Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Phasing protein structures using the group–subgroup relation

**Luigi Di Costanzo, Federico Forneris, Silvano Geremia* and Lucio Randaccio**

Centre of Excellence in Biocrystallography, Department of Chemical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy

Correspondence e-mail: geremia@univ.trieste.it

Received 4 March 2003
Accepted 5 June 2003

**PDB References:** di-$Co^{II}$-DF1-L13A, form 1, 1ovu, r1ovusf; di-$Co^{II}$-DF1-L13A, form 2, 1ovv, r1ovvsf.

Diffraction data from two non-isomorphous crystals (forms 1 and 2) of an artificial protein with a four-helix bundle motif, di-$Co^{II}$-DF1-L13A, have been collected using synchrotron radiation. The phase of form 1 has been assigned using the group and minimal non-isomorphic supergroup relation between the space group of the previously determined di-$Mn^{II}$-DF1-L13G structure and the space group of this form. This unconventional method of solving the phase problem has also been tested with form 2 using a reverse relation. The structure of the latter form has been solved using the group and maximal non-isomorphic subgroup relation with the space group of form 2 of the analogous dimanganese protein. This application has shown that this phasing method can be used for solving the protein structures of polymorphic crystals as an alternative to the molecular-replacement method.

## 1. Introduction

The behaviour of macromolecules in the crystallization process is complex and rather unpredictable owing to their varied shapes and polyvalent surface character. Many proteins can be crystallized in several different unit cells and space groups. Polymorphism is quite common for biological macromolecules and different crystal forms may also develop in the same crystallizing drop. Frequently, polymorphism is associated with a different number of macromolecules being present in the asymmetric unit (in nine lamprey haemoglobin crystal forms the number of crystallographically independent monomers ranges from 1 to 16; Hendrickson *et al.*, 1968). In crystal forms containing more than one independent monomer, the macromolecules with very similar conformation (protein molecules in different environments are never identical) are related by non-crystallographic symmetry (NCS). NCS may be used to improve the electron density and structural refinement and in the most favourable cases may allow direct structure determination (Jack, 1973). Nevertheless, when the structure of one crystal form is known, molecular replacement is the conventional method used for the structure determination of non-isomorphic crystals of the same or similar molecules. However, in the presence of a large number of crystallographically independent symmetric homomultimers, the search for the rototranslation matrixes can be problematic because of the combination of crystallographic and non-crystallographic symmetries.

Different crystal forms of the same or analogous molecules having similar crystal packing may be derived by transformation of a crystallographic symmetry operation to NCS. Typical examples are temperature-dependent phase transi-

tions, which are not uncommon for inorganic materials (Igartua *et al.*, 1996) and have also been reported for several protein crystals (Campobasso *et al.*, 1998). The crystallization of heavy-atom derivatives can also occur in undesirable non-isomorphous forms related to the native protein form by the presence or absence of pseudo-symmetries (Poulsen *et al.*, 2001). The loss of some crystallographic symmetry operators produces changes in the space group and unit cell; the group–subgroup relations between space groups introduced by Hermann (1929) are very important in the study of these order–disorder problems.

Recently, we have undertaken the crystal structure analysis of DF1, a four-helix bundle designed protein (Lombardi *et al.*, 2000). DF1 is a non-covalently associated homodimer of two helix–loop–helix hairpin motifs with a dimetal site in the centre of the four-helix-bundle structure. In particular, the crystal structure of the zinc derivative, di-Zn-DF1 (Lombardi *et al.*, 2000), and of the manganese derivatives of some variants of DF1, di-$Mn^{II}$-DF1-L13A (form 1 and 2; Di Costanzo *et al.*, 2001) and di-$Mn^{II}$-DF1-L13G (DeGrado *et al.*, 2003), have been determined. In order to investigate the influence of other metal centres, we have undertaken the structural characterization of the cobalt derivative di-$Co^{II}$-DF1-L13A. Two crystal forms were obtained in the same crystallizing drop and the phase problem was solved by investigation of the relations between unit-cell parameters and between space groups of polymorphic protein crystals. Here, we present this unconventional method used for solving these structures, which represents a further resource to solve the phase problem when the molecular-replacement method fails.

## 2. Materials and methods

### 2.1. Crystallization

Crystals of di-$Co^{II}$-DF1-L13A were grown at 277 K by the hanging-drop vapour-diffusion technique. The pure powder of the 48 amino-acid synthetic peptide is not completely soluble in water and was dissolved in DMSO to obtain a solution of 100 mg $ml^{-1}$ concentration. This solution was diluted tenfold with water and centrifuged to remove undissolved materials. The drops were prepared by adding 2 µl of the peptide solution (1.5 m*M*) to 2 µl of a reservoir solution containing PEG 400, 30 m*M* $Co(CH_3COO)_2$, 0.1 *M* Tris–HCl pH 7.5. The drops were equilibrated against 1.0 ml of reservoir solution. The crystallization conditions were optimized to give large three-dimensional crystals. Diamond-shaped crystals of di-$Co^{II}$-DF1-L13A crystals grew to typical dimensions of 0.3 × 0.3 × 0.1 mm after one month from a reservoir containing 43%(*w*/*v*) PEG 400. The 20-fold excess of $Co(CH_3COO)_2$ was crucial to obtain crystals of di-$Co^{II}$-DF1-L13A suitable for X-ray diffraction experiments.

### 2.2. Data collection and processing

X-ray diffraction experiments were carried out at the Elettra synchrotron. Data were collected using monochromatic radiation with wavelength 1.200 Å and a MAR Research 345 mm imaging plate as a detector. The crystals were harvested into mother liquor with a small loop of fine rayon fibre and flash-frozen in a stream of $N_2$ at 100 K. Two full data sets were collected from two non-isomorphous crystals of di-$Co^{II}$-DF1-L13A (forms 1 and 2) grown in the same crystallizing drop.

**Table 1**
X-ray data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | Form 1 | Form 2 |
|---|---|---|
| X-ray source | Elettra | Elettra |
| Resolution range (Å) | 43.2–3.1 (3.27–3.1) | 33.3–2.9 (3.06–2.9) |
| Total reflections | 17041 | 37159 |
| Unique reflections | 4679 | 6492 |
| $I/\sigma(I)$ | 5.9 (1.9) | 11.2 (3.1) |
| Completeness (%) | 97.7 (97.7) | 96.8 (89.0) |
| Multiplicity | 3.6 (3.8) | 5.7 (5.4) |
| $R_{merge}(I)$ (%) | 23.6 (51.5) | 10.9 (57.3) |
| Refinement | | |
| *R* factor (%) | 24.7 | 26.7 |
| $R_{free}$ (%) | 30.4 | 32.0 |
| R.m.s. deviations from ideal geometry | | |
| Bond lengths (Å) | 0.025 | 0.028 |
| Bond angle distances (°) | 2.88 | 2.52 |
| Average *B* factors (Å$^2$) | 58.8 | 61.5 |

The determination of unit-cell parameters, integration of reflection intensities and data scaling were performed using *MOSFLM* and *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994) (Table 1).

The diffraction pattern of the form 1 crystal of di-$Co^{II}$-DF1-L13A was indexed using a *C*-centred orthorhombic unit cell ($a$ = 89.78, $b$ = 147.72, $c$ = 37.60 Å) with systematic absences in agreement with the $C222_1$ space group. Analysis of the diffraction data from form 2 reveals the crystal to have a primitive orthorhombic Bravais lattice ($a$ = 36.92, $b$ = 80.05, $c$ = 96.62 Å) with systematic absences in agreement with the $P2_12_12_1$ space group.

### 2.3. Phasing process

The unit-cell parameters and space group of crystal form 1 of di-$Co^{II}$-DF1-L13A produce a high uncertainty in the establishment of the number of monomers in the asymmetric unit. The crystal volume per protein mass ($V_M$) in previously determined DF1 structures ranges from 2.65 to 1.82 Å$^3$ $Da^{-1}$ (Table 2). Form 1 can accommodate four to six crystallographic independent monomers (corresponding $V_M$ values vary from 2.66 to 1.77 Å$^3$ $Da^{-1}$). The undetermined number of independent monomers, together with the presence of twofold axes in the $C222_1$ space group which could permit the location of homodimers on crystallographic symmetries and the relatively low resolution of diffraction data (Table 1), complicated the phasing process of this structure. Several attempts to solve the phasing problem of form 1 through the conventional molecular-replacement method failed.

The comparison of unit-cell dimensions with previously determined DF1 structures (Table 2) revealed a close rela-

**Table 2**
Summary of unit cells and space groups in DF structures.

| Structure | Space group | $a$ (Å) | $b$ (Å) | $c$ (Å) | Independent monomers | $V_M$ (Å$^3$ Da$^{-1}$) | PDB code |
|---|---|---|---|---|---|---|---|
| di-Zn$^{II}$-DF1 | $C222_1$ | 36.07 | 89.16 | 79.89 | 3 | 1.82 | 1ec5 |
| di-Mn$^{II}$-DF1-L13A (form 1) | $P2_12_12_1$ | 37.38 | 80.12 | 99.93 | 6 | 2.13 | 1jm0 |
| di-Mn$^{II}$-DF1-L13A (form 2) | $C222_1$ | 37.12 | 112.45 | 79.88 | 3 | 2.37 | 1jmb |
| di-Mn$^{II}$-DF1-L13G | $P2_12_12_1$ | 38.20 | 89.30 | 146.40 | 8 | 2.68 | 1lt1 |
| di-Co$^{II}$-DF1-L13A (form 1) | $C222_1$ | 89.78 | 147.72 | 37.60 | 4 | 2.66 | 1ovu |
| di-Co$^{II}$-DF1-L13A (form 2) | $P2_12_12_1$ | 36.92 | 80.05 | 96.62 | 6 | 2.03 | 1ovv |

tionship between the form 1 and the di-Mn$^{II}$-DF1-L13G ($a$ = 38.20, $b$ = 89.30, $c$ = 146.40 Å) structures. Thus, we decided to exploit the group–subgroup relation between the space groups of these two structures.

**2.3.1. Phasing form 1 by group and minimal non-isomorphic supergroup relation**. The $C222_1$ space group of di-Co$^{II}$-DF1-L13A form 1 is a minimal non-isomorphic supergroup of the $P2_12_12_1$ space group of the di-Mn$^{II}$-DF1-L13G structure. Space-group diagrams are particularly useful to find the possible relation between these two structures (Fig. 1*a*). To transform the $P2_12_12_1$ space group of the di-Mn$^{II}$-DF1-L13G structure to the minimal non-isomorphic supergroup $C222_1$ of di-Co$^{II}$-DF1-L13A form 1, three steps are necessary: axes permutation, origin shift and acquisition of twofold symmetry operators (Fig. 1*a*). For the last operation, it is necessary to convert pseudo-symmetry operators of the di-Mn$^{II}$-DF1-L13G structure into crystallographic symmetry operators. In particular, two pseudo-$C_2$ symmetry axes passing through two homodimers of the independent unit of di-Mn$^{II}$-DF1-L13G are transformed into crystallographic twofold axes in the new structure (Fig. 2*a*). As a consequence, the eight crystallographically independent monomers in di-Mn$^{II}$-DF1-L13G structure are reduced to four in the new structure and the 'isolated' monomers of this new asymmetric unit are located near the crystallographic twofold axes in such a way as to form four-helix bundles. Half of the di-Mn$^{II}$-DF1-L13G asymmetric unit was opportunely transformed (see Fig. 2*a*) and used as starting model for the di-Co$^{II}$-DF1-L13A form 1 structure. The $R$ factor of 0.353 obtained in the rigid-body refinement of the four monomers, each treated as an independent unit, confirmed the success of this phasing procedure.

**2.3.2. Phasing form 2 by group and maximal non-isomorphic subgroup relation**. The absence of twofold rotation axes in the $P2_12_12_1$ space group of di-Co$^{II}$-DF1-L13A form 2 simplifies the problem of the number of molecules in the asymmetric unit, since an odd number of monomers in the asymmetric unit is not compatible with the presence of homodimers in the crystal. Six monomers per asymmetric unit are present in di-Co$^{II}$-DF1-L13A form 2 crystals ($V_M$ = 2.03 Å$^3$ Da$^{-1}$). This structure is isomorphic with di-Mn$^{II}$-DF1-L13A form 1 ($a$ = 37.38, $b$ = 80.12, $c$ = 99.93 Å; space group $P2_12_12_1$) and was initially solved by the conventional molecular-replacement method. However, the unit-cell parameters and space group of di-Co$^{II}$-DF1-L13A form 2 revealed also a close relationship with the di-Mn$^{II}$-DF1-L13A form 2 structure ($a$ = 37.12, $b$ = 112.45, $c$ = 79.88 Å; space group $C222_1$). This suggested application of the group–subgroup relation to solve the phase problem for this crystal form also. The $P2_12_12_1$ space group of di-Co$^{II}$-DF1-L13A form 2 is a maximal non-isomorphic subgroup of the $C222_1$ space group of the di-Mn$^{II}$-DF1-L13A form 2 structure. The transformation of the $C222_1$ space group of di-Mn$^{II}$-DF1-L13A form 2 in the maximal non-isomorphic subgroup $P2_12_12_1$ of di-Co$^{II}$-DF1-L13A form 2 requires swapping of the $b$ and $c$ axes (changing the sense of one axis to maintain the right-handed system), an origin shift and the loss of twofold symmetry operators (Fig. 1*b*).

As a consequence, the number of crystallographically independent monomers doubles from three in di-Mn$^{II}$-DF1-L13A form 2 to six in the new structure (Fig. 2*b*). The asymmetric unit of di-Mn$^{II}$-DF1-L13A form 2 was opportunely doubled by application of a twofold symmetry operator and transformed by the operations reported in Fig. 2(*b*). These atomic coordinates were used as a starting model for phasing the di-Co$^{II}$-DF1-L13A form 2 structure. The 15% shrinkage of the $c$ axis produced a distorted molecule with r.m.s. deviations from the idealized bond lengths and angles of 0.09 Å and 5.6°, respectively. The significant difference in the unit cell between the two structures suggested a gradual refinement strategy starting with data at low resolution (which are less affected by a bad starting model) and a coarse refinement of the entire model. The rigid-body refinement of the whole model with data in the 15–4 Å resolution range gave an initial $R$ factor of 0.523 that decreased to 0.493 after refinement of the three dimers as independent rigid bodies using all data. The successive rigid-body refinement of the six monomers, each one treated as an independent unit, gave an $R$ factor of 0.456 that dropped to 0.346 in the three successive cycles of restraint refinement. These results confirmed the success of the phasing procedure.

### 2.4. Refinement

Forms 1 and 2 of the di-Co$^{II}$-DF1-L13A complex were refined using the program *REFMAC* (Murshudov *et al.*, 1997). Models were visualized and modified based on $2F_o - F_c$ and $F_o - F_c$ residual electron-density maps using the graphics program *O* (Jones *et al.*, 1991). The initial di-cobalt ions of forms 1 and 2 were positioned in the $2F_o - F_c$ electron-density map with a $10\sigma$ cutoff. After refinement, the Debye–Waller parameters of the cobalt ions, which are similar to those of the surrounding residues, indicate full occupancy of these sites in both structures. The residual $F_o - F_c$ electron-density maps clearly show the presence of several cobalt ions coordinated to the side chains of surface residues. In particular, three and one 'external' crystallographically independent cobalt ions were detected in forms 1 and 2, respectively. Ten and eight water molecules complete the asymmetric unit of forms 1 and 2, respectively. Refinement characteristics are gathered in Table 1.

## 3. Conclusions

The cobalt derivative of the variant DF1-L13A, a designed four-helix bundle protein, has been crystallized. Diffraction data have shown the presence of two non-isomorphous crystals grown in the same crystallizing drop. Several attempts to solve the phasing problem of form 1 through the conventional molecular-replacement method failed. The phase problem has been solved using an alternative method which exploits the group–subgroup relation between space groups of non-isomorphic structures. In particular, the group and minimal non-isomorphic supergroup relation between the space group of the previously determined analogous di-$Mn^{II}$-DF1-L13G structure and the space group of this new crystalline form has been used to prepare a starting model for the di-$Co^{II}$-DF1-L13A form 1 structure. This unconventional method of solving the phase problem has also been applied to form 2. In this case, the structure of this form was solved using a reverse relationship between space groups (with respect to the first case). In fact, the space group of form 2 is a maximal non-isomorphic subgroup of the space group of the analogous manganese-derivative structure (form 2). Such application has demonstrated that this phasing method can be used for solving protein structures of polymorphic crystals as an alternative to molecular-replacement method.

The five principal stages in phasing a structure using this method can be summarized as follows.

(i) Searching for a metric relationship between the unit cells of the unknown and a known structure.

(ii) Searching for a group–subgroup relation between the space groups of these two structures. [Volume A of *International Tables for Crystallography* (2002) reports for each space group the maximal non-isomorphic subgroups and the minimal non-isomorphic supergroups. To extend the analysis to all subgroups or supergroups special programs such as *SUPERGROUPS* (Ivantchev *et al.*, 2002) are necessary.]

(iii) Transformation of the atomic coordinates of the reference structure to that of the unknown structure. For this step, it is necessary to consider axis permutations [if only conventional cells (Mighell, 2001) are used] and change of the origin (for example, in the studied

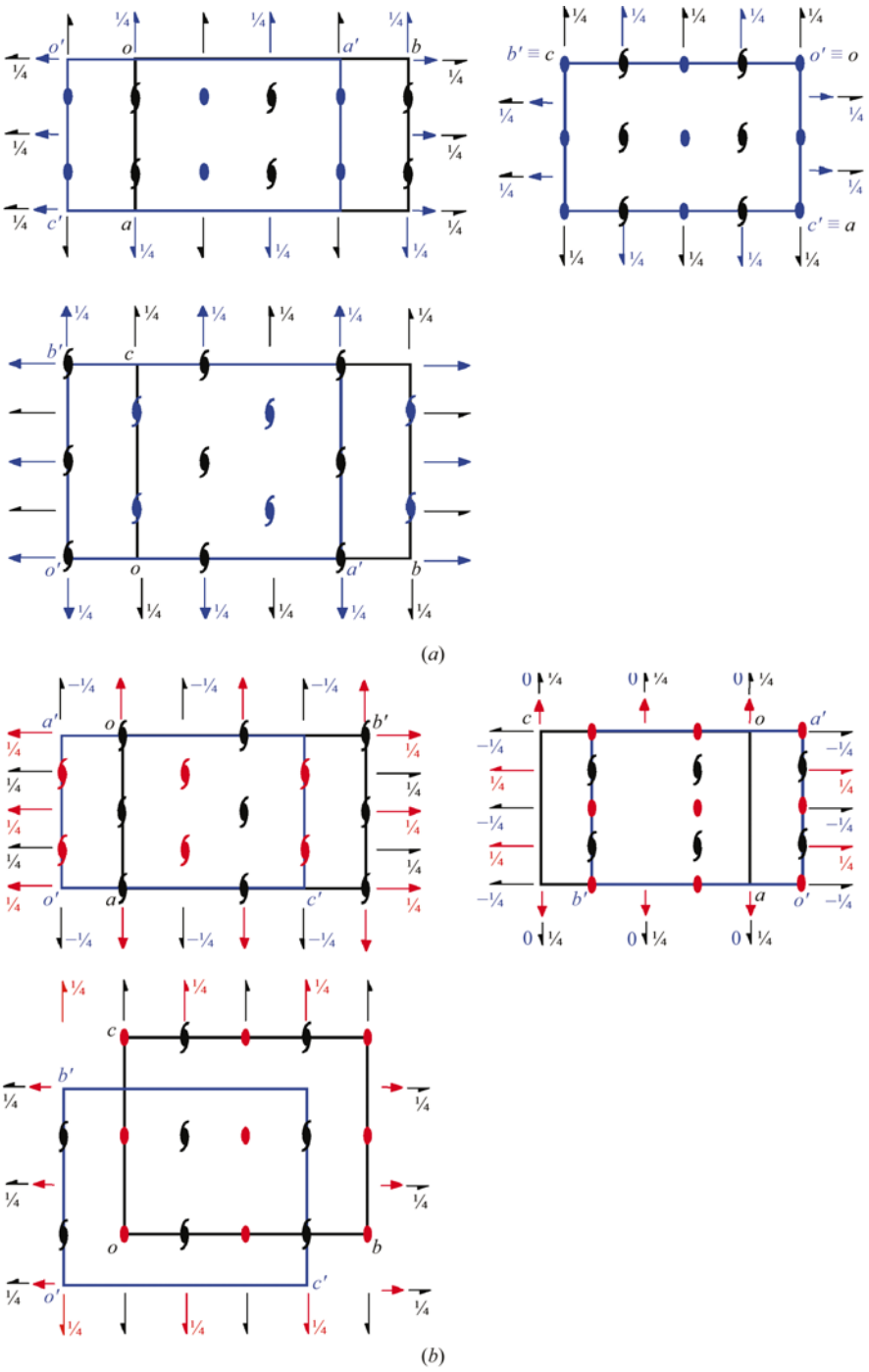


**Figure 1**
(*a*) Superposition of the $P2_12_12_1$ space-group diagram of the di-$Mn^{II}$-DF1-L13G structure (black) with that of the minimal non-isomorphic supergroup $C222_1$ of the di-$Co^{II}$-DF1-L13A form 1 (blue) shows the origin shift ($\frac{1}{4}$, 0, 0), axes permutation (*c*, *a*, *b*) and the acquired symmetry operators (blue symbols) to transform the structure of the di-$Mn^{II}$-DF1-L13G to the that of di-$Co^{II}$-DF1-L13A form 1. (*b*) Superposition of the $C222_1$ space-group diagram of the di-$Mn^{II}$-DF1-L13A form 2 structure (black) with the space-group diagram of the maximal non-isomorphic sub-groups $P2_12_12_1$ of the di-$Co^{II}$-DF1-L13A form 2 (blue) shows the origin shift (0, $\frac{1}{4}$, $\frac{1}{4}$), the axes swapping (−*a*, *c*, *b*) and the missing symmetry operators (red symbols) to transform the structure of di-$Mn^{II}$-DF1-L13A form 2 to that of di-$Co^{II}$-DF1-L13A form 2.

case the origin of the $P2_12_12_1$ space group is at the midpoint of three non-intersecting pairs of parallel $2_1$ axes, while the origin of the $C222_1$ space group is at the intersection of a twofold axis parallel to *a* with a screw axis). Space-group diagrams (*International Tables for Crystallography*, 2002) are quite useful for finding the relationship between the two space groups and the transformation of the coordinates (see Fig. 1). The coordinates of the general equivalent positions which are common to both group and subgroup (maintained equivalent position in the maximal non-isomorphic subgroups are also reported in Volume A of *International Tables for Crystallography*, 2002) can be also used in an equations system to find the translation vector.

(iv) Redefinition of the new asymmetric unit. The coordinates of the asymmetric unit of the reference structure have to be doubled using a lost symmetry operator in the space group/subgroup transformation (or halved using a new symmetry operator of the new space group in the case of group/supergroup relation).

(v) Phase assignment to new structure using rigid-body refinement of the new asymmetric unit.

In conclusion, for the application of this method, knowledge of the crystal structure of a non-isomorphic form of the same or similar protein is required (this situation is similar to that of the molecular-replacement method; however, for the present method it is important to know both the model of a single molecule and the crystal packing). In addition, this known crystal structure must have a space group with a group/subgroup (or group/supergroup) relation with the space group of the unknown structure. This method is particularly useful in the presence of a large number of crystallographically independent symmetric homomultimers when the structure solution by molecular replacement could be problematic because of the combination of crystallographic and non-crystallographic symmetries. Furthermore, in addition to the phasing process, this method can provide insight into both phase change between the polymorphic forms and protein packing relation between non-isomorphic forms.

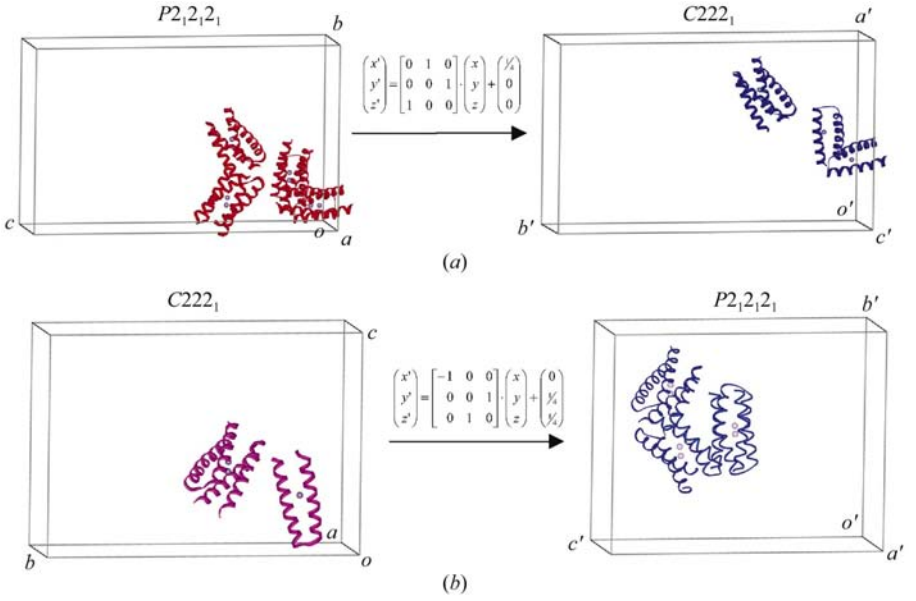


**Figure 2**
Transformation of the asymmetric unit (represented by ribbons) of (*a*) the di-$Mn^{II}$-DF1-L13G structure (space group $P2_12_12_1$) in that of di-$Co^{II}$-DF1-L13A form 1 (space group $C222_1$) and (*b*) the di-$Mn^{II}$-DF1-L13A form 2 (space group $C222_1$) in that of di-$Co^{II}$-DF1-L13A form 2 (space group $P2_12_12_1$). The isolated monomers are located near a crystallographic twofold axis in such a way as to form four-helix bundles. The transformation matrixes and the translation vectors of the atomic coordinates are also reported.

We are grateful for the help of the beamline scientists at the Elettra synchrotron. This work was supported by the MIUR (CofinLab). We thank Louise Johnson for critical reading of the manuscript.

## References

Campobasso, N., Begun, J., Costello, C. A., Begley, T. P. & Ealick, S. E. (1998). *Acta Cryst.* D**54**, 448–450.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeGrado, W. F., Di Costanzo, L., Geremia, S., Lombardi, A., Pavone, V. & Randaccio, L. (2003). *Angew. Chem. Int. Ed. Engl.* **42**, 417–420.
Di Costanzo, L., Wade, H., Geremia, S., Randaccio, L., Pavone, V., DeGrado, W. F. & Lombardi, A. (2001). *J. Am. Chem. Soc.* **123**, 12749–12757.
Hendrickson, W. A., Love, W. E. & Murray, G. C. (1968). *J. Mol. Biol.* **33**, 829–42.
Hermann, C. (1929). *Z. Kristallogr.* **69**, 533–555.
Igartua, J. M., Aroyo, M. I. & PerezMato, J. M. (1996). *Phys. Rev. B*, **54**, 12744–12752.
*International Tables for Crystallography* (2002). Vol. A, *Space Group Symmetry*, 5th ed., edited by T. Hahn. Dordrecht: Kluwer Academic Publishers.
Ivantchev, S., Kroumova, E., Aroyo, M. I., Perez-Mato, J. M., Igartua, J. M., Madariaga, G. & Wondratschek, H. (2002). *J. Appl. Cryst.* **35**, 511–512.
Jack, A. (1973). *Acta Cryst.* A**29**, 545–554.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Lombardi, A., Summa, C. M., Geremia, S., Randaccio, L., Pavone, V. & DeGrado, W. F. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 6298–6305.
Mighell, A. D. (2001). *J. Res. Natl Inst. Stand. Technol.* **106**, 983–995.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Poulsen, J.-C. N., Harris, P., Jensen, K. F. & Larsen, S. (2001). *Acta Cryst.* D**57**, 1251–1259.