Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A corrected space group for *Sulfolobus sulfataricus* 5′-deoxy-5′-methylthioadenosine phosphorylase II

Yang Zhang,[a] Peter H. Zwart[b] and Steven E. Ealick[a]*

[a]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA, and [b]Lawrence Berkeley Laboratory, 1 Cyclotron Road, BLDG 64R0121, Berkeley, CA 94720-8235, USA

Correspondence e-mail: see3@cornell.edu

Received 4 August 2011
Accepted 30 November 2011

5′-Deoxy-5′-methylthioadenosine phosphorylase (MTAP) catalyzes the phosphorolytic cleavage of 5′-deoxy-5′-methylthioadenosine (MTA), a byproduct of polyamine biosynthesis. The *Sulfolobus sulfataricus* genome encodes two MTAPs. SsMTAP I has broad substrate specifity, accepting guanosine, inosine, adenosine and MTA, while SsMTAP II is specific for MTA. SsMTAP I forms a donut-shaped hexamer, while SsMTAP II is a hexamer formed from trimers packed face to face. The structure of SsMTAP II was originally determined in space group *P*1 (PDB entry 2a8y) and showed *R*32 pseudo-symmetry. Post-analysis using *phenix.xtriage* showed that the correct space group is *C*2. Here, the structure refined in space group *C*2 is reported and the factors that initially led to the incorrect space-group assignment are discussed.

1. Introduction

5′-Deoxy-5′-methylthioadenosine phosphorylase (MTAP) catalyzes the reversible phosphorolysis of 5′-deoxy-5′-methylthioadenosine (MTA) to adenine and 5′-deoxy-5′-methylribose 1-phosphate, which is recycled to form methionine in the methionine-salvage pathway (Cacciapuoti *et al.*, 1994). MTA is generated as a byproduct of polyamine biosynthesis (Pegg & Williams-Ashman, 1969). *Sulfolobus sulfataricus* contains two MTAPs: SsMTAP I has broad substrate specificity and is able to cleave adenosine, inosine and guanosine in addition to MTA, while SsMTAP II is specific for MTA. As part of a long-term study on nucleoside phosphorylases, structures of both types of SsMTAP have been determined. SsMTAP I (Appleby *et al.*, 2001) is structurally homologous to *Escherichia coli* purine nucleoside phosphorylase (PNP; Mao *et al.*, 1997); it exists as a donut-shaped hexamer with 32 symmetry and can be thought of as a trimer of dimers. SsMTAP II (Zhang *et al.*, 2006) is also hexameric with 32 symmetry, but exists as a close-packed face-to-face dimer of trimers. Each trimer is structurally homologous to human PNP (Ealick *et al.*, 1990) and human MTAP (Appleby *et al.*, 1999). Despite the difference in oligomeric structure and low sequence identity (∼17%), the monomers of both forms of SsMTAP belong to the nucleoside phophorylase I family (Pugmire & Ealick, 2002).

Crystals of SsMTAP II diffracted to 1.45 Å resolution with unit-cell parameters $a = 96.60$, $b = 96.56$, $c = 96.63$ Å, $\alpha = 91.57$, $\beta = 91.23$, $\gamma = 91.52^\circ$ (Zhang *et al.*, 2006). The similarities in the lengths of the unit-cell axes and in the values of the unit-cell angles suggested a rhombohedral space group; however, we were unable to obtain reasonable data-processing statistics

using *HKL*-2000 (Otwinowski & Minor, 1997) for either space group *R*32 or *R*3 (Fig. 1). Likewise, we were unable to achieve acceptable data-processing statistics for space group *C*2, which is a maximal non-isomorphic subgroup of *R*32. Consequently, we determined the structure in space group *P*1 with two complete hexamers in the unit cell.

Analysis of the crystal packing showed that the arrangement of hexamers approximated *R*32 space-group symmetry; however, the noncrystallographic threefold axis of the hexamer was tilted by 1.1° with respect to the unit-cell body diagonal, which would correspond to the crystallographic threefold axis in the rhombohedral setting of space group *R*32. The noncrystallographic twofold axes of the hexamer were perpendicular to the noncrystallographic threefold axis within experimental error. The separation between the centers of the two hexamers was approximately 4 Å less than half the length of the body diagonal. The structure was refined to an R of 18.0% and an $R_{\rm free}$ of 20.5%.

2. Experimental procedures

2.1. Post-analysis and determination of the correct space group

After the structure of SsMTAP II was published in 2006 (Zhang *et al.*, 2006), *phenix.xtriage*, a computer program for the detection of twinning, pseudosymmetry and other crystal pathologies, became routinely available (Zwart *et al.*, 2005*a*,*b*).

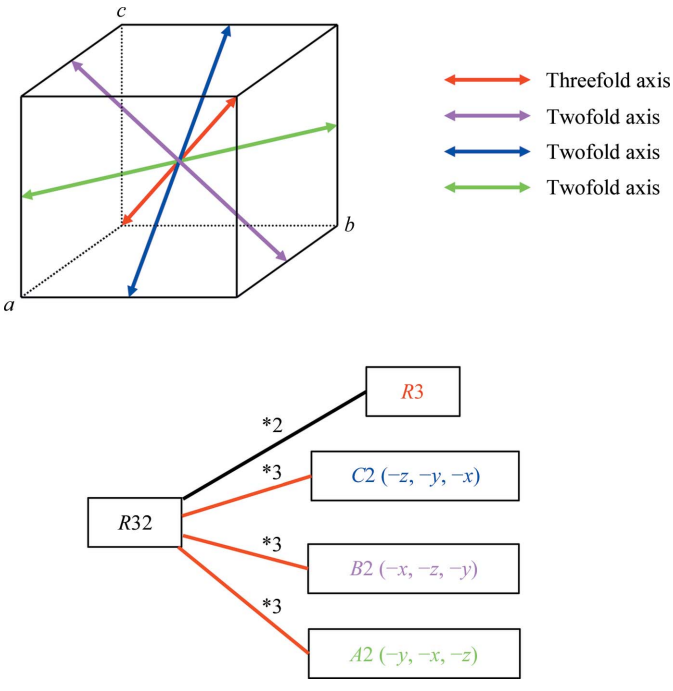

Figure 1
Relationship between space group *R*32 and the maximal non-isomorphic subgroups *R*3 and *C*2. The lattice type (*A*, *B* or *C*) of centered monoclinic cell depends on which twofold axis is retained. The conventional *C*-centered lattice can be generated by transformation of the *A*- or *B*-centered cells. Symmetry operators for the three possible twofold axes (shown in parentheses) are described with respect to the rhombohedral setting of space group *R*32.

The structure-factor file of PDB entry 2a8y (space group *P*1) was downloaded and subjected to *phenix.xtriage* analysis. The analysis showed that the reflections in space group *P*1 were related by the operator $(-l, -k, -h)$ with an $R_{\rm merge}$ of about 4%, indicating that the unit cell contained a crystallographic twofold axis. This operator corresponds to one of three possible crystallographic twofold axes in the rhombohedral space group *R*32 (Fig. 2). Together, these observations demonstrate that the correct space group is *C*2, that the correct unit-cell parameters are a = 135.16, b = 138.09, c = 96.56 Å, β = 92.21° and that the asymmetric unit contains two independent half-hexamers, the other half of each of which is generated by a crystallographic twofold axis.

2.2. Conversion of the diffraction data to space group *C*2

The structure-factor file of the deposited SsMTAP II structure PDB entry 2a8y in space group *P*1 was downloaded from the Protein Data Bank and converted to the standard MTZ format using the program *VARIOUS2MTZ* from the *CCP*4 suite (Winn *et al.*, 2011). The converted structure-factor file was re-indexed in space group *C*2 using the *CCP*4 program *REINDEX* and providing a transformation matrix corresponding to

$$h_{\rm C} = h_{\rm P} + l_{\rm P}, k_{\rm C} = h_{\rm P} - l_{\rm P}, l_{\rm C} = k_{\rm P}, \qquad (1)$$

where the subscript C designates indices in the *C*-centered monoclinic space group and the subscript P designates indices in the primitive (pseudo-rhombohedral) triclinic cell. The *CCP*4 program *COMBAT* was subsequently used to convert the standard MTZ format to multirecord MTZ, after which the equivalent reflections were merged with an overall $R_{\rm merge}$ of 4.6% using the *CCP*4 program *SCALA*. The original $R_{\rm free}$ flags in space group *P*1 were kept for one set of *hkl* equivalents in space group *C*2.

2.3. Re-refinement in space group *C*2

The relationship between the *P*1 and *C*2 unit cells is illustrated in Fig. 3. The fractional coordinates for the *P*1 structure were transformed to space group *C*2 using

$$x_{\rm C} + \tfrac{1}{2}x_{\rm P} + \tfrac{1}{2}z_{\rm P}, y_{\rm C} = \tfrac{1}{2}x_{\rm P} - \tfrac{1}{2}z_{\rm P}, \text{ and } z_{\rm C} = y_{\rm P}, \qquad (2)$$

where the subscript C designates fractional coordinates in space group *C*2 and the subscript P designates fractional coordinates in space group *P*1. The matrix for transforming the coordinates is generated by inverting and then transposing the matrix used for transforming the indices from space group *P*1 to *C*2. Prior to applying the transformation, the center of mass of the hexamer was translated to $x = 0$, $y = 0$ and $z = 0$ to correct for the fact that the origin is arbitrary in space group *P*1 while the origin must lie on the twofold axis in space group *C*2. One half of each hexamer was then discarded because the asymmetric unit in space group *C*2 contains two independent half-hexamers, with the other half generated by the crystallographic twofold axis. All water molecules from the *P*1 structure were discarded prior to beginning the refinement. The structure was refined by iterative cycles of manual rebuilding

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Space group	*P*1	*C*2
Unit-cell parameters (Å, °)	a = 96.60, b = 96.56, c = 96.63, α = 91.57, β = 91.23, γ = 91.52	a = 135.16, b = 138.09, c = 96.56, α = 90, β = 92.21, γ = 90
Resolution (Å)	40–1.45	40–1.45
Completeness (%)	99.7 (98.5)	98.5 (93.7)
No. of reflections	579262	310466
R_{work}/R_{free}	0.180/0.205	0.169/0.185
Contents of asymmetric unit	2 hexamers	2 half-hexamers
No. of atoms†		
Protein	24202	13140
Ligand/ion	300	185
Water	3665	1277
No. of residues with alternate conformations	192	129
B factors (Å^2)		
Protein	16.4	25.1
Ligand/ion	17.6	21.0
Water	33.1	36.4
R.m.s.d.		
Bond lengths (Å)	0.015	0.006
Bond angles (°)	1.50	1.20
Refinement program	*CNS*	*REFMAC*5, *PHENIX*

† Numbers include alternate conformations.

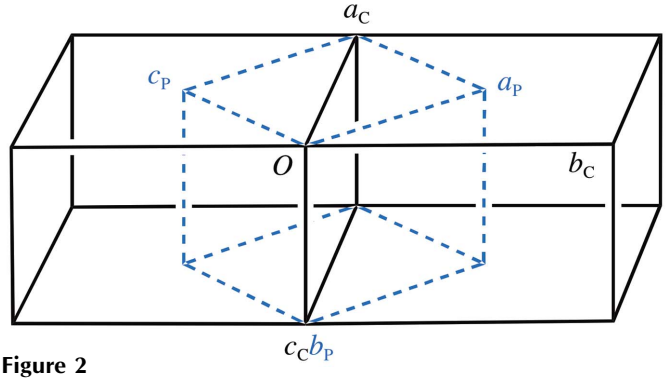

Figure 2
Transformation between the primitive rhombohedral cell and the *C*-centered monoclinic cell.

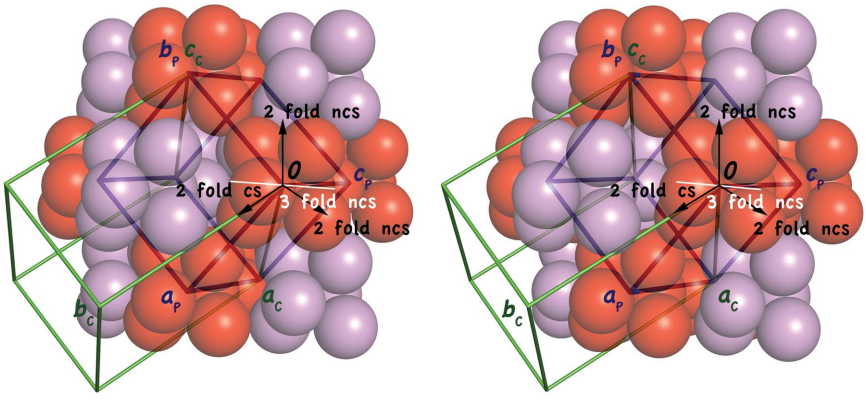

Figure 3
Stereoview of SsMTAP II and its crystal packing. The *P*1 cell and the *C*2 cell are shown in blue and green, respectively. The crystallographic twofold axis is along the *b* axis of the *C*2 cell, and the noncrystallographic twofold and threefold axes are labeled. The half-hexamers related by pseudotranslational symmetry are shown as red and violet spheres, where the spheres represent monomers. The complete SsMTAP II hexamers are generated by twofold crystallographic symmetry.

using the program *Coot* (Emsley *et al.*, 2010) and the refinement programs *REFMAC*5 (Murshudov *et al.*, 2011) initially and *phenix.refine* (Afonine *et al.*, 2005) for the later rounds. TLS parameters were included in the final rounds of refinement, with the initial TLS parameters being determined using the *TLSMD* server (Painter & Merritt, 2006). The test set used for R_{free} calculation was partially inherited from the initial data in space group *P*1 in order to minimize model bias. The reflections included in the test set account for 6.6% of the total reflections. The structure was validated using the program *MolProbity* (Chen *et al.*, 2010).

3. Results

3.1. Description of the overall structure

SsMTAP II is a homohexamer that can be thought of as a close-packed dimer of trimers, with one active site per monomer. For the deposited structure (PDB entry 2a8y) in space group *P*1, the unit cell contains two complete hexamers, each with 32 noncrystallographic symmetry. For the correct *C*2 unit cell, the asymmetric unit contains two half-hexamers and the hexamers are generated by crystallographic twofold symmetry. The monomers within each half-hexamer are related by a noncrystallographic threefold axis, which is perpendicular to the crystallographic twofold axis within experimental error. The two crystallographically independent half-hexamers are related by a shift that deviates by approximately 4 Å from the translation ($\frac{1}{2}$, 0, $\frac{1}{2}$), which is a consequence of pseudo-rhombohedral symmetry.

3.2. Comparison to the original structure

Using *PHENIX* with TLS refinement, the structure of SsMTAP II in space group *C*2 refined to better *R* and R_{free} values of 16.9% and 18.5%, respectively. The improvement in the structure of SsMTAP II during re-refinement arises in part from the use of improved X-ray methods, as has been demonstrated for other deposited structures (Joosten *et al.*, 2009). The electron density was improved and allowed the building of residues 255–261, which were missing from the deposited structure, resulting in an average *B* factor of 62.6 Å^2 for the loop compared with 25.1 Å^2 for all protein atoms. In the final structure 8% of the residues have alternate conformations, compared with 6% in the original structure. Proportionately fewer water molecules were included in the *C*2 structure; the ratio of water molecules to residues is approximately 0.8 for the *C*2 structure and 1.2 for the original structure. The root-mean-square deviations (r.m.s.d.s) from ideal geometry are comparable for the two structures. Overall, there are no significant differences between the two structures when

they are superimposed. The refinement statistics are compared in Table 1.

4. Discussion: pseudo-rhombohedral symmetry

SsMTAP II crystallizes in an apparent primitive unit cell with parameters $a = 96.60$, $b = 96.56$, $c = 96.63$ Å, $\alpha = 91.57$, $\beta = 91.23$, $\gamma = 91.52^\circ$. While the unit-cell parameters were consistent with rhombohedral symmetry, the diffraction data were not, indicating a pseudo-rhombohedral cell. Space group *C*2 is a maximal nonisomorphic subgroup of *R*32 and, because the diffraction data showed twofold symmetry, it is the likely candidate for the correct space group. If the true symmetry were *P*1 then the analysis would be straightforward. However, space group *C*2 can be generated from the primitive cell in three ways corresponding to the three possible choices of the monoclinic twofold axis (Fig. 3). That is, if the true space group is *C*2, one of the hexameric twofold axes will be crystallographic (along the *b* axis by convention) while the other two will be noncrystallographic, because the threefold axis is tilted by 1.1° with respect to the body diagonal of the primitive unit cell.

The data, which were originally processed using the program *HKL*-2000 (Otwinowski & Minor, 1997), failed to scale in space group *C*2 because the axes for the centered monoclinic cell can be chosen in three ways, yet the indexing program arbitrarily chose only one for examination (Fig. 1). Thus, the chance of obtaining the correct result was one in three and in this case this choice turned out to be incorrect. The true symmetry was only revealed later when the computer program *phenix.xtriage* (Zwart *et al.*, 2005*a*,*b*) became routinely available. Nevertheless, the similarity of the unit-cell parameters suggests that *A*-centering, *B*-centering and *C*-centering are all possibilities. These three choices also correspond to the three possible crystallographic twofold axes. The condition for *A*-centering (*c* axis unique) is $b_P = c_P$ and $\beta_P = \gamma_P$, that for *B*-centering (*a* axis unique) is $a_P = c_P$ and $\alpha_P = \gamma_P$ and that for *C*-centering (*b* axis unique) is $a_P = b_P$ and $\alpha_P = \beta_P$, where the subscript P designates unit-cell parameters in the primitive unit cell. All three of the conditions are approximately satisfied in the case of pseudo-rhombohedral symmetry, but only one generates a cell with the correct centered monoclinic symmetry. Once the correct unit cell has been identified, it can of course be transformed to the standard setting (*C*-centered, *b* axis unique) by permutation of the axes. Although programs that analyze structure-factor data for twinning, pseudosymmetry, *etc.*, such as *phenix.xtriage* and *POINTLESS* (Evans, 2006), should be routinely used, a more desirable solution would be for all data-processing programs to test alternate settings during the scaling and merging processes. It should be noted that when the same raw data were indexed, integrated and scaled in *MOSFLM* (Leslie, 1997), the correct space group was found because all centered monoclinic cells were considered possibilities by this program.

5. Conclusion

Pseudosymmetry commonly occurs in protein crystallography and can be a source of difficulty during structure determination and refinement. Therefore, detection and proper treatment of pseudosymmetry is important. In the case of SsMTAP II, the failure to detect a crystallographic twofold axis did not create serious problems; however, the additional effort required for building and adjusting 3240 residues in space group *P*1 is significant compared with building 1620 residues in space group *C*2. Overall, the structures of SsMTAP II described in space group *P*1 and in space group *C*2 are indistinguishable. This is largely because the resolution was high (1.45 Å); therefore, the loss of a twofold crystallographic restriction had only minor impact. However, robust approaches for identification of the correct symmetry in cases of pseudosymmetry could be critical in obtaining accurate structures where the resolution limit is more modest.

References

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **42**, 43–49.

Appleby, T. C., Erion, M. D. & Ealick, S. E. (1999). *Structure*, **7**, 629–641.

Appleby, T. C., Mathews, I. I., Porcelli, M., Cacciapuoti, G. & Ealick, S. E. (2001). *J. Biol. Chem.* **276**, 39232–39242.

Cacciapuoti, G., Porcelli, M., Bertoldo, C., De Rosa, M. & Zappia, V. (1994). *J. Biol. Chem.* **269**, 24762–24769.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Ealick, S. E., Rule, S. A., Carter, D. C., Greenhough, T. J., Babu, Y. S., Cook, W. J., Habash, J., Helliwell, J. R., Stoeckler, J. D., Parks, R. E. Jr, Chen, S. & Bugg, C. E. (1990). *J. Biol. Chem.* **265**, 1812–1820.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Joosten, R. P., Womack, T., Vriend, G. & Bricogne, G. (2009). *Acta Cryst.* D**65**, 176–185.

Leslie, A. G. W. (1997). *MOSFLM* v.5.40. MRC Laboratory of Molecular Biology, Cambridge.

Mao, C., Cook, W. J., Zhou, M., Koszalka, G. W., Krenitsky, T. A. & Ealick, S. E. (1997). *Structure*, **5**, 1373–1383.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Painter, J. & Merritt, E. A. (2006). *J. Appl. Cryst.* **39**, 109–111.

Pegg, A. E. & Williams-Ashman, H. G. (1969). *J. Biol. Chem.* **244**, 682–693.

Pugmire, M. J. & Ealick, S. E. (2002). *Biochem. J.* **361**, 1–25.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Zhang, Y., Porcelli, M., Cacciapuoti, G. & Ealick, S. E. (2006). *J. Mol. Biol.* **357**, 252–262.

Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005*a*). *CCP4 Newsl. Protein Crystallogr.* **42**, 58–67.

Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005*b*). *CCP4 Newsl. Protein Crystallogr.* **43**, 26–35.