JPP JOURNAL OF PROTEINS AND PROTEOMICS
8(2), 2017, pp. 127-132

Research Communication

CLONING, EXPRESSION, PURIFICATION AND CRYSTALLIZATION OF A NOVEL GLCNAC METABOLIC PROTEIN, GIG2 (DUF1479) FROM PATHOGENIC FUNGUS *CANDIDA ALBICANS*

Priya Rani[1], Gunjan Gautam[2], Kongara Hanumantha Rao[3], Swagata Ghosh[4], Samudrala Gourinath[2], Suman Kumar Dhar[1] and Asis Datta[3]

[1] *Special Centre for Molecular Medicine, Jawaharlal Nehru University, New Delhi 110067*
[2] *School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067*
[3] *National Institute of Plant Genome Research, New Delhi 110067*
[4] *Department of Molecular Biology and Biotechnology, University of Kalyani, West Bengal 741235, India*

Abstract: N-acetylglucosamine (GlcNAc), an alternative sugar, is emerging as an important molecule having a multifarious role in *Candida albicans* including a major role in signaling. GlcNAc Inducible Gene 2, *GIG2* is one of the highly upregulated genes in GlcNAc grown cells in *C. albicans*. Our earlier studies show the involvement of Gig2 in the formation of N-acetylneuraminic (NANA) acid from GlcNAc-6-phosphate through an understudied route. The crystal structure of Gig2 would help us in determining the exact reaction that this enzyme catalyzes. Here the cloning, expression, purification and crystallization of this protein are reported along with preliminary X-ray crystallographic analysis at 2.4Å resolution. The crystal belonged to $P2_1$ space group, with unit cell parameters a=59.59, b= 54.43, c= 73.29Å; $\alpha = 90°$, $\beta = 102.7°$ and $\gamma = 90°$. The structure was solved using PDB ID 2CSG as a template which has only 27% identity. Molecular replacement yielded a solution with LLG score of 87. The structure is currently under further refinement.

Keywords: N-acetylglucosamine (GlcNAc); GlcNAc Inducible Gene 2 (*GIG2*); DUF (Domains of Unknown Function) family of proteins; Crystallography.

Introduction

Metabolic flexibility is a prerequisite for free-living microorganisms to maximize the use of available nutrients. The opportunistic fungal pathogen *C. albicans*, a common member of human microbiota often use a limiting nutritional component as a signal to maintain cellular homeostasis. Alternative carbon sources such as GlcNAc and other polysaccharides obtained from infection sites are often used as primary energy and carbon sources. Apart from acting as aprimary carbon source for *C. albicans* within mammalian host GlcNAc acts as a signaling molecule regulating many biological programs such as morphological transition (Shepherd and Sullivan, 1983), virulence (Singh *et al*., 2001) and Cell death (Du *et al*., 2015). *C. albicans* causes serious systemic and superficial infections in humans. Extensive efforts have been put into the discovery of therapeutic strategies, keeping in mind the increase in the frequency of such fungal infections worldwide.

N-acetylglucosamine-6-phosphate which is formed by the action of Hxk1 on GlcNAc appears to occupy a nexus for several metabolic processes. The routes emanating from GlcNAc-6-phosphate are poorly characterized. Our earlier report based on protein-metabolite complex purification and

Corresponding Author: **Asis Datta and Suman Kumar Dhar**
E-mail: asis_datta@rediffmail.com; skdhar@mail.jnu.ac.in; skdhar2002@yahoo.co.in

Received: June 6, 2017
Accepted: June 21, 2017
Published: June 23, 2017

metabolite identification by using UPLC coupled ESI-MS and solution state NMR, hypothesizes the involvement of Gig2 in a route where N-acetylneuraminic acid (NANA) is formed from GlcNAc-6-phosphate (Ghosh *et al*., 2014). The formation of NANA from GlcNAc-6-phosphate is quite important from a pathological aspect since NANA formation has been reported in many cases to modulate cell interactions and carbohydrate-dependent physiological or pathophysiological responses (Alviano *et al*., 1999). Our earlier studies also indicated decreased virulence for *GIG2* deletion mutants when compared to wild type *C. albicans* in mouse infection model (Ghosh *et al*., 2014).

Gig2, a DUF 1479 family member has not been characterized in any organism previously; no PDB homolog with more than 30% identity is known. The crystal structure of this protein will help us in predicting the exact function of Gig2 and its homologs in other organisms. Once the function has been determined, it would be easier to formulate hypotheses about the biological function of the DUF 1479 family (Jaroszewski *et al*., 2009). This would also help us in an improved understanding of the importance of intermediate products of GlcNAc metabolism. The structural information can be implied in designing inhibitory molecules against this enzyme's activity and aid in treatment of candidiasis and other fungal or enterobacterial infections.

Materials and methods

Macromolecule production

The orf 19.4783 (1425bp) encoding gene *GIG2* was PCR amplified from genomic DNA of *Candida albicans* strain SC5314 using gene-specific primers (Table 1). The amplified fragment flanked with BamHI and SacI sites was cloned into MCS of vector pET28a(+) by restriction-ligation. The resultant construct pET28a(+)-GIG2 was verified by sequencing (NIPGR, New Delhi). Verified pET28a(+)- GIG2 was transformed in *E.coli* BL21-codon plus for over-expression. Cells were grown in Luria-Bertani medium supplemented with 35µg/ml chloramphenicol and 50 µg/ml kanamycin at 37°C and180 rpm till 0.6 OD at 600nm, then were induced with 0.5mM IPTG and further grown at 22°C for 8 hours. Cells were harvested by centrifugation for 10 min at 10,000rpm and were then frozen at -80°C. After thawing the pellet was resuspended in lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% NP40, 10mM imidazole, 0.5mg/ml lysozyme) and subjected to repeated cycles of freeze-thaw for cell lysis and then was sonicated at 25% amplitude 30s pulse 30s rest (6-7cycles). The lysate was then centrifuged at 20,000rpm for 30mins at 4°C to remove cell debris. Protein purification from the lysate was done using His-tag affinity chromatography. The filtered supernatant of lysate was passed through 8ml Ni-NTA resin column pre-equilibrated in lysis buffer at 4°C. The protein-bound beads were washed five times bed volume with wash buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 30mM imidazole). Protein was eluted in elution buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 200mM imidazole). After the recombinant Gig2 protein was purified by Ni-NTA resin, it was concentrated using Amicon ultracentrifugal filters(Millipore) upto 20mg/ml and was loaded on a gel filtrationchromatography Superdex 200 10/300 GL column (GE Healthcare), which was pre-equilibrated with buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 5% glycerol). Fractions of the peak from 80- 90 ml were collected checked on 10% SDS-PAGE gel and was concentrated up to 8-10mg/ml. The purified concentrated protein was then used for crystallization.

Crystallization

Purified protein concentrated up to 8-10 mg/ml was used for crystallization experiments. Different commercially available crystallization screens (Hampton Research and Molecular Dimensions) were set up using Mosquito crystallization robot (TTP Labtech) using Greiner 96 well crystallization plates (350 nl drop with protein solution and reservoir solution in 1:1 ratio) and the plates were kept at 16°C. Crystals were observed after 3-4 days in Morpheus screen (Molecular Dimensions) in conditions; A4 [12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD ,0.03 M of each divalent cation $MgCl_2$ and $CaCl_2$, 0.1 M MES/imidazole pH 6.5] (Fig. 2A), A8 [12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD, 0.03 M of each divalent cation, 0.1 M MOPS/HEPES-Na pH 7.5] and A12 [12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD, 0.03 M of each divalent cation,0.1 M bicine/Trizma base pH 8.5].Further crystallisation optimisation experiments were done manually using hanging drop method in 24 well plates (Corning). A4 and A8 crystals were very small rods while A12 crystals were small fine needles. These crystals were used as seeds for macro-seeding in

Table 1
Macromolecule production information.

Source organism	*Candida albicans SC5314*
DNA source	*Candida albicans* SC5314 genomic DNA
Forward primer	ACTGATGGATCCATGTCTCCTTCCAAATTAT
Reverse primer	TATCATGAGCTCCTTAGCAGCATGGTGA
Cloning vector	pET28a(+) (Novagen)
Expression vector	pET28a(+) (Novagen)
Expression host	*E. coli* BL21-codon plus
Complete amino acid sequence of the construct produced	MGSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMSPSKLSNDETPP DLDFRFTGIKQRLIKSENVK QVTASWKRLLVEINKEFTEIAKIGPSYVPK CDFIDIKDNKLPQQVSELFKQRGCLMIENVIDVDRIDIWFNELVEFCKTH PETAGYTFPNPTSWYNVFWSKPQTEARFHPNMKAIFKAMSKEFYVED KENCLIDLDTQLVYGDRIRIREPGKAAALPLHLDSSSIERWEDIMYSEVY KSIFEGDWENWDAFKLDERTYSKENLYKDEDDTGGKSTICSSFRTLQG WLALSNNKSGEGTLRVLPSLKLSMAYIMLRPFFWKDPESGNIDDYEIDL ITPKFPGTVPGTGQLFLDKFYPHLHQGIISIPDVKKGSFVFWHCDLPHEV DREHNGNGHSSVLYYGQTPLSITNIQTLLDTRDAFLKNISPADYRSQLN EEEKQKEFQGANIDDLKNDIDSKRSMGLEEFEKPENMSGGQAKIRSIAN QALKSSGFNVDKYIHHAAKELRRQACGRTRAPPPPPLRSGC

Table 2
Crystallization conditions

Method	*Hanging Drop*
Plate type	24 well plates (corning)
Temperature (K)	289
Protein concentration	8 mg/ml
Buffer composition of protein solution	50mM Tris-HCl pH 7.5, 150mM NaCl, 5% glycerol
Composition of reservoir solution	11% w/v PEG 1000, 14% w/v PEG 3350, 12.5% v/v MPD , 0.01 M of each divalent cation $MgCl_2$ and $CaCl_2$, 0.1 M MES/imidazole pH 6.9
Volume and ratio of drop	2µl, 1:1
Volume of reservoir	500

both A4 and A8 conditions by varying pH from 6.5 to 7.5 by a unit of 0.2 and by varying the percentage-ratio of PEG 3350 /PEG 1000 from 12.5% / 12.5% to 14% /11%; 15% / 10% and 16% / 9%. Best diffractable crystals were obtained in the condition; 11% w/v PEG 1000, 14% w/v PEG 3350, 12.5% v/v MPD , 0.01 M of each $Mgcl_2$ and $CaCl_2$, 0.1 M MES/ imidazole pH 6.9 (Fig. 2B). The crystals obtained were slightly bigger than the initial crystals obtained without seeding.

Data collection and processing

Gig2 crystals from several drops were mounted on 0.2mm Nylon loops and were soaked in cryoprotectant which was mother liquor itself. These soaked crystals were quickly flash frozen in liquid nitrogen. Data was collected on BM14 beamline at the ESRF (Grenoble, France) at a wavelength of 0.97Å. The distance between the detector and the crystal was 267.9mm. The best crystals diffracted at 2.4Åresolution. A total of 380 images were collected with 0.5 Å oscillation using CCD detector (Marresearch) and the data was indexed and scaled using HKL 2000 (Otwinowski and Minor, 1997). SCALEPACK2MTZ from the CCP4 suite (Winn et al., 2011) was run to convert the scaled reflections to MTZ files and these reflections were further used for molecular-replacement trials. The data statistics are given in Table 3.

Table 3
Data collection and processing Values for the outer shell are given in parentheses

Diffraction source	*Beamline BM14, ESRF*
Wavelength (Å)	0.97
Temperature (K)	100
Detector	CCD, Marresearch
Crystal-detector distance (mm)	267.91
Rotation range per image (°)	0.5
Total rotation range (°)	190
Exposure time per image (s)	7.8
Space group	$P2_1$
a, *b*, *c* (Å)	59.59, 54.43, 73.29
α, β, γ (°)	90, 102.70, 90
Mosaicity (°)	0.3
Resolution range (Å)	50.00–2.46 (2.50–2.46)
Total No. of reflections	68853
No. of unique reflections	16747
Completeness (%)	99.5 (92.7)
Redundancy	3.8 (3.0)
$\langle I/\sigma(I)^* \rangle$	7.4 (1.3)
$R_{r.i.m.}$	0.262 (0.950)
Overall *B* factor from Wilson plot (Å^2)	25.92

mean $I/\sigma(I)$ is < 2 at 2.46Å.

Structure solution and refinement

Gig2 crystals diffracted in $P2_1$ space group with unit cell parameters a=59.59, b= 54.43, c=73.29 and β = 102.70Å. The calculated Mathews coefficient was 1.92 Å^3/Da with a solvent content of 35.92%. This corresponded to the presence of a single monomer of Gig2 protein in an asymmetric unit. No structure with more than 30% identity was available in PDB. During structure solution the data was cut off at 2.7 Å resolution which has an $\langle I/\sigma(I)^* \rangle$ equal to 2. The phase problem was solved by molecular replacement using PHASER from the CCP4 suite, using putative oxidoreductase from *Salmonella typhimurium LT2* (PDB ID 2CSG), as a search model which had a sequences identity of only 27%. The template PDB was trimmed by chainsaw and used for molecular replacement. PHASER yielded a top solution with LLG score of 87, RFZ= 9.2 and TFZ= 6.4. First refinement cycle using REFMAC5 from CCP4 suite resulted in an R/R_{free} of 0.47/0.50. The structure was refined to an R/R_{free} of 0.37/0.47 by undergoing 7 cycles of refinement, each followed by manual building in COOT. The structure is currently under further refinement.

Results and Discussion

Full-length *GIG2* gene was amplified from genomic DNA and cloned in pET28a (+) vector to produce recombinant protein with 6XHis-tag both at N- and C-terminal ends. Tagging at both ends of the protein helped in high affinity binding to Ni-NTA column during initial purification steps, which can withstand high stringency washings at 40mM imidazole concentration. Over expression of the recombinant protein was done in *E. coli* BL21-codon plus cells. Optimum expression was obtained when mid log phase cells were induced with 0.5mM IPTG at 22°C for 8hours.The his-tag was used for affinity purification of the protein from the cell lysate. The protein was purified by employing his-tag affinity purification to ***> 70%*** electrophoretic homogeneity based on Coomassie Blue stain of SDS PAGE gels. The protein migrated on SDS-PAGE with an apparent molecular weight of 55 kDa, which is in agreement with the value calculated from the amino acid sequence (54.58 kDa). It eluted as a single peak during size-exclusion chromatography (SEC).

Protein purified using Ni-NTA affinity column was further purified by gel filtration using a Superdex 200 10/300 column (Fig. 1). Approximate yield of purified protein was nearly1mg/l of LB bacterial culture. Purified protein was concentrated upto 8-10mg/ml and used for crystallization trials.

The initial screening yielded crystals under many conditions in Morpheus crystallization screen (Molecular Dimensions) but the diffractable crystals were obtained in 11% w/v PEG 1000, 14% w/v PEG 3350, 12.5% v/v MPD, 0.01 M of each divalent cation, 0.1 M MES/imidazole pH 6.9. The crystals diffracted at 2.4 Å resolution at ESRF (Table 3) in $P2_1$ space group. Matthews's coefficient calculations depicted the presence of one monomer in the unit cell. Molecular replacement using putative oxidoreductase from *Salmonella typhimurium*LT2 (PDB ID 2CSG) as the template was done using PHASER from CCP4 suite resulted in a single solution. Gig2 crystal structure is currently under refinement. Density could not be built for residues in the loop region. Further efforts are being made for a higher resolution data using multiple isomorphous replacement dispersion methods.

The high-resolution crystal structure will provide crucial information about the structure and function of Gig2 protein from *Candida albicans*, currently placed under DUF protein family (Domain of unknown function). Moreover, this

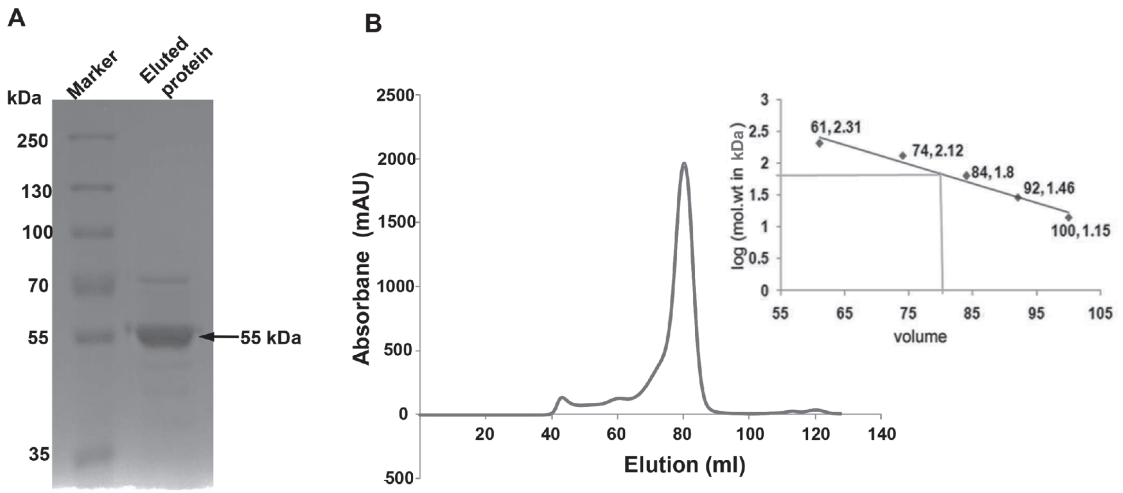

Figure 1: **Expression and purification of Gig2 (A) Ni-NTA purified Gig2 protein on a Coomassie Blue-stained SDS–PAGE gel (12%). Molecular-weight marker (labelled in kDa). (B) Size-exclusion chromatography elution profile of Gig2 purification. Fractions 80–90ml containing purified Gig2 were pooled, concentrated and used for crystallization. Inset: Standard curve for Superdex 200 10/300 GL column showing elution volume of different Molecular weight protein markers**

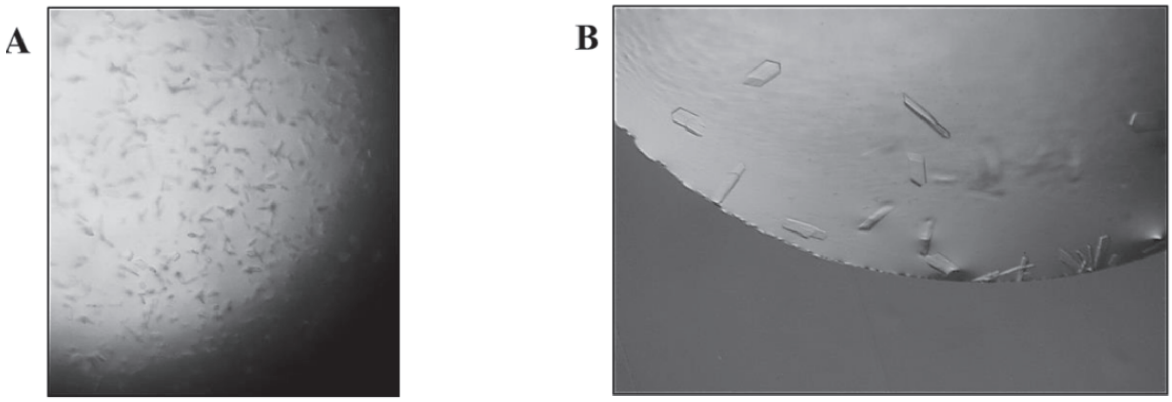

Figure 2: **Crystal images of Gig2 crystals: (A) Initial hits of Gig2 obtained in crystallization condition A8 (12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD, 0.03 M of each $Mgcl_2$ and $CaCl_2$, 0.1 M MOPS/HEPES-Na pH 6.5) of Morpheus Screen from Molecular Dimensions. (B) Examples of Gig2 crystal obtained after macro-seeding with initial small crystals in crystallization condition; 11% w/v PEG 1000, 14% w/v PEG 3350, 12.5% v/v MPD , 0.01 M of each divalent cation $MgCl_2$ and $CaCl_2$, 0.1 M MES/ imidazole pH 6.9. One of these crystals was used for data collection at BM14 beamline, ESRF, Grenoble, France**

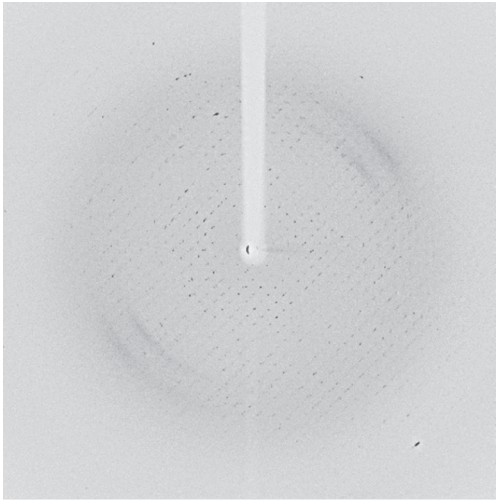

Figure 3: **Gig2 diffraction image obtained using a Mar research CCD detector on BM14 beamline at ESRF**

structure-function analysis will further our understanding about the role of DUF 1479 domain conserved amongst many other fungi and *Enterobacteriaceae* members too. Understanding the crystal structure would be an important step towards elucidation of at least one of the enzymatic conversions that take place during the formation of different metabolic intermediates arising out of GlcNAc-6-phosphate.

Acknowledgement

We thank Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India for funding. We thank UGC-RNW, UGC-SAP, UGC-UPOE-II, DST-FIST, and DST-PURSE for funding central instrumentation facility and for extending institutional funding. PR and GG thank UGC and CSIR for fellowship.

We also thank Department of Biotechnology (DBT) and staffs of BM14, ESRF for helping us with the collection of high-resolution X-ray data.

Abbreviations

GlcNAc, N-acetyl glucosamine; Gig2, GlcNAc inducible gene 2; NANA, N- acetyl neuraminic acid; DUF, Domain of Unkonwn function; LLG score, log-likelihood gain score; PDB, Protein Data Bank.

Conflict of interest

The authors do not have any conflict of interest with the contents of this manuscript.

References

Alviano, C.S., Travassos, L.R., and Schauer, R. (1999). Sialic acids in fungi: A minireview. Glycoconj. J. *16*, 545–554.

Du, H., Guan, G., Li, X., Gulati, M., Tao, L., Cao, C., Johnson, A.D., Nobile, C.J., and Huang, G. (2015). N-Acetylglucosamine-induced cell death in candida albicans and its implications for adaptive mechanisms of nutrient sensing in yeasts. MBio *6*.

Ghosh, S., Rao, K.H., Bhavesh, N.S., Das, G., Dwivedi, V.P., and Datta, A. (2014). N-acetylglucosamine (GlcNAc)-inducible gene GIG2 Is a novel component of GlcNAc metabolism in Candida albicans. Eukaryot. Cell *13*, 66–76.

Jaroszewski, L., Li, Z., Krishna, S.S., Bakolitsa, C., Wooley, J., Deacon, A.M., Wilson, I.A., and Godzik, A. (2009). Exploration of uncharted regions of the protein universe. PLoS Biol. *7*.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Shepherd, M.G., and Sullivan, P.A. (1983). Candida albicans germ-tube formation with immobilized GlcNAc. FEMS Microbiol. Lett. *17*, 167–170.

Singh, P., Ghosh, S., and Datta, A. (2001). Attenuation of virulence and changes in morphology in Candida albicans by disruption of the N-acetylglucosamine catabolic pathway. Infect. Immun. *69*, 7898–7903.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 235–242.