate complexes of the chymotrypsin-like serine proteinases (5, 57), it is possible to rationalize the experimentally observed sequence preferences of the HAV 3C proteinase (29) with the refined crystal structure.

Tetrapeptide substrates with sequence preferences for the S_4 to S_1 sites of a chymotrypsin-like proteinase assume a common, defined conformation when bound in the proteinase active sites (17, 28, 56). The P_4 and P_3 residues adopt a β-sheet conformation. The main-chain dihedral angles of the P_2 residue take on values of approximately $\Phi = -60°$ and $\Psi = 130°$ (polyproline conformation), and those of the P_1 residue take on values of approximately $\Phi = -110°$ and $\Psi = 40°$. These main-chain conformations cause the side chains to point into preformed specificity pockets in the active site of the enzyme.

The S_1-binding pocket of HAV 3C is a groove, formed by the main-chain atoms of Ala A193, Gly A194, Gly A167, and Leu A168 (Fig. 7). The side chains of Leu A199 and Met A171 contribute to the S_1-binding pocket and render it rather hydrophobic. The imidazole of His A191 is centrally located in the pocket, where it could interact with the carbonyl oxygen of the side chain of a glutamine residue in the P_1 position of a substrate. The S_1 pocket of HAV 3C is a hydrophobic cleft rather than a deep pocket, and there is no potential hydrogen-bonding partner for an interaction with the amide nitrogen of the carbonylamide of the P_1 glutamine residue of a substrate (Fig. 7). This correlates well with the fact that an *N*,*N*-dimethylglutamine residue as the P_1 residue is almost equally well recognized as a substrate by the HAV 3C proteinase as is the natural P_1 residue, glutamine (39).

Interacting with the buried edge of the imidazole ring of His A191 ($N^{\varepsilon 1}$) are two buried water molecules, in turn hydrogen bonded to the carboxyl group of the buried side chain of Glu A132. The residues surrounding the two water molecules and the Glu A132 side chain are hydrophobic: Leu A119, Ile A130, Leu A135, Leu A176, and Leu A199. There is electron density for two water molecules in both molecules in the triclinic unit cell of the HAV 3C structure. The two water molecules (Wat V4 and Wat V9) do not exclude each other. (The distance between Wat V4 and Wat V9 is 2.83 Å). However, in both 3C molecules, one water molecule shows much stronger electron density than the other does. Even though the two molecules do not exclude each other, it is not possible to decide with the available diffraction data whether there are two water molecules with different occupancies or two differently occupied positions for one buried water molecule.

There is an important distinction to be considered when charged amino acid residues are buried in the hydrophobic interior of a protein. This was pointed out recently by Qasim et al. (55). Deprotonation of a residue that is neutral in its protonated form (e.g., glutamic acid) creates a negatively charged carboxylate and a proton. Deprotonation of a residue that is positively charged in its protonated form (e.g., histidine) merely shifts the positive charge from the imidazolium ring to the dissociated proton. Transferring the former reaction from water into a hydrophobic environment is associated with a larger and more predictably unfavorable free energy change. The transfer free energy associated with the deprotonation of a histidine is, by comparison, smaller and less predictable (reference 55 and references therein). Therefore, a significantly larger, unfavorable free energy is associated with burying a glutamate side chain inside a protein domain. We propose that the interaction with the buried side chain of Glu 132 associates the protonation of His 191 with a larger energy barrier. As a result, the side chain of His 191 would be locked into a neutral, single tautomeric form over a wide range of pH (with a proton at the $N^{\varepsilon 2}$ of His 191). The trapped water molecules between the side chains of His 191 and Glu 132 serve to prevent the formation of a salt bridge by direct transfer of a proton from Glu 132 to His 191. Thus, Glu 132 provides a larger energy barrier for the protonation of the neutral His 191 than would the burial of the histidine side chain by itself and is crucial for the P_1 glutamine specificity of HAV 3C.

This is in contrast to the crystal structure of the glutamate-

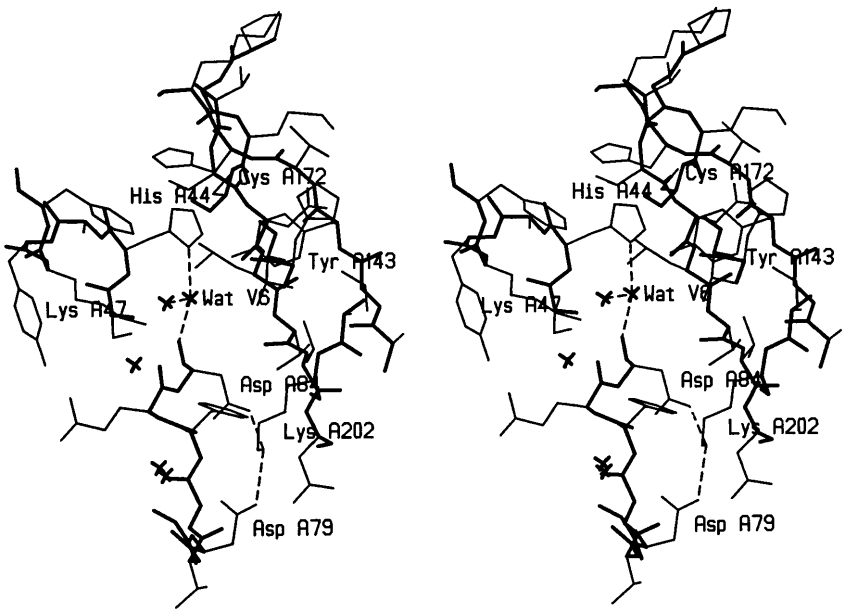

FIG. 6. Stereo view of an all-atom representation of the active-site residues of HAV 3C. The view is from the left side of the molecule relative to that in Fig. 1. The protein main chain is drawn with thick lines, and the side chains are drawn with thinner lines. Water molecules are represented by crosses. Selected hydrogen bonds are represented by broken lines. The general acid-base catalyst His A44 is in the center of the figure. The ionic interactions involving Asp A84, Lys A202, and Asp A79 are also shown.

specific proteinase from *Streptomyces griseus*, which has a histidine in a similar position to His 191 of HAV 3C to recognize the glutamate residue in the P_1 position of a substrate (49). In *S. griseus* glutamate-specific proteinase a buried triad of histidines presumably stabilizes a positive charge by delocalizing the charge over three imidazole ring systems.

The hydrophobic environment of His 191 is also very different from the very charged environment of the general acid-base catalyst His 44 in HAV 3C. The electrostatic environment of His 44 is provided by the many charged side chains in its proximity (Fig. 6). Presumably, this serves to stabilize a positive charge on His 44.

The active site of HAV 3C has little available space for the side chain of the P_2 residue. The side chains of Tyr 143, His 145, and Ala 193 form an S_2 pocket that can accommodate only a relatively small side chain. All confirmed P_2 positions in the HAV polyprotein are occupied by either serine or threonine (18, 23, 41, 64). The imidazole of the side chain of His 145

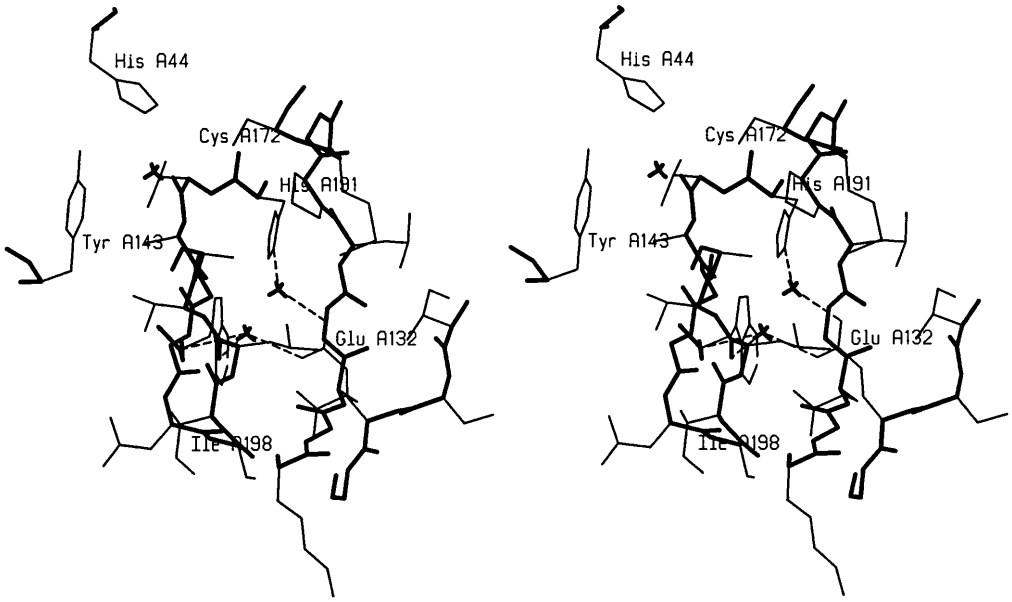

FIG. 7. Stereo view of an all-atom representation of the S_1-binding pocket of HAV 3C. The representation of main chain, side chain, water molecules, and hydrogen bonds is as in Fig. 6. In the center of the figure is the imidazole side chain of His A191. The pocket is formed by β-strand β-eII, the residues of the oxyanion hole, and several hydrophobic side chains (Met A171 and Leu A199).

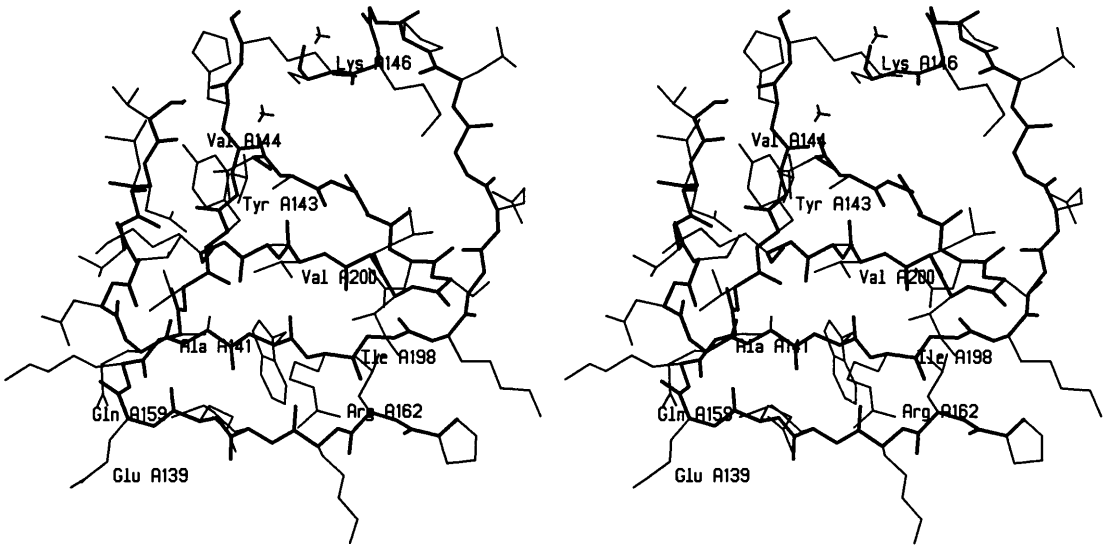

FIG. 8. Stereo view of an all-atom representation of the S_4-binding pocket of molecule A of HAV 3C. The representation of main chain, side chain, water molecules, and hydrogen bonds is as in Fig. 6. The pocket is formed in the cleft between β-strands β-b2II, β-eII, and β-fII and is lined mostly by hydrophobic residues.

would be in a position to form a hydrogen bond with the P2 residue of a bound substrate and could thereby explain the preference for either serine or threonine (Fig. 8 and 9).

The side chain in the P_3 position of a substrate points out of the active site, directly opposite from β-strand eII (Fig. 1b, 8, and 9). HAV 3C is not very selective for the P_3 residue in a peptide substrate (29), but several cleavage sites of the HAV polyprotein have phenylalanine, tryptophan, or glutamate in the P_3 position (23, 41, 61, 64). Only residues from β-strand β-b2II could potentially interact with the side chain in the P_3 position of a substrate (Val A144 and Lys A146 [Fig. 8 and 9]). In the crystal structure of HAV 3C, the side chain of Lys A146 forms a water-mediated intermolecular contact with Glu B139. Lys A146 would, however, be positioned to interact with the side chain in the P_3 position of a substrate (Fig. 8 and 9). This might explain the occurrence of aromatic residues (Phe and Trp) in the P_3 position of HAV 3C substrates: favorable interactions of the nucleophilic faces of aromatic rings with positive charges (e.g., lysine or arginine side chains) are a common feature in protein structures (10).

The S_4-binding pocket of HAV 3C is presumably formed by the cleft between β-strands β-b2II and β-eII in domain II (Fig. 1b), which is lined by mostly hydrophobic residues (Fig. 8). The side chains of Ala A141, Val A144, Ile A198, and Val A200 form the hydrophobic surface of this cleft. HAV 3C shows a strong preference for large, hydrophobic side chains in the P_4 position of its substrates (29), and the P_4 residues in all confirmed cleavage sites in the HAV polyprotein are either leucine, isoleucine, or valine (12, 18, 23, 41, 64).

Whereas it is possible to rationalize the cleavage preferences of HAV 3C for the residues amino terminal to the scissile bond (P_1 to P_4), this is very difficult to do for residues that are carboxy terminal to it (46). First, the P′ residues of serine proteinase substrates are generally less tightly bound because they constitute the leaving group following the catalytic cleavage (53). Second, HAV 3C is much less specific for the residues in the P′ positions of a substrate (54). Furthermore, there are no obvious binding sites or specificity pockets detectable in the crystal structure of the HAV 3C (C24S). On the contrary, the residues following the general acid-base catalyst His 44 in helix B and, more importantly, the residues forming a β-bulge in strand β-bI (Val 28 and Met 29) leave little room for the P′ residues of a substrate. This correlates well with the fact that the preference for a small residue in the P_1' position of a substrate is the most obvious, detectable specificity requirement of HAV 3C, which is C terminal to the scissile bond (54).

Based on models of the binding of protein inhibitors to chymotrypsin-like proteinases, we can identify potential interactions and binding pockets for a substrate of HAV 3C from the P_4 to the P_1 position. Our structure now allows successful interpretation of the specificity requirements of HAV 3C in the context of both a peptide substrate (29, 54) and the sequence of the polyprotein (12, 23, 41, 61). Figure 9 shows a model of a pentapeptide, GLFSQ, bound in the active site of the HAV 3C proteinase. This sequence corresponds to the P_5 to P_1 residues of the 2A-2B cleavage site of the HAV polyprotein, which was recently established by amino-terminal sequencing (41). The model in Fig. 9 summarizes our interpretation of the specificity determinants in the active site of HAV 3C.

The fact that some of the parts of the HAV 3C structure which form the active site are flexible (β-b2II and β-cII) suggests that the cleavage efficiency of HAV 3C on the polyprotein sites could be altered by conformational changes that involve some of the determinants of specificity (S_2 to S_5). The elucidation of these will have to await structural work on protein complexes of 3C with other subunits of the viral polyprotein.

Autoproteolysis. The HAV 3C proteinase can cleave itself out of the polyproteins, both in *cis* (when it is part of the same polyprotein) and in *trans* (23). Kinetic studies of the polyprotein processing of encephalomyocarditis virus suggested that the cleavage events which liberate 3C are characterized by two distinct kinetics: either intermolecular with bimolecular kinetics or intramolecular at low enzyme concentrations (51). In HAV, polyprotein processing yields the 3C subunit and no other 3C-containing stable intermediate or product (41). The amino and carboxy termini of 3C are far apart (Fig. 1). This is an unusual feature in protein domains or subunits (58) and suggests that a protein subunit has been proteolytically processed after the initial folding. Furthermore, both the amino and carboxy termini of 3C are far away from the active site. Any intramolecular cleavage event would have to be followed

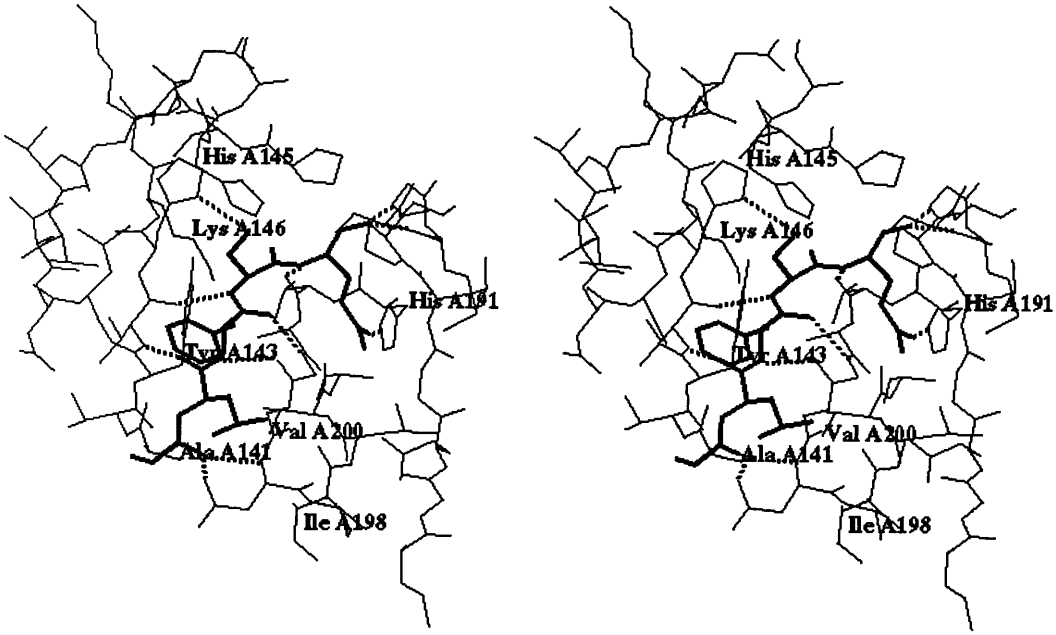

FIG. 9. Stereo view of an all-atom representation of a model of a pentapeptide inhibitor (Gly-Leu-Phe-Ser-Gln aldehyde) fitted into the active site of HAV 3C. The pentapeptide model is represented by thick lines, the protein is shown as thin lines, and hydrogen bonds are shown as broken lines. The model was built and regularized with the program O (31, 32) and is based on the conformation of inhibitor complexes of chymotrypsin-like serine proteinases (17, 49, 56). The sequence of the pentapeptide corresponds to the P1-2A/2B cleavage site of the viral polyprotein (30, 41, 61).

by secondary folding of the newly formed terminus out of the active site.

The carboxy terminus of HAV 3C is formed by a 3_{10} helix, and the last residues are disordered and flexible. The carboxy-terminal helix (H) is preceded by one of the central strands (β-fII) of domain II. If the conformation of the residues forming helix H was indeed the result of a secondary folding event, those residues would have to be initially wrapped around the carboxy-terminal domain (II). The folding of this intermediate is very difficult to imagine, and its structure would not conform to the general rules of the folding of protein domains (58).

It is much more straightforward to rationalize the folding of the amino-terminal helix (A) as a postproteolytic event. It would require an intermediate in which the last turn of helix A is retained as a reverse turn and the first two turns of helix A are in an extended conformation, with residues 1 to 9 reaching back into the active site through the gap between β-strands β-bI and β-aII. After an intramolecular cleavage, the residues of the new amino terminus would presumably fold out of the active site and form helix A. This model of processing of 3C is similar to the one proposed by Matthews et al. (42) and could explain the kinetic data of Palmenberg and Rueckert (51), as well as the rapid polyprotein processing at the amino terminus of poliovirus 3C observed by Hanecak et al. (22).

Matthews et al. (42) report that the amino-terminal helix in the HRV14 3C is flexible. This is not the case for the corresponding residues in the HAV 3C structure, as judged by the refined temperature factors (Fig. 3). The amino-terminal helix A in HAV 3C is a well-ordered and apparently stable amphipathic helix. This is not a contradiction to the model of amino-terminal, intramolecular cleavage of 3C presented above. We suggest that it is exactly the stability of this helix which is necessary to fold the new amino terminus out of the active site, thereby preventing product inhibition of 3C by its own amino terminus.

The kinetics of the polyprotein processing of encephalomyocarditis virus reported by Palmenberg and Rueckert (51) also suggest an intramolecular cleavage at the carboxy terminus of the 3C gene product. The three-dimensional structure makes it very difficult to envision such an intramolecular cleavage event. An alternative explanation for the observed kinetics is cleavage of the 3CD junction by another 3C subunit within a dimer or larger polymer. The existence of such a dimer in the poliovirus system was recently suggested (24, 66). Kinetic data of this kind are not available for HAV polyprotein processing, and it is not known whether HAV 3C can form dimers in solution. Several possible dimers formed by the two molecules are observed in the crystal structure of HAV 3C. Among these, the dimer that corresponds to the crystallographic unit cell provides the most attractive model for 3CD cleavage in a multisubunit complex. An unfolded carboxy terminus of molecule B could reach into the active site of molecule A. Neither the putative RNA recognition site nor our proposed 3C-3D interface (see below), in either molecule A or B, would be obstructed by such dimerization. Finally, the putative RNA recognition sites of the two 3C molecules would be very close, almost contiguous.

The regions of the structure which form the interface of this crystallographic dimer are primarily β-strand β-b1II and the preceding turn and the coil region following β-cII in molecule A and one face of helix F and a part of β-strand β-b2II in molecule B. This would leave the putative RNA recognition site, as well as our proposed 3C-3D interface, unobstructed in both molecules. The heterologous dimer interface we are proposing would also allow the formation of larger polymers of 3CD.

RNA recognition site. A stretch of five highly conserved residues, with the sequence KFRDI (residues 95 to 99 in HAV, and residues 82 to 86 in poliovirus and rhinovirus), forms part of the domain connection on the opposite side of the HAV 3C molecule from the proteolytic active site (Fig. 4a and 10a). Mutational studies on poliovirus 3C initially implicated these residues as having a function in the picornavirus life cycle,

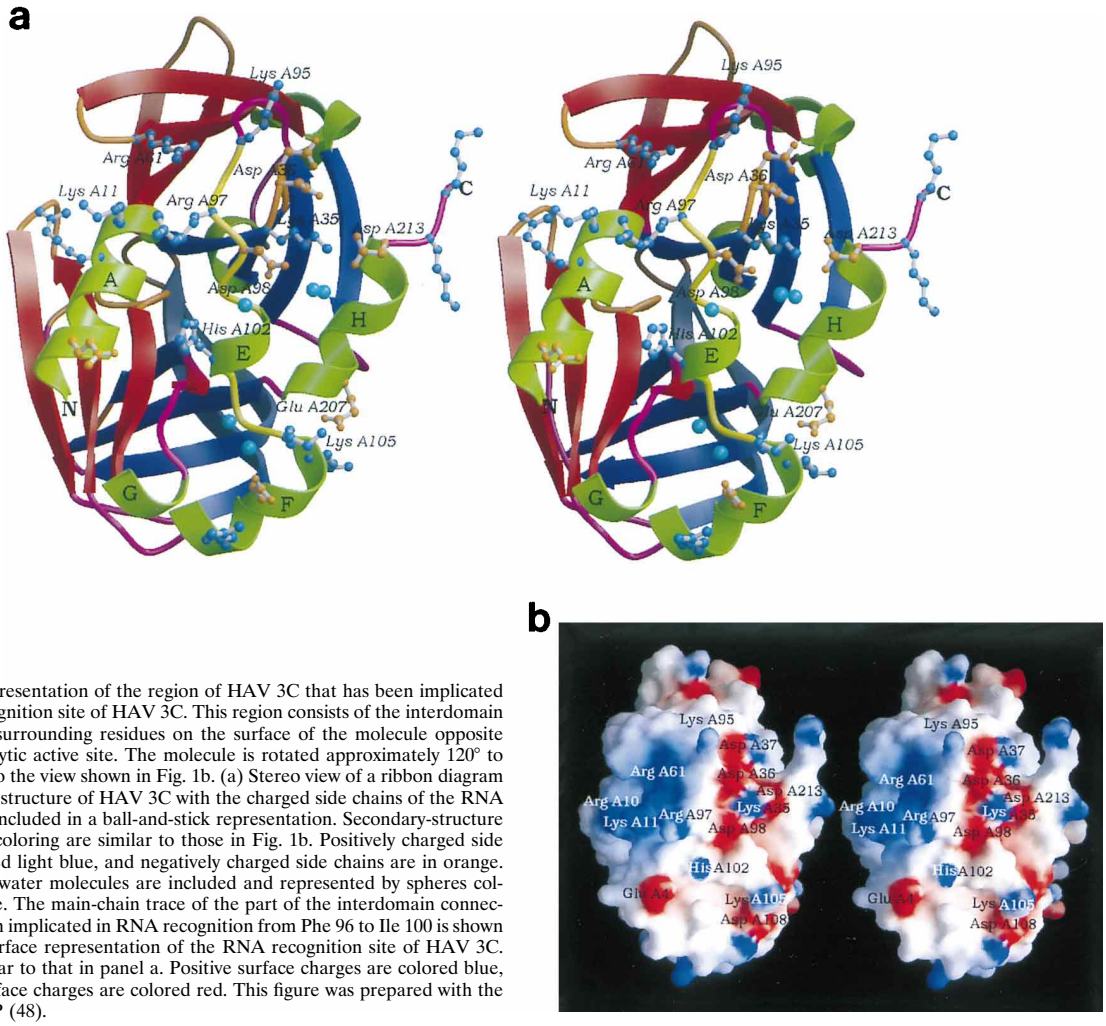

FIG. 10. Representation of the region of HAV 3C that has been implicated as the RNA recognition site of HAV 3C. This region consists of the interdomain connection and surrounding residues on the surface of the molecule opposite from the proteolytic active site. The molecule is rotated approximately 120° to the left relative to the view shown in Fig. 1b. (a) Stereo view of a ribbon diagram of the secondary structure of HAV 3C with the charged side chains of the RNA recognition site included in a ball-and-stick representation. Secondary-structure assignment and coloring are similar to those in Fig. 1b. Positively charged side chains are colored light blue, and negatively charged side chains are in orange. Several ordered water molecules are included and represented by spheres colored in cyan blue. The main-chain trace of the part of the interdomain connection that has been implicated in RNA recognition from Phe 96 to Ile 100 is shown in yellow. (b) Surface representation of the RNA recognition site of HAV 3C. The view is similar to that in panel a. Positive surface charges are colored blue, and negative surface charges are colored red. This figure was prepared with the program GRASP (48).

distinct from the proteolytic activity of 3C (21). Evidence is accumulating that the surface on the backside of the 3C subunit forms an RNA recognition site in several picornavirus systems (2, 24, 37, 66). Recognition of the nontranslated regions of the viral RNA genome (5′ and 3′ NTR) by the 3C subunit of a protein complex appears to be important for the initiation of RNA replication in picornaviruses, but experimental evidence for the HAV system is not available.

The proposed RNA recognition site of HAV 3C is formed by the domain connection, one face of each of the amino- and carboxy-terminal helices (helix A and H), and several of the small helices and reverse turns that connect strands β-bI and β-cI (residues 35 to 38), β-dI and β-e1I (residues 61 to 64), and β-dII and β-eII (residues 181 to 184, helix G). Whereas the 3C gene product is primarily a β-sheet-type protein, most of the RNA recognition site is formed by regions having helical secondary structure. Many charged side chains contribute to this surface, and several well-ordered water molecules are tightly bound in preformed cervices (Fig. 4a and 10a). A large, positively charged pit is formed by the side chains of residues Arg 10, Lys 11, Arg 61, and Arg 97 (Fig. 10b). Whereas most of the charged and hydrophilic residues are in positions, and display conformations, that make it likely that they are directly involved in interactions with the RNA, several conserved, hydrophobic residues are buried in the interior of the protein. These residues (e.g., Phe 96, Ile 99, Phe 103, and Ile 104 in HAV 3C) contribute to the packing of the interior and maintain the conformation of the RNA recognition site. Some of them are also strictly conserved throughout the picornaviruses and are therefore required for RNA recognition. These structural details help to explain the results of mutational studies of the RNA recognition site of poliovirus 3C (21). Besides the residues directly involved in the binding of RNA, residues which maintain the shape and integrity of the RNA recognition site are deemed to be of equal importance. Some residues may have both functions. Arg 61 reaches across from the reverse turn following strand β-dI to the domain connection. The positively charged guanidinium group of its side chain packs against the nucleophilic aromatic side chain of Phe 96 and is hydrogen bonded to the main chain of the domain connection (peptide carbonyl oxygen of Pro 94). The other side of its guanidinium group is hydrogen bonded to an ordered water molecule in the structure of HAV 3C, but it could interact with RNA directly or via this water. Similarly, the highly conserved Asp 98 forms a strong hydrogen bond to the Thr 100 side chain, probably stabilizing the single helical turn from residues Ile 99 to His 102 (Fig. 4a). The other carboxylate oxygen is hydrogen bonded to an ordered water molecule and is in a position to interact directly with RNA.

The RNA recognition site of HAV 3C is a preformed stable

surface on the opposite side of the molecule from the proteolytic active site. The two-domain β-barrel core structure of HAV 3C is rather rigid and does not allow for conformational changes. It appears unlikely from the structure that RNA binding and proteolytic activity will be directly influenced by each other.

Protein complexes. The 3C gene products of picornaviruses form parts of several suggested complexes with other viral and cellular proteins, as well as with RNA (2, 24, 66). Both the proteolytic activity and the RNA binding activity of poliovirus 3C are modulated by the other protein subunits of the complex (44). Whether this is the same in other picornaviruses is not clear. We suggest that the other subunits of the 3C-containing protein complexes interact with 3C at a surface different from either the proteolytic active site or the RNA recognition site. Most probably, this surface would be formed by residues from the face of the carboxy-terminal helix (H), opposite from the RNA recognition site; helices B, C, and D; β-strand β-e2I and the following flexible Ω-loop (residues 79 to 83); and one side of the antiparallel β-ribbon formed by strands β-b2II and β-cII (the top left side of the molecule in the view in Fig. 1b). In particular, the Ω-loop and the antiparallel β-ribbon are not part of the core barrel structure of 3C, and their conformations are quite different in the two molecules in the unit cell of the HAV 3C structure (Fig. 5a). These regions of the molecule contribute to the proteolytic active site, and conformational changes here could influence both the proteolytic activity and the substrate specificity. For example, Asp 84, the topological equivalent of a third member of a catalytic triad, is locked in an electrostatic interaction with Lys 202 and Asp 79 (Fig. 6). Any interaction of 3C with another protein domain would very probably change the conformation of the loop between Asp 79 and Asp 84 and potentially disturb this interaction, possibly allowing Asp 84 to move closer to the proteolytic active site and to interact directly with His 44. Similarly, a conformational change involving the antiparallel β-ribbon formed by residues 139 to 159 could alter the recognition of a substrate by the 3C proteinase which presumably involves His 145, Lys 146, and other side chains in this region.

Conclusion. The refined crystal structure of HAV 3C (C24S) displays the details of the catalytic machinery and substrate recognition sites (S_4 to S_1) of the proteolytic activity. It also shows the details of a well-defined, preformed RNA recognition site which is completely separate from the proteolytic active site. Together with available biochemical data, it also allows one to speculate on the formation of complexes with other protein subunits and their effect on the 3C activities.

The crystal structure will also be a valuable tool for the design of specific proteinase inhibitors with a potential as antiviral drugs. However, the major determinant of substrate specificity is the glutamine residue in the P_1 position of a substrate. Inhibitors based solely on this primary specificity are not sufficiently effective (29, 39). Therefore, it will be necessary to exploit minor determinants of specificity, which contribute much less to the binding energy of a substrate, may be different for the proteolytic activity of 3C in a multisubunit complex, and are less stringently conserved in the polyprotein. Designing effective inhibitors of the 3C proteinase activity which are based on minor determinants of specificity but are very effective and not easily overcome by a virus will certainly prove to be a serious challenge. On the other hand, the fact that 3C has essential functions in picornavirus gene replication, distinct from its proteolytic activity, opens up the possibility to target other inhibitors against those functions. The available structural details of the RNA recognition site provide an important basis for the design of 3C inhibitors that target 3C-RNA complex formation. Eventually, it may even be possible to design inhibitors targeted against subunit interactions in larger protein complexes that include the 3C subunit, although this goal will have to await structural work on larger protein complexes of 3C.

ACKNOWLEDGMENTS

We thank Doug Scraba, Robert Esnouf, Randy Read, and Steven Ness for advice and helpful discussion. The atomic coordinates of the refined crystal structure of HAV 3C have been deposited with the Brookhaven Protein Data Bank and are, until their release, available from the authors upon request.

This research was supported by the Natural Sciences and Engineering Research Council of Canada (CPG PJ 171), the National Institutes of Health (NIH AI 38249), and the Medical Research Council of Canada (DG N013 to BAM).

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. G. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature **369:**72–76.
2. **Andino, R., G. E. Rieckhof, P. L. Achacoso, and D. Baltimore.** 1993. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA. EMBO J. **12:**3587–3598.
3. **Bacon, D. J., and W. F. Anderson.** 1988. A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graphics **6:**219–220.
4. **Bazan, J. F., and R. J. Fletterick.** 1988. Viral cysteine proteinases are homologous to the trypsin-like family of serine proteinases: structural and functional implications. Proc. Natl. Acad. Sci. USA **85:**7872–7876.
5. **Bode, W., and R. Huber.** 1992. Natural protein proteinase inhibitors and their interactions with proteinases. Eur. J. Biochem. **204:**433–451.
6. **Bränden, C.-I., and T. A. Jones.** 1990. Between objectivity and subjectivity. Nature **343:**687–689.
7. **Brünger, A. T.** 1993. X-Plor: a system for X-ray crystallography and NMR. Yale University Press, New Haven, Conn.
8. **Brünger, A. T.** 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature **355:**472–475.
9. **Brünger, A. T.** 1991. Simulated annealing in crystallography. Annu. Rev. Phys. Chem. **42:**197–223.
10. **Burley, S. K., and G. A. Petsko.** 1986. Amino-aromatic interactions in proteins. FEBS Lett. **203:**139–143.
11. **Chothia, C.** 1984. Principles that determine the structure of proteins. Annu. Rev. Biochem. **53:**537–572.
12. **Cohen, J. I., J. R. Ticehurst, R. H. Purcell, A. Buckler-White, and B. M. Baroudy.** 1987. Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. J. Virol. **61:**50–59.
13. **Collaborative Computational Project Number 4.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. **D50:**760–763.
14. **Dougherty, W. G., and B. L. Semler.** 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. **57:**781–822.
15. **Engh, R. A., and R. Huber.** 1991. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. **A47:**392–400.
16. **Fujinaga, M., L. T. J. Delbaere, G. D. Brayer, and M. N. G. James.** 1985. Refined structure of α-lytic protease at 1.7Å resolution. J. Mol. Biol. **184:** 479–502.
17. **Fujinaga, M., A. R. Sielecki, R. Read, W. Ardelt, M. Laskowski, Jr., and M. N. G. James.** 1987. Crystal and molecular structure of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8Å resolution. J. Mol. Biol. **195:**397–418.
18. **Gauss-Müller, V., D. Jürgensen, and R. Deutzmann.** 1991. Autoproteolytic cleavage of recombinant 3C proteinase of hepatitis A virus. Virology **182:** 861–864.
19. **Gorbalenya, A. E., V. M. Blinov, and A. P. Donchenko.** 1986. Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. FEBS Lett. **243:**103–114.
20. **Gorbalenya, A. E., A. P. Donchenko, V. M. Blinov, and E. V. Koonin.** 1989. Cysteine proteinases of positive strand RNA viruses and chymotrypsin-like serine proteinases. A distinct protein superfamily with a common structural fold. FEBS Lett. **243:**103–114.
21. **Hämmerle, T., A. Molla, and E. Wimmer.** 1992. Mutational analysis of the proposed FG loop of poliovirus proteinase 3C identifies amino acids that are necessary for 3CD cleavage and might be determinants of a function distinct from proteolytic activity. J. Virol. **66:**6028–6034.
22. **Hanecak, R., B. L. Semler, H. Ariga, C. W. Anderson, and E. Wimmer.** 1984. Expression of a cloned gene segment of poliovirus in *E. coli*: evidence for autocatalytic production of the viral proteinase. Cell **37:**1063–1073.

23. **Harmon, S. A., W. Updike, J. Xi-Ju, D. F. Summers, and E. Ehrenfeld.** 1992. Polyprotein processing in *cis* and in *trans* by hepatitis A virus 3C protease cloned and expressed in *E. coli*. J. Virol. **66:**5242–5247.
24. **Harris, K. S., W. Xiang, L. Alexander, W. S. Lane, A. V. Paul, and E. Wimmer.** 1994. Interaction of poliovirus polypeptide $3CD^{Pro}$ with the 5′ and 3′ termini of the poliovirus genome. J. Biol. Chem. **269:**27004–27014.
25. **Heinz, B. A., J. Tang, J. M. Labus, F. W. Chadwell, S. W. Kaldor, and M. Hammond.** 1996. Simple in vitro translation assay to analyze inhibitors of retrovirus proteases. Antimicrob. Agents Chemother. **40:**267–270.
26. **Hutchinson, E. G., and J. M. Thornton.** 1996. Promotif—a program to identify and analyze structural motifs in proteins. Protein Sci. **5:**212–220.
27. **Ivanoff, L. A., T. Towatari, J. Ray, B. D. Korant, and S. R. Petteway.** 1986. Expression and site-specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83:**5392–5396.
28. **James, M. N. G., A. R. Sielecki, G. D. Brayer, L. T. J. Delbaere, and C.-A. Bauer.** 1980. Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 resolution: a model for serine protease catalysis. J. Mol. Biol. **144:**43–89.
29. **Jewell, D. A., W. Swietnicki, B. M. Dunn, and B. A. Malcolm.** 1992. Hepatitis A virus 3C proteinase substrate specificity. Biochemistry **31:**7862–7869.
30. **Jia, X.-Y., D. F. Summers, and E. Ehrenfeld.** 1993. Primary cleavage of the HAV capsid precursor in the middle of the proposed 2A coding region. Virology **193:**515–519.
31. **Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. **A47:**110–119.
32. **Jones, T. A., and M. Kjeldgaard.** 1995. O—the manual (Version 5.11). Uppsala University, Uppsala, Sweden.
33. **Kraulis, P. J.** 1991. Molscript: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
34. **Kräusslich, H.-G., and E. Wimmer.** 1988. Viral proteinases. Annu. Rev. Biochem. **57:**701–754.
35. **Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. Procheck: a program to check stereochemical quality of protein structures. J. Appl. Crystallogr. **26:**283–291.
36. **Lemon, S. M., and G. H. Robertson.** 1993. Current perspectives in the virology and molecular biology of hepatitis A virus. Semin. Virol. **4:**285–295.
37. **Leong, L. E. C., P. A. Walker, and A. G. Porter.** 1993. Human rhinovirus 14 protease 3C ($3C^{Pro}$) binds specifically to the 5'-noncoding region of the viral RNA. J. Biol. Chem. **268:**25725–25739.
38. **Malcolm, B. A.** Personal communication.
39. **Malcolm, B. A.** 1995. The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase folds. Protein Sci. **4:**1439–1445.
40. **Malcolm, B. A., S. M. Chin, D. A. Jewell, J. R. Stratton-Thomas, K. B. Thudium, R. Ralston, and S. Rosenberg.** 1992. Expression and characterization of recombinant hepatitis A virus 3C proteinase. Biochemistry **31:**3358–3363.
41. **Martin, A., N. Escriou, S.-F. Chao, M. Girard, S. M. Lemon, and C. Wychoswski.** 1995. Identification and site-directed mutagenesis of the primary (2A/2B) cleavage site of the hepatitis A virus polyprotein: functional impact on the infectivity of HAV RNA transcripts. Virology **213:**213–222.
42. **Matthews, D. A., W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E. Villafranca, C. A. Janson, H. E. McElroy, C. L. Gribskov, and S. Worland.** 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site and means for cleaving precursor polyprotein. Cell **77:**761–771.
43. **Merritt, E. A., and M. E. P. Murphy.** 1994. Raster3D version 2.0: a program for photorealistic molecular graphics. Acta Crystallogr. **D50:**869–873.
44. **Molla, A., K. S. Harris, A. V. Paul, S. H. Shin, J. Mugavero, and E. Wimmer.** 1994. Stimulation of poliovirus proteinase $3C^{pro}$-related proteolysis by the genome-linked protein VPg and its precursor 3AB. J. Biol. Chem. **269:**27015–27020.
45. **Morris, A. L., M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton.** 1992. Stereochemical quality of protein structure coordinates. Proteins Struct. Funct. Genet. **12:**345–364.
46. **Morris, T. S., S. Frormann, S. Shechosky, C. Lowe, M. J. Lall, V. Gauss-Müller, R. H. Purcell, S. U. Emerson, J. C. Vederas, and B. A. Malcolm.** *In vitro* and *ex vivo* inhibition of hepatitis A virus 3C proteinase by a peptidyl monofluoromethyl ketone. Submitted for publication.
47. **Navaza, J.** 1994. AMoRe: an automated package for molecular replacement. Acta Crystallogr. **A50:**157–163.
48. **Nicholls, A., K. A. Sharp, and B. Honig.** 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct. Funct. Genet. **11:**281–296.
49. **Nienaber, V. L., K. Breddam, and J. J. Birktoft.** 1993. A glutamic acid specific serine proteinase utilizes a novel histidine triad in substrate binding. Biochemistry **32:**11469–11475.
50. **Palmenberg, A. C.** 1990. Proteolytic processing of picornaviral polyprotein. Annu. Rev. Microbiol. **44:**603–623.
51. **Palmenberg, A. C., and R. Rueckert.** 1982. Evidence for intramolecular self-cleavage of picornaviral replicase precursors. J. Virol. **41:**244–249.
52. **Pannu, N. S., and R. J. Read.** 1996. Improved structure refinement with maximum likelihood. Acta Crystallogr. **A50:**659–668.
53. **Perona, J. J., and C. S. Craik.** 1995. Structural basis of substrate specificity in the serine proteinases. Protein Sci. **4:**337–360.
54. **Petithory, J. R., F. R. Masiarz, J. F. Kirsch, D. V. Santi, and B. A. Malcolm.** 1991. A rapid method for determination of endoproteinase substrate specificity: specificity of the 3C proteinase from hepatitis A virus. Proc. Natl. Acad. Sci. USA **88:**11510–11514.
55. **Qasim, M. A., M. R. Ranjbar, R. Wynn, S. Anderson, and M. Laskowski Jr.** 1995. Ionizable P_1 residues in serine proteinase inhibitors undergo large pK shifts on complex formation. J. Biol. Chem. **270:**27419–27422.
56. **Read, R. J., M. Fujinaga, A. R. Sielecki, and M. N. G. James.** 1983. Structure of the complex of *Streptomyces griseus* proteinase B and the third domain of the turkey ovomucoid inhibitor at 1.8Å resolution. Biochemistry **22:**4420–4433.
57. **Read, R. J., and M. N. G. James.** 1986. Introduction to the protein inhibitors: X-ray crystallography, p. 301–336. *In* A. J. Barrett and G. Salvesen (ed.), Proteinase inhibitors. Elsevier Science Publishers, Amsterdam, The Netherlands.
58. **Richardson, J. S., and D. C. Richardson.** 1989. Principles and patterns of protein conformation, p. 1–98. *In* G. D. Fasman (ed.), Prediction of protein structure and the principles of protein conformation. Plenum Press, New York, N.Y.
59. **Rossmann, M. G.** 1990. The molecular replacement method. Acta Crystallogr. **A46:**73–82.
60. **Schechter, T., and A. Berger.** 1967. On the active site of proteases. Biochem. Biophys. Res. Commun. **27:**157–162.
61. **Schultheiss, T., Y. Y. Kusov, and V. Gauss-Müller.** 1994. Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2-P3 at all sites including VP1/2A and 2A/2B. Virology **198:**275–281.
62. **Schultheiss, T., W. Sommergruber, Y. Kusov, and V. Gauss-Müller.** 1995. Cleavage specificity of purified recombinant hepatitis A virus 3C proteinase on natural substrates. J. Virol. **69:**1727–1733.
63. **Storer, A., and R. Ménard.** 1994. Catalytic mechanism in papain family of cysteine peptidases. Methods Enzymol. **244:**486–500.
64. **Tesar, M., I. Pak, X.-Y. Jia, O. C. Richards, D. F. Summers, and E. Ehrenfeld.** 1994. Expression of hepatitis A virus precursor protein P3 *in vivo* and *in vivo*: polyprotein processing of the 3CD cleavage site. Virology **198:**524–533.
65. **Wilmot, C. M., and J. M. Thornton.** 1990. β-Turns and their distortions: a proposed new nomenclature. Protein Eng. **3:**479–493.
66. **Xiang, W., K. S. Harris, L. Alexander, and E. Wimmer.** 1995. Interaction between the 5′-terminal cloverleaf and $3AB/3CD^{Pro}$ of poliovirus is essential for RNA replication. J. Virol. **69:**3658–3667.
67. **Xuong, N. H., C. Nielsen, R. Hamlin, and D. Anderson.** 1985. Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. J. Appl. Crystallogr. **18:**342–350.