Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly

Chao-Zhou Ni[1], Rashid Syed[†], Ramadurgam Kodandapani[1], John Wickersham[1], David S Peabody[2] and Kathryn R Ely[1]*

[1]Cancer Research Center, La Jolla Cancer Research Foundation, La Jolla, CA 92037, USA and [2]Departments of Cell Biology and Biochemistry, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA

Background: The coat protein in RNA bacteriophages binds and encapsidates viral RNA, and also acts as translational repressor of viral replicase by binding to an RNA hairpin in the RNA genome. Because of its dual function, the MS2 coat protein is an interesting candidate for structural studies of protein–RNA interactions and protein–protein interactions. In this study, unassembled MS2 coat protein dimers were selected to analyze repressor activity and virus assembly.

Results: The crystal structure of a mutant MS2 coat protein that is defective in viral assembly yet retains repressor activity has been determined at 2.0 Å resolution. The unassembled dimer is stabilized by interdigitation of α-helices, and the formation of a 10-stranded antiparallel β-sheet across the interface between monomers. The substitution of arginine for tryptophan at residue 82 results in the formation of two new inter-subunit hydrogen bonds that further stabilize the dimer. Residues that influence RNA recognition, identified by molecular genetics, were located across the β-sheet. Two of these residues (Tyr85 and Asn87) are displaced in the unliganded dimer and are located in the same β-strand as the Trp→Arg mutation.

Conclusions: When compared with the structure of the coat protein in the assembled virus, differences in orientation of residues 85 and 87 suggest conformational adjustment on binding RNA in the first step of viral assembly. The substitution at residue 82 may affect virus assembly by imposing conformational restriction on the loop that makes critical inter-subunit contacts in the capsid.

Structure 15 March 1995, **3**:255–263

Key words: bacteriophage coat protein, crystal structure, RNA hairpin, translational repressor, virus assembly

Introduction

The coat protein from the MS2 bacteriophage is a member of a group of small proteins that bind RNA in a multifunctional manner in related RNA bacteriophages. These spherical viruses infect *Escherichia coli*, and their genomic RNA serves as mRNA for synthesis of viral proteins. Coat protein binds and encapsidates the viral RNA but also plays a genetic regulatory role. In the latter capacity, the protein effects translational repression of viral replicase synthesis by binding to the RNA operator of the replicase gene [1]. Because of its dual function, the MS2 coat protein is an interesting candidate for structural studies of protein–protein interactions as well as protein–RNA interactions. The coat protein from the MS2 bacteriophage (M_r=13700) is composed of 129 amino acids and self-aggregates to form an icosahedral shell (180 subunits) which binds and encapsidates a single-stranded RNA genome of 3569 nucleotides. Viral RNA encodes a maturation or A protein, the coat protein protomer, a replicase subunit and a lysis protein. Late during the course of infection, coat protein binds to the translation initiation region of the replicase cistron and prevents ribosomes from initiating translation there. The operator is located within an RNA hairpin. Interactions of coat protein with the RNA 'binding site' for translational repression are well characterized. It has been shown that the active repressor is a dimer and that one RNA operator molecule is bound by a repressor dimer at saturation [2]. This protein–RNA interaction is an important first event in nucleation of virus assembly [2].

The crystal structure of the intact MS2 virus has been determined [3,4]. In order to examine the conformation of isolated viral subunits, we crystallized the MS2 coat protein as a dimer utilizing a mutant molecule that was defective in viral capsid assembly, yet retained repressor activity [5]. Here we report the high-resolution structure of the unassembled dimer. This is the first time that high-resolution crystal structures have been available for a viral protomer in both the isolated and assembled states. Comparison of these structures offers a unique opportunity to address the molecular basis of subunit dimerization, RNA-binding, nucleation and assembly. First, critical functional residues in the dimer are described. For example, amino acids in the RNA-binding site have been defined by molecular genetics [6] and we describe here the structural organization of these residues in the dimer. Protein–protein contacts that stabilize the dimer are identified. The structure presented in this report permits examination of these residues in a conformation that is not influenced by capsid interactions.

The amino acid substitution that disrupts virus assembly in this mutant molecule is discussed as part of a comparison with the subunit in the crystal structure of the intact virus [3,4]. In the viral capsid, two types of dimers are

*Corresponding author. [†]Present address: Amgen, Inc., Thousand Oaks, CA 91320, USA.

formed from three coat protein conformers. Although these conformers are chemically identical, they differ in the configuration of a loop that is involved in five-fold or quasi-six-fold icosahedral contacts [4]. The structure of this loop in the repressor dimer is compared with the viral subunits. Finally, the structure of the dimer is compared with the two types of dimeric 'building blocks' in the capsid to identify conformational changes induced by RNA binding that possibly trigger capsid formation.

Results and discussion

In order to crystallize the MS2 coat protein in dimeric form rather than as virus capsids, we tested mutant proteins (described previously [5]) that were defective in viral assembly. These mutants were isolated using an *in vivo* two-plasmid genetic system in which coat protein expressed from one plasmid represses synthesis of a replicase-β-galactosidase fusion protein encoded by a second plasmid [7]. Thus, this system mimics the translational control of replicase synthesis that occurs during MS2 infection. Because the cloned coat protein assembles into virus-like particles within the bacterium after expression, it is also possible to use this system to test mutants for defects in viral assembly. We found several mutant proteins which formed dimers but were unable to assemble into capsids. These molecules were therefore ideal candidates for crystallization of the isolated dimer. These mutations were originally identified because the failure to assemble causes an elevation in dimer concentration and increased repressor activity. Because these dimers retain the capacity to bind the RNA operator, they are also suitable models for the functional dimeric repressor.

Crystallization

Several of the mutant proteins crystallize but one of them, SU647, crystallized very readily. SU647 contains a single substitution (Trp82→Arg). The protein was expressed in *E. coli* strain CSH41F$^-$, under control of the lac promoter, in a soluble form and purified to homogeneity by ion-exchange and gel-filtration chromatography. In solutions similar to those used for crystallization, the protein eluted from the molecular sizing column in a position corresponding to the molecular weight of a dimer rather than a capsid.

The repressor crystallized at 4°C and at pH 6.2 in citrate-phosphate buffer from solutions of polyethyleneglycol (PEG) 8000 in the orthorhombic space group $P2_12_12$ with a=76.2 Å, b=55.7 Å, and c=28.4 Å. The crystals diffract to 2.0 Å resolution. Calculations of the unit cell volume were consistent with one monomer in the asymmetric unit and a solvent content of 44%.

Quality of the refined model

After least-squares refinement, the R-factor was 0.20 for 962 protein atoms and 111 solvent atoms. The refinement statistics are shown in Table 1. The electron-density map, calculated at 2.0 Å resolution, was clear for all residues except for one loop connecting two β-strands (F and G). This is an important loop which is dicussed later. In this loop, the electron density was weak for main-chain and side-chains atoms from residues 67–81. However, the model was built into this density so that all density was accounted for with the model. The B-values for these atoms were significantly higher than the B-values for the rest of the molecule. A Ramachandran diagram [8,9] for the monomer is presented in Fig. 1, showing that φψ values for all non-glycine residues (except 66, 69 and 70) fall within energetically favored regions, including the residues in the FG loop.

Table 1. Refinement statistics.

	Final	Target
Resolution range (Å)	8.0–2.0	
Crystallographic R-factor (%)	20.00	
No. of reflections used	6668 >3σ(F)	
No. of atoms	1073	
Rms deviation from ideal distances (Å)		
Bond distances	0.021	0.020
Angle distances	0.066	0.045
Planar 1–4 distances	0.051	0.035
Rms deviation from ideal planarity (Å)	0.018	0.020

Structure of the dimer

The unassembled dimer is composed of two monomers related by an exact crystallographic two-fold axis. The dimensions of the dimer are 47 Å×67 Å×39 Å.

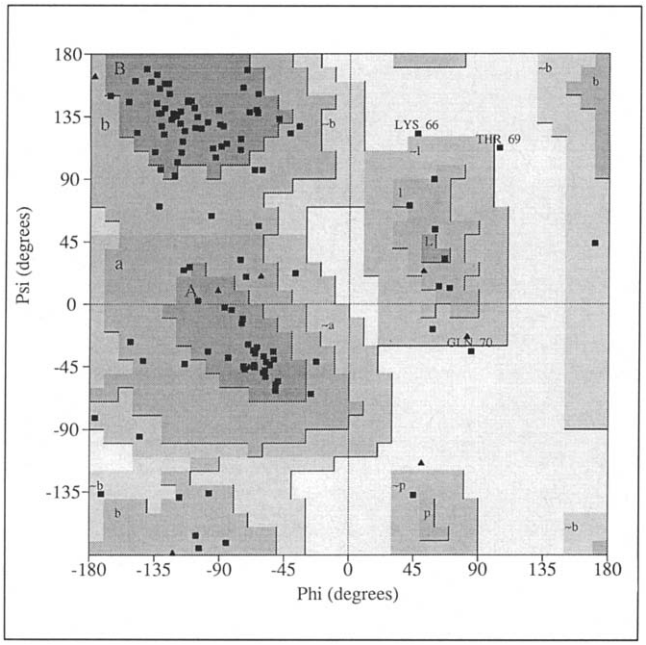

Fig. 1. Ramachandran diagram for the final model of the repressor. This diagram presenting φψ angles [8] was produced using the PROCHECK suite of programs [9]. Glycine residues are represented by triangles. Different regions of the plot are indicated by varying levels of shading with the energetically most favorable φψ angles falling in the darkest shaded areas. Note that only three residues in the FG loop (amino acids 66, 69 and 70) fall outside energetically favored regions.

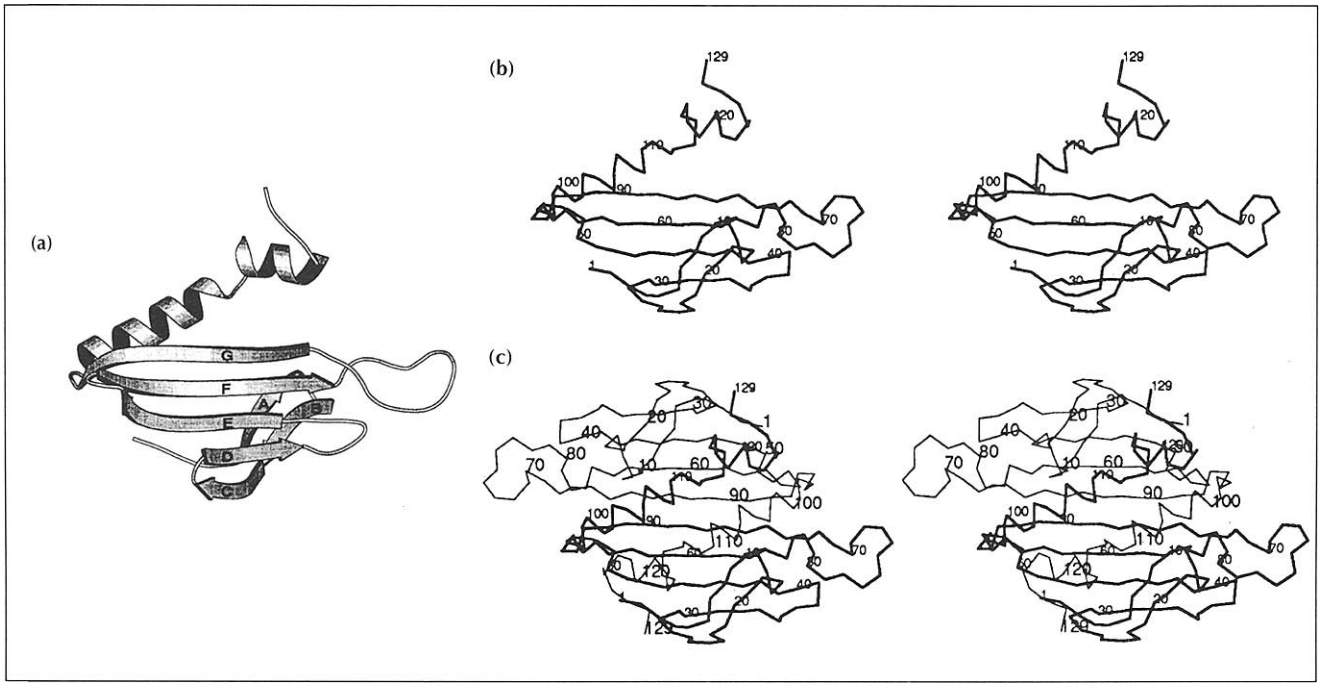

Fig. 2. Folding pattern of the MS2 molecule. **(a)** Ribbon diagram with directional arrows labeled to identify β-strands. (Figure generated with MOLSCRIPT [26].) **(b)** Stereo diagram of the α-carbon backbone of the monomeric asymmetric unit shown in the same orientation as in (a) with every tenth residue labeled. Note that the two helices comprising the C-terminal residues are clearly separated from the rest of the subunit. **(c)** The functional dimer generated from crystallographic two-fold symmetry. Individual monomers are drawn with thick or thin lines and are labeled with two sizes of numbers to aid in the examination of individual polypeptide chains. In the dimer, the α-helices interdigitate to stabilize the dimer, and hydrogen bonding occurs between adjacent antiparallel β-strands at the dimer interface.

As shown in Fig. 2, the monomer folding pattern consists of a β-pleated sheet composed of five antiparallel strands and two α-helices. The α-helices are separated from the β-sheet by a considerable distance (15 Å) and these two helices are nearly co-linear and can be described as a long distorted helix.

The dimer is stabilized by two types of interactions. First, the last β-strands (G) in each monomer are arranged in an antiparallel fashion with hydrogen bonds formed between main-chain atoms in the adjacent strands (see Fig. 3). This interaction forms a 10-stranded antiparallel β-sheet. On the other side of this sheet, the extended α-helices from opposing monomers interdigitate in an antiparallel fashion. Hydrophobic interactions stabilize the tight packing of these helical regions.

Comparison with viral subunits

The overall structure of the monomer is nearly identical to the structure of the viral protomer. The monomer is composed of 129 amino acids. When this monomer is superimposed on the viral subunits A, B and C, the root mean square (rms) differences for 107 pairs of corresponding α-carbons are 0.94 Å, 0.86 Å and 0.87 Å, respectively. These values are comparable (that is, within 0.7–0.9 Å rms difference) to those measured in a comparison of the three viral subunits [4]. The primary difference in the viral subunits and the monomer in our crystals is in the FG loop that assumes three different conformations in the intact virus. Two other loops differ

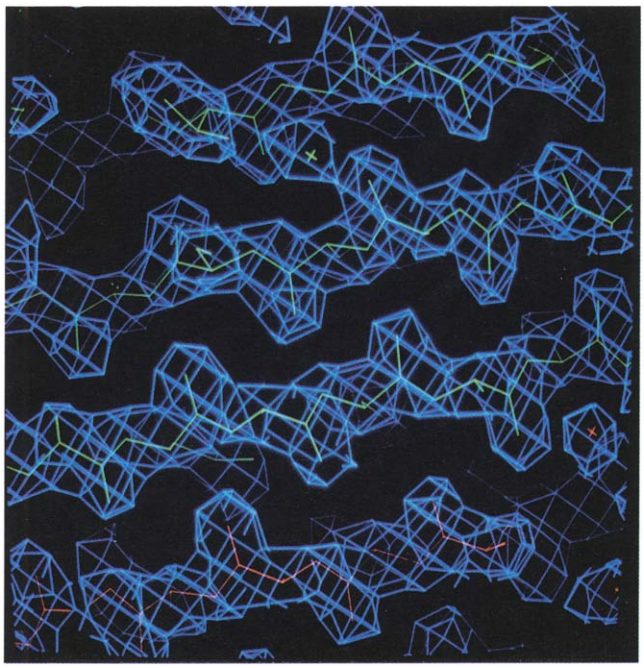

Fig. 3. Electron-density map in the region of the interface between monomeric subunits in the repressor dimer. Contours from the $2F_o–F_c$ electron-density map (2 Å resolution) are drawn at the 1.5σ level and shown with blue lines. The model of one monomer is shown in green in the contour map, with the other monomer in red. At this interface the two β-strands G (residues 82–94) are arranged in an antiparallel fashion with extensive hydrogen-bonded interactions between the two monomers.

less strikingly from the viral subunits. An average difference of 1 Å is measured for α-carbon positions of residues 25–29 in the CD loop and 1–2 Å for backbone atoms of residues 52–54 in the EF loop.

Most of the side chains in the subunit are conformationally similar to corresponding side chains in the viral subunits. As so few differences are observed, side chain R groups that differ significantly in position (for example, by ≥5 Å) may represent important functional differences. Such residues include Ala1 at the N terminus, Thr15 in the AB loop, Gln50 and Tyr85 in the RNA-binding site, Phe95 adjacent to the first α-helix and Tyr129 at the C terminus. The electron density for these residues is clearly defined. The differences in residues 1 and 129 may only reflect flexibility at the ends of the polypeptide chain, although these amino acids are close together in the stable dimeric complex. The differences in residues 85 and 95 are discussed in the following sections.

Mutation at residue 82

The Trp→Arg substitution at residue 82 constitutes an important difference from the viral subunit, and was critical for our success in crystallizing the MS2 protein as a dimer rather than as a capsid. This mutant molecule retains repressor activity [5] and binds RNA, but is defective in capsid formation. In the model of the dimer, Nη2 of Arg82 is located within 2.3 Å of Oϵ1 of Gln54 in the other monomer, in a suitable position to form a hydrogen bond. The Nη1 atom of Arg82 is also positioned to form a hydrogen bond with the carbonyl oxygen of residue 54. These two hydrogen bonds form a strong interaction between the subunits (see Fig. 4). The new hydrogen bonds formed as a result of the substitution of arginine at residue 82 in each subunit apparently stabilize the dimer.

While it is probably true that formation of four new hydrogen bonds between subunits enhances the stability of the dimer, this fact alone does not explain the assembly defect exhibited by these molecules. Dimers are the 'building blocks' of the viral capsid [3,4] and it is not immediately apparent why dimers with hydrogen-bonded subunits involving arginine at position 82 should fail to assemble in the virus. It is more likely that these substitutions and resultant hydrogen bonds affect the conformation of important capsid intersubunit contacts. The FG loop (residues 67–81) makes critical contacts around the five-fold and quasi-six-fold axes in the icosahedral virus [3,4] and therefore the interactions involving this loop are critical for capsid assembly. All of the mutant proteins that are assembly-defective but retain repressor activity involve substitutions within or near the FG loop (see Fig.2). For example, a mutant coat protein, called dlFG, lacking the FG loop (residues 67–81) [5] is unable to form capsids. Another mutation, Leu77→Pro is also located in the FG loop. The substitution Cys101→Arg is near this loop (within 6 Å). Finally, the amino acid substitution Trp82→Arg in the molecule reported here is immediately adjacent to the loop. Careful comparison of the components of the FG loop with the corresponding residues in the viral subunits is critical for assessing whether the substitution at position 82 influences the conformation of the FG loop, and consequently, interactions in the capsid.

The FG loop

The conformation of the FG loop in the unassembled monomer is different from each of the three conformations of this loop seen in the viral coat protein (see Fig. 5). Discussion of this loop must be presented with caution because the model for these amino acids was built into weak electron density. High B-factors indicate flexibility in this region. Although atom-for-atom comparisons cannot be made, the course of the polypeptide backbone can be described and compared with the FG loop conformers in the intact virus. Because the two subunits in the dimer reported here are related by a crystallographic two-fold axis, the conformations of these two FG loops are identical. The C subunits in the viral capsid are also related by exact two-fold symmetry (C-C′) and yet the FG loops in these subunits differ from those seen in the isolated dimer. In one viral subunit, B, Pro78 is in the *cis* configuration. Pro78 in the subunit

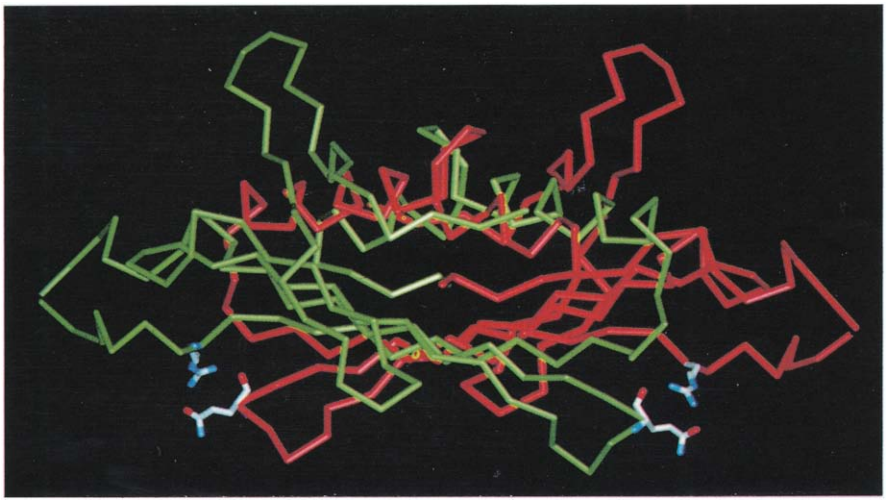

Fig. 4. Schematic model representing the α-carbon backbone of the unassembled dimer. One subunit is shown in red and the other in green. The view of the dimer in this picture is rotated by ~90° relative to the view in Fig. 2c. Individual atoms for Arg82 and Gln54 are shown as white stick models, with nitrogen atoms colored blue and oxygen atoms colored red for easy identification. Note that the guanidinium group of Arg82 is suitably positioned to form two hydrogen bonds with Gln54 (using both side chain and peptide oxygens). Because of the symmetric arrangement of subunits, this stabilizing interaction is duplicated in the dimer.

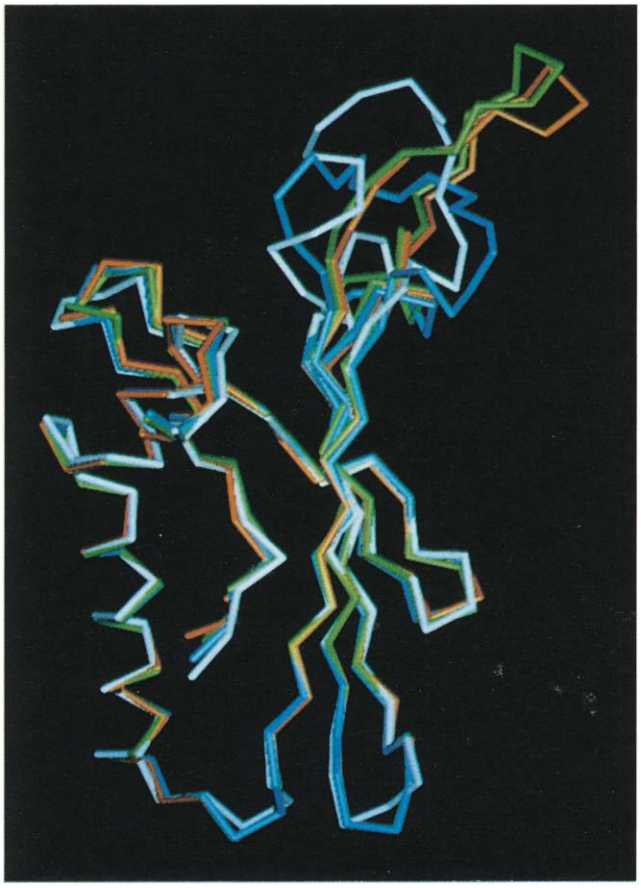

Fig. 5. Comparison of the polypeptide folding pattern of the repressor monomer with the three subunits in the intact MS2 virus. The monomer in this study is shown in white and superimposed on viral subunits A (orange), B (blue) and C (green). From this comparison, it can be seen that the FG loop in white is strikingly different from each of the loops in the individual viral protomers.

presented in this report was modeled in the *trans* configuration, but the loop is different from subunits A and C where Pro78 also exists in a *trans* peptide linkage. Consequently, the conformation of the FG loop in this dimer represents a fourth distinct conformation which cannot be influenced by crystal packing because there are no close lattice contacts to these residues.

The fact that the Leu77→Pro substitution within the loop results in an assembly-defective molecule [5] is of interest because this residue is directly adjacent to Pro78. *Cis-trans* isomerization of Pro78 has been proposed to play a role in formation of the virus protein shell [4]. The mutation at residue 77 introduces another proline next to Pro78 and such a proline–proline sequence would restrict the conformation in this section of the loop and might also affect the switch of Pro78 from *trans* to *cis* configuration. Other substitutions in the FG loop (for example, at residues 67, 74 and 76 [10]) do not affect capsid formation, so the assembly defect that results from the mutation to proline at residue 77 suggests that conformational flexibility of the FG loop is important for assembly.

The Trp82→Arg substitution may affect the proposed conformation of the FG loop in another manner. This residue is the first position in the β-strand following the FG loop. There are two orientations of the side chain of this tryptophan in the viral coat proteins. In the A and C subunits, the side chain is located in a pocket formed by the side chains of residues 97, 100 and 104. However, in the B subunit, the tryptophan side chain emanates from the other side of the polypeptide backbone in hydrophobic contact with residues 67–69 of the FG loop which is bent close to the DE loop of the subunit (see Fig. 12 of [4]). In the mutant molecule, Arg82 is on the same side of the β-strand as the Trp82 side chain in the viral A/C subunits (see Fig. 6). The density at the peptide linkage between residues 81 and 82 is strong, thus the direction of the arginine side chain is clearly defined. When hydrogen bonds are formed between an arginine substituted at position 82 and Gln54, the rotation of ~180° around the ϕ torsion angle of residue 82 that is required for the FG loop to adopt the B subunit conformation is not favored.

The conformation of the FG loop in the unassembled dimer can be viewed as intermediate between the two extremes observed in the capsid; i.e., extended in subunits A and C and bent close to the protomer in subunit B. The tightly bent FG loop in the B subunit packs around the five-fold axis in the capsid, whereas the FG loops in the A and C subunits interact at viral quasi-six-fold axes. If the hydrogen bonds formed between Arg82 and Gln54 in the unassembled dimer prevent rotation around the ϕ torsion angle of residue 82, it is possible that this conformational restriction could affect intersubunit contacts and virus assembly. To test this hypothesis, we superimposed the model of the unassembled dimer on the models of both the AB dimer and the CC′ dimer and thereby positioned the symmetric dimer on the icosahedral lattice of the virus. The resultant 'capsid' was examined for intersubunit contacts. Steric clashes are observed between FG loops at the five-fold axis. The channel between the five protomers is reduced from 16 Å in the virus to 9 Å in this pseudo-capsid. At the quasi-six-fold axis there is ample room to accommodate the FG loops but the interactions between the loops are changed. In the virus, the FG loops in alternating A and C subunits interact at this axis by the formation of hydrogen bonds between residues 71 and 73/74 in adjacent molecules. These stabilizing interactions are not possible when the FG loop is in the 'unassembled' conformation.

Another interesting difference in the quasi-six-fold interactions is a steric clash between Phe95 and Ser37 in the adjacent subunit. When compared with the virus structures, the side chain of Phe95 has moved by ~7 Å (see Fig. 7). This movement results from the Trp82→Arg substitution in the partner subunit of the dimer. The rotation of the Phe95 side chain is necessary to accommodate the arginine side chain, and this significant conformational change sterically affects subunit interaction at the quasi-six-fold axis of the capsid.

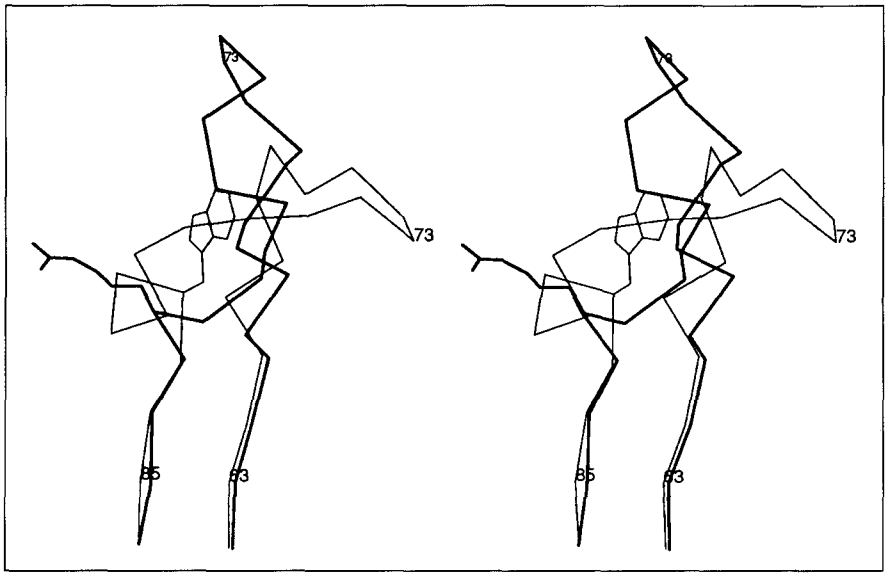

Fig. 6. Stereo diagram comparing the α-carbon backbone near residue 82 in the unassembled dimer and viral subunit B. Residues 62–86 (thick lines) are displayed and superimposed on the same residues from the viral subunit B (thin lines). Note the close similarity in backbone atoms for residues 62–65 and 83–86. Significant differences are evident in the location of residues 68–81 and the side chain of residue 82. Residue 73 at the apex of this loop is labeled to illustrate the differences in these conformations.

The RNA-binding site

Utilizing the two-plasmid genetic system, a variety of mutants have been isolated that help define the binding site for operator RNA. The results of these experiments can now be correlated with the three-dimensional structure of the unassembled dimer and with the recently reported crystal structure of recombinant MS2 capsids in complex with 19-nucleotide RNA fragments [11].

Repressor-defective mutations

Mutant proteins that fail to repress translation but retain the ability to form capsids have been identified [6]. They result from the substitution of amino acids at the following 10 sites which reduce or abolish RNA binding: Ser47, Arg49, Asn55, Lys57, Thr59, Lys61, Tyr85, Asn87, Glu89 and Thr91. The side chains for these residues each protrude from one side of the β-sheet. The β-strands that contain these residues are adjacent to each other, in an antiparallel fashion across the dimeric interface, and because of this arrangement and the consequent two-fold symmetry, six β-strands with 20 residues form the potential RNA-binding site [6]. These residues, shown in Fig. 8, form a concave surface which corresponds to the interior of the particle in the model of the intact virus [4,11]. The dimensions of the concave surface are ~29 Å×29 Å. We also note that the two-fold symmetry within the repressor dimer produces the chemical and structural potential for two equivalent binding sites for the asymmetric RNA hairpin, but only one of these sites is occupied at any one time [2]. The view of the binding site shown in Fig. 8 can be regarded as a 'composite view' of the two overlapping sites. The results of complementation studies using heterodimers of different repressor-defective mutants are consistent with this view (DS Peabody and F Lim, unpublished data).

In the virus–RNA complex [11], 13 nucleotides are well defined in the electron-density map when bound to the AB dimer. In this segment of the RNA hairpin, encompassing nucleotides −12 to +1, direct contact is seen with

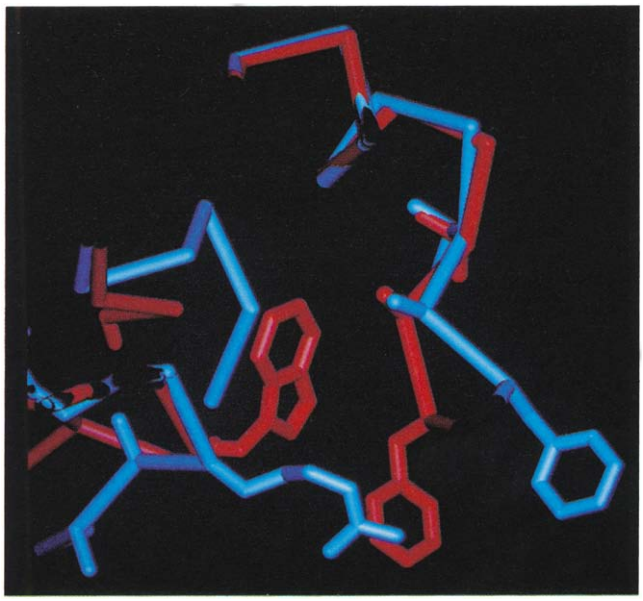

Fig. 7. View comparing the atomic models of amino acid residues near Phe95 in the unassembled dimer (blue) and viral subunit C-C′ (red). The image has been 'clipped' for clarity. The Phe95 ring in the isolated dimer has moved ~7 Å relative to the position in the viral capsid and this conformational change is a result of the Trp82→Arg substitution. The guanidinium group of Arg82 from the opposite subunit in the dimer is accommodated by this rotation of the Phe95 side chain. Note that without this conformational change there would be a severe steric clash of these two residues.

residues Thr45, Ser47, Tyr85 and Asn87 of subunit A and with residues Thr45, Ser47, Arg49, Ser51, Lys57, Thr59 and Lys61 of subunit B. Residues Thr45 and Ser47 in each subunit participate in similar interactions with unpaired adenines.

Mutations that enhance RNA binding

Mutants that bind wild-type and mutant RNAs more tightly than the wild-type repressor have also been isolated and result from substitutions at residues 29, 43, 55

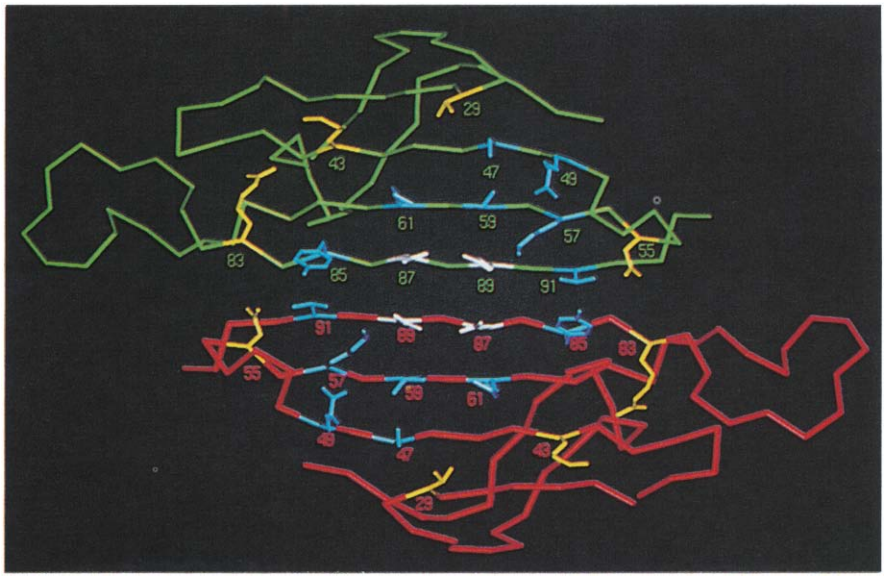

Fig. 8. Amino acid residues in the RNA-binding site of the repressor dimer. The schematic representation of the polypeptide backbone of one monomer is colored green and the other monomer is colored red. Superimposed on the backbone of the dimer are side chains of residues that are critical for RNA recognition as identified by molecular genetics [6]. Residues from both monomers where substitution reduced or abolished RNA binding are colored blue, residues that enhance RNA binding are colored yellow and residues that affect binding specificity are displayed in the center in white. The individual amino acids are numbered in the figure and discussed in the text. Note that these residues are clustered across the 10-stranded antiparallel β-sheet.

and 83 [12,13]. These amino acids are located at the 'edges' of the site defined by the repressor-defective mutants (see Fig. 8). We proposed that mutations at these sites result in additional protein–RNA interactions [12,13]. Consistent with this prediction, the side chain of Val29 in each subunit packs tightly against the purine ring of the unpaired adenines contacted by residues 45 and 47 in the complex [11].

Mutations that affect binding specificity

Mutants have been characterized that alter the RNA-binding specificity of the MS2 repressor. These mutants recognize the operators of other RNA phages such as GA [14] or Qβ (M Spingola and DS Peabody, unpublished data). The mutations involve substitution at residues 87 and/or 89 (see Fig. 8). One of these mutations led us to propose that in the wild-type MS2 repressor–RNA complex, Asn87 interacts with the uridine at position −5 in the operator loop [12]. This site defines specificity differences between MS2 and GA and obviously corresponds to a critical protein–RNA recognition contact. In the complex structure, tight hydrogen bonding of Asn87 (subunit A) was reported [11] with nucleotides −5 and −6. Another important RNA contact seen in the complex was the stacking of the phenolic group of Tyr85 (subunit A) on the base at −5.

The orientations of the side chains of most of the residues in the RNA-binding site are surprisingly similar in the viral subunits and in the unassembled dimer. Therefore it is quite interesting to note differences in residues 85 and 87 (see Fig. 9). The conformation of the side chain of Asn87 is twisted relative to the location in the viral subunits and displaced by 0.5–0.9 Å. A dramatic conformational difference is observed in the position of Tyr85. In the virus, the orientation of the side chain of this residue is very similar in all three subunits. In contrast, in our dimer structure, the phenolic hydroxyl group of this tyrosine is displaced by more than 6 Å relative to its location in the viral protomers. In the virus crystal structure, extra electron density was observed next to some side chains on the virus interior close to residues 45, 61, 63, 85 and 87 [4]. This density probably represented ordered RNA within the capsid. Because the unassembled dimer is unliganded, the conformational differences seen for residues 85 and 87 in the comparison of our structure and the virus may reflect side-chain adjustments that occur on RNA binding. Very little change was observed

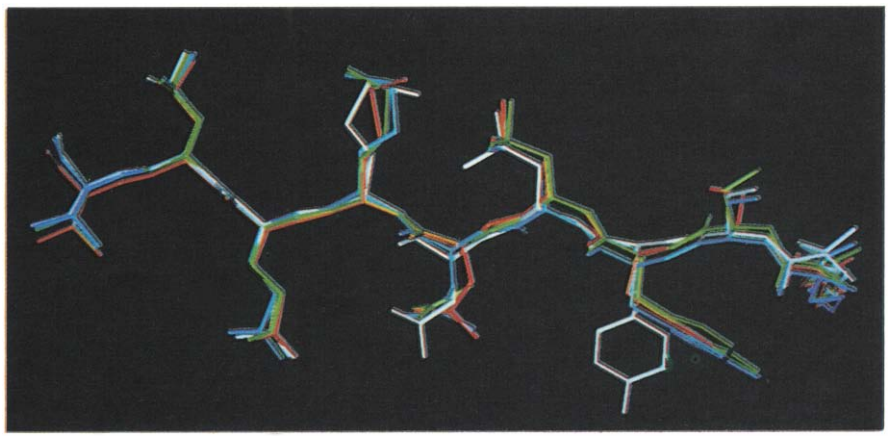

Fig. 9. Comparisons of residues 83–91. The model of the subunit in this study (white) was superimposed on the corresponding regions of the three viral subunits (colored as in Fig. 5). The sequence of this segment is Arg83-Ser84-Tyr85-Leu86-Asn87-Met88-Glu89-Leu90-Thr91 (residue 83 is on the right and residue 91 is on the left). Four of these residues are required for RNA binding: Tyr85, Asn 87, Glu89 and Thr91 [6]. The positions and orientations of main-chain atoms and side-chain atoms of this sequence are very similar except for residues 85, 87 and 88. The striking differences in the orientation of the side chains of Tyr85 and Asn87 may be of functional significance.

in the protein components of the capsid–RNA complex [11] from those seen in the intact virus. Such structural similarity may reflect the fact that the RNA hairpin was bound to assembled coat protein in that study.

Model for virus assembly

In the viral life cycle, encapsidation of genomic RNA is probably initiated by binding of a single dimer to the translational operator. Ultimately, capsid assembly is completed with the addition of 89 dimers. It is intriguing to consider that a conformational change upon RNA binding may trigger a second conformational change that promotes assembly. The conformational differences seen in the unliganded dimer involve two RNA contact residues in the central β-strands of the RNA-binding site. In particular, residue 85 is located in the same strand near the site of mutation that influences FG loop flexibility and capsid assembly. When this mutation and others that disrupt the ability of the dimer to assemble into capsids are considered in the context of the models of unassembled and assembled dimers, conformational flexibility of the FG loop is critical for correct subunit interactions in viral packaging. The fact that the structures of coat protein in contact with genomic RNA [4], RNA hairpins, or even in empty capsids [11] are similar suggests that the stereochemistry of the RNA-binding site is influenced and possibly restricted by intersubunit interactions in the assembled capsid. Therefore the conformational differences noted in the unassembled and unliganded dimer in the present study may be relevant to RNA binding in the first step of the assembly process.

Biological implications

RNA bacteriophages are small spherical viruses that infect *Escherichia coli*. Coat protein (180 copies) binds and encapsidates a single-stranded RNA genome of 3569 nucleotides. The coat protein also plays a genetic regulatory role, acting as a translational repressor of viral replicase synthesis by binding to an RNA hairpin containing the translation initiation codon of the replicase subunit. Coat protein dimers are the 'building blocks' for virus assembly and the dimeric form is active as a repressor. In order to examine the conformation of isolated subunits, we report the crystal structure of a mutant MS2 coat protein that is defective in capsid assembly but retains repressor activity. This is the first time that high-resolution structures have been available for a viral protomer in both the isolated and assembled states.

Residues that affect virus assembly and RNA recognition, identified by molecular genetics, were located on the model. The dimer is stabilized by interdigitation between two α-helices and hydrogen bonding between comparable β-sheets in each subunit. A 10-stranded antiparallel β-sheet is formed across the interface between monomers and critical hydrophilic side chains that mediate RNA binding extend from the large concave surface of the β-sheet. When compared with the structure of the intact virus [4,11], differences in the orientation of two of the RNA recognition residues in the central β-strands are noted in this unliganded dimer. These differences suggest a conformational adjustment on binding RNA in the first step of the assembly process.

The site of the mutation that influences capsid assembly is located at the end of the same β-strand that contains RNA contact residues with conformational differences. The substitution of arginine for tryptophan at this site introduces two new hydrogen bonds between monomers. These hydrogen bonds stabilize the dimer and impose conformational restrictions on an adjacent loop that makes critical interactions at the five-fold and quasi-six-fold axes of the viral capsid. When this mutation and others that disrupt the ability of the coat protein to form capsids are viewed in the model of the unassembled dimer, unrestricted conformational flexibility of this loop is of major importance for precise subunit interactions in viral packaging.

Materials and methods

Crystallization and data collection

The MS2 repressor crystallized in this study was a mutant protein (named SU647) defective in capsid formation but still functional as a repressor by binding to the RNA operator [5]. The mutant protein was expressed in large-scale cultures of *E. coli* strain CSH1F$^-$ cells and purified to homogeneity by ion-exchange chromatographic procedures that will be reported elsewhere. Large crystals of the purified SU647 protein were obtained by vapor diffusion in 3 μl hanging drops containing 10% PEG 8000 with protein at 3 mg ml^{-1}, buffered at pH 6.2 with 100 mM citrate-phosphate.

Diffraction data were collected with two San Diego Multiwire Systems area detectors at 4°C and a Rigaku RU200 rotating anode X-ray generator with a graphite monochromator operating at 50 kV and 150 mA. Data reduction was performed with UCSD area detector data processing programs [15]. The data collection statistics for the SU647 intensity data set are presented in Table 2. The reflection data were 90% complete to 2.03 Å resolution.

Table 2. Summary of data collection statistics.

Resolution limit (Å)	Average intensity (I)	Average I/σ(I)	Number of observations	Number of reflections	R_{sym}[a]
3.48	3348	59.8	9166	1635 (93%)	0.025
2.76	1207	31.1	6667	1648 (99%)	0.038
2.41	486	12.1	3359	1583 (94%)	0.066
2.19	336	8.0	3003	1548 (96%)	0.093
2.03	245	5.7	1740	1033 (65%)	0.107
Totals	1210	25.1	23935	7447	0.031

[a] $R_{sym} = \Sigma_h \Sigma_i |I_{hi} - \bar{I}_h| / \Sigma \bar{I}_h$.

Phase determination by molecular replacement

The molecular replacement method [16] was used to solve the structure of the SU647 dimer. The probe model was a subunit from the MS2 virus structure (coordinates kindly provided by Drs K Valegård and L Liljas, Uppsala University, Sweden). Rotation and translation function searches were calculated using MERLOT [17] initially, and X-PLOR [18]. In the viral capsid, one loop (residues 67–81; the FG loop) assumes three conformations in the three subunits that pack in the viral T=3 icosahedral capsid. Therefore during the refinement, the probe model was subjected to a simulated annealing procedure [19] in X-PLOR in order to permit wide conformational adjustments in this loop. After refinement, the model adjustments were made interactively with FRODO [20,21] on an Evans and Sutherland PS390 color graphics workstation using $2F_o$–F_c and F_o–F_c electron-density maps. OMITMAPS [22] were first calculated omitting 10% of the atoms and used to position atoms that differed significantly from the probe model especially in the FG loop. Refinement was continued to 3.0 Å with B-values fixed at 20 Å^2. When the R-factor had dropped to 0.25 for 8.0–2.6 Å data, restrained least-squares refinement PROLSQ [23,24] was used with the program GPRLSA [25]. Solvent atoms were added to the model and refinement was continued to convergence, including atoms for 129 amino acid residues, and 111 solvent atoms and temperature factors. The final R-factor was 0.20 for 8.0–2.0 Å data.

Color solid image models shown in Figs 4, 5, 7, 8 and 9 were created with the program QUANTA (Molecular Simulations, Inc., Waltham, MA).

The atomic coordinates have been deposited at the Brookhaven Protein Data Bank.

Acknowledgements: This work was supported by grants from the National Institutes of Health GM48295 (KRE) and GM42901 (DSP). The authors are grateful to Kelly Riddle for preparing the manuscript for publication.

References

1. Romaniuk, P.J., Lowary, P., Wu, H.-N. Stormo, G. & Uhlenbeck, O.C. (1987). RNA binding site of R17 coat protein. *Biochemistry* **26**, 1563–1568.
2. Beckett, D. & Uhlenbeck, O.C. (1988). Ribonucleoprotein complexes of R17 coat protein and translational operator analog. *J. Mol. Biol.* **204**, 927–938.
3. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41.
4. Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234**, 620–639.
5. Peabody, D.S. & Ely, K.R. (1992). Control of translational repression by protein–protein interactions. *Nucleic Acids Res.* **20**, 1649–1655.
6. Peabody, D.S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600.
7. Peabody, D.S. (1990). Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. *J. Biol. Chem.* **265**, 5684–5689.
8. Ramachandran, G.N. & Sasiekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–437.
9. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
10. Stockley, P.G., *et al.*, & Valegård, K. (1993). Molecular mechanism of RNA-phage morphogenesis. *Biochem. Soc. Trans.* **21**, 627–633.
11. Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature* **371**, 623–626.
12. Lim, F., Spingola, M. & Peabody, D.S. (1994). Altering the RNA binding specificity of a translational repressor. *J. Biol. Chem.* **269**, 9006–9010.
13. Lim, F. & Peabody, D.S. (1994). Mutations that increase the affinity of a translational repressor for RNA. *Nucleic Acids Res.* **22**, 3748–3752.
14. Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobson, A.B. & Hirashima, A. (1996). The complete nucleotide sequence of the group II RNA coliphage GA. *J. Biochem.* **99**,1169–1180.
15. Howard, A.J., Nielsen, C. & Xuong, N.-h. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 452–472.
16. Rossmann, M.G. (1972). The molecular replacement method. *International Science Review.* Vol. **13**. Gordon and Breach, NY.
17. Fitzgerald, P.M. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.* **21**, 273–278.
18. Brünger, A.T. (1992). *X-PLOR Manual, Version 3.0.* Yale University, New Haven, CT.
19. Brünger, A.T. (1988). Crystallographic refinement by simulated annealing: application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.
20. Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
21. Pflugrath, J.W., Saper, M.A. & Quiocho, F.A. (1984). New generation graphics system for molecular modeling. In *Methods and Applications in Crystallographic Computing.* (Hall, S. and Ashiaka, T., eds), pp. 404–407, Clarendon.Press, Oxford.
22. Bhat, T.N. & Cohen, G.H. (1984). OMITMAP: an electron density map suitable for the examination of errors in a macromolecular model. *J. Appl. Crystallogr.* **17**, 244–248.
23. Hendrickson, W.A. & Konnert, J.H. (1980). Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography.* (Diamond, R., Ramesehan, S. & Venkatesan, K., eds), pp.13.01–13.23, Indian Academy of Sciences, Bangalore, India.
24. Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
25. Furey, W. (1990). Abstracts of the American Crystallographic Association Fortieth Anniversary Meeting. PA33. New Orleans, LA.
26. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.

Received: **12 Dec 1994**; revisions requested: **6 Jan 1995**; revisions received: **16 Jan 1995**. Accepted: **19 Jan 1995**.