Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Iodide-SAD, SIR and SIRAS phasing for structure solution of a nucleosome assembly protein

Manickam Yogavel, Jasmita Gill and Amit Sharma*

Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Road, New Delhi 110 067, India

Correspondence e-mail: amit.icgeb@gmail.com

Received 29 January 2009
Accepted 7 April 2009

PDB References: nucleosome assembly protein, native, 3gyv, r3gyvsf; iodide derivative, 3gyw, r3gywsf.

The crystal structure of *Plasmodium falciparum* nucleosome assembly protein (PfNapL) was determined by iodide-SAD/SIRAS phasing methods using iodide-SAD data to 3.0 Å resolution and native data to 2.4 Å resolution. Halide-derivatized PfNapL crystals were obtained using the quick cryo-soaking method in which the native crystals were soaked in a cryosolution consisting of 500 m*M* NaI for a short period of 30–60 s and data were collected at an in-house X-ray source using Cu $K\alpha$ radiation. Despite a low anomalous signal-to-noise ratio of <1.2 in the >3.5 Å resolution bin, the data were sufficient to determine the structure by SAD/SIR/SIRAS methods using the soaked iodides. Previously, structure solution had failed with both molecular-replacement and selenomethionine-derivatization techniques owing to reasons that are detailed in this work. The phasing at low resolution with three iodides per monomer with high temperature factors was successful using any of the SAD, SIR or SIRAS methods.

1. Introduction

Histone chaperones are responsible for DNA-mediated activities such as transcription, replication, recombination and DNA repair (Bhaumik *et al.*, 2007; De Koning *et al.*, 2007; Taverna *et al.*, 2007; Wang *et al.*, 2007). The multifunctional nucleosome assembly protein (NAP) is a histone chaperone and is conserved from yeast to human. Recently, the structures of the NAP-family members *Saccharomyces cerevisiae* NAP1 (ScNAP1; Park & Luger, 2006; Park, McBryant *et al.*, 2008) *Saccharomyces cerevisiae* Vps75 (ScVps75; Berndsen *et al.*, 2008; Park, Sudhoff *et al.*, 2008; Tang *et al.*, 2008) and *Homo sapiens* SET/TAF-1b (β)/INHAT (HsSET; Muto *et al.*, 2007) have been reported. *Plasmodium falciparum* contains two NAPs, namely PfNapL and PfNapS; the structure and analysis of the former have recently been reported by our group (Chandra *et al.*, 2005; Navadgi *et al.*, 2006; Gill *et al.*, 2009).

PfNapL shares low sequence identity (~20–40%) with previously determined NAP-family crystal structures and the structure solutions obtained using molecular-replacement (MR) phasing methods were unconvincing. This may be the consequence of significant domain offsets in NAP structures. After unsuccessful searches with several standard heavy-atom compounds and failure of selenomethionine derivatization, we decided to use the recently developed halide quick cryo-soaking procedure (Dauter *et al.*, 2000). Indeed, iodides have been used extensively in phasing using SAD/SIRAS/MIRAS methods, as catalogued in Supplementary Table S1.[1] In the present work, we show that iodide-SAD/SIR/SIRAS phasing at low resolution based on three iodides with relatively high *B* factors was successful in determining the crystal structure of PfNapL.

2. Materials and methods

2.1. Crystallization

Crystals were grown by the hanging-drop vapour-diffusion method at room temperature using the conditions described previously (Gill

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5155). Services for accessing this material are described at the back of the journal.

Table 1
Summary of diffraction data and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

Data set	Native	Iodide-SAD
Data collection		
Space group	*C*2	*C*2
Unit-cell parameters (Å, °)	a = 76.24, b = 37.87, c = 79.14, β = 99.7	a = 76.68, b = 37.83, c = 79.31, β = 99.8
Resolution range (Å)	50.0–2.40 (2.49–2.40)	50.0–2.99 (3.10–2.99)
Unique reflections	8943 (892)	4606 (451)
Completeness (%)	100 (100)	99.0 (97.6)
$\langle I/\sigma(I)\rangle$	37.6 (8.1)	45.6 (11.4)
$R_{\rm merge}$	0.040 (0.206)	0.084 (0.331)
Multiplicity	2.4 (2.4)	14.5 (13.2)
Refinement		
Resolution range (Å)	50.0–2.40	20.0–3.0
Reflections in working set	8053	4129
Reflections in test set	889	445
R factor/$R_{\rm free}$ (%)	21.96/26.68	21.68/30.69
Model composition		
No. of protein residues	180	187
Iodides	—	3
Water molecules	43	—
Stereochemistry		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	0.98	1.22
Ramachandran plot, residues in		
Most favoured regions (%)	92.5	89.8
Additional allowed regions (%)	7.5	10.2
Mean B factors (Å^2)		
Protein atoms	37.1	60.0
Iodides	—	73.7
Water molecules	34.1	—

et al., 2009). In brief, the crystallization droplet contained 1 µl protein solution (3 mg ml^{-1} in buffer containing 25 m*M* Tris and 25 m*M* NaCl) and 1 µl reservoir solution (200 m*M* $MgCl_2$ and 20% PEG 3350) and was equilibrated against 200 µl well solution. Plate-shaped crystals appeared after ∼35 d. Halide derivatization of PfNapL crystals was obtained by soaking native crystals for a short period (30–60 s) in a cryoprotectant solution containing 500 m*M* NaI.

2.2. Cryoprotectant optimization

Initially, crystals were tested at room temperature and diffracted to lower than 5 Å resolution even with long exposure times. The crystals were then transferred to various cryoprotectant solutions consisting of 5–20% glycerol/MPD/PEG 400/ethylene glycol in the reservoir solution for periods ranging from 5 s to 10 min. The cryocooled crystals still diffracted to lower than 5 Å resolution with a very high mosaic spread (>2°). The crystals also showed signs of fracture and we noticed the formation of ice ring(s) in some cases. We finally decided to soak the crystals in a higher concentration of the mother liquor for cryoprotection (40% PEG 3350 and 400 m*M* $MgCl_2$; 30 s).

2.3. Data collection and processing

Two forms of PfNapL crystals wre obtained, both of which belonged to the monoclinic space group *C*2 (form I, unit-cell parameters a = 92.41, b = 38.12, c = 80.10 Å, β = 107.9°; form II, unit-cell parameters a = 76.24, b = 37.87, c = 79.14 Å, β = 99.7°). Both form I and II SeMet PfNapL crystals diffracted to lower than 4.0 Å reso-

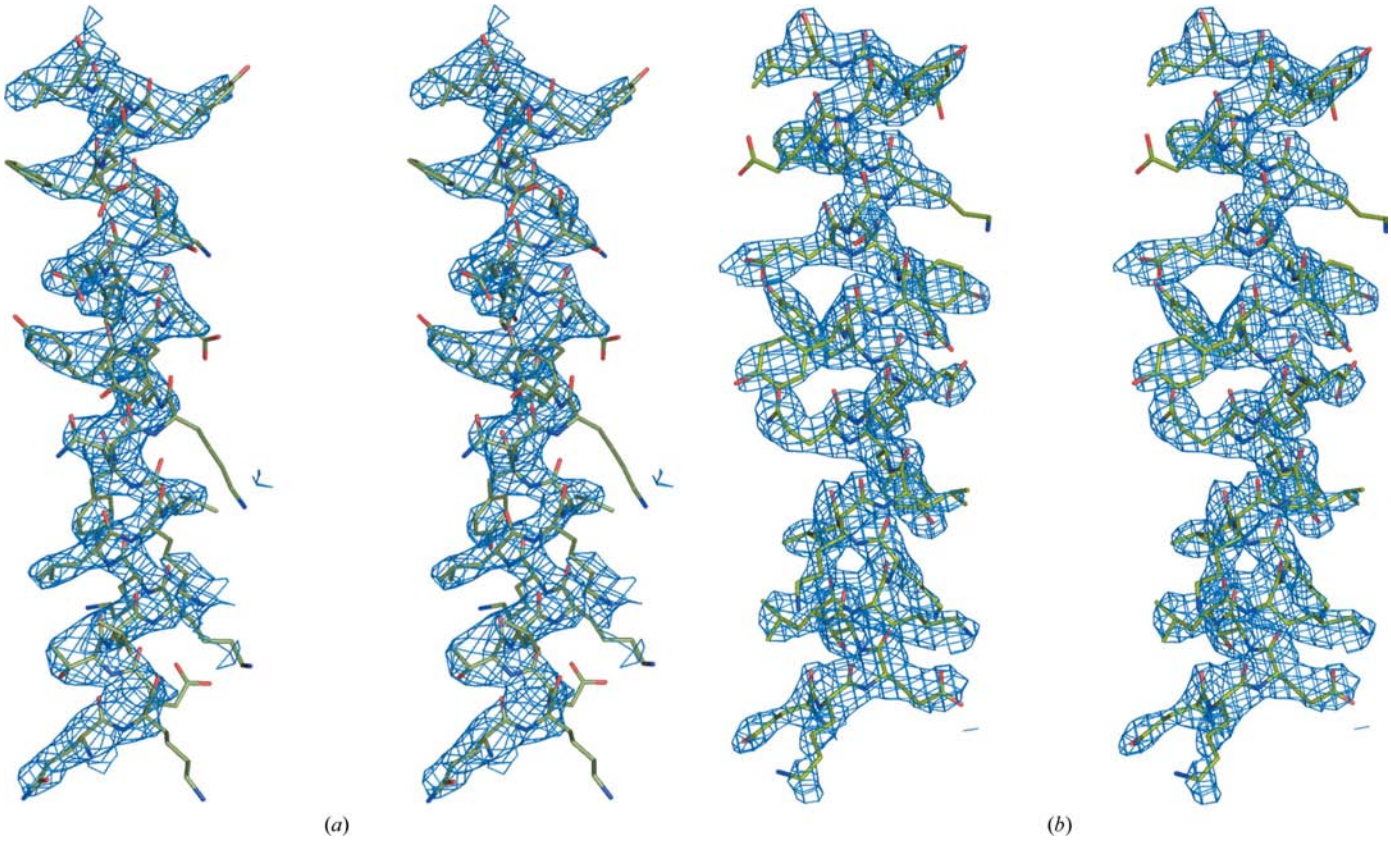

Figure 1
Stereoview of representative regions of electron-density maps after density modifications. (*a*) 3.0 Å resolution iodide-SAD map and (*b*) 2.4 Å resolution iodide-SIRAS map contoured at the 1.5σ level.

Table 2
Statistics of the anomalous signal-to-noise ratio [$\langle\Delta F/\sigma(\Delta F)\rangle$] *versus* resolution.

Values in bold indicate resolution bins with poor anomalous signal.

Resolution (Å)	∝–8.0	8.0–6.0	6.0–5.0	5.0–4.4	4.4–4.2	4.4–4.0	4.0–3.8	3.8–3.6	3.6–3.4	3.4–3.2	3.2–3.0
N (data)	271	341	433	472	203	269	319	397	505	616	780
$\langle I/\sigma(I)\rangle$	84.7	57.1	46.8	51.3	46.0	40.4	34.1	27.8	22.4	15.6	10.6
Completeness (%)	99.6	100.0	99.5	99.4	99.0	99.3	99.1	99.2	98.6	98.6	95.2
$\langle\Delta F/\sigma(\Delta F)\rangle$	3.89	3.11	2.34	1.93	1.58	**1.31**	**1.30**	**1.29**	**1.17**	**1.10**	**1.08**

Table 3
Substructure-solution, phasing and model-building statistics.

FOM, figure of merit. CC, correlation coefficient. HYSS, hybrid substructure search. SKEW, skew of the electron density in the map. $CORR_{RMS}$, the correlation of local r.m.s. density. FP, observed structure-factor amplitudes. FC, calculated structure-factor amplitudes.

Method	SAD	SIR	SIRAS
Bijvoet/isomorphous pairs	4029	7872	7872
Mean anomalous signal ($\langle\Delta F\rangle/\langle F\rangle$)	0.062	—	0.143
Mean signal-to-noise ratio [$\langle\Delta F/\sigma(\Delta F)\rangle$]	1.83	6.02	6.18
AutoSol			
No. of sites	3	3	3
HYSS CC	0.25	0.29	0.29
SKEW	0.11	0.11	0.14
$CORR_{RMS}$	0.83	0.78	0.87
FOM	0.36	0.38	0.41
Estimated map CC	0.39 ± 0.25	0.36 ± 0.24	0.45 ± 0.21
After *DM*			
R factor (FC *versus* FP)	0.29	0.27	0.26
FOM	0.66	0.58	0.58
Map CC	0.61	0.65	0.71
AutoBuild			
Residues built	123	160	165
Side chains placed	11	124	126
R_{work}	42.73	27.41	26.15
R_{free}	48.06	31.74	30.40
Map CC	0.68	0.79	0.81

lution, while native form I PfNapL crystals diffracted to 2.3 Å resolution at a synchrotron source (Gill *et al.*, 2009). Form II native and iodide-derivatized data sets were collected at 100 K using Cu $K\alpha$ radiation ($\lambda = 1.54$ Å) generated by a Rigaku Micro-Max007 rotating-anode X-ray generator operated at 40 kV and 20 mA with Osmic mirrors. Diffraction images were recorded using a MAR345dtb image-plate detector. Initially, the cryocooled crystals diffracted to >5.0 Å resolution with a high mosaic spread of >2°. However, when subjected to few (2–4) rounds of online cryo-annealing, in which the cryogenic gas stream was blocked with a plastic card for periods of 3–30 s, there was a striking improvement in the diffraction of the cryocooled crystals. The online cryo-annealing improved the extent of diffraction from >5.0 to <3.0 Å and also reduced the mosaic spread (∼1°). A native data set to 2.4 Å resolution and an iodide-SAD data set of 720 frames to 3.0 Å resolution were collected using a 1° oscillation step with 180 s exposure per frame after cryo-annealing. The crystal-to-detector distance was set to 220 and 300 mm for the native and derivative data sets, respectively. In order to minimize anomalous signal errors, a high-redundancy (14.5) iodide-SAD data set was also collected. Data were processed using *HKL*-2000 (Otwinowski & Minor, 1997) and relevant statistics are summarized in Table 1. For the iodide-SAD data, Bijvoet mates were treated as equivalent reflections in scaling but were merged separately as I^+ and I^-. The iodide-derivative data were also merged as heavy-atom derivative data (without anomalous flag) and scaled to the native data.

2.4. Substructure solution and phasing

The programs *HKL2MAP* (Pape & Schneider, 2004) and *PHENIX* (Adams *et al.*, 2002) were used for preparing and analyzing the anomalous signal from scaled diffraction data sets. Substructure solution, phasing and model building were carried out for different methods such as SAD, SIR and SIRAS using *AutoSol* and *AutoBuild* in *PHENIX*. In order to compare iodide-site location and phasing we also used *SHELXD* and *SHELXE* (Sheldrick, 2008). *Coot* (Emsley & Cowtan, 2004) was used to build missing residues and side chains into the auto-built models and *REFMAC* (Murshudov *et al.*, 1997) was used for refinement. Figures were generated using *Chimera* (Pettersen *et al.*, 2004) and *PyMOL* (http://www.pymol.org).

3. Results and discussion

3.1. Substructure determination and automatic model building using the iodide-SAD method

Full-length PfNapL contains 347 residues, but the low-complexity N- and C-terminal regions of the PfNapL structure are missing (the model contains residues 33–281). Thus, ∼250 residues were expected in the asymmetric unit, with a solvent fraction of ∼0.40. The mean anomalous signal ($\langle\Delta F\rangle/\langle F\rangle$) was 0.062 and the mean anomalous signal-to-noise ratio [$\Delta F/\sigma(\Delta F)$] was 1.83. The anomalous signal-to-noise ratio was significant only to <3.5 Å resolution (Table 2) and substructure solution in iodide-SAD was surprisingly successful in both *SHELXD* and *PHENIX* (Table 3). Phasing attempts and density modification with *SHELXE* to low resolution (3.0 Å) did not produce interpretable maps, whereas phasing and density modification with *Phaser* and *RESOLVE*, respectively, in *PHENIX* produced reasonably good maps. After density modification, the FOM was 0.66 and this map was submitted for automatic model building by *AutoBuild* in the *PHENIX* program. Approximately 50% of the residues were built, with R_{work} and R_{free} values of 42.7% and 48.1%, respectively. Using both *SHELXE* and *PHENIX*, phasing attempts at a lower resolution of 3.0 Å followed by the extension of phases to 2.4 Å resolution native data produced readily interpretable maps. One of these maps was submitted to *AutoBuild* and 63% of the residues were built, with R_{work} and R_{free} values of 27.6% and 35.1%, respectively.

3.2. Phasing and model building using the SIR/SIRAS method

The iodide-bound and native PfNapL crystals of form II are highly isomorphous and the anomalous signal is good (Table 3). Both *SHELXD* and *PHENIX* located identical sites with high occupancy. The average distance between iodide sites for the SIR and SIRAS methods is <0.2 Å (Table 4). The phases obtained from *SHELXE* and *PHENIX* gave interpretable electron-density maps for both the SIR and SIRAS methods and more than 160 residues with >120 side chains were built automatically by *AutoBuild*. The electron-density maps obtained from the iodide-SAD and SIRAS methods are shown in Fig. 1.

3.3. Structure refinement of iodide-bound and native PfNapL

The partially built iodide-bound and native models from SAD and SIR/SIRAS were subjected to manual model building for missing residues using *REFMAC*. The occupancy values of the bound iodides were refined and manually adjusted based on temperature factors

Table 4
Differences between the positions of iodide sites in the refined model and those located by *PHENIX* and *SHELXD* with the SAD, SIR and SIRAS methods, comparison of the occupancy value for iodide sites with different methods and comparison of refined *B* factors in different methods.

	PHENIX			*SHELXD*		
Method	SAD	SIR	SIRAS	SAD	SIR	SIRAS
R.m.s.d. (Å)						
IOD1	0.279	0.177	0.206	0.316	0.192	0.111
IOD2	0.064	0.192	0.167	0.433	0.132	0.204
IOD3	0.256	0.368	0.326	0.553	0.421	0.436
Comparison of occupancy value						
IOD1 (0.80†)	0.52	0.31	0.30	1.00‡	1.00‡	1.00‡
IOD2 (0.70)	0.49	0.31	0.31	0.94	0.94	0.86
IOD3 (0.50)	0.40	0.28	0.29	0.64	0.82	0.62
Comparison of refined *B* factors (Å^2)						
IOD1 (75§)	54	20	19			
IOD2 (76)	54	28	23			
IOD3 (76)	56	29	32			

† Final refined occupancy value. ‡ The occupancy values from *SHELXD* are normalized to 1.0 for the site. § Final refined *B*-factor value.

and electron density. The average distances between the refined iodide coordinates and those given by *PHENIX* and *SHELXD* are listed in Table 4. In both programs the identified iodide sites were identical with r.m.s.d.s of <0.5 Å. The IOD3 site showed somewhat lower accuracy; this may arise from the lower occupancy and higher *B* factor of this site. The final model of iodide-bound PfNapL contains 187 residues and three iodides with crystallographic R_{work} and R_{free} values of 21.7% and 30.7%, respectively. The average temperature factors for protein atoms and iodides are 60 and 73.7 Å^2, respectively. The final model of native PfNapL contains 180 residues and 43 water molecules, with crystallographic R_{work} and R_{free} values of 22.0% and 26.7%, respectively. The stereochemical quality of the SAD and native structures was assessed using *PROCHECK* (Laskowski *et al.*, 1993). Final refinement statistics are summarized in Table 1.

The asymmetric unit of the PfNapL crystals contains one monomer and the biological dimer molecule was therefore generated using a twofold symmetry operation. PfNapL contains two subdomains (Fig. 2). The overall fold of PfNapL is very similar to the previously determined NAP-family structures from yeast and human such as ScNAP1 (Park & Luger, 2006; Park, McBryant *et al.*, 2008), HsSET (Muto *et al.*, 2007) and ScVps75 (Berndsen *et al.*, 2008; Park, Sudhoff *et al.*, 2008; Tang *et al.*, 2008). Structural comparisons and the biological implications of the PfNapL structure are discussed in our recent publication (Gill *et al.*, 2009).

3.4. Protein residues that chelate the bound iodides

The iodide IOD1 is bound in a shallow hydrophobic cavity between subdomains I and II and has a relatively high occupancy and low *B* factor when compared with the other two sites. This site is surrounded by residues Arg83, Val87, Asn129, Asp130, Arg132 and Tyr151. Iodides IOD2 and IOD3 are bound in surface cavities formed by symmetry-related molecules (Fig. 3*a*). IOD2 is surrounded by three positively charged residues (His232, Lys233 and #Lys47, where # indicates a symmetry-related residue), four hydrophobic residues (Leu251, Ile254, #Leu278 and #Ile280) and two other residues (Ser231 and Gln250), while iodide IOD3 is surrounded by six positively charged residues (Lys198, Arg222, His227, #Lys40, #Lys44 and

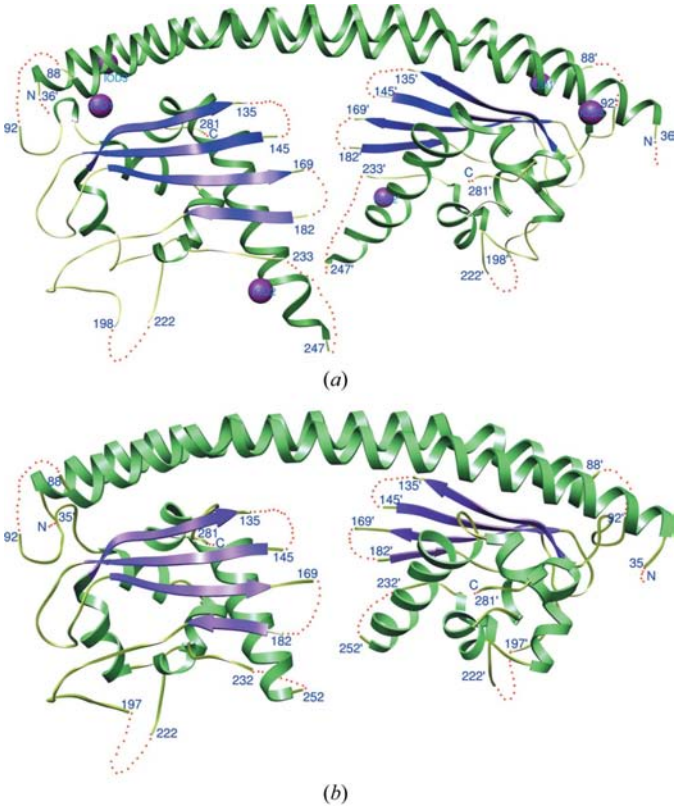

Figure 2
Ribbon diagrams of (*a*) iodide-bound and (*b*) native PfNapL dimer. Bound iodides (magenta) are shown as spheres. The disordered regions are marked as red dots and the protein termini are indicated.

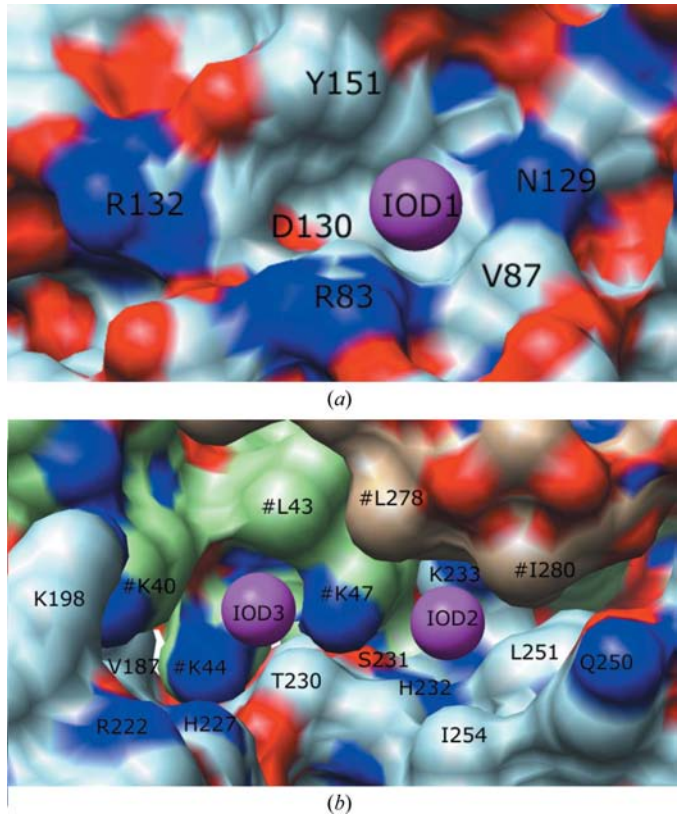

Figure 3
Close-up view of bound iodide sites. (*a*) IOD1 is bound between domains I and II. (*b*) IOD2 is bound between domain II (light blue) and symmetry-related domain II (brown) and IOD3 is bound between domain II (light blue) and symmetry-related domain I (light green). Iodides are shown as magenta spheres and neighbouring residues are labelled.

#Lys47) and three other residues (Val187, Thr230 and #Leu43) (Fig. 3*b*).

4. Conclusion

PfNapL shares low sequence identity (∼20–40%) with previously determined NAP-family crystal structures. Thus, attempts to obtain structure solutions using molecular-replacement techniques failed, possibly owing to inter-domain movements within the NAP structures. Our attempts to solve the structure of PfNapL using selenomethionine derivatization also failed despite ∼100% selenium incorporation and a significant anomalous signal in the fluorescence scan. This was a consequence of the disordered nature of these particular methionines (at positions 169 and 253). As a result, we were unable to solve the crystal structure of PfNapL until we employed iodide phasing methods at the in-house X-ray source. The significant and reproducible improvement we discovered in the extent of diffraction from the PfNapL crystal upon online cryo-annealing clearly contributed tremendously to structure solution. We also show that iodide-SAD/SIR/SIRAS phasing at low resolution based on three iodides with relatively high *B* factors was very successful in determining the crystal structure. In essence, our analysis suggests that as long as higher resolution data (in this case to 2.4 Å) are available for phase extension and model building, there is negligible difference between the SAD, SIR and SIRAS phasing techniques. Phasing by all three of these methods at the low resolution of ∼3 Å results in the location of all three iodide sites equally well. Furthermore, there is no significant difference in the quality of the electron-density maps produced. In each case, when phases were extended to 2.4 Å using a native data set the resulting final models were also essentially of equivalent quality. On the other hand, SAD data alone (to 3 Å) performed poorly in terms of the number of residues and side chains built. Clearly, the use of higher resolution native data helps significantly in cases where phases have only been determined to the low resolution of ∼3.0 Å.

The X-ray facility at ICGEB is funded by the Wellcome Trust, UK. MY and JG are supported by an FP7 grant from the EU called 'Premalstruct'. This work was also supported by a research grant (for nucleosome assembly proteins) and a career development award grant to AS from the Department of Biotechnology, Government of India.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Berndsen, C. E., Tsubota, T., Lindner, S. E., Lee, S., Holton, J. M., Kaufman, P. D., Keck, J. L. & Denu, J. M. (2008). *Nature Struct. Mol. Biol.* **15**, 948–956.
Bhaumik, S. R., Smith, E. & Shilatifard, A. (2007). *Nature Struct. Mol. Biol.* **14**, 1008–1016.
Chandra, B. R., Olivieri, A., Silvestrini, F., Alano, P. & Sharma, A. (2005). *Mol. Biochem. Parasitol.* **142**, 237–247.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
De Koning, L., Corpet, A., Haber, J. E. & Almouzni, G. (2007). *Nature Struct. Mol. Biol.* **14**, 997–1007.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gill, J., Yogavel, M., Kumar, A., Belrhali, H., Jain, S. K., Rug, M., Brown, M., Maier, A. G. & Sharma, A. (2009). *J. Biol. Chem.* **284**, 10076–10087.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Muto, S., Senda, M., Akai, Y., Sato, L., Suzuki, T., Nagai, R., Senda, T. & Horikoshi, M. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 4285–4290.
Navadgi, V. M., Chandra, B. R., Mishra, P. C. & Sharma, A. (2006). *J. Biol. Chem.* **281**, 16978–16984.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Park, Y.-J. & Luger, K. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 1248–1253.
Park, Y.-J., McBryant, S. J. & Luger, K. (2008). *J. Mol. Biol.* **375**, 1076–1085.
Park, Y.-J., Sudhoff, K. B., Andrews, A. J., Stargell, L. A. & Luger, K. (2008). *Nature Struct. Mol. Biol.* **15**, 957–964.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Tang, Y., Meeth, K., Jiang, E., Luo, C. & Marmorstein, R. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 12206–12211.
Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. (2007). *Nature Struct. Mol. Biol.* **14**, 1025–1040.
Wang, G. G., Allis, C. D. & Chi, P. (2007). *Trends Mol. Med.* **13**, 373–380.