SHORT COMMUNICATIONS

Acta Cryst. (1996). D**52**, 858–863

A re-evaluation of the crystal structure of chloromuconate cycloisomerase†

GERARD J. KLEYWEGT,[a]* HELGA HOIER[b] AND T. ALWYN JONES[a] *at* [a]*Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden, and* [b]*Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany*

(*Received 16 March 1995; accepted 15 February 1996*)

Abstract

It is shown here that the reported 3 Å crystal structure of chloromuconate cycloisomerase from *Alcaligenes eutrophus* [Hoier, Schlömann, Hammer, Glusker, Carrell, Goldman, Stezowski & Heinemann (1994). *Acta Cryst.* D**50**, 75–84] was refined in the incorrect space group *I*4. In addition, a stretch of about 25 residues near the N-terminus is out-of-register with the density in the original structure. From the coordinates and structure factors deposited in the Protein Data Bank (PDB), it was possible to determine the correct space group to be *I*422. The structure was then re-refined, using the original data reduced to *I*422, to a crystallographic free *R* factor of 0.264 at 3 Å resolution (conventional *R* factor 0.189). With conservative refinement and rebuilding methods, the errors in the chain tracing could be identified and remedied. Since the two molecules per asymmetric unit in the original structure are actually related by crystallographic symmetry, the observed differences between them are artefacts. In particular, the differences between, and peculiarities of the metal-binding sites are unreal. This case shows the dangers of crystallographic refinement in cases with unfavourable data-to-parameter ratios, and the importance of reducing the number of parameters in such cases to prevent gross errors (for instance, by using NCS constraints). It also demonstrates how the evaluation and monitoring of model quality during the entire refinement and rebuilding process can be used to detect and remedy serious errors. Finally, it presents a strong case in favour of depositing not only model coordinates, but also experimental data (preferably, both merged and unmerged data).

Recently, the crystal structure of chloromuconate cycloisomerase (CMCI) from *Alcaligenes eutrophus* was reported at 3 Å resolution (Hoier *et al.*, 1994). This protein forms homo-octamers, of which two monomers per asymmetric unit were found in the space group *I*4. While evaluating the relation between resolution, refinement practice, and differences between molecules related by non-crystallographic symmetry (NCS) (Kleywegt, 1996), this was one of the structures that had a high root-mean-square difference (RMSD) between NCS-related molecules. Since this structure was one of precious few outliers for which experimental structure-factor amplitudes had been deposited with the Protein Data Bank (Bernstein *et al.*, 1977), we chose it as an example to investigate the possible effects of over-fitting in NCS-related molecules. Despite the low resolution (3.0 Å), the two molecules had been refined without NCS constraints or restraints, with individual isotropic temperature factors, and with alternative conformations for a number of residues. Such a refinement protocol is no different from that used in many structural analyses (Kleywegt & Jones, 1995*a*). However, it meant that the structure had been refined with ~1.5 times as many parameters as there were experimental observations. We believe that this leads to over-modelling, fitting noise and creating artefactual differences between NCS-related molecules (if present). Indeed, in the original model 'significant structural differences' between the two molecules were found, in particular in the coordination of the active-site manganese ion, which showed unusual metal coordination distances, and which had a bound chloride ion in one molecule, but not in the NCS-related molecule.

We started by subjecting the model to a number of standard quality checks with *O* (Jones, Zou, Cowan & Kjeldgaard, 1991; Zou & Mowbray, 1994), *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993; Laskowski, MacArthur & Thornton, 1994) and other programs. Virtually on every criterion used, the structure scored considerably worse than expected for a properly refined model at this resolution, Table 1. Also, the differences between the two monomers were very large: 0.76 Å on Cα atoms, and 1.51 Å for all non-H atoms. A $\Delta\varphi$, $\Delta\psi$ plot (Korn & Rose, 1994; Kleywegt & Jones, 1995*a*; Kleywegt, 1996) of the two monomers revealed that the conformation of the main chain differed considerably in the two molecules, in particular near the N-terminus, Fig. 1. All these factors indicated that there might be something wrong with this model, although initially we assumed that it only suffered from over-fitting. The authors quoted an estimated coordinate error of 0.3 Å based on a Luzzati plot (Luzzati, 1952), but since this plot was made with conventional *R* factors (which can be made arbitrarily low by including far more parameters in the model than is warranted by the quality and quantity of the crystallographic data) this number is likely to be meaningless in low-resolution studies (Kleywegt & Jones, 1995*a*). A more realistic estimate can perhaps be obtained with a free *R* factor (Brünger, 1992*a*, 1993) based Luzzati plot (Kleywegt *et al.*, 1994).

To evaluate the effects of over-fitting, we planned to re-refine the structure with NCS constraints and to use R_{free} to

† Abbreviations: CMCI, chloromuconate cycloisomerase; F_o, F_c = observed and calculated structure-factor amplitudes; MCI, muconate cycloisomerase; MD, molecular dynamics; NCS, non-crystallographic symmetry; PDB, Protein Data Bank; *R*, conventional *R* factor, $\sum_h ||F_o| - |F_c|| / \sum_h |F_o|$; R_{free}, free *R* factor, *R* factor calculated for a small subset of reflections not used in the refinement; R_{merge}, $\sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i |I_{h,i}|$; RMS, root-mean-square; RMSD, root-mean-square difference or deviation.

Table 1. *Comparison of the refinement statistics and model quality of the original, partly mis-traced chloromuconate cycloisomerase structure in the incorrect space group I4, and the corrected structure refined in space group I422*

NR means not reported, NA means not applicable.

	*I*4	*I*422
R_{free}	NR	0.264
R	0.195	0.189
Data and model		
Range of Bragg spacings (Å)	8.0–3.0	8.0–3.0
Number of reflections	16012	9089
Number of atoms ($Z > 1$)	5603	2813
Number of refined parameters	22400	9176
Data-to-parameter ratio	0.71	0.99
Stereochemistry*		
RMSD bond lengths (Å)	0.029	0.007
RMSD bond angles (°)	5.07	1.24
RMSD dihedral angles (°)	NR	23.5
RMSD improper torsion angles (°)	NR	1.20
Temperature factors		
Model	Individual	Grouped
Average temperature factor, all atoms (Å^2)	25.9	28.7
Average temperature factor, Mn^{2+} and Cl^- (Å^2)	23.8	5.6
RMS ΔB bonded atoms (Å^2)	2.2	NA
Ramachandran plot†		
Residues in most favoured regions (%)	75.7	83.1
Residues in additional allowed regions (%)	19.4	15.7
Residues in generously allowed regions (%)	3.4	0.9
Residues in disallowed regions (%)	1.5	0.3
Non-crystallographic symmetry‡		
RMSD NCS Cα atoms (Å)	0.76	NA
RMS ΔB NCS Cα atoms (Å^2)	4.4	NA
RMSD NCS all atoms (Å)	1.51	NA
RMS ΔB NCS all atoms (Å^2)	5.7	NA
RMS $\Delta\varphi$ (°)	38.8	NA
RMS $\Delta\psi$ (°)	38.2	NA
Miscellaneous		
ω dihedral standard deviation (°)†	7.6	1.0
Cα chirality standard deviation (°)†	1.0	1.3
Residues with unusual peptide orientations (%)‡¶	4.5	1.9
Non-rotamer side-chain conformations (%)‡¶	23.0	8.1
Overall G factor †¶	−1.3	+0.3
Average DACA score §¶	−1.2	−0.7

* Calculated with *X-PLOR* (Brünger, 1992*b*) using the Engh & Huber (1991) force field. † Calculated with *PROCHECK* (Laskowski *et al.*, 1993, 1994). ‡ Calculated with *O* (Jones *et al.*, 1991) or *LSQMAN* (Kleywegt, 1996). §Calculated with *WhatIf* (Vriend & Sander, 1993); DACA = directional atomic-contact analysis. ¶ For a good low-resolution model (Kleywegt & Jones, 1995*b*), one would expect to find ~1–2% of the residues to have unusual peptide orientations (Jones *et al.*, 1991), ~5–10% of the side chains to have non-rotamer conformations (Jones *et al.*, 1991), an overall G factor (Laskowski *et al.*, 1994) greater than −0.5, and a DACA score (Vriend & Sander, 1993) greater than −1.0. For a discussion of these quality indicators, see Kleywegt & Jones (1995*b*).

monitor the progress and validity of the various refinement steps. We started by calculating the NCS operator relating the two molecules in the asymmetric unit of space group *I*4. This operator looked suspiciously like a crystallographic operator (approximately X, $1-Y$, $1-Z$). We therefore expanded the contents of the asymmetric unit under the complete space-group symmetry of *I*4, and determined the operators relating the *A* molecule inside the reference asymmetric unit to all crystallographic copies of the *B* molecule inside the unit cell. The result was a set of eight operators which, when added to the operators of space group *I*4, form the set of space-group operators for *I*422. To double-check our assumption that the correct space group was *I*422, we reduced the data from *I*4 to *I*422 by merging reflections *hkl* and *khl*. This reduced 16012 reflections in *I*4 to 9089 reflections in *I*422. The value of R_{merge} (using F^2 as an estimate of I, and including only paired reflections) was 0.067, significantly lower than the value of R_{sym} of 0.101 for the original data reduction in *I*4.

We then partitioned these data into a working set (8647 reflections) and a test set (442 reflections, ~5% of the data) for use in R_{free} calculations. We arbitrarily chose molecule *A* in the original structure as our starting model, but we removed the manganese and chloride ion, removed the alternative conformation of Arg17, and reset all temperature factors to 20.0 Å^2. This starting model in *I*422 had an R factor of 0.301 (R_{free} 0.303). In order to uncouple both R factors, we removed model bias and memory by means of a slow-cooling simulated-annealing protocol (Brünger & Krukowski, 1990) starting from 4000 K (temperature step size −25 K), followed by 50 cycles of energy minimization and 50 cycles of grouped temperature-factor refinement (two temperature factors per residue, one for the main-chain atoms, and one for the side-chain atoms). This yielded a model with vastly improved stereochemistry, owing to the use of the Engh & Huber (1991) force field, but with too large a difference between R (0.223) and R_{free} (0.344). Initially, we assumed that this large difference was due to poor data quality, the excess of adjustable parameters in the model, or both. We rebuilt this model with *O* (Jones *et al.*, 1991) using systematic simulated-annealing omit maps (Hodel, Kim & Brünger, 1992), omitting sequential stretches of ten residues at a time, and rebuilding the omitted residues using their omit map. In addition, these

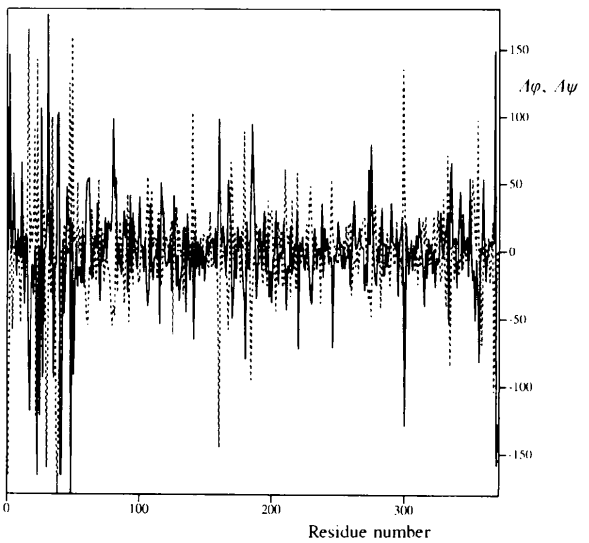

Fig. 1. $\Delta\varphi$, $\Delta\psi$ plot (Korn & Rose, 1994) of the two NCS-related CMCI molecules in the original model in space group *I*4. The solid curve shows the difference between main-chain φ torsion angles of corresponding residues in the two molecules, the dashed curve shows the difference between corresponding main-chain ψ torsion angles.

Table 2. *Comparison of the CMCI model refined in space group I422 (M14) and the original I4 model (molecules A and B)*

	M14 *versus A*	M14 *versus B*	*A versus B*
RMS $\Delta\varphi$ (°)	37.2	36.6	38.8
RMS $\Delta\psi$ (°)	40.5	36.3	38.2
RMSD Cα 1–370 (Å)	1.32	1.38	0.76
RMSD N/Cα/C/O/Cβ 1–370 (Å)	1.44	1.49	0.91
RMSD all atoms 1–370 (Å)	2.20	2.33	1.51
RMSD Cα 15–42 (Å)	3.82	3.96	1.53

maps were used to calculate residue real-space fits (Jones *et al.*, 1991), *i.e.* the real-space equivalent of the free R factor (Brünger, 1992*a*).

During this rebuild, it became apparent that a stretch of ~25 residues was out-of-register with the density. For example, the strand between residues 28 and 36 had hydrophobic and polar side chains alternating the wrong way around. Arg35, for instance, was pointing into a very hydrophobic pocket with the guanidino group surrounded by leucines and isoleucines. The density looked more like that of a valine or a proline, and the refinement program had forced the arginine into a 'pseudo-proline' conformation. Checking the sequence alignment with the muconate cycloisomerase (MCI; Goldman, Ollis & Steitz, 1987) search model (published in the original paper), we found that a deletion in CMCI between residues 37 and 38 had not been accounted for in the proper place in the structure (the structural deletion is in the subsequent turn 39–41). In addition, the loop ~16–27 looked rather unusual. Attempts to rebuild it using fragments from the *O* database (Jones & Thirup, 1986) gave only very poor matches. Since, in addition, several of the outliers in the Ramachandran plot were residues in the range ~15–41, and since many of the residues in this stretch had real-space free R factors greater than 0.4, we decided to trim the model. The whole loop 16–26 was removed from the model, and residues 1–15 and 27–39 were changed to alanines. Since the 13 C-terminal residues also had poor density, these were also omitted from the model.

In subsequent refinement cycles, four different simulated-annealing protocols were used in parallel and the one yielding the model with the lowest value of R_{free} was used for subsequent rebuilding. Typically, two 4000 K conventional slow cools in Cartesian space were executed (one with half and one with one-third weight for the crystallographic pseudo-energy term), one constant high-temperature torsion MD (Rice & Brünger, 1994) calculation (at 5000 or 10 000 K), and one slow-cooling torsion MD calculation (starting from 5000 or 10 000 K). Each of these calculations was followed by energy minimization and grouped temperature-factor refinement (in which, initially, one and, later, two temperature factors per residue were refined). As missing parts of the model became visible in $2F_o - F_c$ and $F_o - F_c$ maps, they were added to the model, initially as alanine residues and later with the side chains corresponding to the sequence. During these cycles, only those residues which violated any of our standard quality criteria were investigated and rebuilt where necessary. After four macrocycles the protein model was complete again, and after an additional slow cool the resulting model was scrutinized residue-by-residue, again using systematic omit maps (this time omitting five residues at a time). This model generally had good to excellent omit density (with the exception of, mainly, residues at the C terminus and in a few loops and turns) and required very little rebuilding. The resulting model was subjected to energy minimization and grouped temperature-factor refinement. At this stage we were sufficiently confident to add a small number of water molecules to the model. Adding 24 waters to the model reduced R_{free} by about 0.01. The final model (M14) constitutes a good 3 Å protein structure which is superior to the original model(s) in most respects, Table 1. For comparison, the Ramachandran plots of the original model and M14 are shown in Fig. 2.

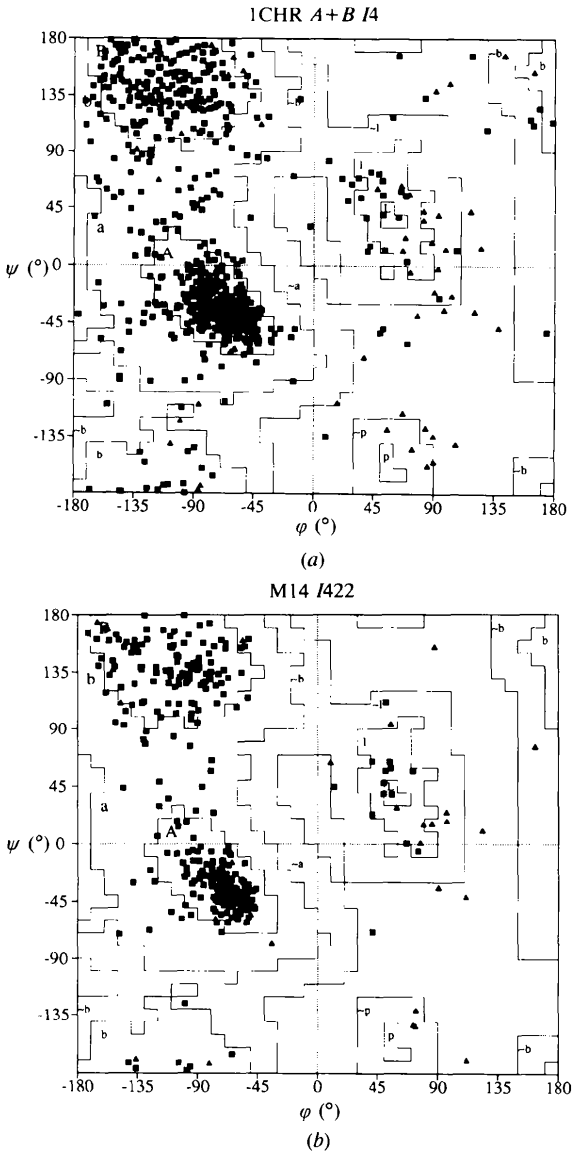

Fig. 2. Ramachandran plot (Ramakrishnan & Ramachandran, 1965; Laskowski *et al.*, 1993, 1994) of (*a*) the original *I*4 model (both molecules), and (*b*) our final model M14 in space group *I*422. Several of the outliers in (*a*) are residues in the incorrectly traced part of the original model.

The structure of residues 13–42 in the two *I*4 molecules, and the tracing in our model M14, are shown in Fig. 3. Table 2 lists coordinate and torsion-angle differences between M14 and the original molecules. This table shows that both original models are approximately equally different from our final model. Residues 15–42 have an RMSD on Cα atoms of ~3.8 Å, which corresponds to one Cα—Cα distance. Fig. 4 shows the structure-based sequence alignment of residues 13–43 of model M14 and the two original models. After residue 15 two residues were skipped in the original model. This register-error is reduced to one residue later by forcing a loop into a very unusual twisted conformation. Moreover, the loop had been built in different ways in both molecules of the *I*4 structure and this, together with weak electron density and high temperature factors, was taken as evidence for the loop being highly flexible. However, in our final model the entire loop has good electron density and lacks excessively high temperature factors. By missing the deletion in the turn 39–41, the original models are back in register at residue 42 and, as far as we could determine, remain so throughout the rest of the structure.

In the original structure, significant differences were observed in the active sites of the two molecules. Since the molecules are actually related by crystallographic symmetry such differences clearly are not significant. The original structure had a chloride ion coordinated to the active-site manganese ion in one of the molecules but not the other. We, however, observe good and consistent density for such a chloride ion. In the original structure, the manganese ions and chloride ion had a temperature factor of ~24 Å^2; in our model the manganese ion has a temperature factor of ~9 Å^2 and the chloride ion of 2 Å^2, which does not support the presence of disorder. Fig. 5 shows the density for the chloride ion and the nearby residues. Another observation made in the original structure was that of an unusually short Mn—O coordination distance of 1.8 Å in one of the molecules. In our model, this distance (which was not restrained during refinement) is ~2.1 Å which is a normal value.

Space-group errors are easy to make [for an excellent discussion on this subject, see the recent review by Marsh (1995)]. If the symmetry of the space group is chosen too low, but the structure is refined with a conservative protocol, the resulting differences will be small and unlikely to tempt authors into making injudicious statements concerning the differences. In Uppsala, and perhaps elsewhere, the advent of electronic and image-plate area detectors has led to a tendency to rely less on film methods for space-group determination. Under these circumstances extra care must be taken in determining the correct space group, both in reciprocal space (analysis of absences and of the symmetry of the diffraction intensities; merging in the highest symmetry point group possible), and in real space (analysis of the operators that relate the individual molecules). In this particular case, the systematic absences are identical for space groups *I*4 and *I*422, but they differ in their Laue symmetry. Note that the free *R* factor cannot be used to detect space-group errors of this type during the refinement of the structure. In the lower symmetry space group, most of the reflections (*hkl*) that are part of the test set will have their symmetry-related counterpart (*khl*) in the work set. Therefore, the refinement implicitly includes most of the test set reflections, which will lead to a deceptively low value of the free *R* factor.

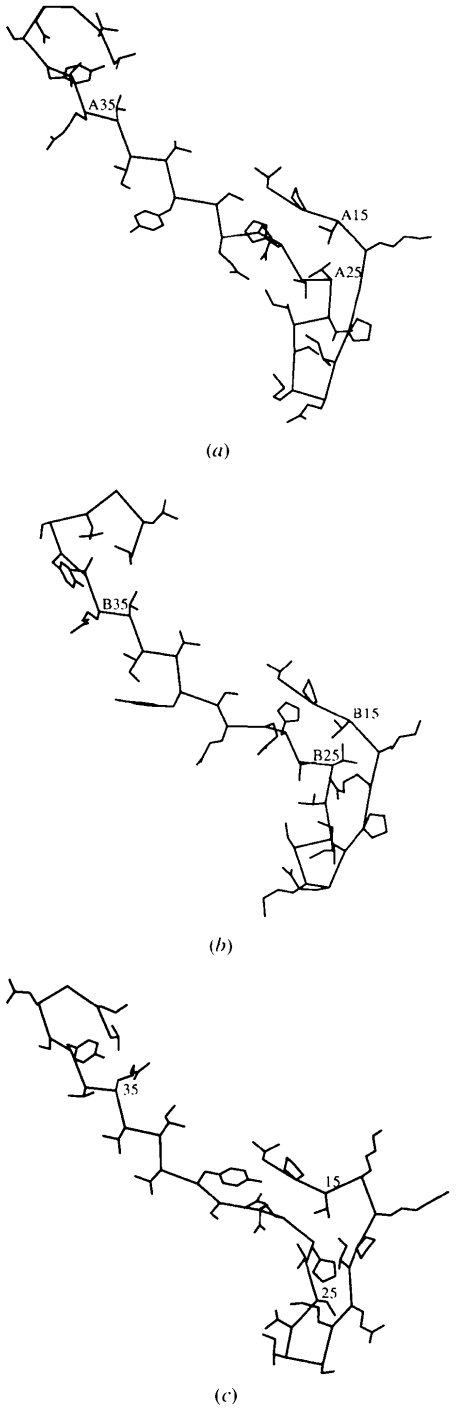

Fig. 3. Comparison of residues 13–42 in the structure of chloromuconate cycloisomerase. The structure of this segment is shown for (*a*) the *A* molecule in the *I*4 structure, (*b*) the *B* molecule in the *I*4 structure, and (*c*) our model M14 in space group *I*422. Note that molecules *A* and *B* have different conformations for the loop. This illustrates the danger of self-fulfilling prophecies with respect to the refinement of, and differences between 'independent' NCS molecules, in particular (but not exclusively) at low resolution.

We believe the tracing error to be at least partly the result of following the molecular replacement search model too closely. The errors occurred in the N-terminal region, where there are three deletions in CMCI compared to MCI, and where the search model lacked a loop of ~14 residues. Of course, the fact that something was wrong, should have been detected by inspection of the model and the structural alignment with the search model. On the other hand, the errors could probably have been avoided if different refinement and rebuilding approaches had been used (use of databases in *O*, use of simulated annealing, omit maps and R_{free} in *X-PLOR*), and if basic quality checks had been carried out after every refinement cycle. (Some of the techniques employed in the re-refinement, such as torsion MD calculations, are not yet generally available; others, such as the use of R_{free} and NCS constraints, were not universally accepted as useful when the original study was carried out.) Instead, the errors were masked by an inappropriate refinement protocol (individual temperature factors, alternative conformations, no NCS constraints). The result was a structure with serious errors and artefacts. It has been shown that the free *R* factor is a good indicator of phase error derived from a model (Brünger, 1992*a*, 1993). Therefore, a refinement protocol driven by the behaviour of R_{free} is likely to yield maps that are easier to interpret and in which errors can be recognised and corrected. On the other hand, phase and amplitude relations between symmetry-related reflections are likely to produce deceptively low R_{free} values; therefore, R_{free} can probably not be used to detect space-group errors (*vide supra*).

```
                    *20          *30          *40
M14          VPTKRPIQM-SITT---VHQ-QSYVIVRVY-SEG--L-VG
A            VPT--KRPI-QMSITTV--H-QQSYVIVRV-YS-EG-LVG
B            VPT--KRPI-QMSI---TTVHQQSYVIVRV-YSE-G-LVG
MCI          LPTIRPPHKLAMHT---MQT-QTLVLIRVRCS--DGV-EG
```

Fig. 4. Structure-based alignment of residues 13–43 of the CMCI model refined in space group *I*422 (M14), the original model (molecules *A* and *B*), and the search model (MCI). Residues were aligned with M14 if the Cα—Cα distance after superpositioning was less than 3 Å. Note that after residue 15 the *I*4 models are two residues out of register, of which one residue is 'recovered' through a strangely twisted loop. Between residues 29 and 41 the original models are still one residue out-of-register.

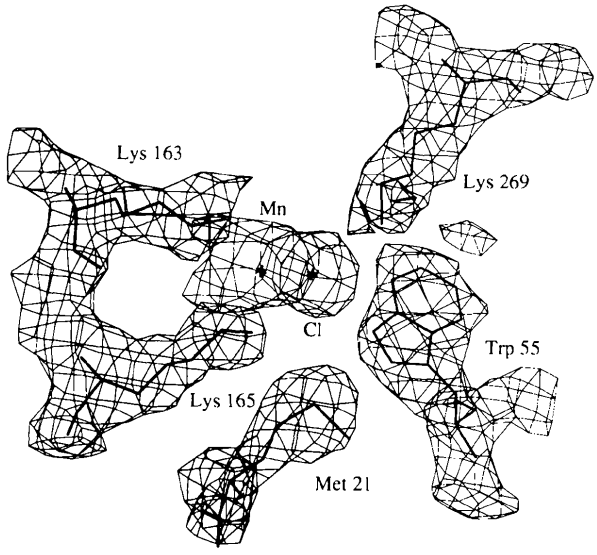

Fig. 5. Density for the chloride ion in the active site of CMCI, and some of the residues that surround it, in the final $2F_o - F_c$ map of model M14 in space group *I*422, contoured at a level of 1σ. In the original structure, a chloride ion was found in only one of the two molecules, and it refined to a temperature factor of ~24 Å^2. In our model, the chloride has good density and a temperature factor of 2 Å^2, the lowest value allowed in the temperature-factor refinement. Also note that the S atom of Met21 at the bottom interacts with the charged N atom of a nearby lysine residue. In the original *A* molecule, Ile23 had been built in the place of this methionine.

Initially, we suspected that most of the observed differences between the NCS-related monomers were artefacts due to over-fitting of the data during refinement. Had the structure been refined with strict NCS constraints, the resulting model might have been better. Also, the two molecules would have been forced to remain identical, and thereby more or less obey the actual crystallographic symmetry. Unfortunately, the liberal approach to refining and rebuilding structures is widespread in the protein crystallographic community. This approach can be characterized by: a fixation on low conventional *R* factors which are clearly no guarantee for a correct model; reluctance to use R_{free} from the very start of the refinement process; over-fitting by individual refinement of NCS-related molecules, individual isotropic temperature factors, alternative conformations and occupancies when this is not warranted by the information contents of the crystallographic data; hesitation in using new software and methodology (high-temperature simulated annealing, free *R* factor, databases for rebuilding).

This example also suggests that some of the cases where 'significant' differences are observed between NCS-related molecules in low-resolution structures may be no more than artefacts due to over-fitting (Kleywegt & Jones, 1995*a*; Kleywegt, 1996). Clearly, if one does not constrain NCS-related molecules, the refinement program will make them different: it becomes a self-fulfilling prophecy. One has to wonder how significant such published differences are, knowing that, in the case of CMCI, two molecules which are necessarily identical (for all practical intents and purposes) because of crystallographic symmetry can be refined to RMSD values of ~0.8 Å on Cα atoms and ~1.5 Å for all atoms, Table 2.

This re-evaluation would not have been possible had not one of us (HH) deposited both coordinates and structure factors with the PDB. This demonstrates that deposition of both the final model and the experimental data (preferably both merged and unmerged) helps maintain the quality and integrity of the PDB. The coordinates of our model M14, and the structure factors reduced in *I*422, have been deposited as well and are available for immediate release (PDB entry code 2CHR).*

This work was supported by Uppsala University. We wish to thank Dr Axel Brünger (Yale University, New Haven, CT)

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2CHR, R2CHRSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: BU0325).

for providing us with a test version of the next release of *X-PLOR* (which includes the torsion MD functionality), and for his help in setting up the torsion MD calculations.

Note from the editor: the following may be useful to some readers. The Laue symmetry is $4/m$ for space group $I4$ which means that $I(hkl) = I(-h-k-l) = I(hk-l)$ but does not equal $I(-hkl)$ which equals $I(h-kl)$. The Laue symmetry is $4/mmm$ for space group $I422$ which means that $I(hkl) = I(-h-k-l) = I(hk-l) = I(-hkl) = I(h-kl)$ and $I(khl) = I(hkl)$. The systematic absences are the same for the two space groups $I4$ and $I422$. A structure in $I4$ has eight asymmetric units in the unit cell whereas in $I422$ there are 16. If the space group is incorrectly assigned as $I4$ this means that two subunits have been put in the asymmetric unit, while in $I422$ there is only one.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

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

Brünger, A. T. (1992*b*). *X-PLOR. A system for crystallography and NMR*, Yale University, New Haven, CT, USA.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Brünger, A. T. & Krukowski, A. (1990). *Acta Cryst.* A**46**, 585–593.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Goldman, A., Ollis, D. L. & Steitz, T. A. (1987). *J. Mol. Biol.* **194**, 143–153.

Hodel, A., Kim, S. H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.

Hoier, H., Schlömann, M., Hammer, A., Glusker, J. P., Carrell, H. L., Goldman, A., Stezowski, J. J. & Heinemann, U. (1994). *Acta Cryst.* D**50**, 75–84.

Jones, T. A. & Thirup, S. (1986). *EMBO J.* **5**, 819–822.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T. A. (1994). *Structure*, **2**, 1241–1258.

Kleywegt, G. J. & Jones, T. A. (1995*a*). *Structure*, **3**, 535–540.

Kleywegt, G. J. & Jones, T. A. (1995*b*). In *Making the Most of Your Model*, edited by W. N. Hunter, J. M. Thornton & S. Bailey, pp. 11–24. Warrington: Daresbury Laboratory.

Korn, A. P. & Rose, D. R. (1994). *Protein Eng.* **7**, 961–967.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. (1994). In *From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. A. Waller, pp. 149–159. Warrington: Daresbury Laboratory.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Marsh, R. E. (1995). *Acta Cryst.* B**51**, 897–907.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Rice, L. M. & Brünger, A. T. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.

Vriend, G. & Sander, C. (1993). *J. Appl. Cryst.* **26**, 47–60.

Zou, J. Y. & Mowbray, S. L. (1994). *Acta Cryst.* D**50**, 237–249.