BMC Biology

Research article

Structure of GlgS from *Escherichia coli* suggests a role in protein–protein interactions

Guennadi Kozlov[1], Demetra Elias[1], Miroslaw Cygler[2] and Kalle Gehring*[1]

Address: [1]Department of Biochemistry, McGill University, Montreal, Quebec H3G 1Y6, Canada and [2]Macromolecular Structure Group, Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada

Email: Guennadi Kozlov - guennadi.kozlov@mcgill.ca; Demetra Elias - demetra_elias@yahoo.com; Miroslaw Cygler - mirek.cygler@bri.nrc.ca; Kalle Gehring* - kalle.gehring@mcgill.ca

* Corresponding author

Abstract

Background: The *Escherichia coli* protein GlgS is up-regulated in response to starvation stress and its overexpression was shown to stimulate glycogen synthesis.

Results: We solved the structure of GlgS from *E. coli*, a member of an enterobacterial protein family. The protein structure represents a bundle of three α-helices with a short hydrophobic helix sandwiched between two long amphipathic helices.

Conclusion: GlgS shows structural homology to Huntingtin, elongation factor 3, protein phosphatase 2A, TOR1 motif domains and tetratricopeptide repeats, suggesting a possible role in protein–protein interactions.

Background

In response to insufficient levels of nitrogen or other nutrients, cells produce and accumulate glycogen when a carbon source is present in the growth medium [1]. The general steps of glycogen synthesis are initiation (priming), elongation and maturation. The priming step consists of a covalent attachment of ADP- or UDP-glucose to an enzyme making $\alpha(1,4)$-glucosidic linkages followed by initial growth of the linear glucopolymer, $\alpha(1,4)$-glucan. In eukaryotes, glycogenin plays this initiating role and is required for glycogen synthesis [2,3]. When the $\alpha(1,4)$-glucan reaches a certain length, it becomes a substrate for the elongation enzyme, glycogen synthase, GlgA, while still attached to glycogenin. This universally conserved enzyme elongates the $\alpha(1,4)$-glucan and releases glycogenin only when the linear glucopolymer reaches its mature length. In the final step of glycogen synthesis, the linear $\alpha(1,4)$-glucan becomes a substrate for the branching enzyme GlgB, which catalyzes the formation of $\alpha(1,6)$ branches to produce mature glycogen [4].

In bacteria, the source of glucose for glycogen synthesis is adenosine 5'-diphosphoglucose (ADP-glucose) rather than uridine 5'-diphosphoglucose (UDP-glucose) as in eukaryotic cells [4]. ADP-glucose is synthesized from glucose 1-phosphate and adenosine triphosphate (ATP) by ADP-glucose pyrophosphorylase, which is encoded by the conserved *glgC* gene (reviewed in [5]). Significantly, there is no known bacterial homolog of the eukaryotic priming protein, glycogenin. Instead, glycogen synthase from *Agrobacterium tumefaciens* was shown to be sufficient for

Table 1: Structural statistics for GlgS.

Restraints for structure calculations	
Total restraints used	421
Total NOE restraints	289
Intraresidual	59
Sequential (|i-j| = 1)	96
Medium range (1<|i-j|<5)	74
Long range (|i-j| ≥ 5)	60
Hydrogen bond restraints	29
Dihedral angles restraints	103
Root mean square deviations from experimental restraints	
Distance deviations (Å)	0.041 ± 0.0020
Dihedral deviations (°)	0.220 ± 0.0382
Deviations from idealized geometry	
Bonds (Å)	0.0040 ± 0.0001
Angles (°)	0.5985 ± 0.0102
Impropers (°)	0.3737 ± 0.0085
Root mean square deviations of the 15 structures from the mean coordinates (Å)	
Backbone (residues Ser4-Met57)	0.52 ± 0.20
Heavy atoms (residues Ser4-Met57)	1.07 ± 0.21
Ramachandran plot statistics for residues Ser4-Met57 (%)	
Residues in most favored regions	93.9
Residues in additional allowed regions	5.6
Residues in generously allowed regions	0.5
Residues in disallowed regions	0.0

both glucan priming and elongation, and this appears to be generally true in eubacteria [6].

One of the regulators of the glycogen synthesis in *Escherichia coli* is the stationary phase sigma factor σ^S encoded by the *rpoS* gene. An increase in the cellular level of σ^S results from changes in environmental conditions such as insufficient nutrients, high osmolarity, pH change, heat shock, and forces cells to switch from growth phase to a stationary phase (for a review, see [7]). *RpoS* binds to the promoter region of RNA polymerase, changing its specificity and activating translation of numerous genes that are important for cellular response to multiple stresses (reviewed in [8,9]). One of these σ^S-induced genes codes for a 7.9 kDa protein named GlgS because it stimulates glycogen biosynthesis when overexpressed in the cell [10]. The search using the basic local alignment search tool (BLAST) [11] reveals that this protein is highly conserved in enterobacteria but not in other bacterial genomes, suggesting differences exist in glycogen synthesis in enterobacteria. This is supported by sequence analysis of other proteins involved in glycogen production. The GlgA and GlgC proteins are highly conserved among enterobacteria with higher than 90% sequence identity but less than 50–60% identical to the homologous proteins from other bacteria. These differences may be related to the σ^S role in cell cycle regulation, because *rpoS* has only been identified in a limited number of bacterial species.

Here, we report the nuclear magnetic resonance (NMR) structure of GlgS from *E. coli* and show that it is similar to two protein–protein interaction modules: Huntingtin, elongation factor 3, protein phosphatase 2A, TOR1 (HEAT) motifs and tetratricopeptide repeat domains.

Results and discussion

Previously reported resonance assignments [10] were used to assign nuclear Overhauser effects (NOEs) from heteronuclear ^{15}N-edited 3D nuclear Overhauser effect spectroscopy (NOESY) and homonuclear 2D NOESY experiments. On average, 7.1 constraints per residue in the GlgS structured region (Ser4-Met57) were used to calculate the GlgS structure. This relatively low number was the result of the tendency of GlgS to aggregate, which severely lowered the sensitivity and resolution of NOESY experiments. The structural statistics are shown in Table 1.

The structure of GlgS comprises a bundle of two parallel amphipathic helices α1 and α3 and a short hydrophobic helix α2 sandwiched between them (figure 1). The protein hydrophobic core is formed by Phe11, Leu14, Phe18, Ile30, Ala32, Val33, Trp44, and Phe45 (figure 1C) and includes almost all residues of the second α-helix. The most structured region of GlgS is the fragment between Asn6 and Met57. The five N-terminal and eight C-terminal amides give low heteronuclear NOE values indicating high mobility and an unstructured C-terminus [12].

Two single-residue mutations S17G and F45S abolish the ability of GlgS to stimulate glycogen accumulation [12]. Phe45 is a part of the hydrophobic core and its substitution with a hydrophilic serine residue is likely to disrupt the protein tertiary structure. Ser17 is positioned in the middle of the α1 helix and is solvent-exposed. Although placing a glycine residue within an α-helix could play some disruptive role, this serine is more likely to have a functional importance, perhaps via its phosphorylation. Analysis of surface charge distribution shows that most of the positively charged residues are located on the side containing Ser17 (figure 1D). Interestingly, this serine is surrounded by several arginine residues (Arg16, Arg20, and Arg26) (figure 1D). This concentration of positive charges is primed for interaction with a negatively charged group such as phosphate.

The protein surface was inspected for other solvent-exposed residues that may have functional significance. Tyrosines, which are often involved in signal transduction pathways through their phosphorylation, are also known to be essential for glycogen synthesis in eukaryotes. Tyr194 of glycogenin serves as a site for initial sugar attachment [13]. Similarly, GlgS contains a single tyrosine residue Tyr49 located on the C-terminal helix. This residue stacks with Phe45, but is not buried. There are also two cysteines in the vicinity of Tyr49. Cys46 and Cys53 are solvent-exposed and quite reactive, as GlgS is easily oxidized and precipitates in the absence of the reducing agent dithiothreitol (DTT).

GlgS-like sequences from the National Center for Biotechnology Information nonredundant database were aligned in order to search for potentially important residues for GlgS function (figure 2). GlgS has no significant sequence similarity to any proteins outside enterobacteria. The most variable region of GlgS is the N-terminal portion. All other regions, including loops, are highly conserved. Interestingly, the unstructured C-terminal sequence LELEH is invariant, suggesting a functional importance.

A structural homology search using the DALI database [14] resulted in several hits. Fragments of HEAT domain from protein phosphatase 2A (PDB code 1B3U) and tetratricopeptide repeat (TPR) domain from protein phosphatase 5 (PDB code 1A17) showed the highest Z-scores of 3.1 and 3.0 (figure 3). Both domains represent an assembly of helical hairpins, with each hairpin accounting for one HEAT [15] or TPR [16] motif. While there is no sequence similarity between HEAT and TPR sequences, functionally both domains act as protein–protein interaction modules.

Interestingly, GlgS has a similar architecture to that of the peripheral subunit-binding domain (PSBD) of dihydrolipoyl acyltransferase (E2 enzyme), an enzyme from the pyruvate dehydrogenase (PDH) multienzyme complex, which is involved in glucose metabolism. PSBD is a protein–protein interaction domain that binds E1 and E3 subunits of PDH with high affinity. In particular, the mechanism of E3 binding involves an electrostatic zipper in which Arg135, Arg139 and Arg156 of PSBD form salt bridges with the aspartate and glutamate of the E3 enzyme [17]. GlgS has a similar pattern of arginine residues (Arg16, Arg20, Arg26 and Arg48), which may function similarly.

Conclusions

Earlier work suggested that GlgS might be a site for primary sugar attachment in the glycogen synthesis pathway [12]. This hypothesis has been weakened by the recent finding that glycogen synthase in *Agrobacterium tumefaciens* does not require additional proteins for glycogen priming [6]. The GlgS structure offers alternative possibilities for its function. The stimulation of glycogen synthesis by GlgS overproduction may be caused by GlgS-mediated interactions between proteins involved in glycogen biosynthesis. A relevant example is the recent discovery of glycogenin interacting protein (GNIP) in the eukaryotic glycogen synthesis [18]. Studies of GlgS involving GlgA and GlgC and further mutagenesis of GlgS would help to test this hypothesis and better define its function in glycogen biosynthesis.
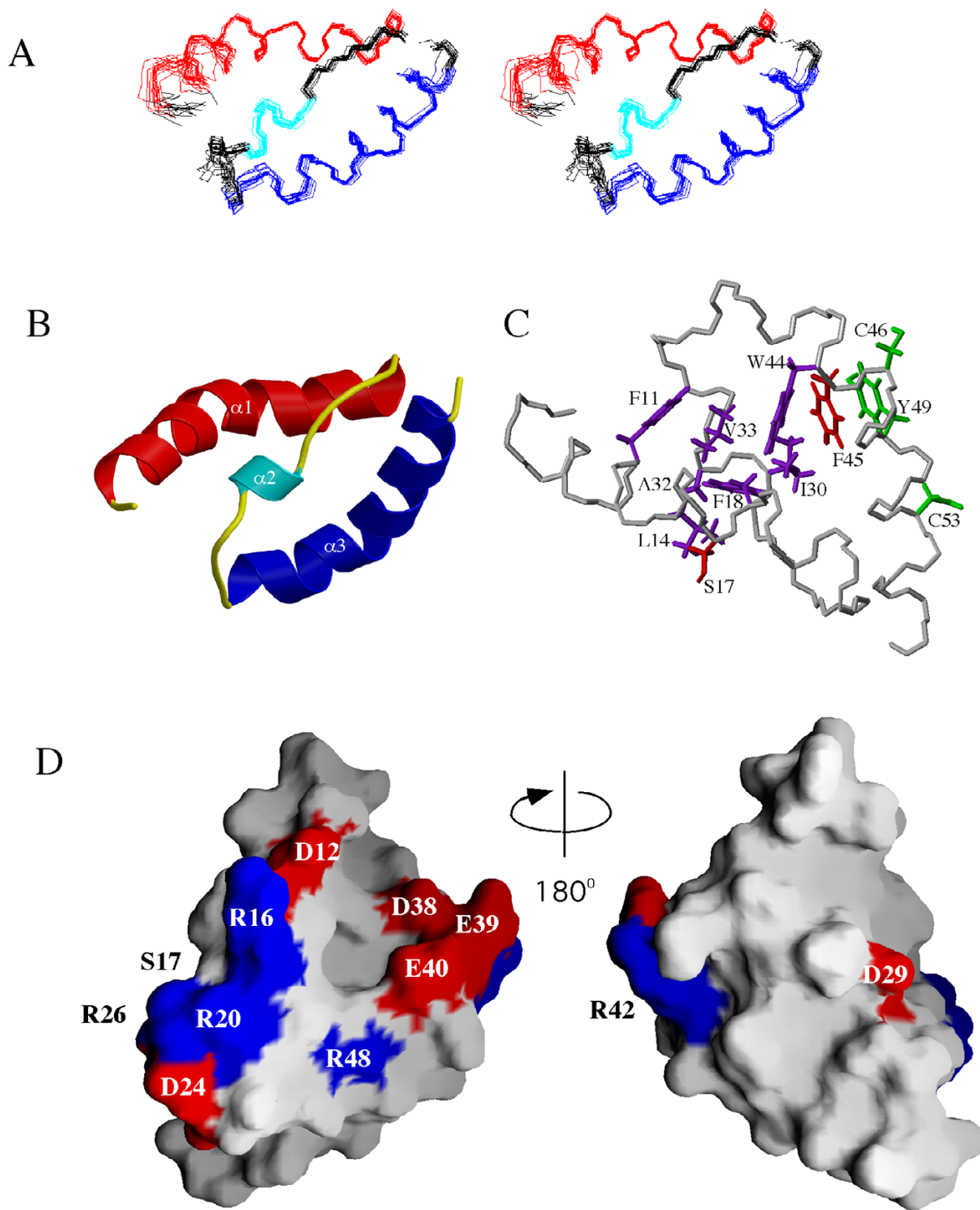
Figure 1 *(see following page)*
Structure of GlgS. (A) The backbone superposition of the 15 lowest-energy structures generated with MOLMOL [26]. The superposition was done using region Ser4-Leu62. (B) Ribbon representation of the GlgS structure generated with MOLSCRIPT [27] and Raster3D [28]. N- and C-termini and α-helices are labeled. (C) Enlarged view of residues that form hydrophobic core (in purple), are potentially involved in the function (in green), and which mutations knocked-out GlgS activity (in red). Phe45 also is a part of the hydrophobic core. The residues shown are labeled with a one-letter name code and a residue number. The figure was generated with MOLMOL [26]. (D) Surface charge distribution suggests importance of charged interactions for GlgS function. The protein side with more charges contains Ser17, which is essential for functionality. Positively charged residues are shown in blue, negatively charged residues in red. The figure was generated with GRASP [29].

Figure 1 *(see legend on previous page)*

Figure 2
GlgS is highly conserved in enterobacteria. The aligned sequences represent GlgS from *Escherichia coli* K12 (*E. coli* K12; gi:1789428), *Escherichia coli* O157 (*E. coli* O157; gi:13363404), *Shigella flexneri* (*Sh. flexneri*; gi:30042655), *Salmonella enterica* serovar Typhi (*S. enterica*; gi:16504274), and *Salmonella typhimurium* (*S. typhimurium*; gi:20138181). The secondary structure elements refer to GlgS from *Escherichia coli*.
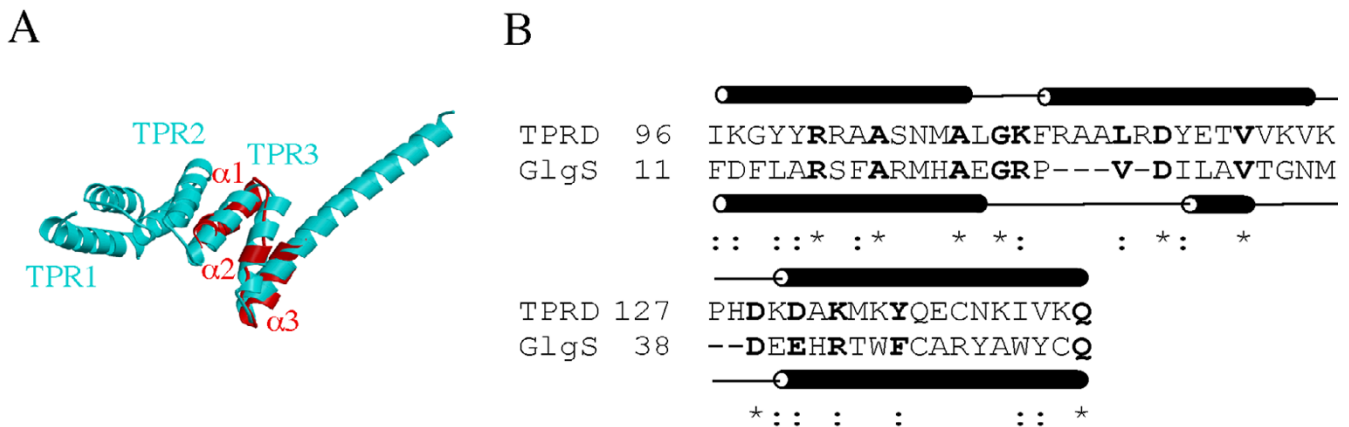

Figure 3
GlgS is structurally similar to a fragment from the tetratricopeptide repeats domain from protein phosphatase 5. (A) Ribbon representation of the backbone superposition of GlgS and TPRD using 38 residues results in root mean square deviation of 2.0 Å. GlgS is shown in red, TPRD in cyan. The figure was generated using MOLSCRIPT [27] and Raster3D [28]. (B) Sequence alignment of TPRD and GlgS with secondary structure elements shown. Homologous residues are in bold.

Methods

Protein expression and purification

The *E. coli glgS* sequence, comprising amino acids 1 to 66, was subcloned into the pET15b vector (Novagen Inc., Madison, WI, USA) and overexpressed in *E. coli* BL21(DE3) as a His-tagged fusion protein. The protein was purified by immobilized metal affinity chromatography on an Ni²⁺-loaded chelating Sepharose column (Amersham Biosciences, Amersham, Bucks, UK). The resulting protein contains the N-terminal His-tag MGSSH-HHHHHSSGLVPRGSH from the vector. Isotopically enriched GlgS was prepared from cells grown on minimal M9 media containing ^{15}N-ammonium chloride with or without ^{13}C$_6$-glucose (Cambridge Isotopes Laboratory, Andover, MA, USA).

NMR spectroscopy

NMR samples with a protein concentration of 1 mM were exchanged into 5 mM potassium phosphate, 1 mM DTT and 0.1 mM sodium azide at pH 6.70. NMR experiments were performed at 298 K. Backbone and side-chain NMR signal assignments of GlgS were determined previously [12]. NOE constraints for the structure determination were obtained from ^{15}N-edited and homonuclear NOESY obtained on a Bruker Avance600 MHz spectrometer. $^{3}J_{H^N-H^\alpha}$ coupling constants were obtained from an HNHA experiment [19]. NMR spectra were processed using XWINNMR (Bruker Biospin Ltd, Milton, ON, Canada) software and analyzed with XEASY [20].

Structure calculations

NOE restraints were obtained from ^{15}N-edited 3D NOESY and 2D homonuclear NOESY experiments. The ϕ and ψ torsion angles were derived from C^α, C^β and H^α chemical shifts using TALOS [21] and compared with experimental ϕ values estimated from an HNHA experiment. Structures were calculated using the CANDID module implemented in the program Cyana [22]. 2D NOESY and 3D ^{15}N-NOESY spectra were used with CANDID to calibrate and assign NOE cross-peaks. The 20 lowest-energy structures obtained after seven cycles of calculations in Cyana were further refined using standard protocols in Xplor-NIH [23] to yield 15 lowest-energy structures comprising the final ensemble. The quality of obtained structures was assessed using PROCHECK [24]. Coordinates were deposited with the Protein Data Bank [25] under PDB ID code 1RRZ.

List of abbreviations used

NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; DTT, dithiothreitol; TPR, tetratricopeptide repeat; PSBD, peripheral subunit-binding domain; PDH, pyruvate dehydrogenase.

Authors' contributions

GK collected the NMR spectra, solved and refined the molecular structure and drafted the manuscript. DE expressed and purified the labeled and unlabeled protein samples and contributed to the written manuscript. MC and KG conceived the study, participated in the coordination, and provided financial support. All authors read and approved the final manuscript.

Acknowledgements

References

1. Okita TW, Rodriguez RL, Preiss J: **Biosynthesis of bacterial glycogen. Cloning of the glycogen biosynthetic enzyme structural genes of *Escherichia coli*.** *J Biol Chem* 1981, **256**:6944-6952.
2. Alonso MD, Lomako J, Lomako WM, Whelan WJ: **A new look at the biogenesis of glycogen.** *FASEB J* 1995, **9**:1126-1137.
3. Smythe C, Cohen P: **The discovery of glycogenin and the priming mechanism for glycogen biogenesis.** *Eur J Biochem* 1991, **200**:625-631.
4. Preiss J, Romeo T: **Physiology, biochemistry and genetics of bacterial glycogen synthesis.** *Adv Microb Physiol* 1989, **30**:183-238.
5. Ballicora MA, Iglesias AA, Preiss J: **ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis.** *Microbiol Mol Biol Rev* 2003, **67**:213-225.
6. Ugalde JE, Parodi AJ, Ugalde RA: *De novo* **synthesis of bacterial glycogen:*Agrobacterium tumefaciens* glycogen synthase is involved in glucan initiation and elongation.** *Proc Natl Acad Sci USA* 2003, **100**:10659-10663.
7. Loewen PC, Hu B, Strutinsky J, Sparling R: **Regulation in the rpoS regulon of *Escherichia coli*.** *Can J Microbiol* 1998, **44**:707-717.
8. Hengge-Aronis R: **Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase.** *Microbiol Mol Biol Rev* 2002, **66**:373-395.
9. Ishihama A: **Functional modulation of *Escherichia coli* RNA polymerase.** *Annu Rev Microbiol* 2000, **54**:499-518.
10. Hengge-Aronis R, Fischer D: **Identification and molecular analysis of glgS, a novel growth-phase-regulated and rpoS-dependent gene involved in glycogen synthesis in *Escherichia coli*.** *Mol Microbiol* 1992, **6**:1877-1886.
11. **BLAST** [http://www.ncbi.nlm.nih.gov/BLAST/]
12. Beglova N, Fischer D, Hengge-Aronis R, Gehring K: **1H, 15N and 13C NMR assignments, secondary structure and overall topology of the *Escherichia coli* GlgS protein.** *Eur J Biochem* 1997, **246**:301-310.
13. Smythe C, Caudwell FB, Ferguson M, Cohen P: **Isolation and structural analysis of a peptide containing the novel tyrosyl-glucose linkage in glycogenin.** *EMBO J* 1988, **7**:2681-2686.
14. **DALI** [http://www.ebi.ac.uk/dali/]
15. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D: **The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs.** *Cell* 1999, **96**:99-110.
16. Das AK, Cohen PW, Barford D: **The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.** *EMBO J* 1998, **17**:1192-1199.
17. Mande SS, Sarfaty S, Allen MD, Perham RN, Hol WG: **Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase.** *Structure* 1996, **4**:277-286.
18. Skurat AV, Dietrich AD, Zhai L, Roach PJ: **GNIP, a novel protein that binds and activates glycogenin, the self-glucosylating initiator of glycogen biosynthesis.** *J Biol Chem* 2002, **277**:19331-19338.
19. Kuboniwa H, Grzesiek S, Delaglio F, Bax A: **Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods.** *J Biomol NMR* 1994, **4**:871-878.
20. Bartels C, Xia T-H, Billeter M, Guntert P, Wuthrich K: **The program XEASY for computer-supported NMR spectral analysis of biological macromolecules.** *J Biomol NMR* 1995, **5**:1-10.
21. Cornilescu G, Delaglio F, Bax A: **Protein backbone angle restraints from searching a database for chemical shift and sequence homology.** *J Biomol NMR* 1999, **13**:289-302.
22. Herrmann T, Guntert P, Wuthrich 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.
23. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM: **The Xplor-NIH NMR molecular structure determination package.** *J Magn Res* 2003, **160**:66-74.
24. Laskowski RA, Rullmann 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.
25. **Protein Data Bank** [http://www.rcsb.org/pdb/]
26. Koradi R, Billeter M, Wuthrich K: **MOLMOL: a program for display and analysis of macromolecular structures.** *J Mol Graph* 1996, **14**:51-55.
27. Kraulis PJ: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.** *J Appl Cryst* 1991, **24**:946-950.

28.   Merritt EA, Murphy MEP: **Raster3D version 2.0: a program for photorealistic molecular graphics.** *Acta Crystallogr D* 1994, **50:**869-873.
29.   Nicholls A, Sharp KA, Honig B: **Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.** *Proteins* 1991, **11:**281-296.