Acta Cryst. (1997). D53, 682–690

Experiences from the Structure Determination of Human Cytomegalovirus Protease

LIANG TONG,[a]* CHUNGENG QIAN,[a] WALTER DAVIDSON,[a] MARIE-JOSÉE MASSARIOL,[b] PIERRE R. BONNEAU,[b] MICHAEL G. CORDINGLEY[b] AND LISETTE LAGACÉ[b]

[a]*Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road/P.O. Box 368, Ridgefield, CT 06877, USA and* [b]*Bio-Méga Research Division, Boehringer Ingelheim (Canada) Ltd, 2100 rue Cunard, Laval, Canada H7S 2G5. E-mail: b4w@cc.purdue.edu*

(Received 10 January 1997; accepted 6 May 1997)

Abstract

Several obstacles were encountered and overcome during the structure determination of human cytomegalovirus protease. Dehydration of crystals, by exposing them to higher concentrations of the precipitant, reduced the mosaicity of the crystals and may have also resolved their microscopic twinning. The initial phase information was obtained with the selenomethionyl multiple-wavelength anomalous diffraction technique. However, site-specific mutagenesis was required to introduce extra Met residues into the protease. The phase information had to be improved by non-crystallographic symmetry averaging, initially among three 'crystal forms'. A change in the composition of the artificial mother liquor led to a significant improvement, from 3.0 and 2.0 Å resolution, in the diffraction quality of the crystals. The experiences reported here may prove useful to structure determination of other proteins.

1. Introduction

Human cytomegalovirus (HCMV), a herpesvirus, is widely distributed in the general population. Most HCMV infections are clinically asymptomatic (Fields & Knipe, 1990). However, in immunosuppressed and immunocompromised individuals (AIDS patients and organ-transplant recipients), HCMV infections can cause severe health problems and even death (Fields & Knipe, 1990). The life-cycle of a herpesvirus requires the correct functioning of the herpesvirus protease which is essential for the proteolytic processing and maturation of the herpesvirus assembly protein (for a review, see Gibson, Welch & Hall, 1995). Herpesvirus mutants that contain a defect in this processing can only produce aberrant empty capsids. HCMV protease, therefore, is an attractive target for therapeutic intervention in the treatment of HCMV infections.

Herpesvirus protease is a serine protease, with Ser132 (HCMV protease numbering) as the active-site nucleophile (Gibson *et al.*, 1995). His63 is believed to be the second member of the putative catalytic triad. Several possible candidates (Asp and Glu residues) have been proposed to be the third member of this triad, based on mutagenesis studies (Gibson *et al.*, 1995). The amino-acid sequences of herpesvirus proteases have no detectable homology to other serine proteases or hydrolases. The crystal structure of HCMV protease has recently been determined in our laboratory by the seleno-methionyl multiple-wavelength anomalous diffraction (MAD) technique (Tong *et al.*, 1996), and in other laboratories by the heavy-atom isomorphous replacement technique (Qiu *et al.*, 1996; Shieh *et al.*, 1996; Chen *et al.*, 1996). It showed that HCMV protease possesses a new polypeptide backbone fold; confirmed that Ser132 and His63 are in close proximity in space; and identified His157 as the possible third member of the catalytic triad.

During the structure determination by the selenomethionyl MAD technique, several major hurdles had to be tackled and overcome. These included treatment of crystals to reduce mosaicity and to remove microscopic twinning, site-specific mutagenesis to introduce extra methionine residues, non-crystallographic symmetry (NCS) averaging among three crystals to improve the electron-density map, and change of artificial mother liquor that improved the diffraction quality of the crystals from 3.0 to 2.0 Å resolution. Our experiences with these problems, and their solutions, may be useful to other crystallographic structure determinations. Therefore, we present here in more detail the structure determination of HCMV protease. A flow chart of the structure determination process is given in Fig. 1.

2. Treatment of crystals

The cloning, expression and purification of HCMV protease will be described elsewhere. The protease samples used in the current study all contain the A143Q mutation (to eliminate one of the internal cleavage sites of HCMV protease) (Pinko *et al.*, 1995). Light-scattering studies showed that the protease was monodisperse in solution and existed as dimers. This observation was made before the appearances of published reports showing the presence of the dimer

Table 1. *Summary of crystallographic information*

Crystal form	Protein sample	Unit cell (a, c) (Å)	Resolution (Å)	No. of observations	No. of reflections	R_{merge} (%)	Dimer twofold orientation (φ, ψ)*	Dimer center (x, y, z)
A	A143Q L229M Se-Met	70.1, 215.4	3.0	313575†	10770	7.7	90.5, 88.5	(0.512, 0.504, 0.125)
B	A143Q	68.5, 211.5	3.0	44026	9432	9.3	101.6, 85.9	(0.453, 0.034, 0.116)
C	A143Q	72.7, 215.3	3.0	157723‡	11714	9.9	84.1, 90.5	(0.508, 0.008, 0.126)
D	A143Q T181M L229M Se-Met	71.0, 209.3	2.0	165253	34045	7.1	99.5, 82.6	(0.447, 0.042, 0.116)

* φ is measured from the unit-cell a axis, and ψ is measured from the c axis. † Reflection data from the four wavelengths were merged. ‡ Reflection data from two crystals were merged.

and suggesting that the dimer is the active form of the enzyme (Darke *et al.*, 1996; Margosiak, Vanderpool, Sisson, Pinko & Kan, 1996). Initial crystallization conditions were found with the sparse-matrix sampling technique (Jancarik & Kim, 1991), using a commercial kit (Hampton Research). A total of six out of 96 conditions gave small crystals. One of these conditions was optimized to produce larger crystals. The reservoir solution contained 0.1 *M* 2-[*N*-morpholino] ethane sulfonic acid (MES) (pH 6.0), 16% PEG 4000, 0.4 *M* LiCl, 10% glycerol, and 5% *t*-butanol. The protein was at about 25 mg ml^{-1} concentration in a solution containing 20 m*M* NaAc (pH 5.0), 80 m*M* NaCl and 1 m*M* dithiothreitol (DTT). The crystals, grown at room temperature, were long tetragonal prisms, the largest of which measured 0.15 × 0.15 × 2.0 mm.

It became clear when these crystals were exposed to X-rays that they were highly mosaic and microscopically twinned. Some of the reflections had an elongated shape or even appeared as two closely spaced spots in the diffraction image. As expected from the crystal morphology, the diffraction pattern suggested that the crystal might be tetragonal. However, the *a* and the *b* axes of the unit cell were given different lengths by auto-indexing because of the twinning problem. Dehydrating these crystals might reduce the mosaicity and help solve the twinning problem. The PEG concentration in the reservoir was raised in a few steps from 16 to 30%, over 3 to 5 d. The crystals were then transferred to an artificial mother liquor containing 0.1 *M* MES (pH 6.0), 30% PEG 4000, 0.4 *M* LiCl, 10% glycerol, and 5% *t*-butanol. Crystals that were treated in this manner were found to have lower mosaicity. In addition, the diffraction pattern from the treated crystals no longer showed any signs of twinning.

Initial X-ray diffraction data were collected on an R-Axis imaging-plate system mounted on a Rigaku RU-200 rotating-anode X-ray generator. The crystals diffracted weakly and anisotropically, with the best diffraction generally extending to about 3.0 Å resolution along c^* and about 3.3 Å along a^* (and equivalently b^*). The treatment with higher PEG concentrations did not remove the anisotropy in the X-ray diffraction of these crystals. Depending on the treatment protocol, the crystals displayed large variation in cell dimensions, with unit-cell volume changes of up to 10% among crystals. The length of the unit-cell *a* axis varied between 68 and 73 Å, and the *c* axis varied between 209 and 215 Å (Table 1). Reflection data from four different types of crystals were used in this structure determination. These different crystals are loosely called 'crystal forms' here (*A*, *B*, *C*, and *D*; see Table 1). The space group is either $P4_12_12$ or $P4_32_12$, with a dimer of the protease in the asymmetric unit. The V_m value for a dimer in the asymmetric unit is about 2.4 Å^3 Da^{-1}. For crystal forms *A* and *C*, pseudo-extinction in the diffraction pattern for $h + k = 2n + 1$ reflections (pseudo *C*-centering) suggested that the twofold axis of the NCS dimer might be aligned with the crystal *a* (or equivalently *b*) direction. This was also seen in the native Patterson map, which contained a large peak at (0.5, 0.5, 0) (not shown).

3. Mutagenesis to introduce extra methionine residues

The search for heavy-atom derivatives for these crystals was hampered by the lack of isomorphism among crystals. The seleno-methionyl MAD technique represented an alternative for this structure solution (Hendrickson, 1991). HCMV protease contains three Met residues. However, residue Met1 is absent in about 50% of the protein molecules, as shown by electrospray mass spectrometry, and residue Met3 is likely to be flexible. This leaves only one Met residue, Met75, for the 256 residues of the protease. This was not expected to produce enough signals for the MAD experiment. Hence, it was necessary to introduce extra methionine residues into the protease.

The aligned amino-acid sequences of herpesvirus proteases were examined to find residues in HCMV protease that could be mutated to Met without

significantly affecting its activity. Two residues, Leu222 and Leu229, were selected for mutagenesis. The basic consideration in this selection was that a hydrophobic residue in HCMV protease could be mutated to Met if it corresponded to a Met residue in other herpesvirus proteases. A hydrophobic residue was chosen so that the resulting Met side chain could be buried and thus ordered. The L229M mutant had about 90% of the catalytic activity of the native protease. The L222M/L229M double mutant, however, was found to have only 10% of the activity. Therefore, the L229M mutant was selected for structural work. To obtain larger signals from the MAD experiments, a second position for mutation was later found at residue 181. The T181M/L229M double mutant had about 60% of the catalytic activity.

4. Seleno-methionyl MAD phasing with the L229M mutant

The L229M mutant was used for seleno-methionyl MAD phasing first. The seleno-methionyl protease was produced in a Met-auxotropic *E. coli* cell line and purified using the same protocol as the native protease. Quantitative incorporation of Se-Met residues into the protease was confirmed by electrospray mass spectrometry. Crystals of the seleno-methionyl protease were obtained under similar conditions as those for the native protease. Sodium bisulfite (5 m*M*) was added to the reservoir as an antioxidant to protect the seleno-methionyl residues.

The X-ray diffraction data were collected at beamline X4A at the Brookhaven National Laboratory. The crystal was flash-frozen, enabling diffraction data at four wavelengths (λ1:0.9919 Å, λ2:0.9793 Å, λ3:0.9792 Å, λ4:0.9724 Å) to be collected (crystal form *A*, Table 1). The crystal was aligned before the start of the data collection. A total of 34 images (2.2° per image, with 0.7° overlap between images) were collected around the c^* axis, starting from the a^*c^* zone. Then, 16 images (1.7° per image, with 0.7° overlap between images) were collected around the b^* axis to cover the missing cone. The crystal diffracted to about 3.0 Å resolution along c^* and about 3.3 Å along a^*. The crystal-to-detector distance was 300 mm. The diffraction pattern was recorded on Fuji image plates which were scanned with 0.1 mm raster step. The diffraction images were processed with *DENZO* and scaled with *SCALEPACK* (Otwinowski, 1993). The partial observations were discarded and the Friedel observations were treated separately during scaling.

The individual observations from the four wavelengths were loaded into the *MADSYS* package (Hendrickson, 1991) (Table 2). A Patterson map was calculated based on the F_A values and interpreted with the *PATSOL* program (Tong & Rossmann, 1993). Two strong sites were found in the Patterson map. They were related by a relative translation of (0.5, 0.5, 0), in agreement with the diffraction pattern which showed

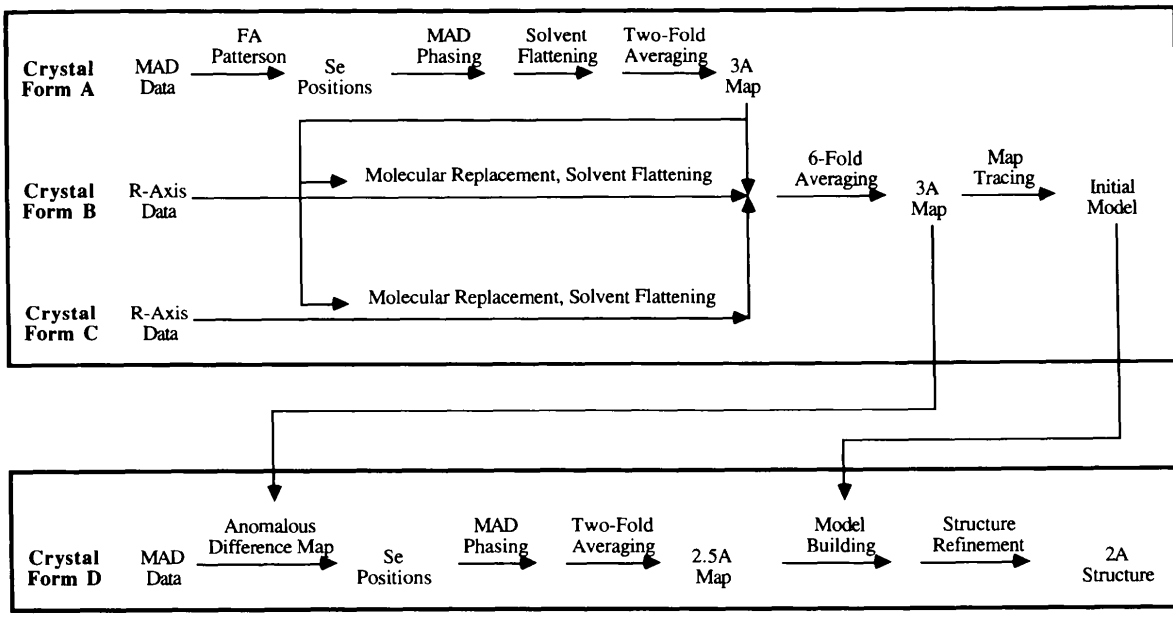

Fig. 1. Flow chart of the structure determination of HCMV protease. The L229M seleno-methionyl MAD phase information (crystal form *A*) was improved by solvent flattening and used to solve crystal forms *B* and *C* by molecular replacement. Subsequent sixfold averaging among the three crystal forms produced further improvement in the electron-density map. The phase information from this map was crucial to locate the Se positions in the T181M/L229M seleno-methionyl crystal (form *D*). Twofold averaging within this crystal form produced a readily interpretable electron-density map at 2.5 Å resolution. See text for detailed descriptions.

Table 2. *Summary of MAD phasing statistics for the L229M mutant*

Observed ratios are Bijvoet difference ratios (diagonal elements) at each wavelength (values in parentheses are for centric reflections) and dispersive difference ratios (off-diagonal elements) between pairs of wavelengths. $R(F_A)$ is the residual between the calculated structure factors based on the Se positions and those obtained from the MAD analysis. $\Delta(\Delta\varphi)$ is the average difference between independent determinations of $\Delta\varphi$ from the MAD analysis. For reflections between 15 and 3 Å resolution, $R(F_A) = 40\%$, $\Delta(\Delta\varphi) = 58°$.

Observed ratios (15–4 Å)	λ1	λ2	λ3	λ4	f' (e)	f'' (e)
λ1	0.039 (0.029)	0.045	0.041	0.036	−3.6	1.1
λ2		0.051 (0.033)	0.035	0.044	−11.0	3.6
λ3			0.065 (0.031)	0.038	−9.1	5.7
λ4				0.053 (0.031)	−4.0	4.0

pseudo *C*-centering. The Se sites were also solved by direct methods, with the program *SHELX* (Sheldrick, 1990). This revealed two additional minor sites, also related by a relative translation of (0.5, 0.5, 0). One of these minor sites was present in the F_A Patterson interpretation as well. MAD phasing was carried out using the four sites, in space groups $P4_12_12$ and $P4_32_12$. The resulting electron-density map for $P4_12_12$ had clearer solvent boundaries and was pursued further.

The electron-density map was of very poor quality, with no recognizable secondary structure elements. Initial attempts at improving the map by solvent flattening (Wang, 1985) were not successful. In the phase-combination step, the phases after solvent flattening were restrained to the original MAD phases and the overall phase shift by solvent flattening was about 60°. However, as the MAD phases were of very poor quality, it might be better not to restrain the phases in this case. Therefore, the solvent flattening was carried out without any phase restraints. The overall phase shift was about 70°. The electron-density map improved recognizably. A few α-helices could be seen in the map.

5. NCS averaging among three crystals

Although some secondary-structure elements could be recognized in the electron-density map after solvent flattening, the map could not be traced as the connectivity between these elements was poor. Moreover, not all the secondary-structure elements were visible in the map. NCS averaging was then attempted to improve the phases. As the seleno-methionyl crystal exhibited pseudo *C*-centering, the NCS twofold was assumed to lie along the *b* axis of the unit cell. A translation function (Tong, 1993) based on the overlap of NCS-related electron-density values clearly showed the position of this twofold axis in the unit cell (Fig. 2). The parameters for the orientation and position of this twofold axis were then determined more accurately, again by electron-density overlap (Tong, Choi, Minor & Rossmann, 1992; Tong & Rossmann, 1997). Twofold NCS averaging within this crystal, using a locally written program (Tong, unpublished work), produced only minor improvements in the electron-density map.

As mentioned above, various data sets on HCMV protease crystals had been collected on the R-Axis, representing different unit-cell parameters and volumes (Table 1). Performing NCS averaging among these different 'crystal forms' could produce greater improvement in the electron density. For this purpose, two additional crystal forms (*B* and *C* in Table 1) were chosen. The electron density for the dimer in the seleno-methionyl crystal (crystal form *A* in Table 1) was used as the model to solve these two crystal forms by molecular replacement, with the *Replace* program package (Tong, 1993) (Fig. 1). After solvent flattening, the twofold axes of the dimers in the three crystal forms were correlated with each other by electron-density overlap. In agreement with the large differences in unit-cell parameters, the orientation and position of the dimers showed significant differences among the crystal forms as well (Table 1). This was originally recognized by the large scaling *R* factors among the three data sets, which varied between 51 and 58%.

The NCS averaging across the three crystal forms was carried out using a locally written program (Tong, unpublished work). The correlation coefficients between observed and calculated structure factors for the three crystal forms increased from 0.75 to 0.85. The average phase change for each crystal form was about 60°. The resulting electron-density map could be tentatively traced manually (Fig. 3). The two Se positions in each monomer were used for the tracing. The ordered site was found to correspond to the L299M mutation position, whereas the mostly disordered site was found to be Met75. Several data sets (in crystal form *B*) collected earlier for crystals soaked in mercury compounds [$K_2Hg(SCN)_4$, K_2HgI_4 and others] were analyzed by difference electron-density maps. The best data set showed the presence of six weak Hg sites, obeying the NCS twofold axis. The three Hg sites in each monomer were also used for the tracing. They were found to be associated with Cys84, Cys87 and Cys161, respectively. Finally, the biochemical knowledge that Ser132 and His63 are probably close in space was also used in the tracing. An atomic model was built based on this trace with the *FRODO* program (Jones, 1978).

At this time, the T181M/L229M double mutant was produced and the diffraction data on the seleno-methionyl protein were first collected on the R-Axis. A difference electron-density map between this mutant and the native protein showed three sites for each

monomer, two of which corresponded to Met75 and L229M, respectively. The third site was within 3 Å of the Thr181 position in this initial atomic model, giving further support that the trace developed was probably correct.

6. Improved X-ray diffraction by a change of artificial mother liquor

The activity of herpesvirus proteases is enhanced by the presence of high concentrations of salt, such as 0.5–1 *M* sodium sulfate (Burck *et al.*, 1994; Hall & Darke, 1995; Darke *et al.*, 1996; Margosiak *et al.*, 1996). The artificial mother liquor for the crystals contained only 0.4 *M* LiCl, and hence was in the low-salt condition. In order to determine whether there are any conformational changes in HCMV protease in the high-salt condition, the LiCl component in the artificial mother liquor was replaced with sodium sulfate. However, only 0.15 *M* Na_2SO_4 could be used as higher concentrations were not soluble. Surprisingly, crystals of HCMV protease that were soaked in this new artificial mother liquor for two days exhibited a significant improvement in the diffraction quality. Diffraction beyond 2.5 Å was recognized on the R-Axis, and the anisotropic pattern in the diffraction was also absent in crystals treated with sodium sulfate.

Seleno-methionyl MAD data were collected on the T181M/L229M double mutant on the X25 beamline at the Brookhaven National Laboratory (crystal form *D* in Table 1). The crystal diffracted to 2.0 Å resolution and the diffraction images were recorded on a Mar Research imaging-plate system. The crystal was not aligned prior to data collection and a total of 25 images (with another 25 images at the inverse-beam position) were collected at each of the four wavelengths. The oscillation range per image was 2.0° and the crystal-to-detector distance was 270 mm. The diffraction images were processed with *DENZO* and scaled with *SCALEPACK*. The partial observations were used. The data processing was

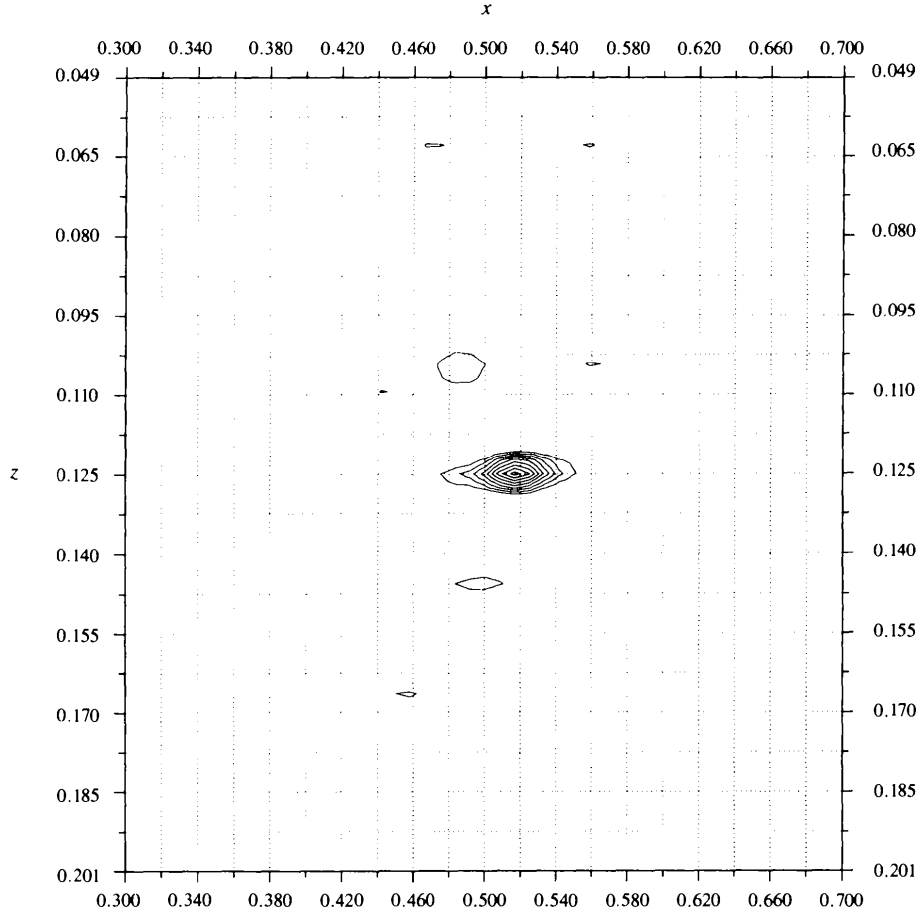

Fig. 2. Contour plot for a region on the $y = 0.5$ section of the special translation function to locate the positions of the NCS twofold axis. Reflection data between 10 and 3.5 Å resolution were used in the calculation, with MAD phases based on the L229M mutant (crystal form *A*) that had been improved by solvent flattening. The NCS twofold was assumed to be parallel to the *b* axis of the unit cell, as the diffraction pattern showed pseudo *C*-centering. The translation function showed that the NCS twofold axis is located roughly at (0.515, 0.5, 0.125).

carried out in two ways. First, data to 2.5 Å resolution were processed for all four wavelengths, using an overload value of 45 000 in *DENZO*. These data were used in the MAD phasing. Second, data to 2.0 Å resolution were processed for diffraction images collected at $\lambda 1$, using the default overload value of 105 000 in *DENZO*. This data set was used in subsequent phase improvement and structure refinement.

The phasing statistics to 2.5 Å resolution from the *MADSYS* package (Hendrickson, 1991) were rather poor (Table 3). It was difficult to locate the Se positions from the F_A Patterson map. The electron density for the dimer after NCS averaging among the three crystal forms was used as a model to solve the structure of this new crystal by molecular replacement. The resulting phases were used in an anomalous difference electron-density map for the data set at $\lambda 3$, which revealed the six Se positions. MAD phasing was also carried out with *X-PLOR* 4.0 (Brünger, 1992*a*), with $\lambda 4$ as the reference data set. The resulting electron-density map, from either *MADSYS* or *X-PLOR*, was not interpretable. The average difference between phases from *MADSYS* and *X-PLOR* was 60°.

To improve the phases, twofold averaging was carried out within this crystal form, starting from either

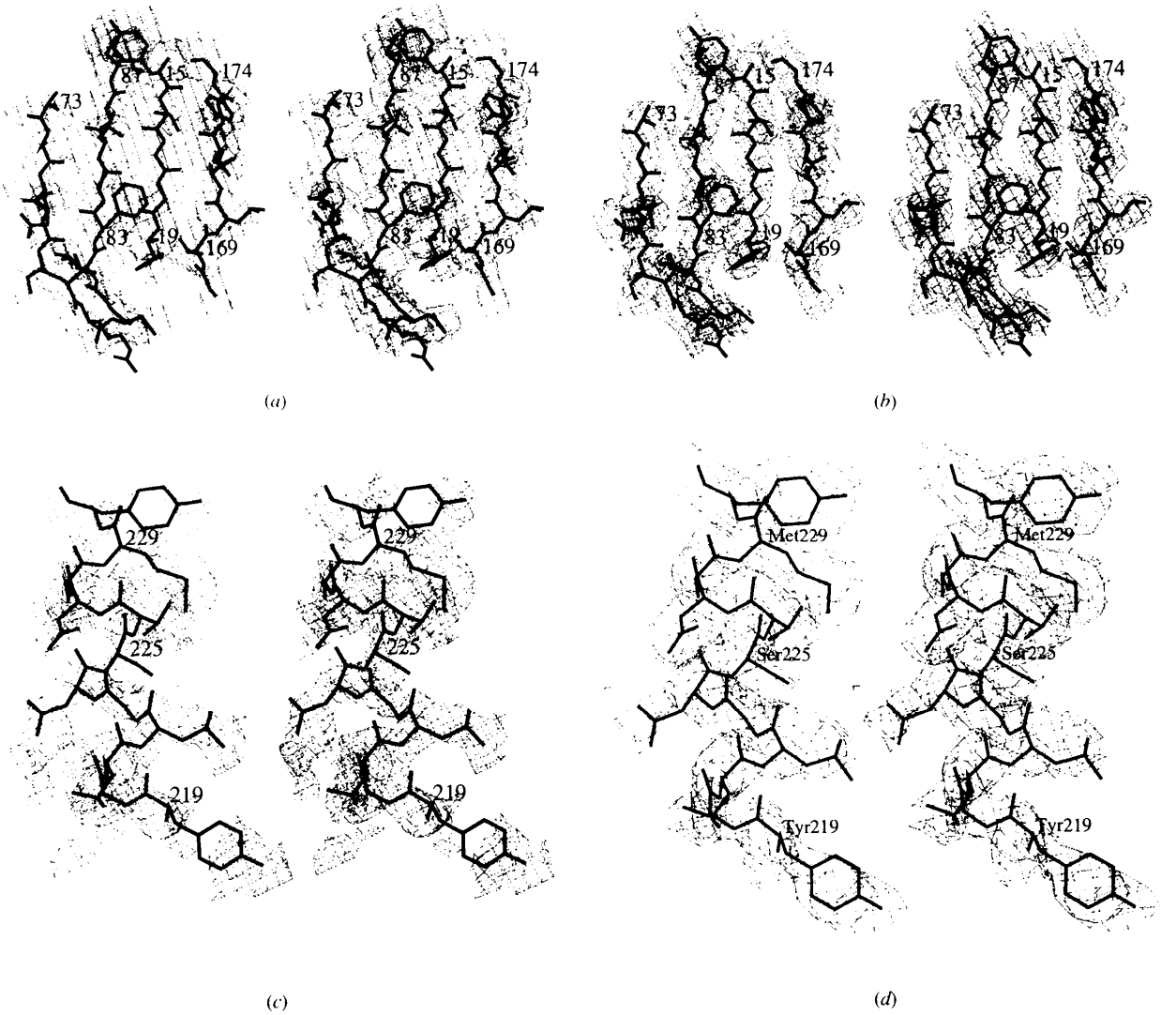

Fig. 3. Stereoscopic plot showing the electron density for some residues of HCMV protease at various stages of the structure determination. All contour levels are at 1σ. (*a*). Electron density at 3.0 Å resolution for residues 15–19, 73–87 and 169–174 after averaging among crystal forms *A*, *B* and *C*. The β-strands are not completely resolved due to the poor phase information and poor reflection data. (*b*) The electron density at 2.5 Å resolution for residues in (*a*) after twofold averaging in crystal form *D*. (*c*). The electron density at 2.5 Å resolution for residues 219–230 based on the MAD phases derived from the T181M/L229M mutant (crystal form *D*). (*d*). The electron density at 2.5 Å resolution for residues in (*c*) after twofold averaging in crystal form *D*.

Table 3. *Summary of MAD phasing statistics for the T181M/L229M mutant*

(*a*) From *MADSYS*

Observed ratios (15–3.5 Å)	λ1	λ2	λ3	λ4	f' (e)	f'' (e)
λ1	0.069 (0.055)	0.054	0.050	0.052	−3.6	1.1
λ2		0.080 (0.058)	0.041	0.056	−11.2	4.2
λ3			0.090 (0.058)	0.050	−8.8	4.8
λ4				0.081 (0.060)	−4.0	4.1

For reflections between 15 and 2.5 Å resolution, $R(F_A) = 45\%$, $\Delta(\Delta\varphi) = 61°$.

(*b*) From *X-PLOR*

Resolution (Å)	5.0	4.0	3.5	3.2	2.9	2.8	2.6	2.5	Overall
(Figure-of-merit)	0.80	0.73	0.69	0.65	0.60	0.57	0.54	0.53	0.64

the *MADSYS* or the *X-PLOR* phases. This led to a phase change of about 65° for each phase set. The average difference between the *MADSYS* and the *X-PLOR* phases was reduced to 40° by this averaging. The two electron-density maps after the averaging were similar and of very good quality (Fig. 3). The average main-chain electron-density value was raised from about 1σ to about 3σ by this averaging (Fig. 4). The side chains of amino-acid residues and the main-chain carbonyl O atoms for many residues could be clearly seen in the electron-density map (Fig. 3). The initial atomic model was found to be essentially correct and could be easily placed into the new electron-density map.

However, there was electron density for only about 190 amino acids out of the 256 total in HCMV protease. The atomic model was used in a partial model phase combination, with the *SIGMAA* program (Read, 1986), and a new helix was seen in the electron density. This helix, for residues 34–45, in the two monomers showed significant deviation from the twofold symmetry of the dimer. This may explain its weak electron density in the averaged map (Fig. 4). The current atomic model, after structure refinement at 2.0 Å resolution (Table 4), is still missing about 40 residues of each monomer of the protease. These residues are in regions of very low sequence homology among the herpesvirus proteases and may be disordered in structure (Tong *et al.*, 1996).

7. Discussion

The packing of HCMV protease molecules in these crystals is rather loose. The apparent V_m value of 2.4 Å^3 Da^{-1} seems normal. However, if the disordered residues are excluded from the calculation, the V_m value becomes 2.9 Å^3 Da^{-1}. The higher mosaicity of the original crystals was probably due to this loose packing

Table 4. *Summary of structure refinement statistics*

Resolution range for refinement (Å)	6.0–2.0
No. of reflections in refinement ($F > 2\sigma$)	30822
Reflection data completeness (%)	90
R factor (%)	22.1
Free R factor (Brünger, 1992*b*) (%)	28.8
R.m.s. deviation in bond lengths (Å)	0.010
R.m.s. deviation in bond angles (°)	1.8
No. of protein residues	419
No. of solvent molecules	1 sulfate + 249 waters

of the protease molecules. By increasing the PEG concentration in the mother liquor and dehydrating the crystals, a tighter packing arrangement was achieved. This process seemed to have also successfully removed a microscopic twinning problem with these crystals. The introduction of 0.15 *M* sodium sulfate led to a dramatic improvement in the diffraction quality of these crystals. Two possible reasons could be identified based on this crystal structure. First, a sulfate ion was found at the interface between two molecules (not from the same dimer) in the crystal. This could lead to a tighter packing of the protease molecules. Second, sodium sulfate caused a small change in the organization of the HCMV protease dimer in these crystals (Tong *et al.*, 1996). This might also have an impact on the crystal packing, as evidenced by the fact that crystal form *D* has the shortest *c*-axis length. Other possible effects of sodium sulfate on the structure of HCMV protease (for example, conformational rigidification of the monomer, stabilization of the dimer) may also play a role in this observed improvement in the diffraction. Further

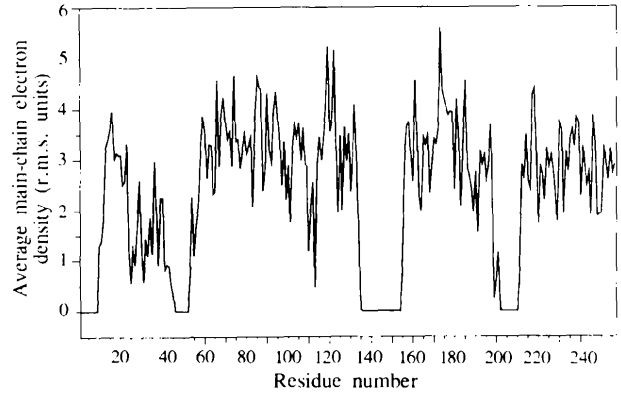

Fig. 4. Plot of the average main-chain electron density value (in r.m.s. units) (Tong *et al.*, 1992) for the second monomer in crystal form *D*, for the phases obtained directly from MAD phasing with the *X-PLOR* program (thin gray lines) and for the corresponding phases after NCS twofold averaging (thick solid lines). The electron-density values at the N, C_α and C positions, based on the current refined atomic model, were averaged for each residue. Very similar results were obtained when phases from the *MADSYS* package were used. Residues 25–45, which deviate from the twofold symmetry, did not show any improvement by the averaging. Those residues that are still missing in the current atomic model are given values of 0.

experiments showed that a sodium sulfate concentration of 25 m*M* was enough to produce the diffraction enhancement in these crystals.

In crystal forms *A* and *C*, the dimer twofold axis is close to being aligned with the crystal *b* (or equivalently *a*) direction. This led to a pseudo *C*-centered cell. If the dimer twofold axis were perfectly aligned with *b* and centered at (0.5, 0.5, 0.125), the unit cell would be perfectly *C*-centered and a smaller unit cell, with 50 Å axes along *a* and *b*, could be chosen (Fig. 5). In other words, the dimer along the $a+b$ direction of the smaller cell deviated from the crystallographic symmetry and became a non-crystallographic dimer in the current unit cell. The center of this dimer was found to be located near (0.5, 0.5, 0.125) or (0.5, 0, 0.125) (Table 1). These two positions are generally not equivalent in the larger unit cell. However, in the smaller unit cell, these two positions are related by an alternative origin translation of (0.5, 0.5, 0) and would therefore be equivalent. The deviation of the dimer twofold axis would lead to the differentiation between the two positions. For crystals that do not show pseudo *C*-centering, it would be expected that the deviation should be large enough such that it would be easy to distinguish between the two possible centers. This was found to be the case. If the dimer electron density was used as a model to solve crystal forms *B* and *D* by molecular replacement, only one of the centers appeared as the solution from the translation function. In contrast, for crystal forms *A* and *C*, both centers appeared as possible solutions, with similar peak heights.

Seleno-methionyl MAD phasing represents a useful alternative for protein structure determination (Hendrickson, 1991), especially for cases, like this one, where the search for heavy-atom derivatives proves difficult. The success of the current structure determination did rely on the introduction of extra methionine residues into the protein. The method that was used here for selecting potential sites for mutagenesis into methionines should be applicable to other proteins.

NCS averaging is a powerful tool for phase improvement for cases with high local symmetry (Rossmann, 1990). With only a twofold NCS, the power of averaging is generally weaker. Introduction of additional crystal forms into the averaging process could help the phase improvement. In this case, crystals belonging to the same space group but having significantly different unit-cell parameters were utilized as additional crystal forms. The large scaling *R* factors among the three data sets suggested that they probably represented sufficiently different sampling of the molecular transform to be treated as separate crystal forms. A recognizable improvement in the electron-density map was achieved through this averaging, enabling a tentative trace of the polypeptide backbone to be established. Greater improvement in the electron density was probably hampered by conformational differences among the crystals and, perhaps more importantly in this case, by the poor quality of the X-ray diffraction data used in this averaging. In comparison, twofold averaging within crystal form *D*, using a much better diffraction data set, successfully produced an excellent electron-density map at 2.5 Å resolution.

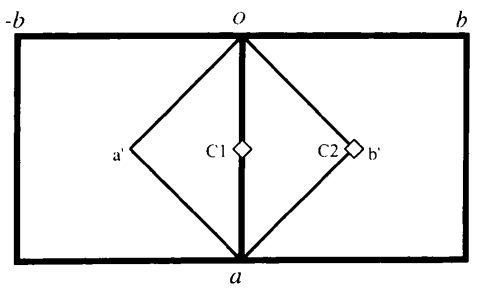

Fig. 5. The pseudo *C*-centering of the HCMV protease crystals was due to the alignment of the dimer twofold axis with the *a* direction of the unit cell. In a truly *C*-centered cell, a smaller unit cell can be chosen, with *a'* and *b'* as the cell edges. The dimer centers at C1 (observed for crystal forms *B*, *C*, and *D*) and C2 (observed for crystal form *A*) are equivalent in this smaller cell, but they are generally not equivalent in the larger cell.

We thank Hao Wu for discussions and help with data collection at the Brookhaven National Laboratory; Craig Ogata and David Cook at the X4A beamline, and Lonny Berman, Malcolm Capel and Zhijian Yin at the X25 beamline for their help with the data collection; Wayne Hendrickson for the *MADSYS* package; Axel Brünger for a pre-release version of *X-PLOR* 4.0; Randy Read for the *SIGMAA* program; and Sina Ghaemmaghami for his contribution in the production of seleno-methionyl HCMV protease mutants.

References

Brünger, A. T. (1992*a*). *The X-PLOR manual*, version 3.0. Yale University, New Haven, CT, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–475.

Burck, P. J., Berg, D. H., Luk, T. P., Sassmannshausen, L. M., Wakulchik, M., Smith, D. P., Hsiung, H. M., Becker, G. W., Gibson, W. & Villarreal, E. C. (1994). *J. Virol.* **68**, 2937–2946.

Chen, P., Tsuge, H., Almassy, R. J., Gribskov, C. L., Katoh, S., Vanderpool, D. L., Margosiak, S. A., Pinko, C., Matthews, D. A. & Kan, C.-C. (1996). *Cell*, **86**, 835–843.

Darke, P. L., Cole, J. L., Waxman, L., Hall, D. L., Sardana, M. K. & Kuo, L. C. (1996). *J. Biol. Chem.* **271**, 7445–7449.

Fields, B. N. & Knipe, D. M. (1990). Editors. *Virology*, Vol. 2, ch. 64–73. New York: Raven Press.

Gibson, W., Welch, A. R. & Hall, M. R. T. (1995). *Perspect. Drug Discov. Design*, **2**, 413–426.

Hall, D. L. & Darke, P. L. (1995). *J. Biol. Chem.* **270**, 22697–22700.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Margosiak, S. A., Vanderpool, D. L., Sisson, W., Pinko, C. & Kan, C.-C. (1996). *Biochemistry*, **35**, 5300–5307.
Otwinowski, Z. (1993). In *Data Collection and Processing* edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Pinko, C., Margosiak, S. A., Vanderpool, D., Gutowski, J. C., Condon, B. & Kan, C.-C. (1995). *J. Biol. Chem.* **270**, 23634–23640.
Qiu, X., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W. & Abdel-Meguid, S. S. (1996). *Nature (London)*, **383**, 275–279.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Shieh, H.-S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C. & Stallings, W. C. (1996). *Nature (London)*, **383**, 279–282.
Tong, L. (1993). *J. Appl. Cryst.* **26**, 748–751.
Tong, L., Choi, H.-K., Minor, W. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 430–442.
Tong, L., Qian, C., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G. & Lagacé, L. (1996). *Nature (London)*, **383**, 272–275.
Tong, L. & Rossmann, M. G. (1993). *J Appl. Cryst.* **26**, 15–21.
Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
Wang, B.-C. (1985). *Method Enzymol.* **115**, 90–112.