Solution-state NMR structure of the putative morphogene protein BolA (PFE0790c) from *Plasmodium falciparum*

Garry W. Buchko,^{a,b,*} Adelinda Yee,^c Anthony Semesi,^c Peter J. Myler,^{a,d,e} Cheryl H. Arrowsmith^{c,f} and Raymond Hui^{c,f}

Received 18 November 2014

Accepted 23 December 2014

Edited by W. N. Hunter, University of Dundee, Scotland

Keywords: *Plasmodium falciparum*; BolA; PFE0790c.

PDB reference: PFE0790c, 2kdn

Supporting information: this article has supporting information at journals.iucr.org/f

^aSeattle Structural Genomics Center for Infectious Disease, USA, ^bBiological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, USA, ^cDepartment of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, ^dSeattle BioMed, Seattle, Washington, USA, ^eDepartment of Medical Education and Biomedical Informatics and Department of Global Health, University of Washington, Seattle, Washington, USA, and ^fStructural Genomics Consortium, England. *Correspondence e-mail: garry.buchko@pnsl.gov

Protozoa of the genus *Plasmodium* are responsible for malaria, which is perhaps the most important parasitic disease to infect mankind. The emergence of *Plasmodium* strains resistant to current therapeutics and prophylactics makes the development of new treatment strategies urgent. Among the potential targets for new antimalarial drugs is the BolA-like protein PFE0790c from *Plasmodium falciparum* (*Pf*-BolA). While the function of BolA is unknown, it has been linked to cell morphology by regulating transcription in response to stress. Using an NMR-based method, an ensemble of 20 structures of *Pf*-BolA was determined and deposited in the PDB (PDB entry 2kdn). The overall topology of the *Pf*-BolA structure, $\alpha 1\text{--}\beta 1\text{--}\beta 2\text{--}\eta 1\text{--}\alpha 2/\eta 2\text{--}\beta 3\text{--}\alpha 3$, with the β -strands forming a mixed β -sheet, is similar to the fold observed in other BolA structures. A helix–turn–helix motif similar to the class II KH fold associated with nucleic acid-binding proteins is present, but contains an FXGXXXL signature sequence that differs from the GXXG signature sequence present in class II KH folds, suggesting that the BolA family of proteins may use a novel protein–nucleic acid interface. A well conserved arginine residue, Arg50, hypothesized to play a role in governing the formation of the C-terminal α -helix in the BolA family of proteins, is too distant to form polar contacts with any side chains in this α -helix in *Pf*-BolA, suggesting that this conserved arginine may only serve a role in guiding the orientation of this C-terminal helix in some BolA proteins. A survey of BolA structures suggests that the C-terminal helix may not have a functional role and that the third helix ($\alpha 2/\eta 2$) has a ‘kink’ that appears to be conserved among the BolA protein structures. Circular dichroism spectroscopy shows that *Pf*-BolA is fairly robust, partially unfolding when heated to 353 K and refolding upon cooling to 298 K.

1. Introduction

Almost half of the world’s human population is at risk of contracting malaria, an infectious disease responsible for 473 000–789 000 deaths in 2012 (World Health Organization, 2013), with the majority of these deaths being children under the age of five. Although this devastating disease was once prevalent worldwide, today it is confined to tropical countries, with 85% of the cases occurring in sub-Saharan Africa. The etiological agents of human malaria are parasitic protozoa from the genus *Plasmodium* that are spread between hosts by female mosquitoes of the genus *Anopheles* (White *et al.*, 2014). Of the five *Plasmodium* species responsible for human malaria (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*), *P. falciparum* and *P. vivax* are the most important, with the former species causing the greatest mortality and the latter species causing the greatest morbidity (Anstey *et al.*, 2009).

Currently, the first line of treatment against malaria caused by all *Plasmodium* species is artemisinin-based combination therapies (Fidock, 2010). Unfortunately, there is evidence that this second-generation and third-generation defense against malaria is beginning to crumble, with reports of artemisinin-resistant *P. falciparum* in Western Cambodia and the Thailand–Myanmar border (Dondorp *et al.*, 2009; Phyo *et al.*, 2012), making the development of new antimalarial drugs urgent (D'Alessandro, 2009).

The genome of *P. falciparum* was sequenced in 2002 to assist with the development of new therapeutic strategies against malaria (Gardner *et al.*, 2002). Among the approximately 5300 genes encoded in the 14 chromosomes of the organism is the gene for an 84-residue BolA-like protein, PFE0790c. BolA proteins are present in organisms from all three domains of life, are highly conserved and range in length from 76 to 123 amino acids (Singh & Montgomery, 2014). While the molecular function of BolA proteins is unknown, their expression under different stressors induces a spherical morphology in *Escherichia coli* and other bacterial systems by inhibiting the transcription of the cytoskeleton protein MreB (Aldea *et al.*, 1989; Santos *et al.*, 1999; Freire *et al.*, 2009). BolA overexpression has also been reported to induce biofilm development (Vieira *et al.*, 2004), to modulate cell permeability, to increase the expression of the penicillin-binding proteins PBP5 and PBP6 (Guinote *et al.*, 2011) and to play a role in iron metabolism (Li *et al.*, 2012; Roret *et al.*, 2014). Given the multiple functions associated with BolA, the protein may be a potential drug target (Myler *et al.*, 2009). Here, we report the NMR-derived solution structure of *P. falciparum* BolA (PFE0790c; *Pf*-BolA) and describe some of its biophysical properties.

2. Materials and methods

2.1. Cloning, expression and purification

The *Pf*-BolA gene (PFE0790c/XP_001351715.1) was PCR-amplified from the genomic DNA of *P. falciparum* strain 3D7 using Platinum Pfx DNA polymerase (Invitrogen, Carlsbad, California, USA). The amplified *Pf*-BolA gene was then inserted into a modified pET-15d expression vector with the thrombin protease cleavage site replaced by a TEV cleavage site such that the expressed protein contained a 22-residue N-terminal tag (MGSSHHHHHSSGRENLYFQGH-) and two non-native residues at the C-terminus (-GS). The recombinant plasmid was then transformed into *E. coli* BL21(DE3) cells (Novagen, Madison, Wisconsin, USA) using a heat-shock method. Uniformly ^{15}N , ^{13}C -labeled *Pf*-BolA was obtained by growth and expression in M9 minimal medium containing $^{15}\text{NH}_4\text{Cl}$ and D- $(^{13}\text{C}_6)$ -glucose followed by purification to homogeneity using metal-affinity chromatography (Yee *et al.*, 2002). The purified protein was then concentrated and exchanged into NMR buffer [10 mM sodium phosphate, 200 mM NaCl, 10 mM DTT, 0.01% sodium azide, 1 mM benzamidine, 1× protease-inhibitor cocktail (Roche, Mannheim, Germany) pH 6.5] by ultrafiltration.

2.2. Circular dichroism spectroscopy

An Aviv Model 410 spectropolarimeter (Lakewood, New Jersey, USA) was used to collect circular dichroism data from an ~ 0.05 mM *Pf*-BolA sample in modified NMR buffer (10 mM sodium phosphate, 200 mM NaCl, 10 mM DTT pH 6.5). Duplicate steady-state wavelength spectra were recorded on a single sample in a quartz cell of 0.1 cm path length in 0.5 nm increments between 200 and 250 nm using a bandwidth of 1.0 nm and a time constant of 1.0 s. The average spectra were processed by subtracting a blank spectrum from the *Pf*-BolA spectrum and automatically line-smoothing the difference data using the Aviv software. Steady-state spectra were recorded at 298, 353 and 298 K after the collection of a thermal denaturation curve and at 298 K again after the collection of a second thermal denaturation curve. The thermal denaturation curves were obtained by recording the ellipticity at 210 nm in 2.0 K intervals from 283 to 353 K.

2.3. NMR spectroscopy

The NMR data were acquired from an ~ 1.5 mM uniformly ^{15}N , ^{13}C -labeled *Pf*-BolA sample at 293 K on Varian Inova 800, 750 and 600 spectrometers equipped with $^1\text{H}\{^{13}\text{C},^{15}\text{N}\}$ triple-resonance probes and pulse-field gradients. Standard two-dimensional ^1H - ^{15}N HSQC, ^1H - ^{13}C HSQC, HBCBCGCDHD and HBCBCGCDCHC experiments and three-dimensional HNCACB, CBCA(CO)NH, HNCO and CC-TOCSY-NNH experiments from the Varian *Protein-Pack* pulse program suite were used to assign the ^1H , ^{13}C and ^{15}N chemical shifts of the backbone and side-chain resonances. Using a mixing time of 80 ms, distance restraints for the structure calculations were obtained from three-dimensional, ^{13}C -edited aliphatic and aromatic NOESY-HSQC experiments and a ^{15}N -edited NOESY-HSQC experiment. Backbone hydrogen bonds were qualified by the identification of amide protons slowly exchanging with deuterons by freeze-drying a ^{15}N -labeled NMR sample, redissolving it in 99.8% D_2O and collecting an ^1H - ^{15}N HSQC spectrum ~ 10 min after resuspension. Felix2007 (Felix NMR Inc., San Diego, California, USA) and Sparky (v.3.115; T. D. Goddard & D. G. Kneller, University of California, San Francisco, USA) were used to process and analyze the data, respectively. Indirect methods were used to reference the chemical shifts to DSS (DSS = 0 ppm; Wishart *et al.*, 1995).

2.4. Structure calculations

The backbone and side-chain ^1H , ^{13}C and ^{15}N chemical shifts for *Pf*-BolA were assigned using established protocols (Cavanagh *et al.*, 1996) and deposited in the Biological Magnetic Resonance Bank (BMRB) under accession number 16118. Using these chemical shift assignments and the peak-picked data from ^{13}C -edited and ^{15}N -edited NOESY-HSQC experiments as initial inputs, ensembles of structures were calculated using CYANA (v.3.0; Guntert, 2004). 63 dihedral angle restraints for both φ and ψ were introduced into the calculations on the basis of the elements of secondary structure identified in the early ensembles and TALOS calculations

Table 1

Summary of the structural statistics for *Pf*-BolA.

All statistics are for the 20-structure ensemble deposited in the PDB (PDB entry 2kdn).

Restraints for structure calculations	
Total NOEs	1619
Intraresidue NOEs	425
Sequential ($i, i + 1$) NOEs	476
Medium-range ($i, i + j; 1 < j \leq 4$) NOEs	399
Long-range ($i, i + j; j > 4$) NOEs	319
φ -angle restraints	63
ψ -angle restraints	63
Hydrogen bond restraints	30
Structure calculations	
No. of structures calculated	100
No. of structures used in ensemble	20
Structures with restraint violations	
One distance restraint violation $> 0.05 \text{ \AA}$	4
Dihedral restraint violation $> 1^\circ$	0
RMSD from mean [†] ($\text{\AA}$)	
Backbone N—C ^{α} —C=O atoms	0.28 ± 0.08
All heavy atoms	0.70 ± 0.11
Ramachandran plot summary for all residues using <i>PROCHECK</i>	
Most favored regions (%)	92.7
Additionally allowed regions (%)	7.1
Generously favored regions (%)	0.2
Disallowed (%)	0.0
Global quality scores, all residues: Z-score (raw)	
<i>PROCHECK</i> (all)	$-2.07 (-0.35)$
<i>PROCHECK</i> (φ, ψ)	$0.00 (-0.08)$
<i>MolProbity</i> clashscore	$-2.90 (25.78)$

† Calculated for the ordered residues Ile3–Asp24 and Phe33–Ser83.

(Cornilescu *et al.*, 1999). Towards the end of the iterative process, 30 hydrogen bond restraints (1.8–2.0 and 2.7–3.0 Å for the NH–O and N–O distances, respectively) were introduced into the structure calculations on the basis of proximity in early structure calculations and the observation of slowly exchanging amides in the deuterium-exchange experiment. 20 structures with the lowest target function were selected from the final set of 100 calculated structures and refined with explicit water (Linge & Nilges, 1999) with *CNS* (v.1.1) using force constants of 500, 500 and 700 kcal for the NOE, hydrogen bond and dihedral restraints, respectively. For the water refinement calculations the upper boundary of the *CYANA* distance restraints was increased by 15% and the lower bound was set to the van der Waals limit. This water-refined ensemble of 20 structures was deposited in the Protein Data Bank (PDB; Berman, 2008) under accession code 2kdn. Structural quality was evaluated using the *Protein Structure Validation Suite* (PSVS; v.1.5; Bhattacharya *et al.*, 2007) and these values are included in the structure-statistics summary provided in Table 1.

Note that the amino acid sequence of *Pf*-BolA deposited in the PDB and BMRB is numbered sequentially, Met1–Ser108, starting with the 22 non-native residues at the N-terminus. In the discussions here, the first 22 non-native residues are numbered sequentially with an asterisk (Gly1*–His22*) and the first native residue, Met23, in the PDB and BMRB depositions is called Met1.

3. Results and discussion

3.1. Solution structure of *Pf*-BoIA

Fig. 1 shows a ^1H - ^{15}N HSQC spectrum for uniformly ^{15}N , ^{13}C -labeled *Pf*-BolA. Good chemical shift dispersion in both the proton and nitrogen dimensions is characteristic of a folded protein and facilitates spectral assignments (Yee *et al.*, 2002). The intensity of the amide cross peaks are not uniform, suggesting a range of dynamics along the backbone of the protein. Eight (blue circles) of the observed amide cross peaks in the ^1H - ^{15}N HSQC spectrum of *Pf*-BolA could not be unambiguously assigned; however, they were associated with very few NOE cross peaks, suggesting they were not amides in well structured parts of the protein. A total of 18 backbone amides were unassigned and these were clustered into two regions: the N-terminus (Gle2*–Ser11*) and a region in the first half of the sequence of the protein (Ser26–Ser32).

The amide assignments, labeled in the *Pf*-BolA ^1H - ^{15}N HSQC spectrum in Fig. 1, together with extensive $^{13}\text{C}^\alpha$ and side-chain proton and carbon chemical shift assignments (BMRB ID 16118) were crucial initial experimental inputs for the automatic assignment of the peak-picked NOESY data in iterative structure calculations with *CYANA*. As summarized in the structure-statistics table (Table 1), 1619 NOESY peaks were assigned in the final round of structure calculations that also included 60 hydrogen bond distance restraints and 126 total φ and ψ torsion angle restraints. Each member of the final ensemble of 20 structures satisfied the experimental data

Figure 1

Assigned ^1H - ^{15}N HSQC spectrum of *Pf*-BolA in NMR buffer (10 mM sodium phosphate, 200 mM NaCl, 10 mM DTT, 0.01% sodium azide, 1 mM benzamidinium pH 6.5) collected at 293 K at a ^1H resonance frequency of 800 MHz. Amide cross peak assignments for the N-terminal 22-residue tag and the C-terminal two non-native residues are colored magenta, while the assigned native residues are colored blue. The assigned side-chain resonances are colored red and are identified with a red horizontal dashed line. Amide resonances that could not be unambiguously identified are circled (cyan) and cross peaks below the contour limit are identified with a $\times$.

with no torsion angle violation greater than 1.0° and with only four structures containing a single upper limit distance violation greater than $0.05 \text{ \AA}$. Ensemble analysis with the *PSVS* validation-software suite confirmed a quality set of final structures (Bhattacharya *et al.*, 2007): Ramachandran statistics for the ϕ/ψ pairs of all of the residues in the ensemble were overwhelmingly in acceptable space (92.7% in most favored regions and 7.1% in additionally allowed regions) and all of the structure-quality Z-scores were greater than -5 . Fig. 2(a)

shows a superposition of the ordered regions of the final ensemble of 20 calculated structures upon the average structure and illustrates that the final ensemble of calculated structures converges well in all regions of the native protein except the region between Lys25 and Ser32, a region for which amide resonances were not identified unambiguously in the ^1H - ^{15}N HSQC spectrum. The good convergence is reflected in the RMSD values for the structured core (Ile3–Asp24, Phe33–Ser83) of the ensemble from the mean structure: $0.28 \pm 0.08 \text{ \AA}$

Figure 2

(a) Backbone superposition of the ensemble of 20 structures calculated for *Pf-BolA* (PDB entry 2kdn) on the average structure. The 22-residue N-terminal tag is disordered and has been removed for clarity. (b) Cartoon representation of the *Pf-BolA* structure closest to the average structure in the ensemble in (a). (c) Secondary structure diagram for *Pf-BolA*. Solid pink lines between β -strand residues indicate dual hydrogen bonds between two residues in an antiparallel β -sheet. The α -helices are colored red, the β -strands are colored blue and the 3_{10} -helices are colored magenta. The regions between secondary structure elements (black lines) are not drawn to scale.

Figure 3

ClustalW multiple amino acid sequence alignment of BolA proteins from various organisms: *Pf-BolA*, *P. falciparum* (gi:812972; PDB entry 2kdn); *Bb-BolA*, *Babesia bovis* (gi:5478071, PDB entry 3o2e; Abendroth *et al.*, 2011); *Xc-BolA*, *Xanthomonas campestris* (gi:3729539, PDB entry 1xs3; Chin *et al.*, 2005); *Eh-BolA*, *Ehrlichia chaffeensis* (gi:3927323; PDB entry 2kz0; Seattle Structural Genomics Center for Infectious Disease, unpublished work); *At-BolA*, *Arabidopsis thaliana* (gi:830843; PDB entry 2mm9; Roret *et al.*, 2014); *Ec-BolA*, *E. coli* (gi:12930846; PDB entry 2dhm; RIKEN Structural Genomics/Proteomics Initiative, unpublished work); *Mm-BolA*, *Mus musculus* (gi:66162; PDB entry 1v9j; Kasai *et al.*, 2004); *Hs-BolA*, *Homo sapiens* (gi:46577124). Identical and similar residues are identified in black and gray, respectively. Above the alignment are the elements of secondary structure observed for *Pf-BolA*: α -helix, red; 3_{10} -helix, magenta; β -strand, blue. Below the alignment are the two invariant residues observed in all BolA sequences (red asterisk), a highly conserved glycine observed in the helix–turn–helix (HTH) motif (red caret) and highly conserved hydrophobic residues in the core of the protein (red +).

for the backbone atoms ($\text{N}-\text{C}^\alpha-\text{C}=\text{O}$) and $0.70 \pm 0.11 \text{ \AA}$ for all heavy atoms.

3.2. Comparison to related BolA structures

Fig. 2(b) shows a cartoon representation of a single structure from the 20-residue ensemble of structures in Figs. 2(a) and 2(c) shows a two-dimensional schematic of the individual secondary structure components. The protein is composed of four helices and three β -strands with overall topology $\alpha 1-\beta 1-\beta 2-\eta 1-\alpha 2/\eta 2-\beta 3-\alpha 3$. The second helix after $\beta 2$ is a short 3_{10} -helix ($\eta 1$; colored magenta in Fig. 2) along with $\eta 2$. The latter 3_{10} -helix is an extension at the C-terminus of $\alpha 2$ that together forms one continuous helix ($\alpha 2/\eta 2$). The first three helices ($\alpha 1$, $\eta 1$ and $\alpha 2/\eta 2$) are nestled on one face of a mixed β -sheet, $\beta 1-\beta 2-\beta 3$, with $\beta 2$ antiparallel to $\beta 1$ and parallel to $\beta 3$. These secondary structure elements form the core of the protein, with the C-terminal helix, $\alpha 3$, extending out at an approximate right angle to the core. The core is stabilized by many hydrophobic amino acids, including Ile7, Leu11, Leu15, Phe19, Leu20, Leu22, Phe33, Ile37, Val38, Phe42, Leu47, Leu48, Leu53, Val54, Leu58, Leu62, Ile65 and Phe68, that form the interior of the core. As illustrated in the amino acid sequence alignment of BolA proteins from various organisms in Fig. 3,

these hydrophobic residues are highly conserved in the BolA protein family (Chin *et al.*, 2005).

Structures have been solved of BolA proteins from other organisms (Kasai *et al.*, 2004; Chin *et al.*, 2005; Abendroth *et al.*, 2011; Roret *et al.*, 2014), including seven of the eight species in the sequence alignment in Fig. 3. Fig. 4(a) shows a cartoon representation of the superposition of these seven BolA structures generated using *SuperPose* (Maiti *et al.*, 2004). While the RMSD of the aligned C^α atoms in ordered regions is high ($2.89 \text{ \AA}$), perhaps because six of the seven structures are NMR solution structures, overall the protein adopts a similar topological fold in all seven structures. The survey of these structures indicates a generally conserved 'kink' in the third helix, $\alpha 2/\eta 2$, suggesting that it may have some functional importance.

3.3. The helix–turn–helix (HTH) motif

Another feature in all of the BolA structures is a helix–turn–helix (HTH) motif (encircled in Fig. 4) with structural similarity to the class II KH fold (Kasai *et al.*, 2004). This KH fold is common to nucleic acid-binding proteins (Grishin, 2001) and is highlighted by a conserved surface of basic residues, as shown for *Pf*-BolA in Fig. 4(b), presumably responsible for binding to the negatively charged phosphodiester

Figure 4

(a) Cartoon representation of the superposition of the seven BolA structures with the sequences listed in Fig. 3: *Pf*-BolA (cyan, PDB entry 2kdn), *Bb*-BolA (purple, PDB entry 3o2e), *Xc*-BolA (blue, PDB entry 1xs3), *Eh*-BolA (gray, PDB entry 2kz0), *Ar*-BolA (orange, PDB entry 2mm9), *Ec*-BolA (red, PDB entry 2dhm) and *Mm*-BolA (yellow, PDB entry 1v9j). A 'kink' in $\alpha 2/\eta 2$ is identified with an arrow and the helix–turn–helix (HTH) motif that may bind nucleic acids is identified within a circle. (b) *PyMOL*-generated map of the electrostatic potentials on the solvent-accessible surface of *Pf*-BolA. The protein is rotated $\sim 90^\circ$ counterclockwise about the y axis of the structures shown in Figs. 4(a) and 2(b). Red and blue correspond to negatively and positively charged surfaces, respectively.

backbone of nucleic acids. Such a function would explain the capacity of BolA proteins to act as transcriptional regulators (Freire *et al.*, 2009; Guinote *et al.*, 2011). Unlike the well conserved GXXG sequence usually present in the 'turn' part of the KH fold, only the first Gly residue (red caret in Fig. 3) is well conserved in the associated region of BolA proteins, with the fourth residue and especially the fifth residue often being a leucine. More conserved among species is the phenylalanine two residues before the glycine (red asterisk in Fig. 3; Roret *et al.*, 2014), suggesting an FXGXXXL signature in the BolA family of proteins. While the nucleic acid sequence signature for BolA proteins has not yet been identified, differences in the amino acid sequence signature (FXGXXXL *versus* GXXG) in the BolA proteins suggests that the protein–nucleic acid interaction may differ from known protein–nucleic complexes. Note that the HTH region of the protein has also been associated with glutaredoxin recognition, suggesting that glutaredoxin may modify the nucleic acid-binding capacity of BolA proteins (Roret *et al.*, 2014).

Figure 5
(a) Close-up view of the relative orientation of the side chain of the conserved arginine in $\alpha 2/\eta 2$ of *Pf*-BolA (Arg50) and the side chain of the nearest polar residue (Thr74) in $\alpha 3$. The structure with the closest Arg50–Thr74 distance in the ensemble (PDB entry 2kdn) is shown. (b) Close-up view of the relative orientation of the side chain of the invariant BolA histidine in the crystal structure (PDB entry 3o2e) of *Bb*-BolA (His66) and the side chain of the conserved cysteine in the H/C loop (Cys29). α -Helices are colored red, 3_{10} -helices are colored magenta and β -strands are colored blue.

3.4. The C-terminal α -helix and a conserved arginine residue

Within the second helix ($\alpha 2/\eta 2$) of the HTH region is a highly conserved RHR signature sequence that is present in 41% of BolA sequences (Roret *et al.*, 2014). It has been proposed that the first arginine in the RHR signature sequence is important in governing the formation of the C-terminal helix ($\alpha 3$ in *Pf*-BolA) because of the polar contacts that the arginine makes with a side chain in this helix (Roret *et al.*, 2014). In *Pf*-BolA the equivalent arginine, Arg50, is greater than 7 Å away from the nearest side chain in $\alpha 3$ capable of forming polar contacts, Thr74, as shown in Fig. 5(a). Nevertheless, $\alpha 3$ still forms in *Pf*-BolA. Hence, the first arginine in the RHR signature may not be critical for the formation of the C-terminal helix, but instead may play a role in guiding the orientation of this helix relative to the core of the protein in some BolA structures. Given that amino acid residues are missing for this region of the C-terminus in some BolA sequences, such as *Eh*-BolA (Fig. 3), and that the sequence length and extent of the helix that forms varies in BolA structures (Fig. 4a), the C-terminal helix may not have a functional property.

3.5. The H/C loop

The BolA proteins can form [2Fe–2S]-bridged dimers with glutaredoxins (Li *et al.*, 2012). One of the ligands that make contact with the [2Fe–2S] cluster is likely to be the invariant histidine in the loop between $\alpha 2/\eta 2$ and $\beta 3$ (red asterisk in Fig. 3); however, there is less certainty regarding the other ligand. The candidate ligands are a conserved cysteine or histidine residue in the loop between $\beta 1$ and $\beta 2$ of BolA proteins. This loop is called the H/C loop depending upon the presence of a conserved cysteine (C-group) or histidine

Figure 6
(a) Circular dichroism steady-state wavelength spectra of *Pf*-BolA (~ 0.05 mM) in modified NMR buffer collected at 298 K (black), 353 K (red) and 298 K after heating to 353 K (purple). (b) The CD thermal plots for *Pf*-BolA obtained by measuring the ellipticity at 210 nm in 2.0 K intervals between 283 and 353 K. The profile in black is the original sample and the plot in purple is after heating to 353 K.

(H-group) residue (Roret *et al.*, 2014). In general, the length of the H/C loop in the C-group is smaller, with a narrower range (7 ± 2 residues), and that in the H-group is larger, with a wider range (11 ± 6 residues) (Roret *et al.*, 2014). *Pf*-BolA and four of the other sequences listed in Fig. 3 (*Bb*-BolA, *At*-BolA, *Mm*-BolA and *Hs*-BolA) fall into the C-group (Cys29 in *Pf*-BolA). In the crystal structure of *Bb*-BolA the distance between the cysteine in the H/C loop and the invariant histidine is ~ 4 Å (Fig. 5*b*), which is close enough to accommodate a [2Fe–2S] cluster by both residues. However, in *Pf*-BolA no amide cross peaks were unambiguously identified for the Lys25–Ser32 amides in the H/C loop, suggesting that this region was disordered and was the reason for the ‘hairball’ at the top of the calculated ensemble in Fig. 2(*a*). It has been suggested that backbone flexibility may be a property of the H/C loop that enables this more variable region to bind to a [2Fe–2S] cluster in association with the invariant histidine residue (Roret *et al.*, 2014). Supporting this hypothesis is the observation of backbone mobility on the fast timescale (picoseconds to nanoseconds) for the H/C loop residues of *At*-BolA (Roret *et al.*, 2014).

3.6. Thermostability of *Pf*-BolA

Circular dichroism spectroscopy was used to characterize the structure of *Pf*-BolA and the stability of this structure over a range of temperatures. Fig. 6(*a*) shows the steady-state wavelength spectra for *Pf*-BolA in a modified NMR buffer at two temperatures. The black profile corresponds to a steady-state wavelength spectrum collected at 298 K and is highlighted by a double minimum at 208 and 220 nm, features that are characteristic of a protein with significant α -helical content (Holzwarth & Doty, 1965; Woody, 1974), which is supported by our NMR-based structure. The red profile corresponds to a steady-state wavelength spectrum collected at 353 K and is highlighted by a more positive sharp minimum at 204 nm and a broad minimum at ~ 215 nm, suggesting an overall loss of structure. However, the absence of an extrapolated minimum at ~ 198 nm and a positive maximum at ~ 218 nm indicates that *Pf*-BolA is not fully random coil either (Buchko *et al.*, 2011). The purple profile is the steady-state wavelength spectrum after returning the sample to 298 K. While the intensity of the minima are more positive, overall the shape of the profile is similar to the original spectrum collected at 298 K (black), suggesting that most of the *Pf*-BolA refolded back to its native state after the first heating cycle. Note that upon heating the same sample to 353 K and returning it to 298 K again no further change was observed in the spectrum (data not shown; identical to the purple spectrum in Fig. 6*a*).

To assess the thermal stability of *Pf*-BolA, the ellipticity at 210 nm was measured as a function of temperature between 283 and 353 K. Typically, when a structured protein becomes denatured with increasing temperature, the transition may be followed by a concomitant increase in ellipticity (Buchko *et al.*, 2010; Karantzeni *et al.*, 2003). As shown in the two consecutive thermal melts for the same sample of *Pf*-BolA in Fig. 6(*b*), a rapid increase in ellipticity is observed starting at about 333 K

that begins to level out at 353 K. A third thermal melt not shown in Fig. 6(*b*) was identical to the second thermal melt (purple). While there is an inflection point at approximately 340 K, the steady-state wavelength spectra in Fig. 6(*a*) suggest that the protein has not fully unfolded at 353 K and therefore this does not represent a true melting point for the protein (Buchko *et al.*, 2010). Given that β -strands are generally more robust than α -helices (Creighton, 1993), this inflection point might represent an unraveling of the α -helices in *Pf*-BolA with the mixed β -sheet still largely intact. If true, the intact β -sheet at high temperature may be responsible for the ability of *Pf*-BolA to return to its native structure after heating to 353 K.

4. Conclusions

The solution structure of *Pf*-BolA consists of an $\alpha 1$ – $\beta 1$ – $\beta 2$ – $\eta 1$ – $\alpha 2$ / $\eta 2$ – $\beta 3$ – $\alpha 3$ topology with a mixed three-stranded β -sheet. The topology of the *Pf*-BolA structure is similar to the topology observed in other BolA proteins and is highlighted by a helix–turn–helix motif similar to the class II KH fold associated with nucleic acid-binding proteins. Because the signature sequence of the HTH motif in the BolA proteins, FXGXXXL, differs from the GXXG signature sequence present in class II KH folds, the protein–nucleic acid interface in the BolA family of proteins may differ from the interfaces in known protein–nucleic complexes and may represent a potential pathway for the conception and development of new therapies to treat and control the spread of malaria.

Acknowledgements

This research was funded with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under Contract Nos. HHSN2722001200025C and HHSN272200700057C, the Natural Sciences and Engineering Research Council of Canada and the Structural Genomics Consortium (SGC). The SGC is a registered charity (No. 1097737) that receives funds from AbbVie, Boehringer Ingelheim, the Canada Foundation for Innovation (CFI), the Canadian Institutes of Health Research (CIHR), Genome Canada, Ontario Genomics Institute Grant OGI-055, Glaxo-SmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda and the Wellcome Trust Grant 092809/Z/10/Z. The SSGCID internal ID for *Pf*-BolA is Pf1aA.01650.a. Part of this research was conducted at the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the US Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). Battelle operates PNNL for the US Department of Energy.

References

- Abendroth, J., Gardberg, A. S., Robinson, J. I., Christensen, J. S., Staker, B. L., Myler, P. J., Stewart, L. J. & Edwards, T. E. (2011). *J. Struct. Funct. Genomics*, **12**, 83–95.

- Aldea, M., Garrido, T., Hernández-Chico, C., Vicente, M. & Kushner, S. (1989). *EMBO J.* **8**, 3923–3931.
- Anstey, N. M., Russell, B., Yeo, T. W. & Price, R. N. (2009). *Trends Parasitol.* **25**, 220–227.
- Berman, H. M. (2008). *Acta Cryst.* **A64**, 88–95.
- Bhattacharya, A., Tejero, R. & Montelione, G. (2007). *Proteins*, **66**, 778–795.
- Buchko, G. W., Kim, C.-Y., Terwilliger, T. C. & Myler, P. J. (2010). *Tuberculosis*, **90**, 245–251.
- Buchko, G. W., Phan, I., Myler, P. J., Terwilliger, T. C. & Kim, Y.-C. (2011). *Arch. Biochem. Biophys.* **506**, 150–156.
- Cavanagh, J., Palmer, A. G., Fairbrother, W. & Skelton, N. (1996). *NMR Spectroscopy: Principles and Practice*. San Diego: Academic.
- Chin, K.-H., Lin, F.-Y., Hu, Y.-C., Sze, K. H., Lyu, P.-C. & Chou, S.-H. (2005). *J. Biomol. NMR*, **31**, 167–172.
- Cornilescu, G., Delaglio, F. & Bax, A. (1999). *J. Biomol. NMR*, **13**, 289–302.
- Creighton, T. E. (1993). *Proteins: Structure and Molecular Properties*. New York: W. H. Freeman & Co.
- D'Alessandro, U. (2009). *Expert Opin. Pharmacother.* **10**, 1291–1306.
- Dondorp, A. M. *et al.* (2009). *N. Engl. J. Med.* **361**, 455–467.
- Fidock, D. A. (2010). *Nature (London)*, **465**, 297–298.
- Freire, P., Neves Moreira, R. & Arraiano, C. M. (2009). *J. Mol. Biol.* **385**, 1345–1351.
- Gardner, M. J. *et al.* (2002). *Nature (London)*, **419**, 498–511.
- Grishin, N. V. (2001). *Nucleic Acids Res.* **29**, 638–643.
- Guinote, I. B., Matos, R. G., Freire, P. & Arraiano, C. M. (2011). *J. Microb. Biotechnol.* **21**, 243–251.
- Guntert, P. (2004). *Methods Mol. Biol.* **278**, 353–378.
- Holzwarth, G. M. & Doty, P. (1965). *J. Am. Chem. Soc.* **87**, 218–228.
- Karantzeni, I., Ruiz, C., Liu, C.-C. & Licata, V. J. (2003). *Biochem. J.* **374**, 785–792.
- Kasai, T. *et al.* (2004). *Protein Sci.* **13**, 545–548.
- Li, H., Mapolelo, D. T., Randeniya, S., Johnson, M. K. & Outten, C. E. (2012). *Biochemistry*, **51**, 1687–1696.
- Linge, J. P. & Nilges, M. (1999). *J. Biomol. NMR*, **13**, 51–59.
- Maiti, R., van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). *Nucleic Acids Res.* **23**, W590–W594.
- Myler, P. J., Stacy, R., Stewart, L. J., Staker, B. L., Van Voorhis, W. C., Varani, G. & Buchko, G. W. (2009). *Infect. Disord. Drug Targets*, **9**, 493–506.
- Phyo, A. P., Nkhoma, S., Stepniewska, K., Ashley, E. A., Nair, S., McGready, R., ler Moo, C., Al-Saai, S., Dondorp, A. M., Lwin, K. M., Singhasivanon, P., Day, N. P. J., White, N. J., Anderson, T. J. C. & Nosten, F. (2012). *Lancet*, **379**, 1960–1966.
- Roret, T., Tsan, P., Couturier, J., Zhang, B., Johnson, M. K., Rouhier, N. & Didierjean, C. (2014). *J. Biol. Chem.* **289**, 24588–24598.
- Santos, J. M., Freire, P., Vicente, M. & Arraiano, C. M. (1999). *Mol. Microbiol.* **32**, 789–798.
- Singh, S. P. & Montgomery, B. L. (2014). *Mol. Microbiol.* **93**, 167–182.
- Vieira, H. L. A., Freire, P. & Arraiano, C. M. (2004). *Appl. Environ. Microbiol.* **70**, 5682–5684.
- White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A. & Dondorp, A. M. (2014). *Lancet*, **383**, 723–735.
- Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). *J. Biomol. NMR*, **6**, 135–140.
- Woody, R. W. (1974). *Peptides, Polypeptides and Proteins*, edited by E. R. Blout, F. A. Bovey, M. Goodman & N. Lotan, pp. 338–348. New York: John Wiley & Sons.
- World Health Organization (2013). http://www.who.int/malaria/publications/world_malaria_report_2013/en/.
- Yee, A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1825–1830.