Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization, X-ray diffraction and preliminary structure analysis of *Mycobacterium tuberculosis* chaperonin 10

Michael M. Roberts,[a]* Alun R. Coker,[b] Gianluca Fossati,[c] Paolo Mascagni,[d] Anthony R. M. Coates[a] and Steve P. Wood[b]

[a]Department of Medical Microbiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, England, [b]Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, England, [c]Molecular Immunology Division, NIMR, The Ridgeway, Mill Hill, London NW7 1AA, England, and [d]Italfarmaco Research Center, Via dei Lavoratori 54, 20092 Cinisello B., Milano, Italy

Correspondence e-mail: mroberts@sghms.ac.uk

Received 24 November 1998
Accepted 24 December 1998

The *Mycobacterium tuberculosis* chaperonin 10 (Mtcpn10) has been crystallized by the sitting-drop vapour-diffusion method. The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 76.5$, $b = 87.9$, $c = 124.4$ Å, $\beta = 106.8^\circ$. X-ray diffraction data were collected to 2.8 Å. The self-rotation function and the molecular-replacement solution show that the asymmetric unit contains a dimer of heptamers related by twofold non-crystallographic symmetry. The two heptamers interact through interleaving flexible loops in a similar fashion to *M. leprae* and Gp31 cpn10. In addition to its role in protein folding, Mtcpn10 has unique effects on the growth of host cells and is a major immunogen in tuberculosis infections. The structure determination will permit the analysis of the amino acids identified as important for the protein-folding and cell-signalling activity of Mtcpn10.

1. Introduction

Chaperonins are ubiquitous molecules which are essential for growth (Fayet *et al.*, 1989). They also play an important role by reducing cellular damage caused by stress such as heat shock (Kawata *et al.*, 1994). The chaperonins cpn10 (~10 kDa) and cpn60 (~60 kDa) are found as homologues in numerous organisms (Mande *et al.*, 1996). Mtcpn10 is the homologue of the cpn10 in *Escherichia coli* (GroES), *M. leprae* and the Gp31 protein of the bacteriophage T4, with which it shares 44, 90 and 29% amino-acid sequence identity, respectively. Unlike *E. coli* cpn10, Mtcpn10 is found in large quantities outside the cell in the culture filtrate (Orme *et al.*, 1992). This has important implications for the extracellular activity of Mtcpn10, which has been shown to generate a strong T-cell immune response (Barnes *et al.*, 1992) and also affects the growth of mouse teratocarcinoma cells (Galli *et al.*, 1996). Furthermore, we have shown that Mtcpn10 stimulates bone resorption and is probably an integral component of the pathology of spinal tuberculosis (Meghji *et al.*, 1997).

The cpn10 structures of *E. coli* (resolution 2.8 Å; Hunt *et al.*, 1996), *M. leprae* (resolution 3.5 Å; Mande *et al.*, 1996) and Gp31 (resolution 2.3 Å; Hunt *et al.*, 1997) have been solved by X-ray crystallography as heptamers. Cpn10 complexes with the cpn60 14-mer to facilitate the folding and assembly of other proteins (Ellis, 1994). Cpn10 binds to cpn60 through a mobile loop (Landry *et al.*, 1993) which projects from the outer base of the heptamer (amino acids 16–33 in *E. coli*, 17–34 in *M. leprae*). This interaction is visible in the GroES–GroEL–$(ADP)_7$ crystal structure (Xu *et al.*, 1997). Owing to their high flexibility, these loops are not completely resolved in the cpn10 structures of *E. coli* or *M. leprae*, although they are resolved in the Gp31 structure.

The association between cpn10 monomers is quite labile. Heptameric *E. coli* cpn10 dissociates into a dynamic monomer–heptamer equilibrium below micromolar concentrations (Zondlo *et al.*, 1995). Also, *E. coli* cpn10 is unfolded in the monomeric form but folded in the heptameric form (Boudker *et al.*, 1997). This may explain why *E. coli* cpn10 has the exceptional ability to reassemble into the folded heptameric form after denaturation. *E. coli* cpn60 is unable to refold alone in this way, and requires cpn10 to assemble into a functional protein-folding apparatus (Seale *et al.*, 1996). Similarly, preformed heptameric cpn10 is not required for the formation of an active complex with cpn60 (Seale *et al.*, 1997). This suggests that cpn60 as well as cpn10 can play active roles in assembling each other. The amino-acid sequence differences between homologous cpn10s may also determine how the monomers oligomerize. In a separate study, Mtcpn10 was found to be tetrameric at 5–20 μM concentration. However, in the presence of saturating amounts of divalent cations or at 100 μM protein concentration, Mtcpn10 is heptameric (Fossati *et al.*, 1995). *E. coli* cpn10 was found to be heptameric under all these conditions. The exceptional stability

Table 1
Data collection and processing of Mtcpn10.

Resolution range (Å)	Unique reflections	R_{merge} (%)	Multiplicity	$I > 3\sigma(I)$ (%)	Completeness (%)
30.0–8.0	1639	2.05	3.4	98.8	95.8
8.0–5.1	4772	2.95	3.5	97.3	98.5
5.1–3.8	8816	4.66	3.5	93.0	95.9
3.8–2.8	23094	18.14	3.5	50.0	98.9
Overall	38321	4.99	3.5	67.9	98.0

of *E. coli* cpn10 is reflected in its ability to be purified by heat treatment (Kamireddi *et al.*, 1997).

In this paper, we demonstrate the ability of Mtcpn10 to refold after reversed-phase HPLC purification and lyophilization to form crystals of the Mtcpn10 heptamer. In view of the fact that alternative oligomeric forms of Mtcpn10 may exist in the crystal, it was crucial to examine the self-rotation function for axes of symmetry which define the subunit positions. There have been difficulties in observing the sevenfold symmetry in the self-rotation functions of other heptameric cpn10 structures. It is very weak in GroES (Hunt *et al.*, 1996) and not detected at all in Gp31 (Hunt *et al.*, 1997). The sevenfold axes were not immediately obvious in the Mtcpn10 structure and a careful choice of parameters for the self-rotation function was required to confirm that the Mtcpn10 structure is heptameric.

2. Methods and materials

The 99 amino-acid sequence of Mtcpn10 was overexpressed in *E. coli* (Fossati *et al.*, 1999). Cells were disrupted by sonication and the lysate was clarified by centrifugation. The composition of the lysate was adjusted to 2%(*v*/*v*) acetonitrile and 0.1%(*v*/*v*) trifluoroacetic acid at pH 2.2, whereupon most *E. coli* proteins precipitated. This treatment does not affect Mtcpn10 solubility. Mtcpn10 was purified using a reversed-phase HPLC procedure with an acetonitrile gradient, essentially as described in Legname *et al.* (1995), and then freeze dried. The optimum conditions for Mtcpn10 crystallization were selected from a sparse-matrix screen (Jancarik & Kim, 1991) using Crystal Screen solutions (Hampton Research) by the hanging-drop vapour-diffusion method at 294 K. Lyophilized Mtcpn10 was dissolved in molecular-biology grade water (BDH) and mixed with an equal volume of each Crystal Screen solution. From these preliminary vapour-diffusion experiments, solution conditions were optimized by the sitting-drop method for the growth of crystals of sufficient quality for X-ray diffraction.

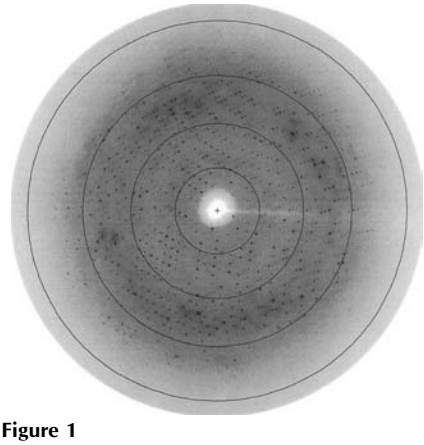

Figure 1
An X-ray diffraction image taken at CCLRC Daresbury Laboratory on station 7.2, recorded on a large MAR Research image plate with an X-ray wavelength of 1.488 and crystal-to-detector distance of 230 mm. The circles indicate diffraction limits of 11.2, 5.6, 3.7 and 2.8 Å, and a rotation angle of 1° was used.

X-ray data collection on Mtcpn10 crystals at room temperature results in a decay in X-ray diffraction with time. Therefore, X-ray data were collected at 100 K. A data set was collected by the rotation method with 1° rotations per frame at an X-ray wavelength of 1.488 Å on station 7.2 at the CCLRC Synchrotron Source, Daresbury Laboratory, using a 30 cm MAR Research image plate. Data were processed and analysed using the *XDS* package (Kabsch, 1988) and the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). Self-rotation functions were calculated using the program *POLARRFN* (*CCP*4 suite); *E* values generated from *ECALC* (*CCP*4 suite) were used rather than *F* values, since the peaks of the self-rotation function were much sharper with the former. All data from 30 to 2.8 Å were used for molecular replacement and rigid-body refinement with *X-PLOR* 3.851 (Brünger, 1996).

To generate the molecular-replacement search model, the GroES structure was numbered according to the homologous amino acids in the Mtcpn10 sequence and modified by deleting amino acids 16–34, 49–57 and 70–72. These regions correspond to conformationally flexible loops or significant differences in sequence to Mtcpn10. All remaining amino-acid differences with Mtcpn10 were mutated to alanines and the overall temperature factor was set to $B = 30$ Å^2. To determine phases for the Mtcpn10 data, a direct rotation search with this modified GroES heptamer search model followed by Patterson correlation (PC) refinement was performed with a 45 Å Patterson vector (Brünger, 1990; DeLano & Brünger, 1995).

After rigid-body refinement of the model, the non-crystallographic symmetry (NCS) operators relating the 14 subunits were calculated with *LSQKAB*. $(2|F_o| - |F_c|, \alpha_c)$ electron-density maps were calculated with the *CCP*4 suite of programs. The superimposition of the partial GroES search model onto the electron-density maps was viewed using the *QUANTA*97 program (Molecular Simulations, Inc.).

3. Results and discussion

3.1. Crystallization

During the initial sparse-matrix screen using the hanging-drop vapour-diffusion method, tiny crystals of Mtcpn10 (<0.1 mm) appeared after two weeks with 5 µl Crystal Screen solution 1 (20 m*M* $CaCl_2$, 0.1 *M* sodium acetate and 30% MPD at pH 5.4) and 5 µl Mtcpn10 (20 mg ml^{-1}). The growth of larger Mtcpn10 crystals was achieved by using the sitting-drop method with Crystal Screen solution 1 diluted to 80% with water and Mtcpn10 (30 mg ml^{-1}). Crystals with a monoclinic morphology and maximum dimensions 0.5 × 1.0 × 1.0 mm were obtained in one week. Crystals were grown under the same conditions from a 35 mg ml^{-1} stock solution (10 µl). This was equilibrated with precipitant (10 µl) for 2 d before seeding with crystals from a previous preparation. Monoclinic crystals of an improved morphology grew to their full size (0.5 × 1.0 × 1.0 mm) in 3 d. The crystal used for X-ray data collection was selected from this preparation.

3.2. Data collection and processing

The crystals diffract to 2.8 Å resolution (Fig. 1) and the unit-cell parameters were determined to be $a = 76.5$, $b = 87.9$, $c = 124.4$ Å, $\beta = 106.8°$, with a consequent cell volume of 0.8×10^6 Å^3. Mtcpn10 has a molecular weight of 10674 Da (Fossati *et al.*, 1995). With four heptamers in the unit cell, this gives a value for V_m of 2.68 Å^3 Da^{-1} and a solvent content of 46%, which lies within the normal range for proteins (Matthews, 1977). The space group was determined to be $P2_1$. Data processing gave a set which was 98% complete in the 30–2.8 Å resolution range, with 134674 measurements of 38321 unique reflections, an overall multiplicity of 3.5 and a merging *R* factor of 5.0% (Table 1).

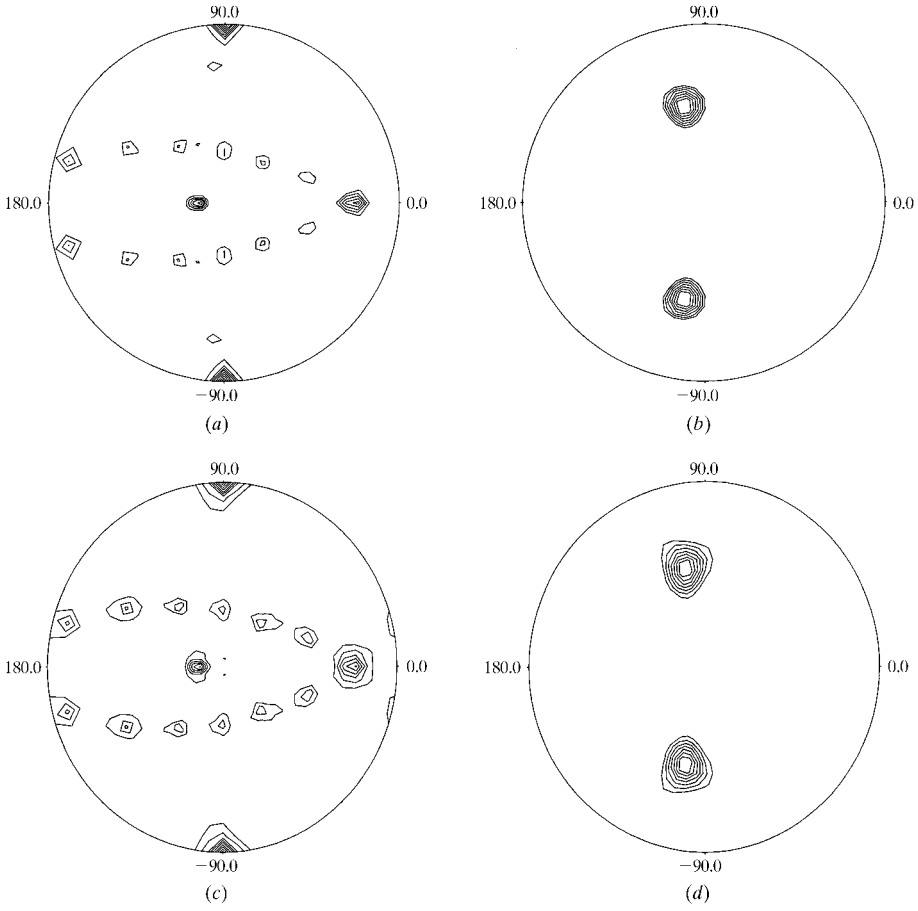

Figure 2
Stereographic projections of self-rotation functions produced by *POLARRFN* with normalized data (*E* values) from 10 to 5.1 Å resolution with a 30 Å integration radius. (*a*) and (*c*) illustrate the twofold axes on $\kappa = 180^\circ$. The peaks corresponding to the crystallographic 2_1 axis can be seen at the perimeter of the plot ($\omega = 90^\circ$) and occur at φ values of 90 and -90°. Two strings of seven unique NCS twofold axes are generated from each double heptamer in the unit cell. These appear as arcs above and below the equator and reflect the tilt in the plane of each double ring with respect to the *xz* plane owing to their positioning about the 2_1 axis. The NCS twofold axes from each heptamer coincide to give a prominent NCS axis on the right-hand side. Together with the crystallographic 2_1 axis, this generates an orthogonal NCS axis as a peak close to the centre of the projection ($\omega = 0^\circ$). (*b*) and (*d*) illustrate the sevenfold axes on $\kappa = 51^\circ$. The displacement of each peak from the perimeter at $\varphi = 90$ and -90° represents the 35° tilt of the NCS sevenfold axis of each double heptamer related by 2_1 symmetry from the *y* axis. (*a*) and (*b*) represent the experimentally collected data, whilst (*c*) and (*d*) represent data back-transformed from the initial molecular-replacement search model rigid-body refined to 2.8 Å.

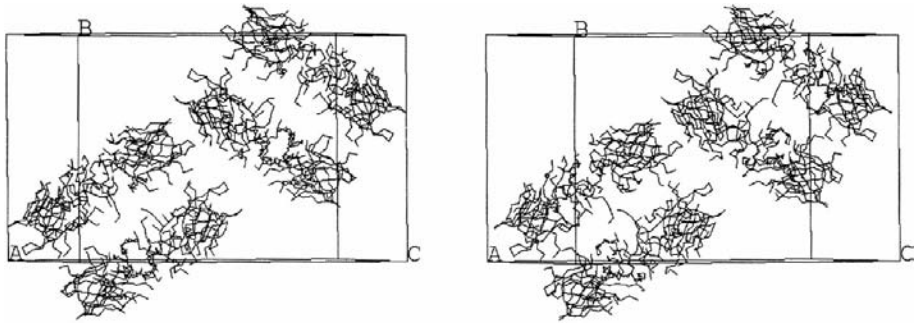

Figure 3
Stereoview of the packing of the rigid-body refined partial GroES search model in the Mtcpn10 unit cell in the direction of one of the major NCS twofold axes which coincides with both double heptamers. The other major NCS twofold axis runs orthogonally in the *xz* plane and is coplanar with the two sevenfold NCS axes. The orientation of the molecules about the 2_1 axis running vertically is clearly seen. Fig. 3 was obtained using *SYBYL* (Tripos Associates, Inc.).

3.3. Self-rotation functions

Self-rotation functions indicated seven twofold axes on the $\kappa = 180^\circ$ plot (Fig. 2*a*) and a sevenfold axis on the $\kappa = 51^\circ$ plot (Fig. 2*b*). The twofold axis at the equator (*xz* plane) combined with the crystallographic twofold screw axis (2_1), generates a mutually perpendicular NCS twofold axis, resulting in local 222 symmetry (Fig. 2*a*).

These peaks only appear with an integration radius of 30 Å and data in the resolution range 10.0–5.1 Å. The high-resolution limit is set by *POLARRFN* to a level permitted by the integration radius.[1] This suggests that Mtcpn10 is in the heptameric form, with two heptamers related by twofold NCS symmetry in the asymmetric unit. The sevenfold NCS axis is tilted by about 35° with respect to the *y* axis (Fig. 2*b*). It has been difficult to observe sevenfold NCS in the self-rotation functions of other cpn10 structures. In this case, the use of *E* rather than *F* values for the calculation of well defined peaks of the self-rotation function has been an important aid in determining the NCS of the Mtcpn10 structure.

3.4. Molecular replacement

The direct rotation search with the partial GroES search model followed by PC refinement produced a number of solutions. The one which gave the best translation function was located at Eulerian angles $\theta_1 = 323.85$, $\theta_2 = 37.41$, $\theta_3 = 345.79^\circ$. At this orientation, the translation function for the first heptamer gave a solution which was 7.6σ above the highest background peak. This solution was combined with the twofold NCS operator to locate the second heptamer. A translational search with the double heptamer gave a unique solution which was 14.2σ above the highest background peak. The packing was checked with *QUANTA*97. The bases of each dome-shaped heptamer are facing each other through a 180° rotation and positioned on a common sevenfold axis. Viewed down the sevenfold axis, the subunits of each heptamer are in a staggered arrangement. Rigid-body refinement of each heptamer followed by the individual subunits with data in increasing shells of resolution from 30–10 Å to 30–2.8 Å gave an initial *R* value of 0.52. To confirm that the molecular-replacement solution was correct, a self-rotation function was calculated using data back-transformed from this rigid-body-refined molecular-replacement solution.

[1] The coupling between the integration radius and the maximum resolution used is $R_{MAX} = I/5.83$.

This showed close agreement with the experimental data (Fig. 2) and gives a strong indication that the molecular-replacement solution serves as a good initial model for the NCS symmetry relating the Mtcpn10 subunit positions. The unit-cell packing in Fig. 3 illustrates the orientations of the symmetry axes relating the Mtcpn10 molecules, and a direct comparison can be made with the stereographic projections in Fig. 2, where all the peaks can be accounted for by the Mtcpn10 search model.

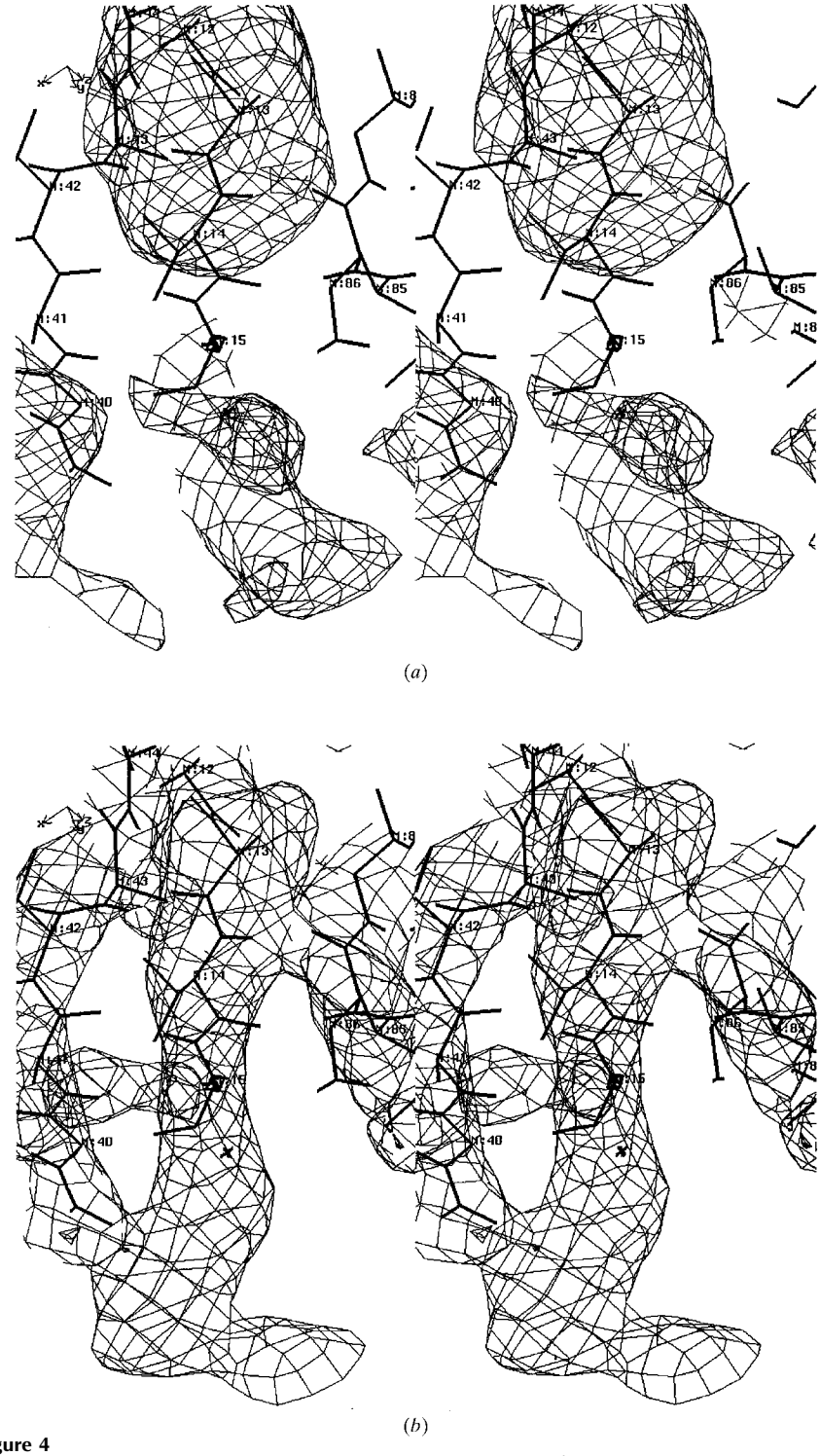

Figure 4
Stereoviews of the partial GroES search model rigid-body refined to 2.8 Å at amino acid 15 before the flexible loop deletion (amino acids 16–34) superimposed on $(2|F_o| - |F_c|, \alpha_c)$ electron-density maps contoured at 1σ level. (*a*) Non-averaged map calculated with model phases to 2.8 Å. (*b*) NCS-averaged map calculated by phase extension from 8 to 2.8 Å using NCS matrices. Amino acids corresponding to Leu13, Gln15 and Val43 in Mtcpn10 are truncated to alanines, but electron density arising from these side chains is clearly visible. Figure produced with *SETOR* (Evans, 1990).

3.5. Electron-density maps

The initial $(2|F_o| - |F_c|, \alpha_c)$ map calculated from the search model rigid-body refined to 2.8 Å was poor and did not show good agreement with the model (Fig. 4*a*). To overcome this, the *DM* program was used to remove model phase bias from the $2|F_o| - |F_c|, \alpha_c$ map by phase extension from 8 to 2.8 Å, using the model rigid-body refined with 30–8 Å data and the NCS relating the Mtcpn10 subunits, with solvent flattening and histogram matching. This gave a considerable improvement in the resulting NCS-averaged map (Fig. 4*b*), which matched the model rigid-body refined to 2.8 Å better than the model rigid-body refined to 8 Å, indicating that the phase extension had converged toward the 2.8 Å resolution NCS matrices. In this phase-extended map there was contiguous electron density which showed good agreement with the model and extended into the gaps that were deleted in the model. In particular, the appearance of a significant portion of the flexible loop in the map which was not present in the initial search model is encouraging (Fig. 4*b*). The flexible loop density extends to the neighbouring heptamer where it interacts with the subunit interface. Therefore, the dimer of heptamers interacts through 14 interleaving flexible loops. Also, in many cases there is visible electron density in the NCS-averaged phase-extended map corresponding to the side chains of Mtcpn10 at positions in the GroES model where differing amino acids were truncated to alanines (Fig. 4*b*). This will allow model building of the Mtcpn10 structure to proceed.

The matching self-rotation functions of processed X-ray data and back-transformed data from the search model and the power of NCS in phase extension and map averaging illustrates an example of the use of additional criteria to confirm the validity of the molecular-replacement search model in a situation where the initial *R* value and model-derived electron-density maps appear unconvincing.

We thank Dr John Hunt for the supply of the GroES coordinates for molecular replacement, Dr Gareth Lewis for computing support and Drs Jon Cooper, John Reid, Darren Thompson and Kevin Cowtan for their helpful advice on running the programs.

References

Barnes, P. F., Mehra, V., Rivoire, B., Fong, S.-J., Brennan, P. J., Voegtline, M. S., Minden, P., Houghten, R. A., Bloom, B. R. & Modlin, R. L. (1992). *J. Immunol.* **148**, 1835–1840.

Boudker, O., Todd, M. J. & Freire, E. (1997). *J. Mol. Biol.* **272**, 770–779.

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T. (1996). *X-PLOR Version* 3.851. Yale University, New Haven, Connecticut, USA.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.

Ellis, R. J. (1994). *Curr. Biol.* **4**, 633–635.

Evans, S. V. (1990). *J. Mol. Graph.* **11**, 134–138.

Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. P. (1989). *J. Bacteriol.* **136**, 955–964.

Fossati, G., Lucietto, P., Giuliani, P., Coates, A. R. M., Harding, S., Cölfen, H., Legname, G., Chan, E., Zaliani, A. & Mascagni, P. (1995). *J. Biol. Chem.* **270**, 26159–26167.

Fossati, G., Lucietto, P., Legname, G., Giuliani, P., Ball, H. L., Leoni, F., Modena, D., Coates, A. R. M. & Mascagni, P. (1999). Submitted.

Galli, G., Ghezzi, P., Mascagni, P., Marcucci, F. & Fratelli, M. (1996). *In Vitro Cell. Dev. Biol.* **32**, 446–450.

Hunt, J. F., van der Vies, S. M., Henry, L. & Deisenhofer, J. (1997). *Cell*, **90**, 361–371.

Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. M. & Deisenhofer, J. (1996). *Nature (London)*, **379**, 37–45.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Kamireddi, M., Eisenstein, E. & Reddy, P. (1997). *Protein Expr. Purif.* **11**, 47–52.

Kawata, Y., Nosaka, K., Hongo, K., Mizobata, T. & Nagai, J. (1994). *FEBS Lett.* **345**, 229–232.

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. (1993). *Nature (London)*, **364**, 255–258.

Legname, G., Fossati, G., Gromo, G., Monzini, N., Marcucci, F. & Modena, D. (1995). *FEBS Lett.* **361**, 211–214.

Mande, S. C., Mehra, V., Bloom, B. R. & Hol, W. G. (1996). *Science*, **271**, 203–207.

Matthews, B. W. (1977). *The Proteins*, 3rd ed., Vol. 3, edited by H. Neurath & R. L. Hill, pp. 468–477. New York: Academic Press.

Meghji, S., White, P. A., Nair, S. P., Reddi, K., Heron, K., Henderson, B., Zaliani, A., Fossati, G., Mascagni, P., Hunt, J. F., Roberts, M. M. & Coates, A. R. M. (1997). *J. Exp. Med.* **186**, 1241–1246.

Orme, I. M., Miller, E. S., Roberts, A. D., Furney, S. K., Griffin, J. P., Dobos, K. M., Chi, D., Rivoire, B. & Brennan, P. J. (1992). *J. Immunol.* **148**, 189–196.

Seale, J. W., Chirgwin, J. M., Demeler, B. & Horowitz, P. M. (1997). *J. Protein Chem.* **16**, 661–668.

Seale, J. W., Gorovits, B. M., Ybarra, J. & Horowitz, P. M. (1996). *Biochemistry*, **35**, 4079–4083.

Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). *Nature (London)*, **388**, 741–749.

Zondlo, J., Fisher, K. E., Lin, Z., Ducote, K. R. & Eisenstein, E. (1995). *Biochemistry*, **34**, 10334–10339.