WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

The structure of flavin-dependent tryptophan 7-halogenase RebH

Eduard Bitto,[1] Yu Huang,[2,3] Craig A. Bingman,[1] Shanteri Singh,[2,3] Jon S. Thorson,[2,3] and George N. Phillips Jr.[1]*

[1] Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706

[2] Division of Pharmaceutical Sciences, School of Pharmacy, University of Wisconsin-Madison, Madison, Wisconsin 53705

[3] University of Wisconsin National Drug Discovery Group, University of Wisconsin-Madison, Madison, Wisconsin 53705

Key words: rebeccamycin synthesis; regiospecific halogenation.

INTRODUCTION

Enzyme catalyzed regio- and stereo-specific halogenations influence the biological activity of a diverse array of therapeutically important natural products, including the antibiotics vancomycin and chloramphenicol as well as the anticancer agents calicheamicin and rebeccamycin.[1–5] The major class of enzymes responsible for this challenging synthetic reaction, the flavin-dependent halogenases, catalyzes the formation of carbon-halogen bonds using flavin, a halide ion (Cl^-, Br^- or I^-), and O_2.[6] Recent mechanistic and structural advances achieved with the model flavin-dependent tryptophan 7-halogenases PrnA and RebH[7–10] have greatly enhanced the level of understanding of this unique reaction. According to these studies, the mechanism for tryptophan halogenation proceeds via FAD(C4a)—OOH activation of a chloride ion into the transient chlorinating species HOCl.[11–14] The key evidence for the requirement of a transient chlorinating species is the discovery that a ~10-Å-long tunnel separates FAD and tryptophan in the ligand-bound form of PrnA.[12] In a recent compelling study to elucidate the strategy by which RebH controls this highly reactive and indiscriminant oxidant, a Lys79-εNH-Cl chloramine intermediate was implicated as the actual chlorinating species within RebH and a structural investigation of RebH was reported.[10] Here we report our independent structural analysis of *Lechevalieria aerocolonigenes* RebH (UniProt accession number Q8KHZ8, 530 amino acids) in its apo-form as well as in a complex with both tryptophan and FAD.

MATERIALS AND METHODS

Cloning and expression

The *rebH* gene was amplified from Cosmid DNA pJST2301[9] using the primer pair 5′-GGTACGTCATATGTCCGGCAAGA-3′ and 5′-GACGTAAGCTTCCGTCTGTCAGC-3′. After restriction digestion with *NdeI*/*HindIII* (Promega, Madison, WI), the PCR products were ligated into *NdeI*/*HindIII*-linearized pET28a (Novagen, San Diego, CA) containing an amino-terminal His_{10}-tag. The cloned expression plasmid was confirmed by sequencing and subsequently transformed into *E. coli* strain BL21(DE3) (Stratagene, La Jolla, CA) for expression. Cells were grown at 37°C until reaching an optical density at 600 nm of 0.9 and then induced with 1 m*M* isopropyl-β-D-thiogalactopyranoside and grown for additional 40 h at 16°C.

Protein purification

Cells were resuspended in binding buffer (50 m*M* sodium phosphate, 300 m*M* NaCl, 20 m*M* imidazole pH 8) and disrupted by sonication. Insoluble debris was removed by centrifugation at 10,000*g* for 1 h, and the supernatant with the His_{10}-fusion construct was captured

*Correspondence to: George N. Phillips Jr., Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706. E-mail: phillips@biochem.wisc.edu
Received 10 April 2007; Accepted 1 May 2007
Published online 17 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21627

on a HiTrap Chelating HP column charged with Ni^{2+} (Amersham Biosciences, Piscataway, NJ). The recombinant protein was eluted with a linear 20–500 m*M* imidazole gradient, desalted using a PD-10 column (Amersham Biosciences, Piscataway, NJ). The final storage buffer contained 50 m*M* potassium phosphate, pH 8, 0.2 m*M* dithiothreitol (DTT) and 20% glycerol. Protein used for crystallization trials was dialyzed into the solution containing 50 m*M* NaCl, 10 m*M* tris(hydroxymethyl) aminomethane (TRIS) pH 8.0. Protein concentrations were determined by Bradford protein assay (Bio-Rad, Hercules, CA) using bovine serum albumin as standard.

Activity assay

Typical assays were conducted at 303 K in 100 μL total volume 20 m*M* potassium phosphate buffer (pH 8.0) containing 0.6 m*M* L-Trp, 0.2 mg/mL RebH, 50 μ*M* FAD and 100 m*M* halide (NaCl or NaBr) and an in situ flavin reduction system. In situ flavin reduction was accomplished using either 20 m*M* NADH and 0.2 unit/mL NADH oxidase (Sigma-Aldrich) or 20 m*M* DTT. At given time points, a reaction aliquot was removed, diluted with an equal volume of MeOH, centrifuged to remove precipitated protein, and analyzed by reverse-phase HPLC (Phenomenex LUNA C18, 100 Å, 4.6 × 250 mm; 1 mL/min; isocratic for the first 0–10 min, 15% B followed by a gradient of 15–80% B from 10–17 min; A, H_2O/0.1% TFA; B, acetonitrile; A_{280}). RebH specific activity was consistent with previous reports[14] and all halogenated products were confirmed by LC-MS.

Protein crystallization

Crystals of native RebH were grown at 277 K by the hanging drop method from a 18 mg mL^{-1} protein solution in a buffer (50 m*M* NaCl, 10 m*M* tris(hydroxymethyl)aminomethane (TRIS) pH 8.0) mixed with an equal amount of the well solution (0.5 *M* NaH_2PO_4, 0.9 *M* K_2HPO_4). Crystals with shape of trigonal bi-pyramids and dimensions up to 300 × 40 × 40 μm grew from a heavy precipitate in several months. Crystals were cryoprotected at 277 K by soaking in well solutions containing 10, 20, and 30% (v/v) glycerol and were flash frozen in a stream of cryogenic nitrogen gas at 100 K. To prepare the complex, the RebH crystals were placed for 22 h in the soaking solution containing 0.3 *M* NaH_2PO_4, 0.6 *M* K_2HPO_4, ≈5 m*M* FAD, ≈3 m*M* tryptophan, and 30 m*M* NaCl. Crystals were cryoprotected by stepwise addition of 100% (v/v) glycerol to the soaking solution until ≈30% (v/v) final concentration of glycerol was achieved.

Diffraction data collection

X-ray diffraction data for both the apo-structure and the substrate complex were collected at the General Medicine and Cancer Institute Collaborative Access Team (GM/CA-CAT) 23-ID-D beamline at the Advanced Photon Source at Argonne National Laboratory. Each of the 200 diffraction images for the apo-structure was collected at a crystal-to-detector distance of 325 mm and exposed for 4 s with 100-fold attenuation of the incident beam. The data were collected in a single pass with 1° oscillations per frame. Each of the 360 diffraction images for the complex was collected at a crystal-to-detector distance of 250 mm and exposed for 6 s with 200-fold attenuation of the incident beam. The data were collected by inverse-beam strategy with 30° wedges and 1° oscillation per frame. The diffraction images were integrated and scaled using HKL2000.[15] Crystals belong to the space group $P6_2$ with unit cell parameters $a = b = 114.8$ Å, $c = 230.6$ Å.

Structure determination

The apo-structure of RebH was solved by molecular replacement in MOLREP[16] using the structure of *Pseudomonas fluorescens* PrnA as a search model (PDB ID 2aqj, 55% identity). The outstanding solution with R-factor of 0.46 and a correlation coefficient of 0.49 was obtained after two molecules were placed in the asymmetric unit. Prerefinement of the molecular replacement solution without noncrystallographic symmetry constraints in REFMAC5[17] resulted in a model with $R = 32.5\%$ ($R_{free} = 37.5\%$). To reduce the model bias and improve the initial map quality, σ_A-weighted model phases were density modified in DM[18] with 2-fold noncrystallographic symmetry averaging constraints. The atomic model was then built based on the resulting phases using an automatic building procedure implemented in ARP/WARP.[19] The initial model obtained from this procedure had $R = 21.2\%$ ($R_{free} = 26.3\%$) and contained 978 residues of which 956 had side-chains assigned. The structure was completed with multiple cycles of manual building in COOT[20] and refinement in REFMAC5.[17] Final refinement protocol included TLS refinement[21] with 10 TLS-groups per molecule[22] and used medium noncrystallographic symmetry restraints to relate the main-chain atoms of the two molecules in the asymmetric unit. The final refined model has $R = 16.1\%$ ($R_{free} = 20.8\%$). The structure of the complex was rebuilt by ARP/WARP from the apo-structure model and refined using the protocol described for the apo-structure.

RESULTS AND DISCUSSION

Structure quality

The structure of *Lechevalieria aerocolonigenes* RebH was determined by molecular replacement in MOLREP using the structure of the PrnA monomer[12] as a search model (PDB ID 2aqj) and was refined to a resolution of 2.5 Å. The structure of the complex of RebH with bound

DOI 10.1002/prot

Table I
Crystal Parameters, Data Collection, Phasing, and Refinement Statistics

	Apo-RebH	RebH complex
Space group	$P6_2$	$P6_2$
Unit-cell parameters (Å; °)	$a = b = 114.8$, $c = 230.6$; $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 114.5$, $c = 231.9$; $\alpha = \beta = 90$, $\gamma = 120$
Data collection statistics		
Wavelength (Å)	0.97919	0.97931
Energy (eV)	12,662	12,660
Resolution range (Å)	29.13–2.50 (2.59–2.50)	49.58–2.15 (2.20–2.15)
No. of reflections (measured/unique)[a]	762,860/59,677	1,881,583/92,078
Completeness (%)	100.0 (100.0)	98.8 (90.9)
R_{merge}[b]	0.131 (0.432)	0.102 (0.534)
Redundancy	12.8 (10.3)	20.4 (9.4)
Mean *I*/sigma(*I*)	24.5 (8.1)	24.9 (3.6)
Phasing statistics[c]		
Correlation coefficient	0.49	
R factor	0.46	
Refinement and model statistics		
Resolution range	29.13–2.49 (2.56–2.49)	19.95–2.15 (2.21–2.15)
No. of reflections (work/test)	56,611/3008	87,284/4630
R_{cryst}[d]	0.161 (0.221)	0.152 (0.259)
R_{free}[e]	0.208 (0.267)	0.194 (0.362)
r.m.s.d. bonds (Å)	0.014	0.015
r.m.s.d. angles (°)	1.407	1.412
ESU from R_{free} (Å)	0.209	0.141
B factor (Å^2): Wilson/average[f]	43.2/31.9	35.2/33.4
No. of protein molecules/all atoms	2/8902	2/9627
No. of waters/ions	455/2 phosphates	1056/0
No. of substrates and/or cofactors		1 chloride, 2 tryptophans, 1 FAD, 1 adenosine portion of FAD
Ramachandran plot by PROCHECK (%)		
Most favorable region	91.5	91.3
Additional allowed region	8.5	8.7
Generously allowed region	0.4	0.0
PBD code	2o9z	2oa1

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection.
[c]Phasing by molecular replacement in MOLREP using the PrnA monomer as a search model (PDB ID 2aqj).
[d]$R_{cryst} = \sum_h ||F_{obs}| - |F_{calc}|| / \sum_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.
[e]R_{free} was calculated as R_{cryst} using ~5.0% of the randomly selected unique reflections that were omitted from structure refinement.
[f]Reported values for models refined in REFMAC5 by restrained refinement with no TLS-refinement.

tryptophan, FAD, and chloride was refined to a resolution of 2.15 Å. Data collection, phasing, and refinement statistics for both the apo- and complex structures are summarized in Table I. The final model of apo-RebH contains residues 3–526 of chains A and residues 3–528 of chain B; terminal residues were not observed in the electron density and therefore not included in the model. Similarly, the final model of the RebH complex contains residues 2–528 of chains A and residues 2–528 of chain B. The chloride anion was modeled in the active site of molecule A of the RebH complex based on the structural similarity to the PrnA-FAD-chloride complex.[12] This interpretation is further supported by the existence of a 4.5σ-peak in the σ_A-weighted 2*Fo-Fc* electron density map at the expected position, compared to 2.0σ, 2.2σ, 2.6σ, 3.4σ peaks of the nearby water molecules. Clear electron density was observed for bound tryptophans in both modeled molecules. In addition, well defined electron density for FAD was observed in molecule A. However, the electron density corresponding to the cofactor in molecule B showed a significant disorder and only allowed for unambiguous modeling of the adenosine portion of FAD.

Comparison of RebH and PrnA structures

RebH and PrnA share 55% sequence identity and their structures align closely with a root mean square deviation (rmsd) of 0.68 Å for 3238 structurally equivalent atoms; this value is consistent with these proteins having a very similar fold. Similarly to PrnA, the crystal structure of RebH revealed that this protein forms a dimer with a buried surface area of 1630 Å^2. In addition to van der Waals contacts, 14 hydrogen bonds and a single salt

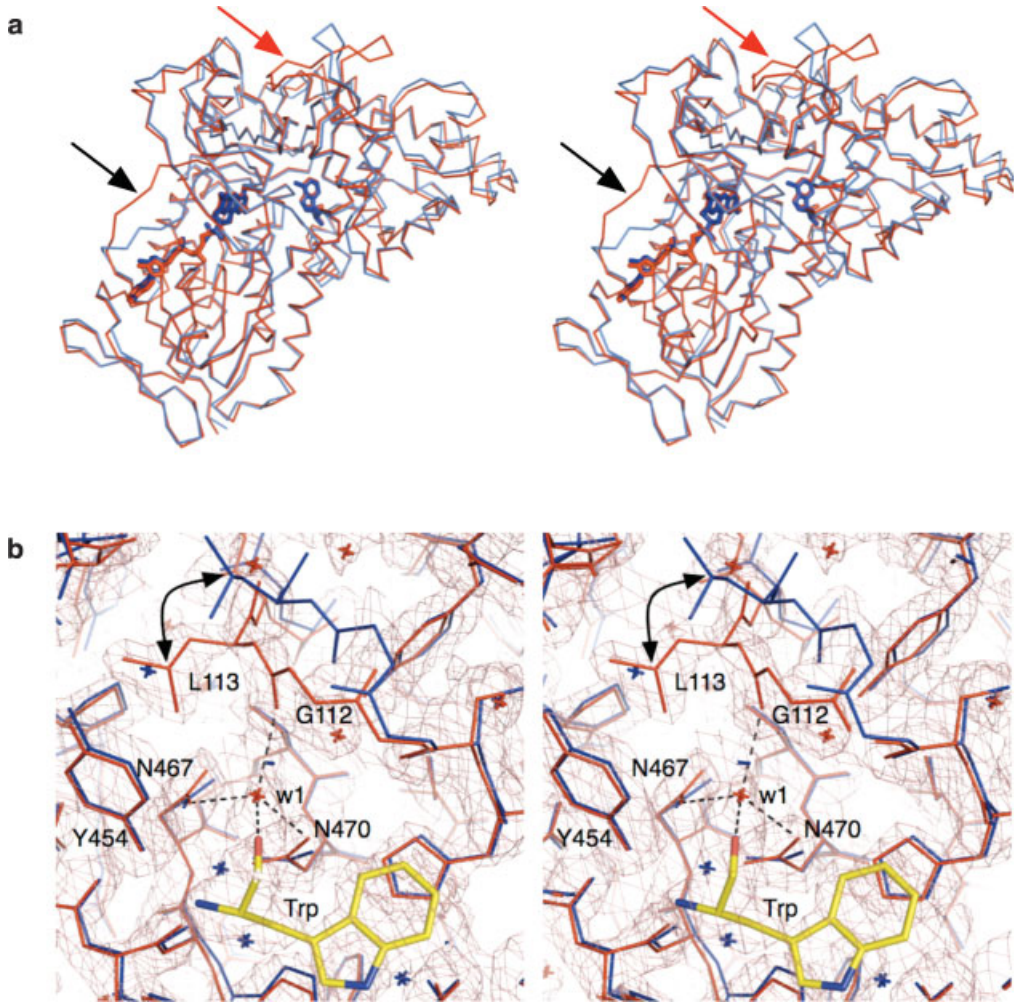

Figure 1

*Structure of RebH. (**a**) Stereo image of the C_α-trace of structurally superposed RebH (red) and PrnA (cyan, PDB ID 2ar8) complexes with FAD and tryptophan (sticks). (**b**) Stereo image of the tryptophan-binding cavity of the RebH complex with bound tryptophan (yellow sticks). The loop harboring Gly112 and Leu113 undergoes a conformational change upon tryptophan binding as seen from the difference between the structures of apo-RebH (blue) and the RebH-complex (red). Total omit map[23] of the RebH complex (salmon) is shown at contour level of 1.0σ. Hydrogen bonds to a stabilizing water molecule (w1) are shown as black dashed lines.*

bridge (Arg387–Glu432) stabilize the dimer interface. Each monomer of RebH folds into a single domain with a complex topology described previously for PrnA.[12] As in the case of the PrnA structure, flavin and tryptophan binding sites are separated by ≈10 Å [Fig. 1(a)].

Two notable structural differences exist between structures of these proteins: (i) a surface-exposed loop that spans residues 86–105 in RebH harbors an eight-residue insertion [Fig. 1(a), red arrow] compared to the analogous loop in PrnA (residues 87–97), (ii) the conserved loop that spans residues 40–48 in both RebH and PrnA adopts entirely different conformation [Fig. 1(a), black arrow]. In PrnA, the loop forms multiple direct hydrogen bonds to FAD as well as hydrogen bonds that are part of the hydrogen bond networks involving structurally conserved water molecules in contact with FAD. In RebH, the same loop adopts an "open" conformation, which is not in contact with the cofactor. As a result, the flavin binding site as observed in the RebH complex remains solvent exposed. The electron density for loop 40–48 was of lower quality in both apo-RebH and the RebH complex structures than for the rest of protein, indicating that this loop is quite flexible. The loop was partially stabilized in the observed conformation by the crystallographic contact with the RebH molecule related by a noncrystallographic twofold axis located in the proximity of Pro43. Despite the structural differences in this loop, the conformation of both the cofactor and the bound tryptophan are very similar in both RebH and PrnA [Fig. 1(a), sticks].

DOI 10.1002/prot

Comparison of apo-RebH and the RebH complex

The apo-RebH structure and the RebH complex align with a rmsd of 0.216 Å for 3866 structurally corresponding atoms from chains A and a rmsd of 0.187 Å for 3836 structurally corresponding atoms from chains B. The dimers of apo-RebH and the RebH complex align with rmsd of 0.276 for 7752 structurally corresponding atoms. These low values confirm that no large-scale changes in tertiary and/or quaternary structure took place upon binding of substrate and cofactor. The only notable change between the structures of apo-RebH and the RebH complex is related to the ordering of the loop comprising residues 111–114, which are involved in formation of the tryptophan binding site. In apo-RebH crystals, the electron density corresponding to this loop is of poor quality, indicating that the loop is flexible. Upon tryptophan binding, the loop undergoes a conformational change that flips the peptide carbonyl of Gly112 and displaces sidechain of Leu113 by as much as 5 Å. Substrate binding results in an apparent stabilization of the loop as judged by the well-defined electron density [Fig. 1(b)]. Upon the conformational change, the bound tryptophan becomes completely buried within the interior cavity of RebH. In addition to tryptophan, a well ordered water molecule becomes trapped in the substrate-binding cavity; this water molecule participates in a hydrogen bond network that further stabilizes the bound tryptophan [Fig. 1(b)].

ACKNOWLEDGMENTS

We acknowledge financial support from the Wisconsin Alumni Research Foundation, and NIH grants U54 GM074901 (E.B., C.A.B., and G.N.P.) and R01 CA084374 and U19 CA113297 (Y.H., S.S., and J.S.T.). J.S.T is a UW HI Romnes Fellow. We thank the Center for Eukaryotic Structural Genomics for research infrastructure and the UW School of Pharmacy Analytical Instrumentation Center for analytical support. GM/CA-CAT has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract no.W-31-109-ENG-38.

REFERENCES

1. Harris CM, Kannan R, Kopecka H, Harris TM. The role of the chlorine substituents in the antibiotic vancomycin—preparation and characterization of monodechlorovancomycin and didechlorovancomycin. J Am Chem Soc 1985;107:6652–6658.
2. Piraee M, White RL, Vining LC. Biosynthesis of the dichloroacetyl component of chloramphenicol in *Streptomyces venezuelae* ISP5230: genes required for halogenation. Microbiology Sgm 2004;150:85–94.
3. Li TH, Zeng ZJ, Estevez VA, Baldenius KU, Nicolaou KC, Joyce GF. Carbohydrate minor-groove interactions in the binding of calicheamicin γ-1(I) to duplex DNA. J Am Chem Soc 1994;116:3709–3715.
4. Lam KS, Schroeder DR, Veitch JM, Colson KL, Matson JA, Rose WC, Doyle TW, Forenza S. Production, isolation and structure determination of novel fluoroindolocarbazoles from Saccharothrix aercolonigenes ATCC 39243. J Antibiotics 2001;54:1–9.
5. Lam KS, Schroeder DR, Veitch JM, Matson JA, Forenza S. Isolation of a bromo analog of rebeccamycin from saccharothrix-aerocolonigenes. J Antibiotics 1991;44:934–939.
6. Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. Nature's inventory of halogenation catalysts: oxidative strategies predominate. Chem Rev 2006;106:3364–3378.
7. Hammer PE, Hill DS, Lam ST, VanPee KH, Ligon JM. Four genes from Pseudomonas fluorescens that encode the biosynthesis of pyrrolnitrin. Appl Environ Microbiol 1997;63:2147–2154.
8. Sanchez C, Butovich IA, Brana AF, Rohr J, Mendez C, Salas JA. The biosynthetic gene cluster for the antitumor rebeccamycin: characterization and generation of indolocarbazole derivatives. Chem Biol 2002;9:519–531.
9. Hyun CG, Bililign T, Liao JC, Thorson JS. The biosynthesis of indolocarbazoles in a heterologous *E-coli* host. Chembiochem 2003;4:114–117.
10. Yeh E, Blasiak LC, Koglin A, Drennan CL, Walsh CT. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. Biochemistry 2007;46:1284–1292.
11. Keller S, Wage T, Hohaus K, Holzer M, Eichhorn E, van Pee KH. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. Angewandte Chemie Int Ed 2000;39:2300–2302.
12. Dong CJ, Flecks S, Unversucht S, Haupt C, van Pee KH, Naismith JH. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. Science 2005;309:2216–2219.
13. Yeh E, Cole LJ, Barr EW, Bollinger JM, Ballou DP, Walsh CT. Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. Biochemistry 2006;45:7904–7912.
14. Yeh E, Garneau S, Walsh CT. Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. Proc Natl Acad Sci USA 2005;102:3960–3965.
15. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol 1997;276:307–326.
16. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997;30:1022–1025.
17. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53 (Part 3):240–255.
18. Cowtan K, Main P. Miscellaneous algorithms for density modification. Acta Crystallogr D Biol Crystallogr 1998;54 (Part 4):487–493.
19. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463.
20. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132.
21. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 2003;374:300–321.
22. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 2006;62 (Part 4):439–450.
23. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 1999;55 (Part 1):191–205.