J Biomol NMR (2007) 38:353–358
DOI 10.1007/s10858-007-9172-8

NMR STRUCTURE NOTE

Hypothetical protein AF2241 from *Archaeoglobus fulgidus* adopts a cyclophilin-like fold

Xuanjun Ai · Lei Li · Anthony Semesi · Adelinda Yee · Cheryl H. Arrowsmith · Shawn S. C. Li · Wing-Yiu Choy

Received: 21 March 2007 / Accepted: 29 May 2007 / Published online: 4 July 2007

Abstract AF2241 is a hypothetical protein from *Archaeoglobus fulgidus* and it belongs to the PFam domain of unknown function 369 (DUF369). NMR structural determination reveals that AF2241 adopts a cyclophilin-like fold, with a β-barrel core composed of eight β-strands, one α-helix, and one 3_{10} helix located at each end of the barrel. The protein displays a high structural similarity to TM1367, another member of DUF369 whose structure has been determined recently by X-ray crystallography. Structural similarity search shows that AF2241 also has a high similarity to human cyclophilin A, however, sequence alignment and electrostatic potential analysis reveal that the residues in the PPIase catalytic site of human cyclophilin A are not conserved in AF2241 or TM1367. Instead, a putative active site of AF2241 maps to a negatively charged pocket composed of 9 conserved residues. Our results suggest that although AF2241 adopts the same fold as the human cyclophilin A, it may have distinct biological function.

Keywords *Archaeoglobus fulgidus* · Cyclophilin-like fold · DUF369 domain · NMR · Protein structure · Structural proteomics

Biological context

AF2241 is a hypothetical protein from *Archaeoglobus fulgidus*, a sulphur-metabolizing organism grows at extremely high temperatures. The 131-residue protein AF2241 has been classified as a member of the DUF369 (PFam 04126) family, a protein family composed of mainly hypothetical proteins from bacteria and archaea (Bateman et al. 2004), and these proteins share sequence identities varying from 31% to 57%. Recently, the structure of TM1367 from *Thermotoga maritima*, another member of the DUF369 family, has been determined by X-ray crystallography (Jin et al. 2006). This protein shows significant structural similarity to human cyclophilin A (hCypA) that functions as a peptidylprolyl isomerase (PPIase, Takahashi et al. 1989) and is known to bind to calcineurin in the form of a hCypA-CsA (cyclosporine A) complex (Huai et al. 2002; Jin et al. 2006). However, structural comparison of TM1367 and hCypA reveals that the active site of hCypA is not conserved in TM1367 (Jin et al. 2006). To date, no functional annotation has been made for the DUF369 proteins.

In this work, we determined the solution structure of AF2241 using nuclear magnetic resonance (NMR) spectroscopy as part of a structural proteomics effort to understand the function of proteins in the DUF369 family. Similar to the X-ray structure of TM1367, the solution structure of AF2241 adopts a cyclophilin-like fold that is characterized by an eight-strand closed β-barrel, an α-helix, and a 3_{10} helix. Sequence alignment and electrostatic

Electronic supplementary material The online version of this article (doi:10.1007/s10858-007-9172-8) contains supplementary material, which is available to authorized users.

X. Ai · L. Li · S. S. C. Li (✉) · W.-Y. Choy
Department of Biochemistry, The University of Western Ontario, London, ON, Canada N6A 5C1
e-mail: sli@uwo.ca

W.-Y. Choy (✉)
e-mail: jchoy4@uwo.ca

A. Semesi · A. Yee · C. H. Arrowsmith
Ontario Centre for Structural Proteomics, University Health Network, Toronto, ON, Canada M5G 2C4

potential analysis indicate that the active site of AF2241 may locate in a negatively charged pocket composed of nine conserved residues.

Methods and results

Protein expression and purification

The *AF2241* gene was subcloned from genomic DNA into a pET derived vector with N-terminal hexa-histidine followed by a TEV protease cleavage site (with the sequence MGTS(H)6SSGRENLYFQGH). The expression vector was transformed into the *Escherichia coli* strain BL21-Gold(DE3). Cells were grown at 37°C in M9 minimal medium containing ^{13}C-glucose and $^{15}NH_4Cl$ and supplemented with $ZnCl_2$, thiamine, and biotin. Protein expression was induced by 1 mM IPTG at OD600 ~1.0. The cells were then grown overnight at 15°C before harvesting. Frozen cell pellets were thawed and lysed by sonication in 500 mM NaCl, 20 mM Tris, 5 mM imidazole at pH 8. The His-tagged AF2241 was extracted from the soluble fraction by batch Ni^{2+} affinity chromatography, washed with 500 mM NaCl, 20 mM Tris, 30 mM imidazole (pH 8) to remove contaminants, and eluted with 500 mM imidazole in the same buffer. The purified proteins were concentrated, and exchanged into the final NMR buffer containing 10 mM MOPS, 450 mM NaCl, 10 μM Zn^{2+}, 10 mM DTT, 1 mM benzamidine, 1 × inhibitor mixture (Roche Molecular Biochemicals) and 0.01% NaN_3 at pH 6.5. The 22-residue non-native leader sequence was retained in the NMR sample. Result of size-exclusion chromatography confirms that AF2241 is monomeric in solution.

NMR spectroscopy

Three-dimensional heteronuclear NMR experiments were performed on a uniformly ^{13}C-, ^{15}N-labeled AF2241 sample. Spectra were acquired with a Varian Inova 600 MHz spectrometer at 25°C. Backbone and side-chain chemical shift assignments were obtained using the following heteronuclear NMR experiments: HNCACB, CBCA(CO)NH, C(CO)NH, HNCO, HCCH-TOCSY, HC(CO)NH. In addition, two-dimensional (HB)CB(CGCD)HD and (HB)CB (CGCDCE)HE were recorded for the assignment of aromatic ring resonances. NOE distance restraints were derived from the ^{13}C NOESY-HSQC and ^{15}N NOESY-HSQC NMR data recorded with a mixing time of 80 ms. All data sets were processed using NMRPipe (Delaglio et al. 1995) and the spectra were analyzed with NMRView 5.2.2 (Johnson and Blevins 1994). A nearly complete backbone and side-chain chemical shift assignments have been obtained with the exception of residues 1–3 (N-terminal residues), 80–81, 92, 95–96, 107 and 113–116, for which most of them are in the loop or turn regions in the structure that we calculated. The data has been deposited in BioMagResBank (BMRB) with the accession number 7271 (Ai et al. 2007).

NMR structure calculation

Structure calculation was performed using CNS version 1.1 (Brunger et al. 1998). Experimental restraints used in the NMR structure calculation of AF2241 are summarized in Table 1. A total of 1313 non-redundant NOE distance restraints (458 intra-residue, 361 sequential, 129 medium-range, and 365 long-range) were used together with 150 dihedral angle restraints derived from the assigned ^{15}N, $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C'$ chemical shifts using the programs TALOS (Cornilescu et al. 1999) and CSI (Wishart et al. 1994). The qualities of the 20 lowest

Table 1 Structural statistics for the 20 lowest energy structures of AF2241

Distance restraints	
All	1313
Intra-residue	45
Sequential ($\|i-j\| = 1$)	361
Medium range ($1 < \|i-j\| \leq 4$)	129
Long range ($\|i-j\| > 4$)	365
Dihedral angle restraints	
All (ϕ,ψ)	150 (75, 75)
Deviation from idealized geometry	
Bonds (Å)	0.0029 ± 0.00005[a]
Angles (°)	0.458 ± 0.0046[a]
Impropers (°)	0.301 ± 0.0056[a]
Ramachandran map analysis[b]	
Most favored regions (%)	65.4 ± 2.0[b]
Additional allowed regions (%)	27.7 ± 2.3[b]
Generously allowed regions (%)	4.4 ± 1.0[b]
Disallowed regions (%)	2.5 ± 0.8[b]
Pairwise atomic rmsd (Å)	
Backbone atoms (residues 8–130)	0.56 ± 0.14
Heavy atoms (residues 8–130)	1.33 ± 0.16
Backbone atoms (residues 8–90, 98–111, 117–130)	0.32 ± 0.05
Heavy atoms (residues 8–90, 98–111, 117–130)	1.08 ± 0.09
NOE violations (>0.2 Å)	0[c]
Dihedral angle violations (>2°)	0[c]

[a] The averaged deviations from the idealized geometry of the 20 lowest energy structures evaluated by CNS

[b] Calculated with PROCHECK-NMR (Laskowski et al. 1996)

[c] The total number of violations in the 20 lowest energy structures

energy structures (out of 200 structures calculated) were analyzed using PROCHECK-NMR (Laskowski et al. 1996) and the resulting structural statistics is reported in Table 1. All experimental restraints are satisfied in this final set of structures, which all have good covalent geometry. Backbone dihedral angle analysis using the program PROCHECK-NMR (Laskowski et al. 1996) indicates that, on average, 97.5% of the residues are in the most favored or allowed regions of the Ramachandran plot. The averaged pairwise RMSD between structures in this final set are 0.56(±0.14) Å and 1.33(±0.16) Å for backbone and heavy atoms, respectively. It should be noted that the first 7 residues (Met1-Glu7) from the N-terminus and the C-terminal Ala131 were excluded in the quality evaluation of the structures due to their flexibility and the limited numbers of NOEs that have been assigned to these residues. Besides the N-terminus, there are two other regions, namely Thr91-Lys97 and Asn112-Gly116, which also show relatively poor convergence in the calculated structures (Fig. 1a). The former is part of the loop connecting two β-strands (β6 and β7, see below), while the latter includes residues that are unassigned (Ser113-Glu115) or have no observed/assigned NOEs. Atomic coordinates of the calculated structures and the experimental restraints used have been deposited in the Protein Data Bank (PDB 2NNZ).

Solution structure of AF2241

Figure 1a shows the superimposition of the 20 lowest energy structures (selected out of 200 calculated structures). The solution structures determined by NMR reveal that the 131-residue AF2241 adopts a cyclophilin-like fold. The protein is composed of a closed β-barrel with eight β-strands, an α-helix, and a 3_{10} helix. As shown in Fig. 1b, the eight β-strands of the closed barrel are arranged in the order of 1-2-7-5-6-4-3-8, including Leu10-Phe14 (β1), Ala17-Val21 (β2), Ile41-Trp48 (β3), Glu51-Phe54 (β4), Asp74-Trp78 (β5), Ala83-Phe87 (β6), Val104-Arg109 (β7) and Glu124-Glu129 (β8). The α-helix (Glu30-Asp37) and the 3_{10} helix (Leu117-Gly119) are located at either end of the barrel between (β2, β3) and (β7, β8), respectively, and are denoted by the symbols α and 3_{10} in Fig. 1b accordingly.

Discussion and conclusions

Protein structural comparison

To search for the structural homologues of AF2241, the representative of the final set of structures were compared to those in the Protein Data Bank (PDB) using the DALI server (http://www.ebi.ac.uk/dali/index.html, Holm et al.

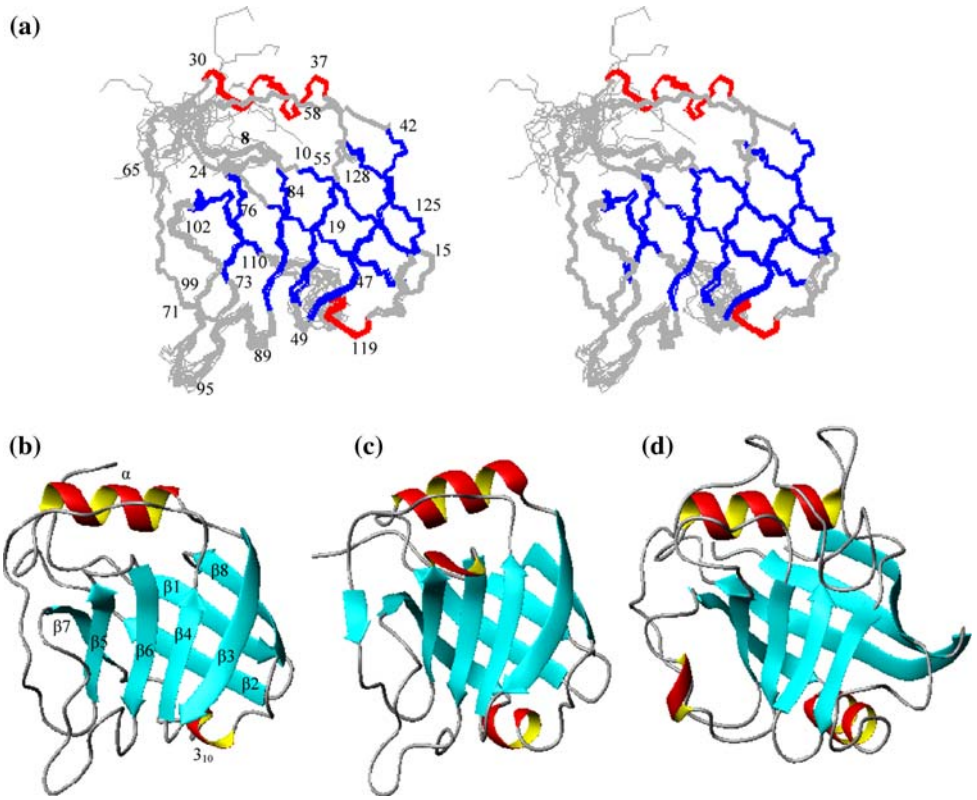

Fig. 1 (**a**) Stereoview of the superimpositions of the 20 lowest energy structures of AF2241. The numeric labels denote residue numbers. (**b**) Ribbon representation of the representative structure of AF2241. The secondary structure elements were labeled according to their definition in the text. (**c**) and (**d**) are the ribbon representations of the X-ray structures of TM1367 (PDB 1ZX8-A) and human cyclophilin A (PDB 2CPL), respectively. The Figure was generated using the program MOLMOL (Koradi et al. 1996)

1995). The search reveals that the hypothetical protein TM1367 from *T. maritima* (PDB 1ZX8-A; DALI Z-score 16.4; Jin et al. 2006) and human cyclophilin A (hCypA; PDB 2CPL; DALI Z-score 8.8; Takahashi et al. 1989) are the two nearest structural neighbors of AF2241. Ribbon representations of these two protein structures are shown in Fig. 1c, d, respectively, and their superimpositions with AF2241 are illustrated in Fig. S1 in the Supplementary Material. The fact that TM1367 is among the top hits in the DALI search for structural homologues of AF2241 is expected since these two hypothetical proteins belong to the same DUF369 family and share a 55% sequence identity. Structural alignment of AF2241 and TM1367 shows a RMSD of 2.0 Å (over 125 C_α atoms) in the rigid-

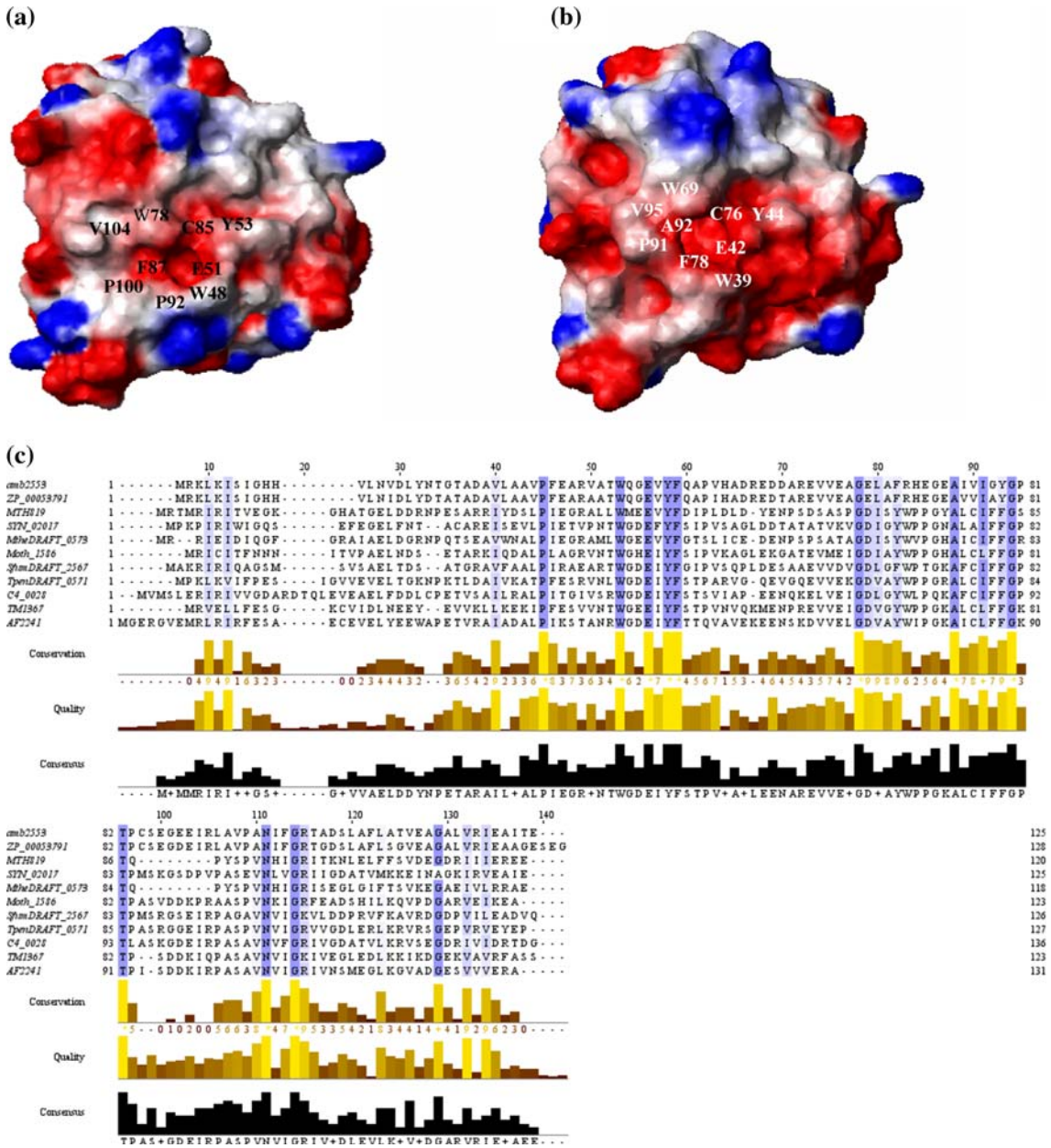

Fig. 2 Electrostatic potential surfaces of (**a**) AF2241 and (**b**) chain C of TM1367 in the same orientation as the ribbon diagrams in Fig. 1b, c, respectively. Negative and positive charges were shown in red and blue, respectively, while the neutral surface was displayed in white. The surfaces were calculated using the program MOLMOL (Koradi et al. 1996). (**c**) Sequence alignments of the 11 proteins in the DUF369 family based on the PSI-BLAST analysis. Multiple sequence alignment was performed using the program ClustalX (Thompson et. al. 1994) and the diagram was generated using Jalview (Clamp et al. 2004). "Conservation" is a numerical index calculated by Jalview base on the method proposed by Livingstone and Barton (Livingstone and Barton 1993). It reflects the conservation of physiochemical properties in the alignment. Highly conserved residues (conservation ≥9; * and + indicate fully conserved and a conservation between 9 and fully conserved, respectively) are highlighted in the sequences

body superimposition of these two structures. On the other hand, despite the low sequence identity (11%) between hCypA and AF2241, the superimposition of these two structures (over 117 C_α atoms) has a RMSD of only 2.9 Å. The high structural similarity between AF2241, TM1367, and hCypA agrees with the classification of AF2241 (as well as TM1367) to the cyclophilin-like fold by the Structural Classification of Proteins Server (SCOP; http://scop.mrc-lmb.cam.ac.uk/scop/; Murzin et al. 1995).

Potential active site and function

It is well known that hCypA has a function of a peptidyl-prolyl isomerase (PPIase, E.C. 5.2.1.8) (Takahashi et al. 1989). In the presence of cyclosporin (CsA), hCypA can also bind to calcineurin to suppress it from dephosphorylating the transcription factor nuclear factor of activated T cells (Huai et al. 2002). Detailed sequence analysis and structural comparison of TM1367 and hCypA performed by Jin et al. (Jin et al. 2006) revealed that the residues comprised of the PPIase catalytic site and the calcineurin binding site of hCypA were either not conserved or no corresponding residues were identified in TM1367. Structural comparison of AF2241 and hCypA leads to a similar conclusion: nearly all residues in hCypA that are involved in the binding to calcineurin or the PPIase activity are absence in AF2241. Our data together with the result reported by Jin et al. (Jin et al. 2006) strongly suggest that the proteins in the DUF369 family may have very distinct function from cyclophilin A.

Figure 2c shows the sequence alignment of proteins in the DUF369 family. "Conservation" is a numerical index calculated by Jalview (Clamp et al. 2004) based on the method proposed by Livingstone and Barton (Livingstone and Barton 1993). It reflects the conservation of physiochemical properties of residues in the alignment. Based on this calculation, 23 residues are found to be highly conserved (with conservation value ≥9) in AF2241 and 8 of them are on the surface of the protein.

The CASTp (http://sts.bioengr.uic.edu/castp/; Binkowski et al. 2003) server was used to search for potential binding pockets on the surfaces of AF2241 and TM1367. This analysis, together with the multiple sequence alignment suggest that three highly conserved residues, namely Trp48, Glu51 and Tyr53, and an additional six residues (Trp78, Cys85, Phe87, Pro92, Pro100, and Val104) with conservation values between 5 and 8 form a negatively charged pocket on the surface of AF2241 (Fig. 2a). As shown in Fig. 2b, eight of these nine residues are identified in TM1367 (i.e., Trp39, Glu42, Tyr44, Trp69, Cys76, Phe78, Pro91 and Val95) that form a similar pocket with residue Ala92 (conservation value of 6). Despite the lack of conservation in active site residues between hCypA and AF2241 or TM1367, it is interesting that the putative active site of the latter protein superimposes with the PPIase catalytic site of hCypA (Jin et al. 2006).

By using a comparative genomic method that combines information from phylogenetic profiling, chromosomal proximity and domain fusion, Yanai and DeLisi predicted that TM1367 is involved in energy production based on its position in a functional network (Yanai et al. 2002). To investigate whether AF2241 has a similar function, we used the STRING server (release 6.3; http://string.embl.de; von Mering et al. 2003), another bioinformatics method, to predict functional associations of AF2241 with other proteins. STRING is a database of known and predicted functional associations between proteins. The information is derived from high-throughput interaction and functional genomics data, as well as from text mining of the literature. Based on the information retrieved from STRING, *AF2241* forms a four-gene clique with *AF2242*(*PurB*), *AF2240*, and *AF2243*. Positioning of AF2241 in the same functional network as PurB (adenylosuccinate lyase, EC 4.3.2.2) and AF2243 (3-ketoacyl-CoA thiolase E.C. 2.3.1.16) suggests that it may have a function related to these two enzymes. It should be pointed out that due to the absence of experimental data of AF2241 in the literature, the functional association predicted is based solely on the genomic neighborhood information. Therefore, the confidence levels are not high. Further investigation using enzymatic assays should help to establish the function of AF2241 and to confirm the location of its active site.

Supplementary material

The file contains two figures. Figure S1a(b) shows the structural superimpositions of AF2241 and TM1367 (human cyclophilin A) in ribbon representations. Figure S2 shows portions of the functional networks of AF2241, TM1367, and MTH819 predicted by STRING.

Acknowledgements We would like to thank Drs. Hsiu-Ju Tseng and Matthew Revington for helpful discussions. This work was supported by the Ontario Research Development Challenge Fund (to CHA and SSCL), Genome Canada, and a CIHR resource Grant.

References

Ai X, Semesi A, Yee A, Arrowsmith CH, Li SSC, Choy WY (2007) Backbone and side chain ^{1}H, ^{13}C, and ^{15}N resonance assignments of AF2241 from *Archaeoglobus fulgidus*. J Biomol NMR 38:183

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst D54:905–921

Binkowski TA, Naghibzadeg S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 31:3352–3355

Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. Bioinformatics 20:426–427

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Holm L, Sander C (1995) Dali – a network tool for protein-structure comparison Trends Biochem Sci 20:478–480

Huai Q, Kim HY, Liu Y, Zhao Y, Mondragon A, Liu JO, Ke H (2002) Crystal structure of calcineurin-cyclophlin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes. Proc Natl Acad Sci USA 99:12037–12042

Jin KK, Krishna SS, Schwarzenbacher R et al (2006) Crystal structure of TM1367 from Thermotoga maritima at 1.90 Å resolution reveals an atypical member of the cyclophilin (Peptidylprolyl Isomerase) fold. Proteins 63:1112–1118

Johnson BA, Blevins RA (1994) NMRView: A computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structure. J Mol Graph 14:51–55

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

Livingstone CD, Barton GJ (1993) Protein sequence alignments: A strategy for the hierarchical analysis of residue conservation. CABIOS 9:745–756

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Takahashi N, Hayano T, Suzuki M (1989) Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. Nature 337:473–475

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) String: a database of predicted functional associations between proteins. Nucleic Acids Res 31:258–261

Wishart DS, Sykes BD (1994) The 13C chemical-shift index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:171–180

Yanai I, DeLisi C (2002) The society of genes: networks of functional links between from comparative genomics. Genome Bio. 3, research0064.1–0064.12