Published in final edited form as:
J Struct Funct Genomics. 2013 March ; 14(1): 19–24. doi:10.1007/s10969-012-9148-0.

Solution NMR structure of the helicase associated domain BVU_0683(627–691) from *Bacteroides vulgatus* provides first structural coverage for protein domain family PF03457 and indicates domain binding to DNA

Jeffrey L. Mills,
Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA

Thomas B. Acton,
Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

Rong Xiao,
Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

John K. Everett,
Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

Gaetano T. Montelione, and
Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA, Department of Biochemistry, Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, USA

Thomas Szyperski
Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA

Jeffrey L. Mills: szypersk@buffalo.edu; Thomas Szyperski: szypersk@buffalo.edu

Abstract

A high-quality NMR structure of the helicase associated (HA) domain comprising residues 627–691 of the 753-residue protein BVU_0683 from *Bacteroides vulgatus* exhibits an all α-helical fold. The structure presented here is the first representative for the large protein domain family PF03457 (currently 742 members) of HA domains. Comparison with structurally similar proteins supports the hypothesis that HA domains bind to DNA and that binding specificity varies greatly within the family of HA domains constituting PF03457.

Correspondence to: Thomas Szyperski, szypersk@buffalo.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s10969-012-9148-0) contains supplementary material, which is available to authorized users.

Keywords

A6KY75_BACV8; BVU_0683; PF03457; Helicase associated domain; Structural genomics; SANT domain

Introduction

753-residue protein BVU_0683 from *B. vulgatus* (UniProt accession number A6KY75) is a putative helicase [1] and contains (1) a type III restriction enzyme domain (residues 105–196), (2) a helicase conserved C-terminal domain (residues 361–439), and (3) four sequentially located helicase associated (HA) domains (residues 501–562, 563–625, 627–689 and 690–751) which exhibit a pairwise sequence identity ranging from 38 to 47 %. The HA domains belong to PFam [2] protein family PF03457 and are predicted to bind DNA [3]. Notably, the presence of four consecutive HA domains in BVU_0683 may well reflect cooperative DNA binding.

PF03457 is a large family currently containing 742 members from a wide range of different bacteria and lower eukaryotes (algae and water molds). The domain BVU_0683(627–691) was selected as a target of the Protein Structure Initiative and assigned to the Northeast Structural Genomics Consortium (NESG; http://www.nesg.org) for structure determination (NESG Target ID BvR106A) as part of a cooperative inter-center effort aimed at providing structural coverage of large, uncharacterized protein domain families [4]. Initial structural representatives of such families exhibit high modeling leverage [5], expand our understanding of protein evolution [6], and generally expand our knowledge of fundamental relationships between protein sequences, three-dimensional structures, and protein function. The NMR structure of BVU_0683(627–691) presented here is the first atomic resolution structure for a member of the PF03457 domain family.

Materials and methods

BVU_0683(627–691) was cloned, expressed, and purified following protocols [7, 8] established by the NESG (see Supplementary Material and http://www.nmr2.buffalo. edu/nesg.wiki for details). The protein included a C-terminal hexaHis tag (LEHHHHHH) to facilitate protein prification. The corresponding pET expression vector (NESG BvR106A-627-691-21.2) has been deposited in the PSI Materials Repository (http://psimr.asu.edu/). Protein samples contained [*U*-^{13}C, ^{15}N]-labeled sample for NMR structure determination (concentration 1.1 mM in 90 % H_2O/10 % D_2O at pH 6.5 and containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 50 μM DSS, 0.02 % NaN_3). An isotropic overall rotational correlation time of about 5 ns was inferred from average ^{15}N spin relaxation times (see Supplementary Material and http://www.nmr2.buffalo.edu/nesg.wiki for details), indicating that the protein is monomeric in solution. This was confirmed with analytical gel-filtration (Agilent Technologies) and static light scattering (Wyatt Technology Co.) (Supplementary Fig. S1).

All NMR data were acquired at 25 °C on a Varian INOVA 750 MHz spectrometer equipped with a cryogenic ^{1}H{^{13}C,^{15}N} probe within 5.6 days. Nearly complete sequence-specific ^{1}H, ^{15}N and ^{13}C resonance assignments (Table 1; Supplementary Fig. S2) were obtained from G-matrix Fourier transform and conventional triple-resonance NMR experiments (Supplementary Material) using the programs AutoAssign 2.3.0 [9, 10] followed by manual assignment of side-chain resonances. Chemical shifts, NO-ESY peak lists, and time domain NMR data were deposited in the BioMagResBank (accession number 16692).

Structure calculations were performed by consensus analysis of ^{1}H-^{1}H NOE-derived upper limit distance constraints from automated NOESY assignment programs CYANA [11, 12] and AutoStructure 2.2.1 [13], and backbone dihedral angle constraints derived from chemical shifts using the program TALOS+ [14] for residues located in well-defined regular structure elements. Stereospecific assignments of methylene protons were performed with the GLOMSA module of CYANA and the final structure calculation was performed with CYANA followed by refinement of selected conformers in an "explicit water bath" [15] using the program CNS 1.2 [16]. Validation of the resulting 20 refined conformers was performed with the Protein Structure Validation Software (PSVS) server 1.4 [17] and the agreement of structures and NOESY peak lists was verified using the AutoStructure/RPF 2.2.1 package [18].

Results and discussion

We obtained a high-quality NMR structure of BVU_0683(627–691) (Fig. 1; Table 1). The coordinates were deposited in the Protein Data Bank [19] on 1/26/2010 (accession code 2KTA). The structure exhibits an all α-helical fold (Fig. 1) consisting of three α-helices: I (633–646), II (658–672), and III (677–685). α-helices I and II are oriented nearly perpendicular to each other with the short C-terminally located α-helix III being tilted approximately 45° relative to α-helix II. While the HA domain family is predicted to be structurally conserved [3], the pairwise sequence identity between BVU_0683(627–691) and the other members of PF03457 ranges from 19 to 100 % (Fig. 2a, b).

BVU_0683(627–691) represents the third HA domain of BVU_0683 and a search of the PDB database for similar structures using the program DALI [20] yielded 29 structurally similar proteins albeit all with rather marginal Z-scores between 3.0 and 4.0. The best scoring hit [Z-score 4.0, root mean square deviation (RMSD) of C^{α} atoms = 4.6 Å for 63 aligned residues with 10 % sequence identity] is with the X-ray crystal structure of the cytoplasmic domain of human transmembrane receptor plexin B1 (3HM6) [21]. However, given (1) the different functions of BVU_0683 and plexin B1, (2) the rather large RMSD of the structurally aligned polypeptide segments, and (3) the insignificant sequence identity, there is no indication that BVU_0683(627–691) and the cytoplasmatic plexin B1 domain are homologues.

Among the remaining DALI hits, only seven proteins (after removing redundant chains from the same PDB entries) had RMSD values ≤3.5 Å for the structurally aligned residues. Of those, five proteins are known to bind DNA [that is, the small subunit of gp1 bacteriophage Sf6 terminase and four SANT (Swi3, Ada2, N-Cor, and TFIIIB) domains from the ISW (Imitation Switch) complex] while the remaining two (herpes virus membrane fusion regulator and a component of a transcriptional co-activator and nuclear pore binding complexes) do not bind DNA. The lowest RMSD value of all DALI hits was obtained for the N-terminal domain of the DNA-recognition component gp1 of bacteriophage Sf6 terminase (best match with chain B of 3HEF: Z-score 3.0, RMSD of C^{α} atoms = 2.4 Å for 47 aligned residues with 13 % sequence identity; Fig. 2c, d). The N-terminal region comprising residues 14–62 [22] contains three α-helices that are quite similarly oriented as those in BVU_0683 (Fig. 2d). Moreover, Zhao et al. [23] have shown that this region binds DNA and modeled the structure of the gp1-DNA complex. According to this model, Asp 19, Asp 20 and Ser 23 of gp1 convey specific binding to the DNA major groove, and Lys 33, Lys 43, and Lys 62 contact the negatively charged phosphate backbone. Intriguingly, in a structure based sequence alignment (Fig. 2c, d) these residues match up with, respectively, Glu 639, Glu 640, and Ser 643, and Lys 653, Lys 668, and Lys 685 in BVU_0683(627–691) (see also resulting electrostatic surface potential; Fig. S3). This finding provides strong support for the predicted DNA binding function of HA domains [3]. However, these

residues are conserved in only seven out of 742 sequences of PF03457 (all of which belong to the genus bacteroides or parabacteroides) and, in particular, they are *not* conserved in any of the other three HA domains of full length BVU_0683 (Fig. 2a). This indicates that the four HA domains of BVU_0683 bind to DNA differently, and that HA domains in general are rather promiscuous DNA binding domains. Furthermore, this view is also supported by the structural comparison of BVU_0683(627–691) with the SANT domain of the chromatin remodeling factor ISW1a(ΔATPase)-DNA complex [24] (best match with chain A of 2Y9Z: Z-score 3.2, RMSD of C^{α} atoms = 3.3 Å for 51 aligned residues with 6 % sequence identity): a structure based sequence alignment of the two domains (Fig. 2c, e) reveals that residues involved in the DNA binding of the SANT domain are not conserved in the HA domain of BVU_0683(627–691). Consistently, Yamada et al. [24]. have shown that the SANT domain of ISW1a does not exhibit pronounced DNA sequence preference for binding. Moreover, the SANT domain interacts through the minor groove while the small subunit of gp1 (see above) binds to the major groove. Inspection of the electrostatic surfaces (Figure S3) shows higher similarity between BVU_0683(627–691) and the gp1 domain when compared with BVU_0683(627–691) and the SANT domain, which suggests that BVU_0683(627–691) might likewise bind to the major groove of a B-DNA helix. In addition, since neither functionally important residues nor core residues are conserved (Fig. 2c, e) the structural similarities between BVU_0683(627–691) and the SANT domain may well be a result of convergent evolution.

The two remaining DALI hits [sus1 (best match with chain D of 3KIK: Z-score 3.1, RMSD of C^{α} atoms = 3.4 Å for 47 aligned residues with 4 % sequence identity) and herpes membrane fusion regulator (best match with chain A of 3M1C: Z-score 3.2, RMSD of C^{α} atoms = 3.5 Å for 53 aligned residues with 4 % sequence identity)] are rather likely not homologues of BVU_0683(627–691). First, the putative DNA binding function of BVU_0683(627–691) is different from the function of these two protein domains exhibiting remote structural similarity. Second, structural alignment of sus1 and BVU_0683(627–691) reveals that α-helix II of BVU_0683 and its sus1 counterpart are rotated nearly 45° relative to each other. Similarly, structural alignments of the herpes virus membrane fusion regulator and BVU_0683(627–691) exhibits significant differences when considering the relative spatial orientation of α-helix III. Finally, identification of modeling families reveals that the novel modeling leverage [5, 25], i.e., the number of protein structures that can be modeled using the experimental structure presented here, is ~ 350. Thus, considering that PF03457 contains 687 unique sequences, the NMR structure presented here provides ~50 % structural coverage of the large family PF03457.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Belote, C. Ciccosanti, K. Hamilton, and G.V.T. Swapna for helpful discussions and technical support. This work was supported by the National Institutes of Health, grant number: U54 GM094597 (T.S. and G.T.M.).

References

1. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007; 76:23–50. [PubMed: 17506634]
2. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. PFam: clans, web tools and services. Nucleic Acids Res. 2006; 34:247–251.

3. Yeats C, Bentley S, Bateman A. New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. BMC Microbiol. 2003; 3:3. [PubMed: 12625841]
4. Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C. PSI-2: structural genomics to cover protein domain family space. Structure. 2009; 17:869–881. [PubMed: 19523904]
5. Liu J, Montelione GT, Rost B. Novel leverage of structural genomics. Nat Biotechnol. 2007; 25:849–851. [PubMed: 17687356]
6. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995; 247:536–540. [PubMed: 7723011]
7. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GV, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. Methods Enzymol. 2005; 394:210–243. [PubMed: 15808222]
8. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. The high-throughput protein sample production platform of the Northeast structural genomics consortium. J Struct Biol. 2010; 172:21–33. [PubMed: 20688167]
9. Moseley HN, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol. 2001; 339:91–108. [PubMed: 11462827]
10. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol. 1997; 269:592–610. [PubMed: 9217263]
11. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol. 1997; 273:283–298. [PubMed: 9367762]
12. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]
13. Huang YJ, Moseley HN, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione GT. An integrated platform for automated analysis of protein NMR structures. Methods Enzymol. 2005; 394:111–141. [PubMed: 15808219]
14. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009; 44:213–223. [PubMed: 19548092]
15. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins. 2003; 50:496–506. [PubMed: 12557191]
16. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macro-molecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54:905–921. [PubMed: 9757107]
17. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007; 66:778–795. [PubMed: 17186527]
18. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc. 2005; 127:1665–1674. [PubMed: 15701001]
19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
20. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci. 1995; 20:478–480. [PubMed: 8578593]

21. Tong Y, Hota PK, Penachioni JY, Hamaneh MB, Kim S, Alviani RS, Shen L, He H, Tempel W, Tamagnone L, Park HW, Buck M. Structure and function of the intracellular region of the plexin-b1 transmembrane receptor. PDB ID: 3HM6. 200910.2210/pdb3hm6/pdb
22. Zhao H, Finch CJ, Sequeira RD, Johnson BA, Johnson JE, Casjens SR, Tang L. Crystal structure of the bacteriophage Sf6 terminase small subunit. PDB ID: 3HEF. 200910.2210/pdb3hef/pdb
23. Zhao H, Finch CJ, Sequeira RD, Johnson BA, Johnson JE, Casjens SR, Tang L. Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. Proc Natl Acad Sci USA. 2010; 107:1971–1976. [PubMed: 20133842]
24. Yamada K, Frouws TD, Angst B, Fitzgerald DJ, DeLuca C, Schimmele K, Sargent DF, Richmond TJ. Structure and mechanism of the chromatin remodelling factor ISW1a. Nature. 2011; 472:448–453. [PubMed: 21525927]
25. Nair R, Liu J, Soong TT, Acton TB, Everett JK, Kouranov A, Fiser A, Godzik A, Jaroszewski L, Orengo C, Montelione GT, Rost B. Structural genomics is the largest contributor of novel structural leverage. J Struct Funct Genomics. 2009; 10:181–191. [PubMed: 19194785]
26. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]
27. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004; 5:113. [PubMed: 15318951]
28. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22:1658–1659. [PubMed: 16731699]
29. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010; 26:680–682. [PubMed: 20053844]
30. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; 38:W529–W533. [PubMed: 20478830]
31. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27:221–224. [PubMed: 19854763]
32. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477–486. [PubMed: 9008363]
33. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: φ, ψ and Cβ deviation. Proteins. 2003; 50:437–450. [PubMed: 12557186]
34. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992; 356:83–85. [PubMed: 1538787]
35. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993; 17:355–362. [PubMed: 8108378]
36. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR. 2004; 28:341–355. [PubMed: 14872126]

Abbreviations

DNA	Deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonate sodium salt
DTT	Dithiothreitol
HA	Helicase associated
ISW	Imitation switch
MES	2-(*N*-morpholino)ethanesulfonic acid

NESG	Northeast structural genomics consortium
NOE	Nuclear overhauser effect
PDB	Protein data bank
RMSD	Root mean square deviation
SANT	SWI3, ADA2, N-CoR, and TFIIIB

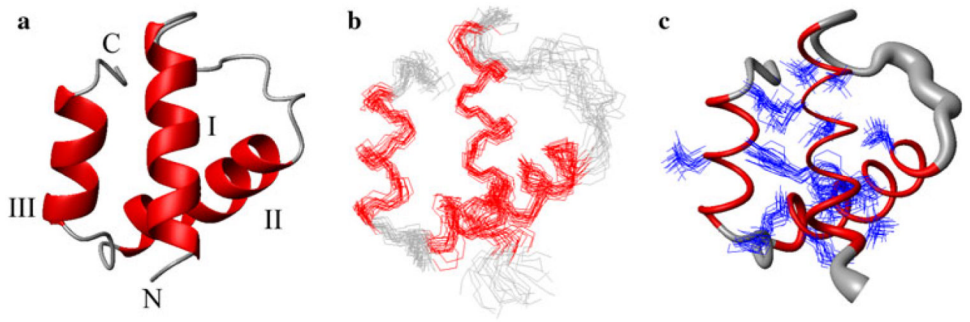

Fig. 1.

NMR Structures of BVU_0683(627–691) (only residues 631–689 are shown for clarity) **a** *Ribbon drawing* of the lowest energy conformer. The α-helices are shown in *red* while other polypeptide segments are shown in *grey*. The N- and C-termini are labeled with "N" and "C", respectively and the *helices* are indicated as *I–III*. **b** *Line* representations of the polypeptide backbone depicting the ensemble of 20 structures after superposition of the backbone heavy atoms of the regular secondary structure elements for minimal RMSD. **c** Sausage representation of backbone and superposition of the core side chains (Table 1). A *spline curve* was drawn through the mean positions of C^{α} atoms with the thickness proportional to the mean global displacement of backbone heavy atoms in the 20 conformers superimposed in **b**

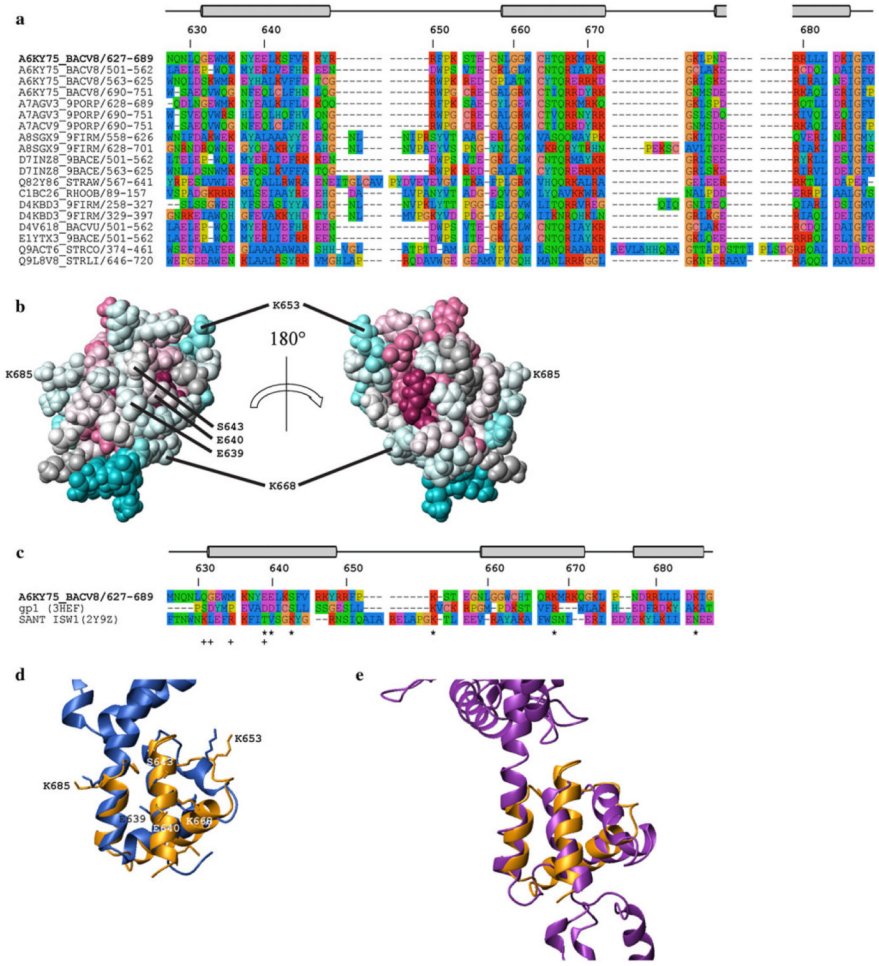

Fig. 2.

a Multiple sequence alignments of selected members from PF03457 generated using the program MUSCLE [26, 27]. The *top* four *rows* represent the four HA domains of protein BVU_0683 and the remaining 16 contain sequences which are representatives of a PF03457 sub-set defined by having <40 % pairwise identity (generated with CD-HIT [28, 29]). Sequences are labeled with their corresponding UniProt accession codes/residue numbers and the code of BVU_0683(627–691) is in *bold*. Residue numbers and regular secondary structure elements are relative to BVU_0683(627–691) **b** *Space-filling* representation of the lowest-energy conformer of BVU_0683(627–689). The degree of residue conservation within protein family PF03457 is *color-coded* according to the default ConSurf [30] scheme: burgundy for the strongest conservation and *cyan* for the highest variability. The *image* on the *left* has the same orientation as in Fig. 1 and the *image* on the *right* is a 180° rotation along the *vertical axis.* **c** Structure-based sequence alignment of BVU_0683(627–689), N-terminal domain of gp1, and SANT domain generated with DALI [20]. The residue numbers and positions of the α-helices of BVU_0683(627–689) are indicated at the *top,* and residues involved in DNA binding are *labeled* below the aligned sequences with (*) for the N-terminal domain of gp1, and with (+) for the SANT domain (see text). Figs. **a, c** were generated with the program Seaview v4.3.0 [31] using the default *color* scheme: *red* = Arg, Lys (*positive*)*; purple* = Asp, Glu (*negative*)*; dark blue* = Ala, Ile, Leu, Met, Phe, Trp, Val (hydrophobic); *green* = Asn, Gln, Ser, Thr (polar); *cyan* = His, Tyr; *yellow* = Pro; and *orange* = Gly. Structural alignment of the lowest energy conformer of BVU_0683(627–691)

(2KTA, *orange*) with polypeptide segments in the crystal structures comprising **d** the N-terminal domain of gp1 (3HEF:B, residues 14–64, *blue*) and **e** the SANT domain of ISW1a(ΔATPase)-DNA complex (2Y9Z:A, residues 888–951 *purple*). In b and **d**, the residues that are conserved in BVU_0683(627–691) and gp1, and which are implicated in DNA binding, are indicated. The orientation of BVU_0683(627–691) is the same as in Fig. 1

Table 1
BVU_0683(627–691) structure statistics

Completeness of resonance assignments[a] [%]	
Backbone/side-chain	97.8/98.5
Completeness of stereospecific assignments[b] [%]	
Val & Leu isopropyl/$\beta CH_2/\alpha CH_2$ of Gly	30/54/17
Conformation-restricting distance constraints	
Intra [i = j]	308
Sequential [|i−j| = 1]	396
Medium range [1 < |i−j| < 5]	336
Long range [|i−j| ≥ 5]	232
Total	1,272
Dihedral angle constraints (φ/ψ)	35/35
Distance constraints per residue (of those, long-range)	19.3 (3.5)
CYANA target function [$Å^2$]	0.38 ± 0.08
Average number of distance constraint violations per conformer [Å]	
0.2–0.5	0.0
>0.5	0.0
Average number of dihedral angle constraint violations per conformer	
>10	0.0
Average RMSD from mean coordinates [Å]	
Regular secondary structure elements[c], backbone heavy atoms	0.67 ± 0.14
Regular secondary structure elements[c], all heavy atoms	1.36 ± 0.15
Ordered residues[d], backbone heavy atoms	1.07 ± 0.19
Ordered residues[d], all heavy atoms	1.70 ± 0.16
Heavy atoms of molecular core including best-defined side chains[e]	0.81 ± 0.10
Global quality scores[f] (raw/Z-score)	
PROCHECK [32] G-factor (φ and ψ)	0.4/1.7
PROCHECK [32] G-factor (all dihedral angles)	−0.1/−0.5
MOLPROBITY [33] clash score	28.6/−3.4
Verify3D [34]	0.3/−2.7
ProsaII [35]	0.3/−1.5
RPF scores [18] [%]	
Recall/Precision/F-Measure	0.984/0.905/0.943
DP-score	0.846
Ramachandran plot summary[d] from Molprobity [33] (%)	
Most favored regions	99.2
Allowed regions	0.8
Disallowed regions	0.0

[a]Calculated with the AVS suite [36] excluding C-terminal tags, N-terminal residue and Lys and Arg side chain amino groups, hydroxyl of Ser, Thr and Tyr, carboxyls of Asp and Glu, and non-protonated aromatic carbons

[b] Relative to pairs with non-degenerate chemical shifts

[c] Residues 633–646, 658–672, and 677–685

[d] Residues 633–648, 658–690

[e] Residues 637, 638, 641–643, 645, 652, 662, 665, 666, 669, 675, 676, 681, 686

[f] Calculated with PSVS 1.4 [17]