Structural and Mechanistic Aspects of 3C Proteases from the Picornavirus Family†

Dorit Arad,* Racheli Kreisberg,‡ and Michael Shokhen‡

Department of Microbiology and Biotechnology, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel

Received September 24, 1992

Picornavirus 3C proteases are prime targets for rational drug design. This viral protease appears in a large number of viruses from the Picornavirus family that cause serious disease syndromes, and it has an important role in the life cycle of the virus, processing the translation product of the Picornavirus genome by progressive co- and posttranslational cleavages. It is, therefore, important to gain structural and mechanistic information about this family of enzymes. We concentrate in this paper on the specific features of the 3C; particularly, we are trying to show that the 3C constitute a new family to enzymes which is neither a serine- nor a cysteine-type protease. General basic theory on the behavior of sulfur nucleophile vs the behavior of oxygen nucleophile regarding the nucleophilic attack on carbonyl compounds and the possible determinants in the structure of 3C viral proteases of rhinovirus 1A are being discussed.

INTRODUCTION

Picornaviruses cause a wide range of disease syndromes in man, among them: fatal paralysis, encephalitis, meningitis, conjunnctivitis, myocarditis, and common cold.[1] The incidence of human infections with Picornaviruses is high, and although most of the syndromes are not fatal, they cause common serious side effects. Since vaccination against more than 70 nonpolio enteroviruses and 110 rhinoviruses is impossible at the present time, the development of a broad range of chemotherapeutic agents is desirable. The primary translation product of the Picornavirus genome is a single large polyprotein which is processed to the mature viral polypeptides by progressive co- and posttranslational cleavages. Replication of the Picornaviruses is, thus, entirely dependent upon the proteolysis of viral precursor proteins. Virus-encoded proteases are, therefore, prime targets for the development of specific inhibitors that may be tailored for use in vivo, and much effort is being put into the study of the structure of the relevant enzymes.

The 3C protease is an enzyme which is highly conserved within the Picornavirus family, and it is responsible for the proteolytic processing (at sites Gln-Gly) of the precursor polypeptide. A lot of effort[2] has been directed toward the study of the molecular clones of the enzymes, and site-directed mutagenesis experiments were performed in order to derive the significant residue for catalysis and binding of 3C. Yet, the enzymatic mechanism of these viral proteases has still not been well-characterized.

The emphasis of this paper lies on the hypothesis that the *3C Protease and other related proteases constitute a new family of enzymes which is neither a cysteine nor a serine family.*

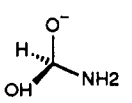
Using homology comparisons and multiple sequence alignments to close-related families of enzymes, we have gained some more insight about the structure-function relationship of the 3C enzyme. Quantum mechanical methods are applied on model structures in order to derive information about the origin of the difference between sulfur and oxygen nucleophile-based enzymes. The results may be directed toward the design of novel types of inhibitors for sulfur nucleophile-based enzymes.

METHODS

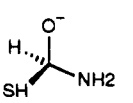
The following computational methods were used: Ab-initio calculations were performed using the program Gaussian 90[3] on the Convex 220. Geometries were optimized using the 6-31 + G^* basis set. Coordinates of proteins with known strucures were extracted from the PDB.[4] Modeling and molecular mechanics calculations were performed on the Silicon Graphics 20/G using Quanta by Polygen.[5] Within Quanta, the homology program was used to construct a model structure for the 3C protease. Multiple sequence alignments were found using the program GCG.[6]

RESULTS AND DISCUSSION

Behavior of Sulfur Nucleophile. It has been suggested by Kollman et al.[7] that the potential surface for sulfhydryl anion reaction with a carbonyl is different than the same reaction with a hydroxyl nucleophile. The main difference is that a charged tetrahedral intermediate (**1**) is a stable species on the reaction potential surface of hydroxyl attack, yet the analogous species which contains a sulfhydryl nucleophile (**2**) is not an intermediate on the reaction potential surface. When the carbonyl is protonated, the calculated reaction path for the nucleophilic attack of sulfhydrile ion resembles the one calculated for hydroxyl, thus the carbonyl has to be protonated before a nucleophilic attack by sulfhydrile. This observation has had a direct implication on the mechanism of proteolytic enzymes whose catalytic site contain oxygen or sulfur as nucleophiles, (e.g., serine and cysteine proteases). This implication will be discussed later, but first one should ask the question: what is the basic reason for the different behavior of sulfur vs oxygen, relative to their nucleophilic attack on a carbonyl compound.



1



2

Why Is the Sulfur Nucleophile So Different? The question of why a sulfur nucleophile behaves differently than oxygen regarding its nucleophilic attack on a carbonyl center[8] has a major importance in relation to the enzymatic mechanism of 3C viral protease. In order to elucidate the reason for the different behavior of sulfur, we have performed high-level ab

* To whom correspondence should be addressed. FAX 972-3-6409407.
† This paper was originally given as an invited lecture in the 10th ICCCRE conference in Jerusalem, July 1992.
‡ Department of Chemistry, Technion, Israel Institute of Technology, Haifa, Israel.

0095-2338/93/1633-0345$04.00/0

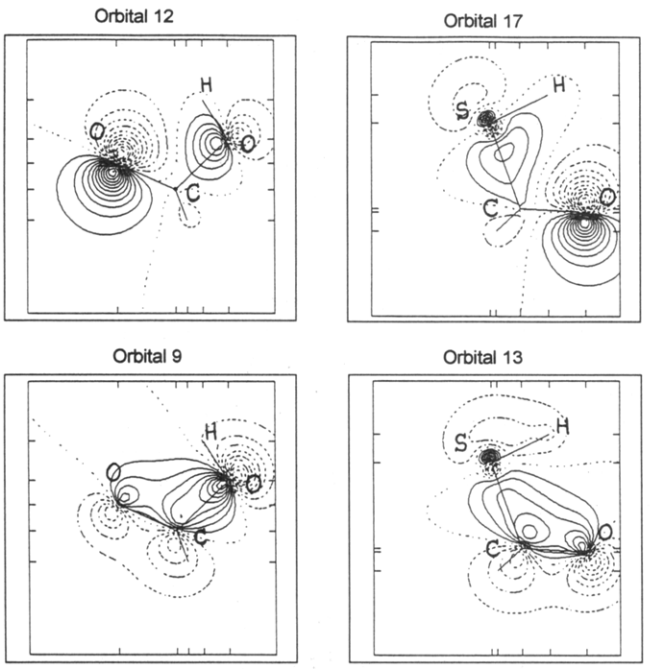


Figure 1. MO's for tetrahedral intermediates formed by HO^- and HS^- addition to CH_2O. These MO's are the result of the interaction between π-symmetry MO's of the reagents. Orbitals 9 (HO^-) and 13 (HS^-) are the main components of the σ-symmetry overlap in the new forming bond. Orbitals 12 (HO^-) and 17 (HS^-) are responsible for donor–acceptor electron transfer.

initio calculations (6 – 31 + G^*, Gaussian 90), calculating the energy, orbital map, and electron distribution of the nucleophiles, OH- and SH-; the substrate, CH_2=O; and the product, tetrahedral complex **1** and **2**.

Analysis of charge density maps for the tetrahedral complexes **1** and **2** revealed that the charge on the central carbon atom is positive when oxygen is the nucleophile (in **1**) yet negative when sulfur attacks (in **2**).

A negative charge distribution on the carbon results in a disfavored interaction between carbon and sulfur and may be one of the reasons for the tetrahedral intermediate **2** not to be formed.

PMO analysis revealed another direction for reasoning the nucleophilic attack of sulfur vs oxygen: Nucleophylic additions are governed by two types of orbitals: (1) π orbitals of reagents that are responsible for the electron transfer from the donor (nucleophile) to the acceptor (C=O bond) and (2) σ orbitals that are responsible for the geometrical reorganization of the reacting molecules to form the product tetrahedral complex (**1** and **2**). The relevant σ and π orbitals are given in Figure 1 (for OH + C=O) and in Figure 2(for SH + C=O).

According to the frontier orbitals theory,[9] electron transfer occurs as a result of π symmetry fragment MO interactions. This interaction may occur in the nucleophilic attack of SH as well as that of OH, as shown by the overlap map (Figures 1 and 2).

Geometrical reorganization is a process that depends on interaction between occupied orbitals, and intuitively it seems that this interaction does not have an effect on the product formation since it does not contribute to the energetic stabilization. A closer analysis into the relevant orbitals shows that actually the σ orbitals do participate in the C-X bond formation and may serve as a key interaction that allows the electron transfer to occur or not. The result of this interaction when a new bond is formed as a partial cleavage of the σ-symmetry component of the carbonyl bond in the acceptor site. The σ MO in Figure 1 reflects the topological picture of this process. A σ-MO's interaction results in geometrical

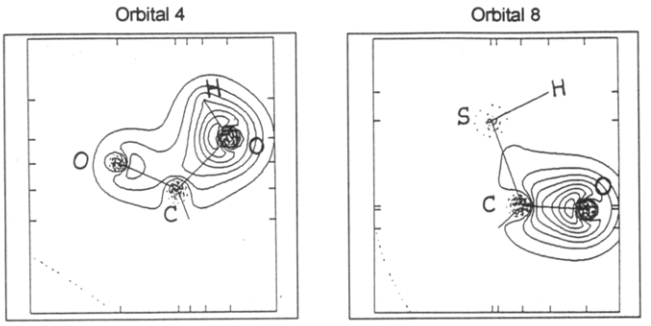
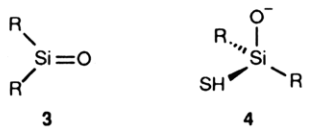


Figure 2. MO's of the tetrahedral intermediate for nucleophiles HO^- and HS^- addition to CH_2O. These MO's are the result of the interaction between σ-symmetry occupied MO's of the reagents. Orbital 4 (HO^-) plays the main role in weakening the σ-bond of the carbonyl group. Orbital 8 (HS^-) conserves the structure of the initial σ-part of the carbonyl group, reflecting luck of interaction with the occupied σ-MO of the nucleophile HS^-.

reorganization that allows the donor–acceptor electron transfer in π-MO's interaction. Since the σ OH orbital is much lower in energy than the σ SH (0.3 AU gap), the σ interaction is expected to be weaker in SH than in OH. Looking at calculated MO's, we observe indeed the lack of overlap interaction (MO8, SH^-, Figure 2) for SH^- attack, and overlap interaction in the newly formed bond for OH^- attack (MO4, OH^-, Figure 2). The conclusion is that the tetrahedral intermediate is not formed in the case of SH attack on carbonyl as a result of a weak interaction between the inner σ orbitals involved in the new bond formation. A more detailed discussion on this subject will be published elsewhere.

Our aim is to design new compounds that may serve as inhibitors of the catalytic reaction of cysteine proteases; thus, if a stable intermediate will be formed instead of the original tetrahedral structure, it could stop the reaction from propagating to the next step. Analyzing the MOs again, our conclusion is that we should look for a compound with an acceptor σ orbital which lies higher in energy than the C=O, so it may allow the σ interaction with σ-SH to occur. A possible candidate may be Si=O or C=S bonds. We have found that the calculated tetrahedral structure intermediate for silanone **4** is indeed stable and constitutes a minima on the reaction potential surface. In principle, there types of compounds could serve as a starting point for drug design; however, there are no stable structures known which contain a free silanone **3**.[10]



3 **4**

Serine and Sulfhydryl Proteases vs Cysteine Viral Proteases. The catalytic triad of serine proteases constitutes three residues: serine, histidine, and aspartic acid. The catalytic triad of cysteine protease differs in two amino acids. Thus it is cysteine instead of serine and histidine and asparagine instead of aspartic acid. The spatial arrangement of the catalytic triad in both enzymes is very similar (see Figure 3), yet their catalytic path is probably different.

As has been suggested before[11] since the sulfur nucleophile reacts in a different manner than oxygen toward a carbonyl center, the mechanisms of the corresponding enzymes are different, particularly at the first stage of the reaction, e.g., in the formation of the tetrahedral intermediate. In cysteine proteases there occurs a protonation of the substrate carbonyl prior to the nucleophilic attack which enables the formation of the tetrahedral intermediate. It has been further claimed[11] that the *catalytic histidine* is the residue that donates a proton

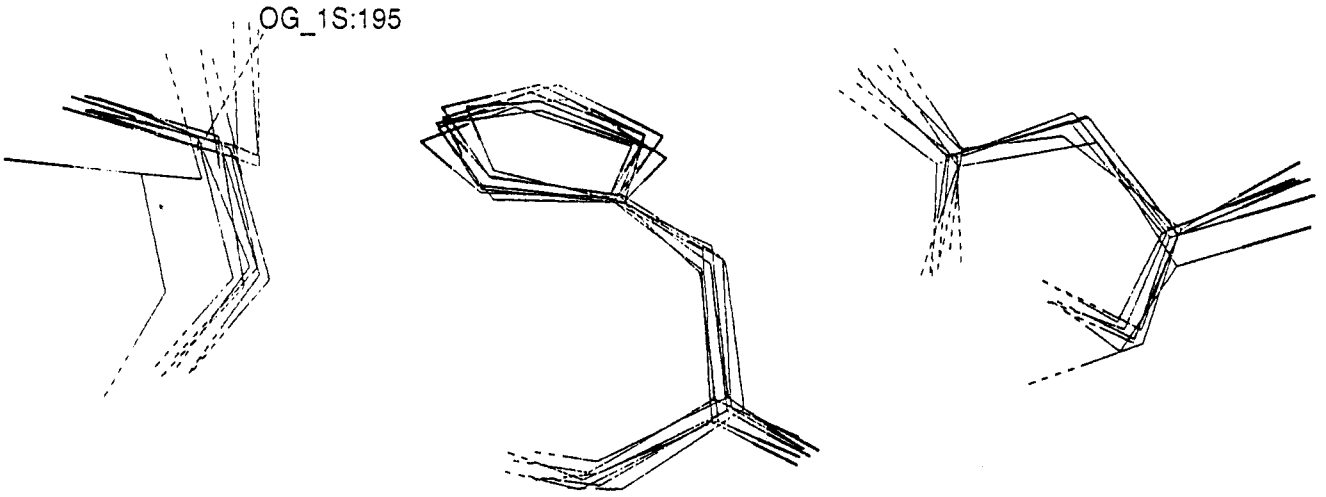


Figure 3. Spatial arrangement of catalytic amino acids of elastase, trypsin, chymotrypsin, sindbis, α-lytic protease, and bacterial trypsin.

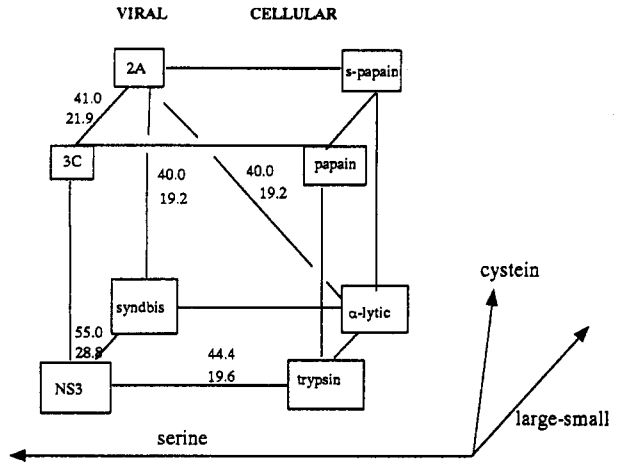


Figure 4. Homologies between different families of serine and cysteine proteases.

to the substrate, to protonate it, prior to nucleophilic attack. Thus, the role of histidine in serine proteases is to abstract a proton from the hydroxyl serine (as a part of the charge relay mechanism), whereas its role in cysteine proteases is to donate a proton to the substrate. This conclusion has raised questions on the role of the third residue member of the catalytic triad: Aspartic acid in serine proteases and asparagine in cysteine proteases. In serine proteases, its role is to extract a proton as a part of the chain reaction of charge relay.[19] In cysteine proteases, its role is to stabilize the charge on histidine and to increase its ability to donate a proton.

According to the same considerations, the catalytic role of histidine is different in both enzymes: In serine proteases, the role of histidine is presumably abstracting a proton from serine as a part of the charge relay mechanism, whereas in sulfhydryl proteases, the role of histidine is to donate a proton to the substrate to enable the nucleophilic attack of the sulfur anion on the carbonyl center.

By analyzing the possible role of the catalytic amino acids that exist in 3C and by implying the above considerations to the mechanism that may operate in the 3C viral cysteine proteases family of enzymes, one derives the conclusion that these enzymes differ mechanistically from serine and cysteine proteases and may constitute a new group of enzymes.

The 3C protease from the Picornavirus family is classified as a "cysteine-like serine protease".[12] It was hypothesized[13] that the enzyme structurally resembles trypsin (by homology) and reacts mechanistically as serine proteases (charge relay mechanism[14]), although its catalytic triad, as found by homology and by site-directed mutagenesis, is not quite the same as that for trypsin since it contains a cysteine nucleophile (Cys 147) instead of a serine nucleophile (Ser 195). Thus, the catalytic triad of 3C constitutes the following residues: cysteine, histidine, and aspartic acid. As explained before, a tetrahedral intermediate cannot be formed with sulfur as a nucleophile, unless histidine participates in the activation state by donating a proton to the substrate. However, in the hybrid-type catalytic site of 3C protease, histidine is not readily available to donate a proton since the third catalytic residue, aspartic acid, reduces its proton donor ability. (Please note that in cellular cysteine proteases, the third catalytic residue is asparagine, which only stabilizes the charge on histidine and does not withdraw a proton as does the more acidic aspartic acid.)

Thus, the catalytic triad hypothesized for 3C, as is, is not very efficient *by itself* for activating and catalyzing the proteolytic process. If the role of histidine is to donate a proton to the substrate as it is in cysteine proteases, then His 40 (of 3C) when part of the His-Asp pair is a poor proton donor, and, thus, its ability to donate a proton decreases. The conclusion is that there should be other important catalytic residues which are responsible for activating the substrate toward the nucleophilic attack, and one should search for this group in the protein sequence using structural considerations.

Classification, Comparative Sequence Alignment, and Model Building of 3C Protease. Bazan and Fletteric have suggested[12] that the 3C proteases from the Picornavirus family are homologous in structure to trypsin. This suggestion has been made by multiple sequence alignment and by analyzing the secondary structures where it was found that in 3C there are 12 β-barrels, as is found in trypsin. Bazan and Fletteric have also compared the 3C proteases to large trypsin and the 2A proteases to α-lytic protease, which is a "small" serine protease, and have found homologies between both families. We have continued the analysis along this line and found some clues which would help us to build a reasonable model.

Homology within the 3C Family. The amino acids that are found by site-directed mutagenesis to be important in determining the activity of 3C proteases are Cys 147, His 40, Asp 86.[15] These amino acids can be well-aligned with the catalytic triad of trypsin, Ser 175, His 57, Asp 102, and so are the regions neighboring to the active site. Other authors have argued the Asp 86 is predicted by their model to be in a connective loop between two domains and, thus, is not located in the active-site region.[16] They suggested that Glu 72 may serve as the third corner of the catalytic triad. We have extended the sequence alignments to 3C enzymes that are not included in the Picornavirus family but are homologous to the

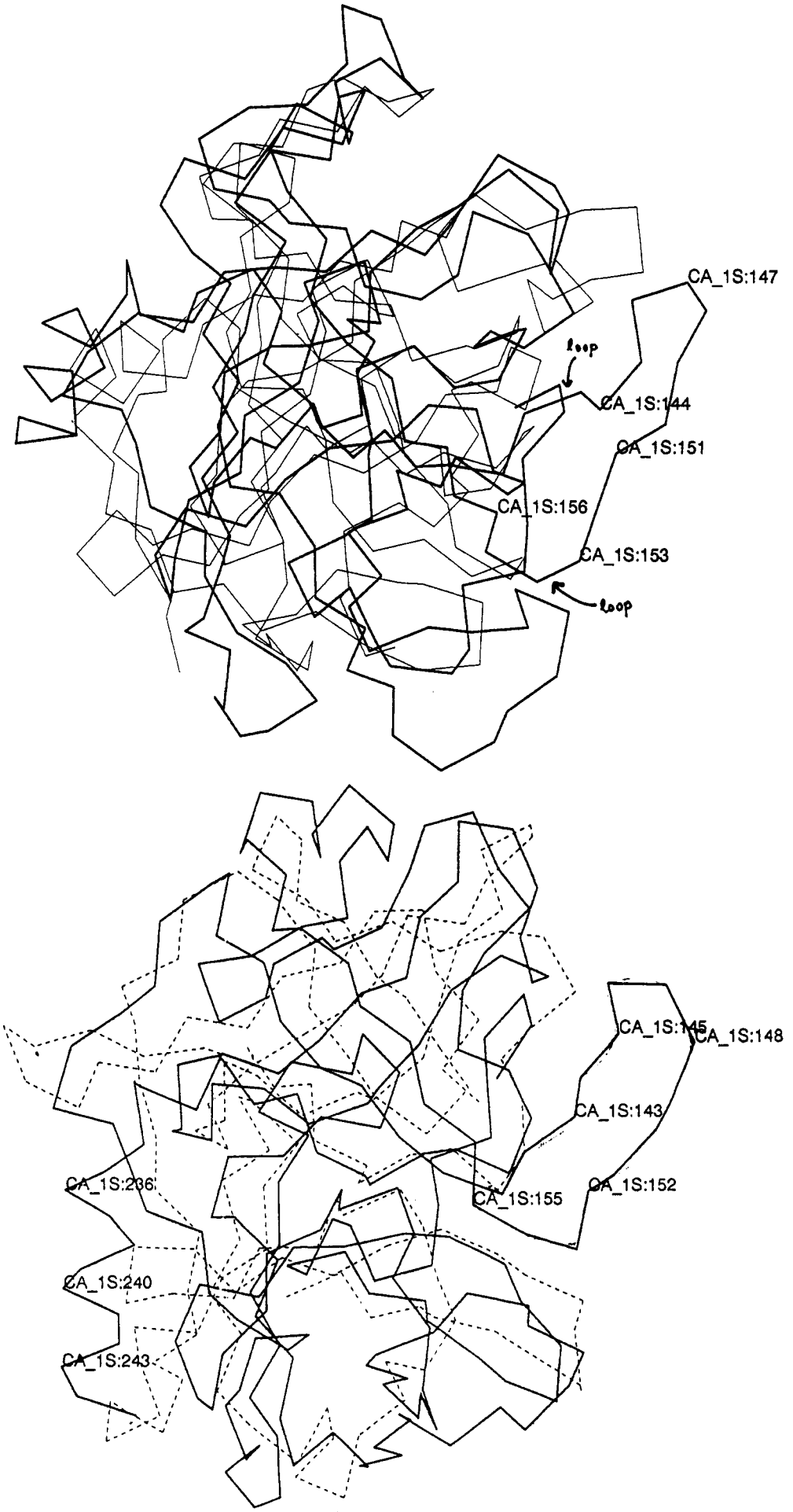


Figure 5. Plots for (a) trypsin (bold line) and α-lytic protease (regular line) and (b) trypsin (bold line) and sindbis (dotted line). Note the additional loop (132–164) in trypsin which is missing in α-lytic protease and sindbis.

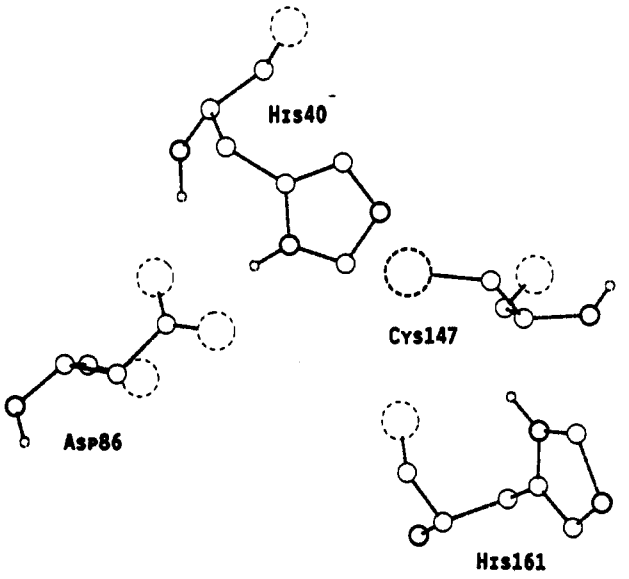


Figure 6. Arrangement of catalytic amino acids in a model structure constructed for 3C enzyme. His 161 is located near to the active site in an opposite position to His 40, relative to a line drawn between the cysteine and aspartic acid.

3C Cys-type proteases (such as NIa) and found that the only residues which are conserved along all the series are Cys 147, His 40, Asp 86, and an additional histidine, His 161. It was shown that mutation of His 161 reduces the enzyme's activity;[17] however, this effect was attributed to the posssible role of His 161 in binding. *Does this additional histidine (161) have a role in catalysis*? The structural information deducted from constructing a model gave some hints for answering this question. Analyzing the sequences of other related families helped in elucidating structural inforamtion (Figure 4). Figure 4 presents the groups of enzymes studied and the homologies between them. A distinction is made between small and large enzymes (trypsin and α-lytic), cellular and viral enzymes, and serine-like vs cysteine-like enzymes. The numbers present the similarity and the degree of identity between the groups.

Model Building. We have constructed a rough model based on homology to trypsin. This model differs from the models that others have constructed[18] in the way that it is based on the similarity of the three catalytic residue regions. The region that has the largest number of identical amino acids is the catalytic Cys region (GQCGG in 3C and 2A vs GDSGG in trypsin). In order to align the coordinates of this region, we chose to remove a loop in the trypsin from residue 132 to residue 164 (Figure 5). This loop, as seen in Figure 5 is present in the large-type serine proteases, as trypsin, but is absent in small-type serine proteases as α-lytic (cellular) or sindbis (viral). Since the size of the 3C comprises with the size of the small serine proteases, the reduction of this particular loop in order to align the active sites seemed to be a reasonable choice.

The aligned regions are displayed, the missing coordinates are being added, and the whole structure is subjected to energy minimization.

The constructed model that is obtained is very rough, and certainly it is not an accurate estimation for the structure of 3C. Yet, the striking observation is that the additional conserved histidine—His 161—is located at the active-site region, opposite the second histidine, His 40. The active-site residues arrangement is shown in Figure 6.

This histidine may have an important role in the catalytic mechanism. This role, either in the catalytic path or in binding, has yet to be determined. Simulations and experiments that are aimed at understanding the mechanism of 3C in general, and the role of His 161 specifically, are now in progress in our laboratory. We are also applying new modeling techniques for constructing a more accurate model that might help to define some other residues that could be important in the catalytic process.

ACKNOWLEDGMENT

The Lady Davis Fellowship Trust, Jerusalem, Israel (for D.A.), the Bertha (Hartz) Axel Fellowship (for R.K.), the Ministry for Science and Technology and Ministry for Absorbing New Emigrants (for M.S.) are acknowledged. The work was supported by the German-Israel cooperation for Biotechnology, to which we are thankful.

REFERENCES AND NOTES

(1) Rueckert, R. R. In *Virology*; Firlds, N., Knipe, D. M., Eds.; Raven: New York, 1989; pp 507–548.
(2) (a) Palmenberg, A. C. *Annu. Rev. Microbiol.* **1990**, *44*, 603–623. (b) Krausslich, H. G.; Nicklin, M. J. H.; Lee, C. K.; Wimmer, E. *Biochimie* **1988**, *70*, 119–130.
(3) Frisch, M. J.; Head-Gordon, M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, D.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *Gaussian 90*; Gaussian, Inc.: Pittsburgh, 1990.
(4) (a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542. (b) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. Protein Data Bank. In *Crystallographic Databases—Information Content, Software Systems, Scientific Application*; Allen, F. H., Bergerhoff, G. Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn, 1987; pp 107–132.
(5) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. I.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217. (b) *QUANTA version 3.1.2.* Poygen Inc.: 1991.
(6) Kristofferson, D. GenBank Stats, Genbank.bio.net, 1992.
(7) Howard, A.; Kollman, P. A. *J. Am. Chem. Soc. 110*, 7195–7200.
(8) Cardy, H.; Loudet, M.; Policard, J.; Oliver, J.; Poquet, E. *Chem. Phys.* **1989**, *131*, 227.
(9) Albright, T. A.; Burdett, J. K.; Whangbo, M.-H. In *Orbital Interactions in Chemistry*; Wiley: New York, 1990; Chapters 3 and 5.
(10) *The Chemistry of Silicon Compounds*; Rappoport, Z., Patai, S., Eds.; Wiley: New York, 1988.
(11) Arad, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491–502.
(12) Bazan, J. F.; Fletterick, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7872–7876.
(13) Blinov, V. M.; Donchenko, A. P.; Gorbalenya, A. P. *Dokl. Acad. Nauk SSSR* **1985**, *281*, 984–987.
(14) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-R. *Biochemistry* **1989**, *28*, 3629–3637.
(15) Cheah, K.-C.; Leong, L. E.-C.; Porter, A. G. *J. Biol. Chem.* **1990**, *265*, 7180–7187.
(16) Gorbalenya, A. E.; Dochenko, A. P.; Koonin, V. M. *FEBS Lett.* **1989**, *243*, 103–114.
(17) (a) Ivanoff, L. A.; Towatari, T.; Ray, J.; Korant, B.; Petteway, S. R., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5392–5396. (b) Baum, E. Z.; Bebernitz, G. A.; Palant, O.; Mueller, T.; Plotch, S. J. *Virology* **1991**, *185*, 140–150.
(18) Bazan, J. F.; Fletterick, R. *FEBS Lett.* **1989**, *249*, 5–7.
(19) Craig, C. S.; Roczniak, S.; Largman, C.; Rutter, W. *Science* **1987**, *237*, 909–913.