Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of the second domain of the *Bacillus subtilis* DEAD-box RNA helicase YxiN

Jonathan M. Caruthers,‡ YaoXiong Hu and David B. McKay*

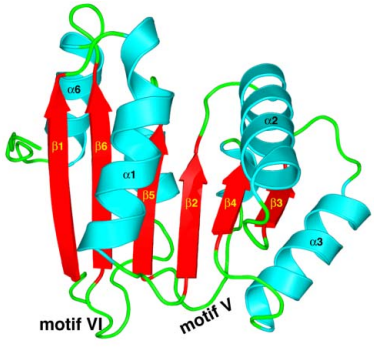
Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA

‡ Current address: Array BioPharma, 3200 Walnut Street, Boulder, Colorado 80301, USA.

Correspondence e-mail: dave.mckay@stanford.edu

Received 11 August 2006
Accepted 25 October 2006

PDB Reference: YxiN(207–368), 2hjv, r2hjvsf.



The *Bacillus subtilis* RNA helicase YxiN is a modular three-domain protein. The first two domains form a conserved helicase core that couples an ATPase activity to an RNA duplex-destabilization activity, while the third domain recognizes a stem-loop of 23S ribosomal RNA with high affinity and specificity. The structure of the second domain, amino-acid residues 207–368, has been solved to 1.95 Å resolution, revealing a parallel $\alpha\beta$-fold. The crystallographic asymmetric unit contains two protomers; superposition shows that they differ substantially in two segments of peptide that overlap the conserved helicase sequence motifs V and VI, while the remainder of the domain is isostructural. The conformational variability of these segments suggests that induced fit is intrinsic to the recognition of ligands (ATP and RNA) and the coupling of the ATPase activity to conformational changes.

1. Introduction

Proteins of the DE*x*(D/H)-box RNA-helicase family are involved in a broad spectrum of activities in which they apparently modulate the structures of RNAs. At one end of the spectrum, some of the activities are relatively nonspecific with respect to the RNA substrate, such as 'melting out' secondary structure in messenger RNAs (mRNA) by eukaryotic initiation factor 4A (eIF4A), thereby allowing translation to proceed unimpeded (Pause & Sonenberg, 1992; Rogers *et al.*, 1999). At the other end of the spectrum are activities that are highly specific with respect to target RNA, as exemplified by the several DE*x*(D/H)-box proteins involved in sequential steps of pre-mRNA splicing in yeast (Staley & Guthrie, 1998). This family of proteins shows a wide variation with respect to size, ranging from fewer than 400 to more than 1200 amino-acid residues. However, the proteins all share a highly conserved fragment that is approximately 400 residues in length which encompasses seven conserved peptide-sequence motifs originally identified as defining the larger family of RNA and DNA helicases (Gorbalenya & Koonin, 1993). The first crystal structure of a DE*x*(D/H)-box helicase, that of yeast eIF4A (Benz *et al.*, 1999; Johnson & McKay, 1999; Caruthers *et al.*, 2000), showed that the conserved ∼400-residue helicase fragment contains two structural domains, each with a parallel $\alpha\beta$-folding topology related to that of the RecA protein (Story & Steitz, 1992), consistent with the broader scheme of helicase structure (for a review, see Caruthers & McKay, 2002).

Subsequent to the structure determination of eIF4A (Caruthers *et al.*, 2000), structures of a putative DEAD-box helicase from *Methanococcus jannaschi* (MjDEAD; Story *et al.*, 2001), the conserved helicase fragments of human UAP56 protein (Shi *et al.*, 2004) and yeast Dhh1p protein (Cheng *et al.*, 2005) and a complex between single-strand U_{10} and the *Drosophila* Vasa protein fragment with adenylyl imidodiphosphate (AMPPNP; Sengoku *et al.*, 2006) have emerged. These molecules or molecular fragments all have two domains similar to those of eIF4A, but vary with respect to the relative disposition of the two domains: eIF4A is an extended 'dumbbell' in the crystal with the two domains connected by an extended peptide linker and no intramolecular interaction between

them, while the other structures are more compact. Furthermore, several of the helicase peptide motifs, particularly motifs V and VI, are poorly defined in many of the structures.

It has been established from biochemical studies that residues in helicase motif VI of the carboxy-terminal domain of the minimal helicase fragment must interact with the amino-terminal domain as a prerequisite for efficient ATP hydrolysis (Pause *et al.*, 1993). For the purpose of suggesting which residues participate in coupling the ATPase activity to interdomain interactions and conformational changes, arguably the most informative structure to date is that of the Vasa fragment complexed with AMPPNP and the single-strand oligonucleotide U_{10} (Sengoku *et al.*, 2006), which displays specific interactions between amino-acid side chains and ligands. Additional structural work on related helicases should amplify and extend the insights derived from the Vasa fragment–AMPPNP–U_{10} structure by comparison.

The YxiN protein of *Bacillus subtilis*, a homolog of *Escherichia coli* DbpA, is a modular three-domain protein that binds a segment of 23S ribosomal RNA (rRNA) with high affinity and specificity (Kossen & Uhlenbeck, 1999; Karginov *et al.*, 2005). The first two domains constitute the minimal helicase fragment common to all DE*x*(D/H)-box helicases, while the third domain is responsible for binding 23S rRNA and has an RNA-recognition motif structure (Karginov *et al.*, 2005; Wang *et al.*, 2006). Here, we present the crystal structure of the second domain, *i.e.* the carboxy-terminal domain of the minimal helicase fragment of YxiN. The crystal contains two protomers in the asymmetric unit; comparison of the two protomers allows us to directly identify peptide segments of the domain that are conformationally variable without the ambiguity introduced by differences in primary structure when comparing the tertiary structures of two different proteins.

2. Materials and methods

2.1. Protein expression, purification, crystallization and data collection

The subcloning of the coding sequence for wild-type YxiN residues 207–368 into the pTWIN1 vector of an intein-based expression system (New England Biolabs) utilized the same protocols as described for the subcloning of other YxiN fragments (Karginov *et al.*, 2005). Briefly, the coding sequence for YxiN(207–368) was amplified using the polymerase chain reaction (PCR) from an expression plasmid for the full-length protein using a forward primer of sequence 5′-GGAATTC**CATATG**GCGGCCGGCCTAACCACAAG and a reverse primer 5′-TTAATT**ACTAGT**GCATCTCCCGTGATGCA-TGCTTCTATTTTTTGGATTTC which included *Nde*I and *Spe*I sites, respectively (bold). Both the PCR product and the pTWIN1 vector were digested with *Nde*I and *Spe*I and the PCR fragment was ligated into the vector. The accuracy of the final construct was confirmed by DNA sequencing. In this construct, the codon for residue 368 of YxiN abuts the first codon of the self-cleaving intein encoded by the vector, so that no extraneous amino acids were retained in the final protein product.

The protein was expressed in *E. coli* BL21(DE3) and purified at 277 K using the protocol described for full-length YxiN protein (Karginov *et al.*, 2005). The vector expressed a fusion construct consisting of the YxiN(207–368) domain, a self-cleaving intein and a chitin-binding domain in tandem. Cells were harvested and lysed in 20 m*M* Tris–HCl, 500 m*M* NaCl pH 7.9 (buffer *A*); after centrifugation, the clarified supernatant was gradually brought to 0.1%(*w*/*v*) in polyethyleneimine (PEI) and precipitate was removed by centrifugation. The supernatant was applied onto a chitin column, washed with several column volumes of buffer *A* and then loaded with 50 m*M* dithiothreitol (DTT) in buffer *A*, after which the flow was stopped overnight to allow self-cleavage of the target–intein fusion. Protein was then eluted from the column, concentrated to a volume of a few millilitres and loaded onto a gel-filtration column (Superdex-75; Pharmacia) pre-equilibrated with 20 m*M* Tris–HCl, 100 m*M* NaCl titrated to pH 7.9 at 293 K (buffer *B*).

Protein in buffer *B* was crystallized at 291 K by vapor diffusion in hanging drops (1 ml precipitant in the well; 5 µl protein plus 5 µl precipitant initially in the drop) using protein at an initial concentration of 10 mg ml^{-1} and a precipitant consisting of 3% polyethylene glycol monomethyl ether of average molecular weight 550 (PEG MME 550) buffered to a pH of approximately 6.5 with 0.06 *M* MES, 0.04 *M* MOPS. Crystals were flash-frozen in liquid nitrogen after a rapid pass through a cryoprotectant solution consisting of mother liquor plus 20% glycerol. Data were collected to 1.95 Å resolution on a crystal cooled to ∼100 K in a stream of cold nitrogen on beamline 8.2.2 of the Advanced Light Source (ALS) using an ADSC 2 × 2 charge-coupled device (CCD) detector. Data were indexed, integrated and scaled with *HKL*-2000 (Otwinowski & Minor, 1997). Data-collection parameters were wavelength 1.000 Å, crystal-to-detector distance 250 mm, exposure time 45 s, oscillation 0.8° per frame, 110 frames. Data completeness was tempered by ice rings in some resolution ranges. Unit-cell parameters and data-collection statistics are summarized in Table 1.

Table 1
Data-collection and refinement statistics for native YxiN(207–368).

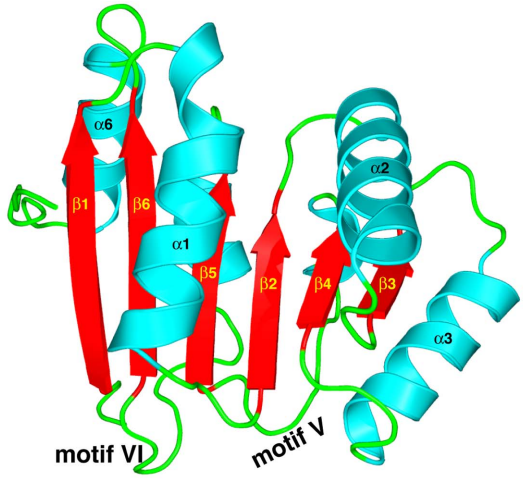
Data-collection statistics were computed in *HKL*-2000 (Otwinowski & Minor, 1997) and refinement statistics were computed in *CNS* (Brünger *et al.*, 1998), as described in §2. Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$P6_1$
Unit-cell parameters (Å)	a = 105.20, c = 66.03
Wavelength (Å)	1.000
Resolution range (Å)	50.0–1.95 (1.98–1.95)
Total reflections	111487
Unique reflections	27397 (1450)
Completeness (%)	90.5 (95.9)
Redundancy	4.1 (4.3)
R_{sym}†	0.069 (0.227)
Refinement	
Resolution range (Å)	30.0–1.95 (2.02–1.95)
R_{cryst}‡	0.209 (0.221)
R_{free}	0.248 (0.227)
No. of reflections (working set)	25391
No. of reflections (test set)	1600
No. of protein atoms	2533
No. of water molecules	163
Average B value (Å^2)	34.9
R.m.s.d. bond lengths (Å)	0.008
R.m.s.d. angles (°)	1.32
Ramachandran analysis	
Residues in most favored regions (%)	95.8
Residues in additional allowed regions (%)	4.2
Residues in generously allowed regions	None
Residues in disallowed regions	None

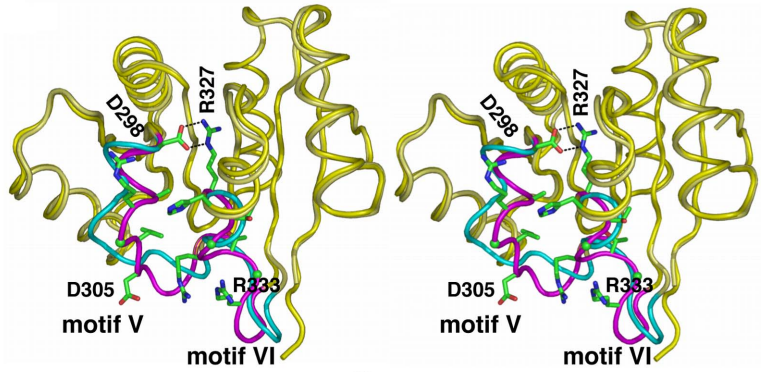
† $R_{sym} = \sum |I_{hkl} - \langle I_{hkl}\rangle| / \sum \langle I_{hkl}\rangle$, where I_{hkl} is a single value of the measured intensity of reflection hkl and $\langle I_{hkl}\rangle$ is the mean of all measured values of the intensity of reflection hkl. ‡ $R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$, where F_{obs} is the observed structure-factor amplitude and F_{calc} is the structure factor calculated from the model. R_{free} is computed in the same manner as R_{cryst} using the test set of reflections.

2.2. Structure determination and refinement

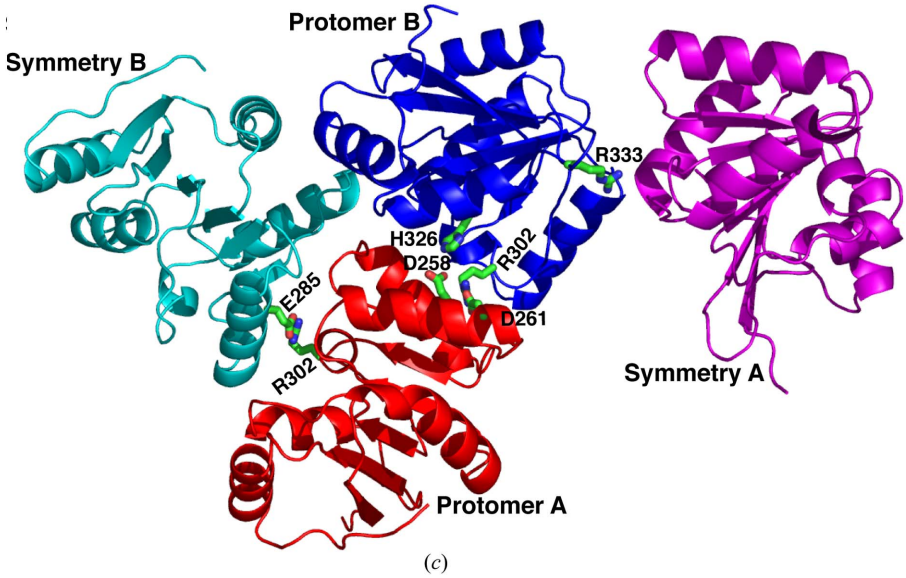
Crystallographic computations were carried out using *CCP*4 (Collaborative Computational Project, Number 4, 1994), *CNS* (Brünger *et al.*, 1998) and *EPMR* (Kissinger *et al.*, 1999, 2001); model



(*a*)



(*b*)



(*c*)

Figure 1
Structure of YxiN(207–368). (*a*) Ribbon diagram of protomer *A*. *β*-Strands and selected *α*-helices are numbered sequentially. The motif V and VI loops are labeled. (*b*) Superposition of the two protomers in the asymmetric unit. A stereo diagram of the C_α trace is shown as a smoothed coil. The backbones of segments of peptide where C_α positions differ by <0.5 Å are shown in yellow for both protomers; segments where C_α positions differ by >0.5 Å are shown in magenta (protomer *A*) and cyan (protomer *B*). The side chains of selected residues of helicase motifs V and VI, as well as Asp298, are shown for protomer *A*. The orientation of the figure is rotated approximately 180° around a vertical axis relative to the schematic drawing in (*a*). (*c*) Crystal-packing interactions involving residues from motifs V and VI. Only the side chains of residues from motifs V and VI that are involved in intermolecular interactions are shown. Protomer *A*, red; protomer *B*, blue; symmetry-related *A*, magenta; symmetry-related *B*, cyan. All figures were produced using *PyMOL* (http://www.pymol.org).

building was performed with the program *O* (Jones, 1978; Jones *et al.*, 1991). The structure was solved by molecular replacement. An initial search model for the domain that incorporated the YxiN amino-acid sequence (Swiss-Prot P42305) was constructed from the equivalent domains of yeast eIF4A (PDB code 1fuu) and MjDEAD (PDB code 1hv8) using the *Swiss-Model* server (http://swissmodel.expasy.org/; Schwede *et al.*, 2003). Molecular searches were carried out with the program *EPMR*. The search for the first protomer in the asymmetric unit yielded a solution with a correlation coefficient of 0.316 and the subsequent search for the second protomer, with the first one fixed, gave a solution with a correlation coefficient of 0.561. Refinement was carried out with *CNS*. Rigid-body refinement of the molecular-replacement solution using data to 2.3 Å resolution gave a model with $R_{\rm cryst}$ = 0.342 and $R_{\rm free}$ = 0.364. Successive cycles of refinement and model rebuilding into simulated-annealing OMIT maps, as well as $F_{\rm o} - F_{\rm c}$ and $2F_{\rm o} - F_{\rm c}$ maps computed using model phases, yielded a model with $R_{\rm cryst}$ = 0.209 and $R_{\rm free}$ = 0.248 (Table 1). Candidate solvent molecules were identified automatically with the 'water_pick' module of *CNS* and were edited manually in *O*. The model includes residues 211–368 of protomer *A*, residues 214–368 of protomer *B* and 163 solvent molecules. The coordinate error estimated from a Luzzati plot is 0.22 Å.

3. Results and discussion

As anticipated (Caruthers & McKay, 2002), YxiN (207–368) has a parallel $\alpha\beta$ architecture similar to that found in the structures of cognate helicase domains (Pfam 00271; Bateman *et al.*, 2000; PDB codes 1fuk, 1hv8, 1xti, 2db3). There are six β-strands and six α-helices, with β-strand topological order 1–6–5–2–4–3 (Fig. 1*a*). The two independent protomers in the asymmetric unit reveal conformational differences which may be of relevance to biochemical function. Superposition of the protomers shows that they are very similar in structure except for two peptide segments. These segments are residues 299–309, which encompass helicase motif V (residues 301–305, sequence ARGID), and residues 328–337, which overlap helicase motif VI (residues 326–333, sequence HRTGRTGR; Rocak & Linder, 2004). When the protomers are superimposed, the root-mean-square difference (r.m.s.d.) in C_α positions excluding these two segments is 0.59 Å; in comparison, the r.m.s.d. of C_α atoms within the two segments is 2.22 Å. The average main-chain *B* factors for these segments are somewhat higher than for the overall structure, consistent with their apparent flexibility; nonetheless, they were generally sufficiently well defined that they could be modeled unambiguously. The main-chain atoms of the first peptide segment, residues 299–309, have an average *B* factor of 44.0 Å^2 in protomer *A* (using the chain-designation convention of the coordinates in PDB entry 2hjv) and 40.0 Å^2 in protomer *B*. These values can be compared with an average main-chain *B* factor of 30.8 Å^2 for the entire structures. The second peptide segment is well defined in protomer *A*, with average main-chain *B* factors of 34.8 Å^2; it is less well defined in protomer *B*, with average main-chain *B* factors of 46.3 Å^2.

These two segments, which are adjacent to each other in the structure, appear to be tethered by a specific side-chain interaction at their amino-terminal ends and backbone interactions of β-strands five and six at the carboxy-terminal ends (Fig. 1*b*). Specifically, Arg327, which is a conserved residue of motif VI, forms a salt bridge with Asp298, which precedes motif V in the amino-acid sequence. The aspartic acid is not rigorously conserved in the amino-acid sequences of the DEAD-box helicases but is found in a large fraction of them; a similar salt bridge is found in the structures of eIF4A (Caruthers *et al.*, 2000) and MjDEAD (Story *et al.*, 2001). These interactions appear to be designed to allow conformational freedom in the segments of peptide that include helicase motifs V and VI that interact with ATP and RNA, while the remainder of the domain has a relatively static structure. Notably, comparison of the structures of eIF4A, MjDEAD, UAP56 and Dhh1p in the absence of ligands to that of Vasa with U_{10} and AMPPNP bound (Sengoku *et al.*, 2006) revealed substantial structural variation in motifs V and VI.

The different conformations of these two peptide segments are stabilized only slightly by crystal-packing interactions of the two protomers. In protomer *B*, Arg302 makes a salt bridge with Asp261 of protomer *A* and the imidazole of His326 of protomer *B* interacts with the carboxyl of Asp258 of protomer *A* (Fig. 1*c*). Additionally, the guanidinium group of Arg333 of protomer *B* is in proximity to, and may form hydrogen bonds with, the peptide backbone of a symmetry-related *A* protomer. In protomer *A*, Arg302 of motif V forms a salt bridge with Glu285 of a symmetry-related *B* protomer. None of the other residues of motifs V and VI in protomer *A* are involved in crystal-packing interactions. The majority of the crystal-packing interactions do not involve the two peptide segments that show conformational variability in the structures. The capability of these segments of peptide to adapt their conformations to accommodate

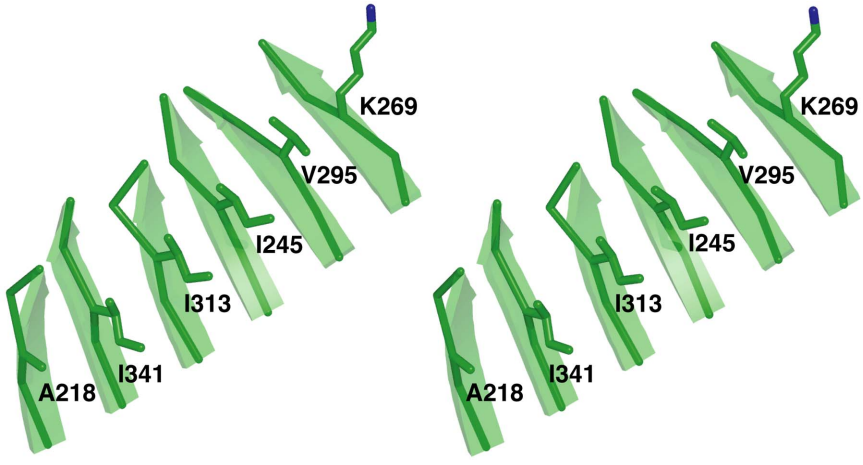


Figure 2
Stereo diagram of an isoleucine/leucine/valine 'stack' of the β-sheet in protomer *A*. The side chains of adjacent residues of one stack, as well as a C_α backbone and ribbon drawing of five-residue segments of the β-strands, are shown.

local interactions suggests a mechanism of induced fit is utilized by motifs V and VI to bind ATP and RNA.

Functions for specific amino acids in these motifs are suggested by comparison with the structure of the *Drosophila* Vasa two-domain helicase fragment complexed to U_{10} (Sengoku *et al.*, 2006). Arg302 of YxiN motif V corresponds to Arg551 of the second helicase domain of Vasa, which interacts trans-domain with the final aspartic acid of the DEAD motif of the first domain. Asp305 of YxiN corresponds to Asp554 of Vasa, which interacts with the ribose of AMPPNP bound to the first domain. Mutagenesis of either of these residues to alanine in the Vasa protein uncouples the ATPase and RNA-unwinding activities. Arg330 of YxiN motif VI corresponds to Arg579 of Vasa, the guanidinium group of which interacts with the γ-phosphate of AMPPNP, while Arg333 of YxiN corresponds to Arg582 of Vasa, which bridges the α- and γ-phosphates of the nucleotide. Although mutational analysis of these specific residues was not carried out in the Vasa protein, it was noted that mutagenesis of the equivalent residues in eIF4A impaired the ATPase activity (Pause *et al.*, 1993).

An additional interesting structural feature of this domain can be described as an isoleucine/leucine/valine 'stack' in the β-sheet (Fig. 2). This stack is formed by adjacent residues on the central four strands of the six-strand β-sheet, where the side chains of isoleucine or leucine, or less frequently valine or alanine, stack in a uniform conformation. The residues on the two edge strands are variable and are often hydrophilic, since they may interface with solvent. For example, on one side of the β-sheet of the YxiN domain the adjacent residues are (edge)–Ala218–Ile341–Ile313–Ile245–Val295–Lys269–(edge) (Fig. 1*c*), while on the other side a less well defined stack of residues His217–Ala340–Val312–Ile244–Leu294–Asp268 occurs. In the isostructural positions on the eIF4A carboxy-terminal domain (Caruthers *et al.*, 2000), the amino-acid stacks are Phe236–Ile360–Ile332–Ile264–Ile312–Ala288 and Gln235–Ala359–Leu331–Leu263–Leu313–Ser287. Regular stacks of hydrophobic residues have been noted in proteins that have repeating parallel β-strands. For example, the parallel β-roll of the alkaline protease of *Pseudomonas aeruginosa* is formed by a nine-residue repeat of consensus sequence GG*x*G*x*D*x*L*x*, where *x* represents a variable residue (Baumann *et al.*, 1993); the leucines of successive repeats stack on the interior of the β-roll. However, we have not found any note of such regular stacks in parallel $\alpha\beta$ protein domains. It is probable that the stacking of the side chains influences the folding of the domain; by implication, this motif may be helpful in threading sequences of unknown structures onto template folding topologies.

We would like to thank the members of the Roger Kornberg laboratory for participation in data collection. This work was supported by National Institutes of Health (NIH) research grant GM-71696 to DBM. Portions of this research were conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy, Office of Basic Energy Sciences. The Berkeley Center for Structural Biology is supported in part by the Department of Energy, Office of Biological and Environmental Research and by the NIH National Institute of General Medical Sciences.

References

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). *Nucleic Acids Res.* **28**, 263–266.
Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. (1993). *EMBO J.* **12**, 3357–3364.
Benz, J., Trachsel, H. & Baumann, U. (1999). *Structure*, **7**, 671–679.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
Caruthers, J. M. & McKay, D. B. (2002). *Curr. Opin. Struct. Biol.* **12**, 123–133.
Cheng, H., Coller, J., Parker, R. & Song, H. (2005). *RNA*, **11**, 1258–1270.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Gorbalenya, A. E. & Koonin, E. V. (1993). *Curr. Opin. Struct. Biol.* **3**, 419–429.
Johnson, E. R. & McKay, D. B. (1999). *RNA*, **5**, 1526–1534.
Jones, A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Jones, T. A., Zhou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Karginov, F. V., Caruthers, J. M., Hu, Y., McKay, D. B. & Uhlenbeck, O. C. (2005). *J. Biol. Chem.* **280**, 35499–35505.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kissinger, C. R., Smith, B. A., Gehlhaar, D. K. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.
Kossen, K. & Uhlenbeck, O. C. (1999). *Nucleic Acids Res.* **27**, 3811–3820.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pause, A., Methot, N. & Sonenberg, N. (1993). *Mol. Cell. Biol.* **13**, 6789–6798.
Pause, A. & Sonenberg, N. (1992). *EMBO J.* **11**, 2643–2654.
Rocak, S. & Linder, P. (2004). *Nature Rev. Mol. Cell Biol.* **5**, 232–241.
Rogers, G. W. Jr, Richter, N. J. & Merrick, W. C. (1999). *J. Biol. Chem.* **274**, 12236–12244.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2006). *Cell*, **125**, 287–300.
Shi, H., Cordin, O., Minder, C. M., Linder, P. & Xu, R. M. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 17628–17633.
Staley, J. P. & Guthrie, C. (1998). *Cell*, **92**, 315–326.
Story, R. M., Li, H. & Abelson, J. N. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
Story, R. M. & Steitz, T. A. (1992). *Nature (London)*, **355**, 374–376.
Wang, S., Hu, Y., Overgaard, M. T., Karginov, F. V., Uhlenbeck, O. C. & McKay, D. B. (2006). *RNA*, **12**, 959–967.