Journal of Biomolecular NMR (2006) 35: 149–154
DOI 10.1007/s10858-006-0029-3

NMR structure note

MTH187 from *Methanobacterium thermoautotrophicum* has three HEAT-like repeats★

Olivier Julien[a,†], Isabelle Gignac[a,†], Anna Hutton[b], Adelinda Yee[c], Cheryl H. Arrowsmith[c] & Stéphane M. Gagné[a,*]

[a]*Département de biochimie et de microbiologie and CREFSIP, Université Laval, 3255 pav. Marchand, Québec G1K 7P4, Canada;* [b]*Institute for Biomolecular Design, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada;* [c]*Division of Molecular and Structural Biology, Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, Ontario M5G 2M9, Canada*

Received 27 February 2006; Accepted 7 March 2006

Key words: HEAT-like repeat, *Methanobacterium thermoautotrophicum*, MTH187, NMR, protein structure, structural proteomics

Abstract

With the completion of genome sequencing projects, there are a large number of proteins for which we have little or no functional information. Since protein function is closely related to three-dimensional conformation, structural proteomics is one avenue where the role of proteins with unknown function can be investigated. In the present structural project, the structure of MTH187 has been determined by solution-state NMR spectroscopy. This protein of 12.4 kDa is one of the 424 non-membrane proteins that were cloned and purified for the structural proteomic project of *Methanobacterium thermoautotrophicum* [Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. (2000) *Prog. Biophys. Mol. Biol.*, **73**, 339–345]. *Methanobacterium thermoautotrophicum* is a thermophilic archaeon that grows optimally at 65 °C. A particular characteristic of this microorganism is its ability to generate methane from carbon dioxide and hydrogen [Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N. et al. (1997) *J. Bacteriol.*, **179**, 7135–7155].

Abbreviations: DSS – 2,2-dimethyl-silapentane-5-sulfonic acid; NOE – nuclear Overhauser effect; NOESY – NOE spectroscopy; PSI-BLAST – Position-Specific Iterative Basic Local Alignment Search Tool; RMSD – root-mean-square deviation

MTH187 is a small conserved protein of 111 residues with three known paralogs having about 37% sequence identity; MTH1378, MTH1806 and MTH1715 with 132, 183 and 449 residues, respectively. None of them have a known structure, and no function has been determined experimentally. BLAST and PSI-BLAST searches found less than 45% sequence identity with sequence databases. A conserved domain search suggests that MTH187 potentially belongs to one of these evolutionary related families: Armadillo/β-catenin-like repeats, E-Z type HEAT repeats present in subunits of cyanobacterial phyco-

★Structure data have been deposited at PDB (1TE4) and NMR data at BMRB (5629).

†Both authors contributed equally to this work.
*To whom correspondence should be addressed. E-mail: sgagne@rsvs.ulaval.ca

cyanin lyase, PBS lyase HEAT-like repeats, or COG1413/FOG of HEAT repeats involved in energy production and conversion. A PSI-BLAST search against proteins with known structure in the PDB indicates a maximum of 32% sequence identity over 63 residues of MTH187 with importin-β proteins. These α-helical proteins which are found in various organisms are involved in protein–protein interactions for macromolecular transport through nuclear pores (Vetter et al., 1999). With neither biochemical characterization nor high sequence similarity with known proteins, functional classification remains ambiguous. In the hope of obtaining clues into the function of MTH187 as well as the three paralogs, we have determined its structure by solution-state NMR spectroscopy.

Methods and results

The gene mt0187 amplified by PCR was cloned in the expression vector pET-15b (Novagen) between restriction sites *Nde*I and *Bam*HI. A coding region of 20 residues containing a HIS_6-tag sequence and a thrombin cleavage site was added to the N-terminus. The expression of MTH187 was done in *E. coli* BL21 (λDE3) in M9 minimal medium with $^{15}NH_4Cl$ and/or ^{13}C-Glucose. Finally, the purification was performed on a nickel column and the HIS-tag extension was retained (Christendat et al., 2000).

Two samples uniformly labelled with either ^{15}N or $^{15}N/^{13}C$ were prepared in identical solution conditions. Both were dissolved in 90% H_2O:10% D_2O with 25 mM potassium phosphate buffer at pH 6.2. In addition, 10 mM of NaCl and 0.5 mM of DSS were added. Since MTH187 has a tendency to form weak dimers in solution, 22 mM of CHAPS was added to displace the equilibrium towards the monomeric state.

The ^{15}N-HSQC of MTH187 is well resolved (Figure 1). The first spectra analyzed were a 3D ^{15}N-TOCSY-HSQC and a 3D ^{15}N-NOESY-HSQC, which allowed for the assignment of 76% of $^1H^N$, ^{15}N and $^1H_\alpha$. The greater part of the backbone atoms were assigned with the 3D CBCA(CO)NH and 3D HNCACB spectra. Backbone assignments were obtained for all amino acids, except for the $^1H^N$ and $^1H_\alpha$ of M1 and G67. 3D HNCO, 3D HN(CA)CO, 3D HCA(CO)N, 3D HN(CO)CA, and 2D ^{13}C-HSQC were also used to complete the backbone assignment. Side chain assignments have been completed to 98% for non-aromatic residues using 3D ^{15}N-TOCSY-HSQC, 3D HCCH-TOCSY and CC(CO)NH. A 3D HCCH-COSY was also acquired to assign aromatic residues. The majority of these spectra were acquired at 600 MHz on a Varian INOVA spectrometer at 43.2 °C. The ^{15}N-TOCSY-HSQC (500 MHz, $\tau_m = 100$ ms) and the ^{15}N-NOESY-HSQC (800 MHz, $\tau_m = 100$ ms) were acquired at the Canadian National Center for NMR (NANUC). Finally, a 3D ^{13}C-NOESY-HSQC was acquired in addition to the 3D ^{15}N-NOESY-HSQC to obtain all of the NOE restraints. The NMR spectra were processed with NMRpipe (Delaglio et al., 1995) and analyzed with NMRview 5.0.4 (Bruce A. Johnson, One Moon Scientific Inc). A manual calibration based on intensities was used for 768 NOE peaks of the 3D 15N-NOESY-HSQC, while other NOEs were calibrated by the default ARIA protocol (Linge et al., 2001).

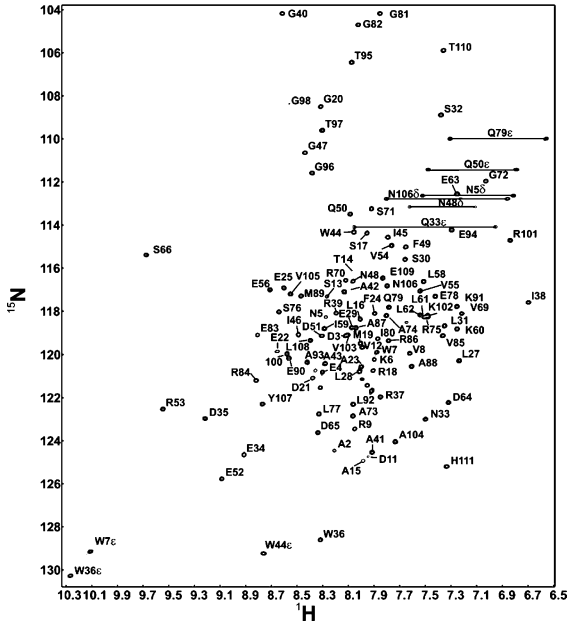

Figure 1. ^{15}N-HSQC spectrum (600 MHz) of MTH187 with peak assignments.

Structure calculations were performed using the ARIA 1.2 protocol for automated NOE assignment (Linge et al., 2001). A total of 3122 NOEs were used to calculate a family of structures (Table 1). The default ARIA settings were used with

Table 1. Characteristics of MTH187 for the 40 lowest energy structures

Number of experimental restraints		
Total restraints used	3305	
Total ambiguous distance restraints	739	
Total unambiguous distance restraints	2383	
Intra-residual	872	
Sequential ($\|i-j\| = 1$)	497	
Medium range ($1 < \|i-j\| < 5$)	490	
Long range ($\|i - j\| = 5$)	524	
Dihedral angle restraints from TALOS (Φ/Ψ)	49/78	
Coupling constants	47	
Structural statistics		
RMSD from idealized geometry		
Bonds (Å)	1.88e-03 ± 8.0e-05	
Angles (°)	0.37 ± 0.01	
Impropers (°)	0.28 ± 0.02	
RMSD from restraints		
Coup (Hz)	0.90 ± 0.05	
NOE (Å)	0.0127 ± 0.0008	
cdih (°)	0.49 ± 0.26	
Residual angle violations		
Number $>2°$	32	
Maximum °	4.96	
NOE violations/structure		
Number >0.1 Å	13.8 ± 2.8	
Final energies (kcal/mol)		
E (bonds)	6.2 ± 0.6	
E (angles)	63.6 ± 4.3	
E (impropers)	10.6 ± 1.4	
E (vdw)	91.1 ± 5.1	
E (noe)	25.4 ± 3.3	
E (cdih)	2.4 ± 3	
Ramachandran plot for residues 24–109[a]		
Φ/Ψ in most favored region	94.2%	
Φ/Ψ in additionally allowed region	5.8%	
Atomic RMSD (Å)	Backbone atoms (N,Ca,C′)	Heavy atoms
Residues 1–111	2.47 ± 0.66	2.97 ± 0.65
Residues 24–109	0.59 ± 0.11	1.15 ± 0.11
Helices (25–32, 37–46, 52–61, 68–80, 83–93, 99–107)	0.54 ± 0.10	1.08 ± 0.12

[a]Calculated from PROCHECK_NMR v.3.5.4 (Laskowski et al., 1996), excluding GLY and PRO.

minor modifications. A total of 100 structures were generated for each of the 9 ARIA iterations. The 40 best structures were selected for the next iteration. The coupling constant energy term was set to a square potential function for the 47 $^{3}J_{HN-H\alpha}$. The TALOS database system (Cornilescu et al., 1999) was used to obtain empirical predictions for dihedral angles based on protein sequence and $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, and ^{15}N chemical shift assignments (Table 1). When a residue had a coupling constant restraint, the TALOS prediction for Φ angle was not used, except for 11 residues. These 11 TALOS constraints were included to disallow the non-favoured region of the Ramachandran plot predicted by the degenerate Karplus relation. Also, according to the rate of amide proton exchange, nine hydrogen bonds were used for structure calculations. Finally, the 40 lowest energy

structures from the last iteration were used to constitute the final representative family (Figure 2a).

The final 40 lowest energy structures were validated using PROCHECK_NMR (Laskowski et al., 1996). All dihedral angles for structured residues are in allowed conformations of the Ramachandran plot (Table 1). However, two residues are found to have Φ angles around $+60°$: GLU50 and ASP65 which are located in the turns between $\alpha2$–$\alpha3$ and $\alpha3$–$\alpha4$, respectively. Several iterations of calculation without restraints for these regions were performed to try to relax these turns. However, no significant changes were observed. Dihedral angle violations were obtained using an in-house PyMOL extension (PyNMR, L. Willard and S. Gagné, Laval University); on average, there is less than one dihedral restraint violation ($>2°$) per structure (Table 1).

The solution structure of MTH187 reveals six anti-parallel α-helices forming a repetition of helix-turn-helix motif (Figure 2b). The turns between the helices are very short and each helix is composed of 8–13 residues: E25–S32 ($\alpha1$), R37–I46 ($\alpha2$), E52–L61 ($\alpha3$), F68–I80 ($\alpha4$), E83–A93 ($\alpha5$), F99–Y107 ($\alpha6$). The tertiary structure of MTH187 is composed of two rows of three parallels α-helices: $\alpha1$, $\alpha3$ and $\alpha5$ on one side and $\alpha2$, $\alpha4$ and $\alpha6$ on the other. The angles between parallel helices are between 4 and 16°. In the five helix-turn-helix motifs, successive helices are related by an inter-helix angle of $42 \pm 5°$. Backbone superposition of residues 24–109 results in a RMSD of 0.59 ± 0.11 Å for the 40 lowest energy final structures. The first 23 N-terminal residues are not well structured and have high RMSDs. The flexibility of the first 23 residues was confirmed by ^{15}N-relaxation data (data not shown).

Discussion and conclusions

The three-dimensional structure reveals several characteristics that are complementary to sequence homology information, and provides clues about the function of MTH187: type of folding, electrostatic surface characteristics and structural homology.

The hydrophobic core of MTH187 is formed by the helix juxtapositions: hydrophobic side chains are compact in the center of the protein and charged residues are mostly at the surface. The distribution of charges at the surface is asymmetric. One side of the protein presents a negative charged distribution, whereas the other is slightly hydrophobic (Figure 2c). T_2 relaxation data of the amides indicates that, in the absence of CHAPS, MTH187 has an apparent molecular weight of approximately 1.5 times the monomeric molecular weight, suggesting monomer–dimer equilibrium in solution. The asymmetric distribution of charges could be the reason for the presence of dimers in solution: the hydrophobic side likely makes contact with other monomers to form dimers in the absence of CHAPS. The structure of MTH187 was determined in the presence of CHAPS.

Structural homologues of MTH187 were found with the server Dali (Holm and Sander, 1993) and were confirmed with the CE server (Shindyalov and Bourne, 1998). The 10 best hits found by each structural homology server were superposed on MTH187. Comparison of these three-dimensional structures with MTH187 revealed similarities to two protein families which are evolutionarily related: the Armadillo repeat and the HEAT repeat (Andrade et al., 2001). These motifs are composed of a tandem repeat of 37–47 amino acids and consist of 3–36 pairs of anti-parallel α-helices forming a rod-like helical structure. Although HEAT-repeat proteins are involved in a great diversity of cellular processes, a function common to many is that of mediating important protein–protein interactions (Andrade et al., 2001). The two proteins most similar to MTH187 are aconitase B from *E. coli* (PDB ID 1L5J; 2.2 Å RMSD) (Williams et al., 2002) and hypothetical protein Z5015 from *E. coli* (PDB ID 1OYZ; 2.8 Å RMSD; Northeast Structural Genomics Consortium target Et31). These proteins present a HEAT-like domain where the helices are shorter than typical HEAT repeats (Andrade et al., 2001). For the same distinctiveness, we class MTH187 in the HEAT-like repeat structural family.

A particular characteristic observed in HEAT repeats concerns dihedral angles; several non-glycine residues in turns between sequential helices have positive Φ angles like MTH187. Moreover, when the aconitase B is superimposed to MTH187 according to the Dali alignment, we observed that two GLUs are present at the same position in the two structures, and both of them have a positive Φ angle. These potentially

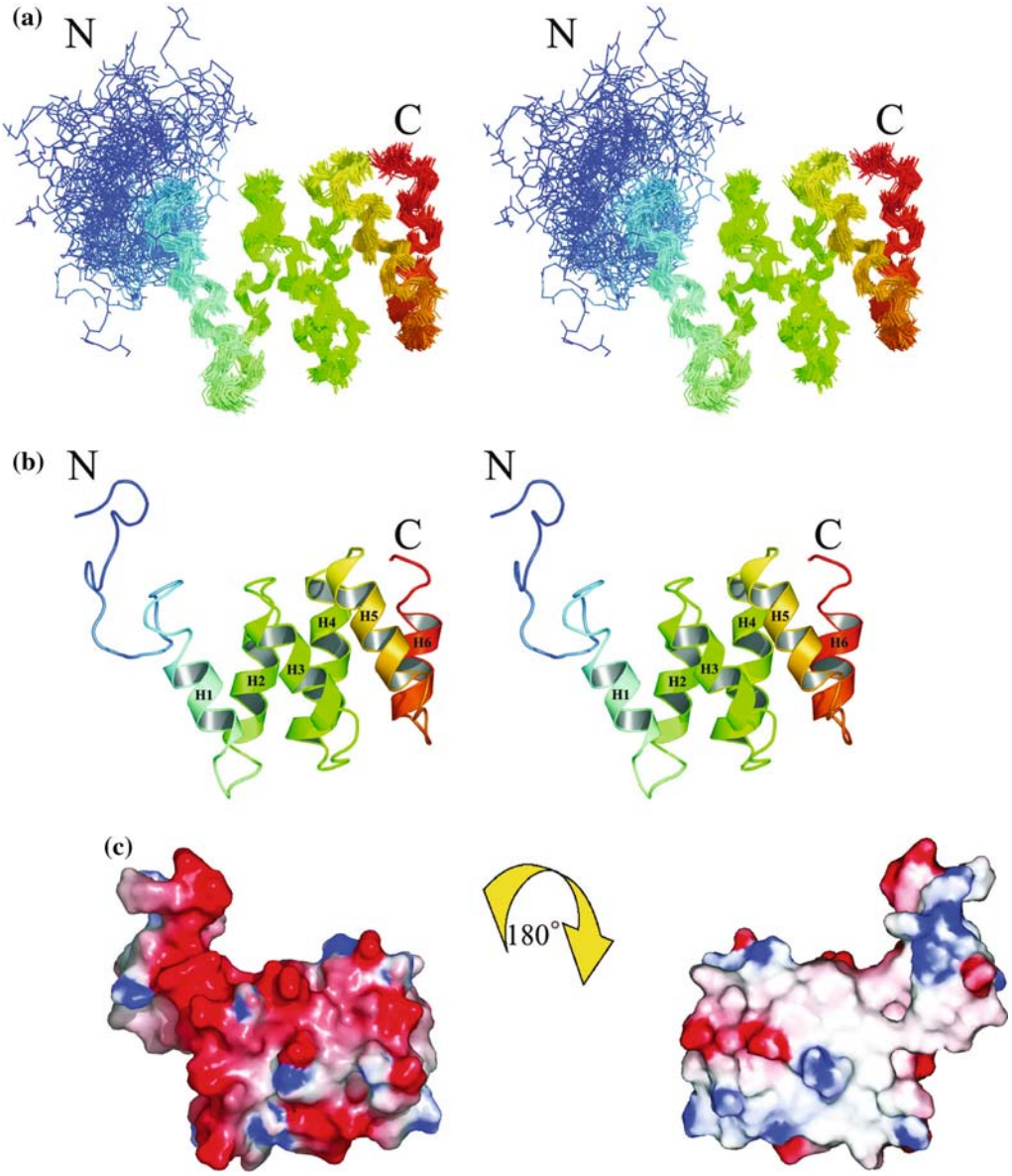

Figure 2. Structures of MTH187. (a) Backbone superimposition of the 40 lowest energy structures obtained by simulated annealing with ARIA/CNS and displayed in stereo view. The backbone RMSD of theses structures is 0.59 Å for residues 24–109. Residues 1–23 are not well defined. (b) Stereo view of a cartoon representation of the representative structure. The six anti-parallels α-helices are composed of 8–13 residues: E25–S32, R37–I46, E52–L61, F68–I80, E83–A93 and F99–Y107. (c) Front and back view of the electrostatic surface of MTH187. Red corresponds to negative charges and blue to positive charges. This figure was generated with PyMOL (Warren L. Delano, Delano Scientific LLC, San Carlos, CA). The surface representation was generated using a combination of Grasp (Nicholls et al., 1991) and PyMOL.

unfavourable conformations are a result of short turns and are compensated by favourable hydrophobic interactions created by helix packing. This observation confirms that unfavourable dihedral angles are allowed in these structural motifs.

The solution structure of MTH187 has provided important information concerning the structures of MTH1378, MTH1715 and MTH1806. Using Modeller 6v2 (Marti-Renom et al., 2000), hypothetical structures for each of the three paralogs were constructed by molecular modeling, with

MTH187 as template. The three models clearly show hydrophobic cores in the middle of the structures; side chains of charged residues are mostly observed at the surface for all models (Supplementary Figure S1b). Charge distribution at the surface is asymmetric for MTH1378, as observed for MTH187 (Supplementary Figure S1c). Moreover, secondary structure prediction was performed with the programs PHD, APSSP2, SAM-T99, SSPRO and PSIPRED for each paralog. The consensus for MTH1378, MTH1806 and the first 300 residues of MTH1715 (residues 300 to 449 have ambiguous consensus) showed a high level of confidence for repeated α-helices connected with short turns. Sequence alignment of each paralog with MTH187 was then performed with T-coffee (Notredame et al., 2000); those alignments confirmed that MTH187 helices fit by length and emplacement to the ones calculated from the secondary structure prediction (Supplementary Figure Sla). Consequently, the homology models are corroborated by secondary structure predictions and electrostatic analyses. There is a strong likelihood that MTH1378, MTH1715 and MTH1806 possess a structure similar to MTH187.

To summarize, the solution structure of MTH187 has been determined by NMR spectroscopy. The three-dimensional structure reveals six anti-parallel α-helices forming three HEAT-like repeats. Based on sequence similarity and structural homology with other proteins, MTH187 could play a role in molecular transport or in protein-protein interaction. However, it is also possible that MTH187 is an artefact or the result of the cleavage of a larger protein in the evolution of *Methanobacterium thermoautotrophicum*. Finally, potential structures for three paralogs were constructed using homology molecular modeling.

Acknowledgements

This project was supported by an NSERC grant. This project is part of the Canadian Structural Genomics Initiative (Cheryl Arrowsmith and Aled Edwards, Ontario Cancer Institute and the University of Toronto, Toronto, ON, Canada). We thank Dr. Brian D. Sykes, University of Alberta, for allowing the continuation of this project at Laval University. We thank Drs. Jeong Yong Suh and John Bagu, University of Alberta, for the expression and purification of the ^{15}N-labelled sample of MTH187. We thank Leigh Willard for maintenance of the computing infrastructure. We would like to thank the Canadian National High Field NMR Centre (NANUC) for the acquisition of a 3D ^{15}N-TOCSY HSQC at 500 MHz and a 3D ^{15}N-NOESY HSQC at 800 MHz. Operation of NANUC is funded by the Canadian Institutes of Health Research, the Natural Science and Engineering Research Council of Canada and the University of Alberta.

Electronic supplementary material is available at http://dx.doi.org/10.1007/s10858-006-0029-3

References

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W. and Bork, P. (2001) *J. Mol. Biol.*, **309**, 1–18.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. (2000) *Prog. Biophys. Mol. Biol.*, **73**, 339–345.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Holm, L. and Sander, C. (1993) *J. Mol. Biol.*, **233**, 123–138.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) *Nucl. Magn. Reson. Biol. Macromol. Pt. B*, **339**, 71–90.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291–325.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) *Proteins*, **11**, 281–296.

Notredame, C., Higgins, D.G. and Heringa, J. (2000) *J. Mol. Biol.*, **302**, 205–217.

Shindyalov, I.N. and Bourne, P.E. (1998) *Protein Eng.*, **11**, 739–747.

Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D. and Reeve, J.N. (1997) *J. Bacteriol.*, **179**, 7135–7155.

Vetter, I.R., Arndt, A., Kutay, U., Gorlich, D. and Wittinghofer, A. (1999) *Cell*, **97**, 635–646.

Williams, C.H., Stillman, T.J., Barynin, V.V., Sedelnikova, S.E., Tang, Y., Green, J., Guest, J.R. and Artymiuk, P.J. (2002) *Nat. Struct. Biol.*, **9**, 447–452.