STRUCTURE NOTE

Crystal Structure of Human SH3BGRL Protein: the First Structure of the Human SH3BGR Family Representing a Novel Class of Thioredoxin Fold Proteins

Lei Yin,[1,2] **Ye Xiang,**[1,2] **De-Yu Zhu,**[1,2] **Na Yan,**[1,2] **Ren-Huai Huang,**[1] **Ying Zhang,**[1] **and Da-Cheng Wang**[1*]
[1]*Center for Structural and Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing, People's Republic of China*
[2]*Graduate School of the Chinese Academy of Sciences, Beijing, People's Republic of China*

Introduction. The human *SH3BGRL* (*h-SH3BGRL*) gene is a member of the human *SH3BGR* family.[1,2] The human *SH3BGR* (*SH3 binding glutamic acid-rich*) gene was cloned and characterized in 1997 in an effort to identify new genes located to chromosome 21. After the characterization of *SH3BGR*, three novel human genes, *h-SH3BGRL*, *h-SH3BGRL2*, and *h-SH3BGRL3*, were identified, showing a high homology to the N-terminal region of the h-SH3BGR protein. They are therefore believed to be a new family of human gene, the *h-SH3BGR* gene family.[2]

The *h-SH3BGRL* gene located to chromosome Xq13.3 encodes for a small protein of 114 amino acids, which is apparently widely expressed in many tissues including liver and blood. The h-SH3BGRL protein features a proline-rich sequence (PLPPQIF), which contains both the SH3 binding (PXXP)[3] and the Homer EVH1 binding (PPXXF)[4] motif. This proline-rich region is highly conserved in h-SH3BGR family and was expected to be functionally important. Although it is known that the SH3BGRL gene is missing in a mentally retarded male patient,[5] the exact functional role of h-SH3BGRL is still needed to be defined.

Here we report the crystal structure of h-SH3BGRL protein determined by the SIRAS method. The structure indicates that SH3BGRL can not bind to the SH3 or Homer EVH1 domain as previously expected due to our finding that the binding motif (PLPPQIF) is buried in the tertiary structure. As the first structure of the human SH3BGR protein family, the tertiary structure of h-SH3BGRL shows a typical thioredoxin fold at the N-terminal part and a helix-loop-helix motif at the C-terminal lobe. Sequence and structure comparisons show that h-SH3BGRL belongs to the thioredoxin fold protein family but it is distinct from all five classes of the thioredoxin fold proteins[6] identified thus far. According to the unique structural and functional characterizations, h-SH3BGRL represents a novel class of the thioredoxin fold proteins.

Materials and Methods. Cloning, expression and purification. The coding sequence for h-SH3BGRL protein was amplified from RT-PCR products of human hemopoietic stem cell using the polymerase chain reaction (PCR) method. The PCR product was restricted with NdeI and XhoI, and ligated into the pET22b(+) vector (Novagen Inc.) with a his tag. The recombinant plasmid was transformed into *Escherichia coli* strain BL21(DE3) for protein expression. The cells were grown in LB medium at 310 K and induced with 1 mM IPTG. The harvested cells were then resuspended and sonicated. The protein was first purified in two steps using a Ni-affinity column and a Superdex 75 gel-filtration column. A further purification step was performed with a Mono Q HR 5/5 column (Pharmacia). The Purified protein was analyzed by SDS-PAGE.

Crystallization: The purified protein was desalted and concentrated to about 30 mg ml^{-1} in ultra pure water for crystallization. All crystallization experiments were performed with the hanging-drop vapor-diffusion method. The drops were formed by mixing 1.5 μl protein solution with 1.5 μl reservoir solution and equilibrated against 400 μl reservoir solution in each well at 293 K. The best crystals were obtained with a reservoir solution (400 μl) containing 1.3 M tris-Sodium Citrate, 0.1 M Tris (pH 8.0) at 293 K.

Data collection and processing: All diffraction data were collected on a Rigaku R-Axis IV++ image plate using Cu *K*α radiation (λ = 1.5418 Å) from a Rotating Anode operating at 40 kV and 20 mA with 0.1 mm cofocus incident beam diameter. All crystals were briefly soaked in paraffin oil (Hampton Research) after being mounted in nylon cryoloops (Hampton Research) and then flash-cooled in a nitrogen-gas stream at 95 K. The native data were collected with a 160 mm crystal-to-detector distance and 1°

*Correspondence to: Da-Cheng Wang, Center for Structural and Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, People's Republic of China. E-mail: dcwang@ibp.ac.cn/wangdcibp@yahoo.com.cn

TABLE I. Data Collection, Phasing and Refinement Statistics of SH3BGRL

	Native	KI
A. Data collection		
Space group (molecules/asu)	P2$_1$2$_1$2$_1$ (1)	P2$_1$2$_1$2$_1$ (1)
a (Å)	28.886	28.968
b (Å)	34.815	35.094
c (Å)	97.965	97.435
Wavelength (Å)	1.5418	1.5418
Resolution range (Å)	32.66–1.90	35.17–2.40
Observation (I/σ (I) > 0)	29470	55067
Unique reflections (I/σ (I) > 0)	8169	4255
Redundancy	3.6	12.9
Last shell (Å)	2.00–1.90	2.53–2.40
R$_{sym}$ (%)^a: overall (last shell)	5.0 (12.1)	11.8 (26.8)
$\langle$I/σ (I)$\rangle$: overall (last shell)	9.0 (4.8)	5.4 (2.6)
Completeness (%): overall (last shell)	98.2 (98.2)	100 (100)
B. Phasing		
Iodine atom sites	7	
Resolution range of data used (Å)	32.66–3.0	
R$_{cullis}$	Iso	ano
	0.764	0.754
Phasing power	Iso	ano
	1.204	1.193
FOM	0.52	
Density modification, FOM	0.79	
C. Structure refinement		
Resolution range (Å)	28.38–1.90	
Sigma cutoff	0.0	
R-factor (%)	16.8	
R$_{free}$^{b} (%)	21.2	
Ramachandran statistics (%)^c		
Most favored regions	94.3	
Additionally allowed regions	4.8	
Generously allowed regions	0	
Disallowed regions	1.0	

^{a}R$_{sym}$ = $\Sigma(|I| - |\langle I \rangle|)/\Sigma|I|$.
^{b}R$_{free}$ calculated using 10% of total reflections omitted from refinement.
^{c}Ramachandran statistics calculated using PROCHECK.

oscillation per frame. Each frame was exposed for 3 min and 110 frames were collected. *MOSFLM* (V6.2.2), *TRUNCATE* and *SCALA* from *CCP4* program suite V.4.2.2[7] were used for processing, reduction, and scaling of the diffraction data.

Structure determination and refinement: The structure of h-SH3BGRL was successfully determined by the SIRAS method. The solution of h-SH3BGRL met difficulty with MAD method because there is only one methionine in the molecule. It was also troublesome in preparation of heavy-atom derivatives in such a crystallization condition containing 1.3 M tris-sodium citrate. So we soaked the native crystals for 72 h in the mother liquor containing the KI reagent at 0.1 M concentration. Seven possible iodine sites were found by ShelxD using the SIRAS method. Calculating phases and density modification were achieved by SHARP.[8] The model was manually built using O.[9] The structure was refined to a resolution of 1.9 Å with a R-factor of 0.168 and R$_{free}$ of 0.212 using CNS version 1.1. The final model contains 113 ordered residues (with the first methionine missing), one citrate molecule, 192 water molecules and a his tag consisting of 6 His residues. There was one molecule in an asymmetric unit. The quality of the structure was checked using PROCHECK. The statistics are listed in Table I.

Results and Discussion. The tertiary structure of h-SH3BGRL displays a N-terminal domain from residue 2 to 88 and a distinct C-terminal lobe from residue 89 to 114 [Fig. 1(A)]. The N-terminal domain consists of a $\beta\alpha\beta$ motif (β_1 3–7, α_1 13–30 and β_2 34–38) connected to a $\beta\beta\alpha$ motif (β_3 68–71, β_4 74–78 and α_4 79–88) through a solvent-exposed region that incorporates another helix α_2 (42–53), and a distinct short helix α_3 (54–58). The four-stranded β-sheet with topology −1x, +2x, +1, where β_1, β_2, and β_4 are antiparallel with β_3, are surrounded by three flanking helices (α_1, α_2, and α_4) [Figs.1(A) and 2 (B)]. This structural feature is generally consistent with the typical thioredoxin fold.[6] The C-terminal lobe consists of two stretches of α-helix, α_5 and α_6, which are linked by a six residue loop to form a helix-loop-helix motif [Fig.1(A)]. This structural character, first observed in thioredoxin fold proteins, is the most striking feature of the h-SH3BGRL structure. The α_6 as a rather long helix is well stabilized in the crystal lattice by five pairs of hydrogen bonds between the helix and three symmetric molecules [Fig.1(B)]. The interactions mainly involve residues in his-tag (his 117, 119 and 121).

So far over 143 thioredoxin folds have been found that belong to the Thioredoxin-like homologs, and they have been classified into 35 sequence superfamilies (CATH version 2.5.1). Compared with their sequences using SSAP, h-SH3BGRL has less than 21% sequence identity to each representative member of the 35 sequence families for at least 60% residues of the smaller aligned protein, which indicates that h-SH3BGRL should be assigned to a new sequence family of the homologous superfamily of Thioredoxin-like.

According to both structure and function differences, thioredoxin folds have been separated into one of five classes: glutaredoxin, thioredoxin, glutathione S-transferase, DsbA, or glutathione peroxidase.[6] The DALI structure similarity search of h-SH3BGRL revealed that the closest structural homolog with definite function was the oxidized bacteriophage T4 glutaredoxin[10] (Grx) (PDB ID, 1ABA; Z score, 8.5; sequence identity, 18% for 76 equivalenced residues; RMSD, 2.7 Å for 76 C$_\alpha$ atoms). In contrast, h-SH3BGRL displays few structural similarities with the other four classes of thioredoxin fold. Compared with the structural characterization of the Grx proteins, h-SH3BGRL is distinct in the following respects. Firstly, in h-SH3BGRL there are three additional helices, α_3, α_5 and α_6 [Fig. 2(A,B)]. The tiny α_3 helix is located in between the conserved α_2 and β_3 and a loop of six residues connects the helix α_5 with the rather long helix α_6 to form a helix–loop–helix motif. In addition, the local structures relevant to the functional performance in Grx proteins are significantly changed in h-SH3BGRL. The Grx proteins display a conserved sequence and structural motif Cys-X-X-Cys (X indicates any amino acid) at the N-terminal of helix α_1 (residue 14-17 according to T4 Grx), which is the active site for glutathione binding and redox reaction.[6] However, this motif Cys-X-X-Cys is completely absent in h-SH3BGRL (Fig. 3).

According to the criteria of the classification of thioredoxin fold proteins,[6] h-SH3BGRL should present a novel
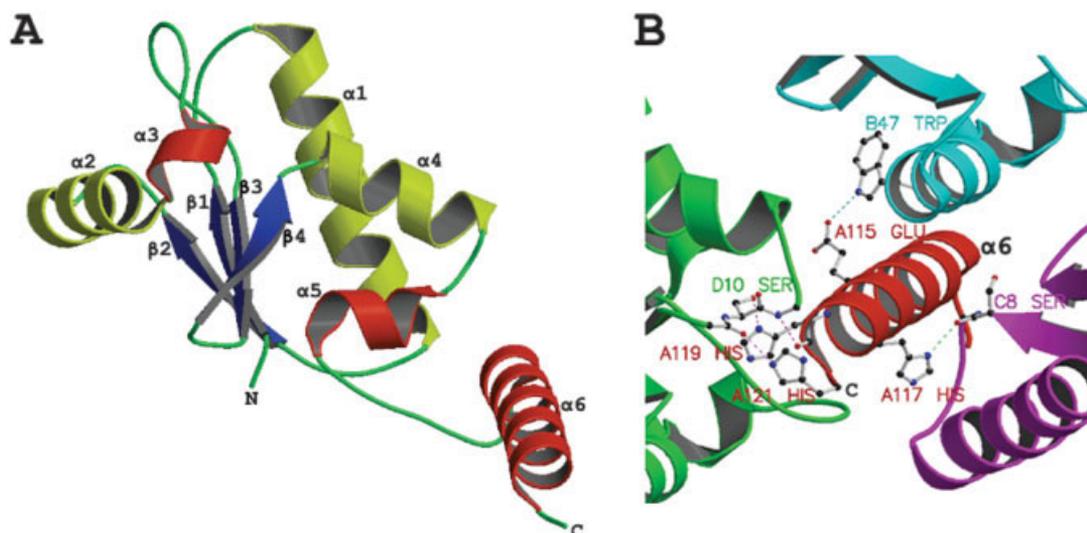
Fig. 1. **A**: Overall structure of h-SH3BGRL. The prototype of thioredoxin fold is shown in blue and yellow (blue for four β sheets and yellow for three α helices). The distinct insertion of secondary structures are shown in red, which include a tiny helix α3 and two helices, α5 and α6, in a helix–loop–helix motif at the C-terminal lobe. **B**: Interactions between the C-terminal helix 6 and three symmetric molecules in crystal lattice. The helix 6 is well stabilized by five pairs of intermolecular hydrogen bonds. Three symmetric molecules are respectively shown in cyan, magenta, and green, and the helix α6 in red.
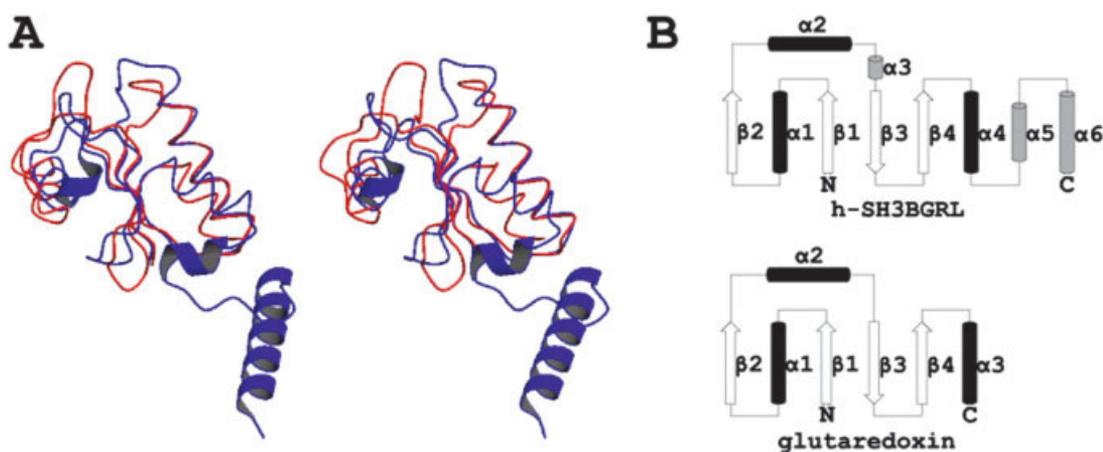
Fig. 2. **A**: Structural comparison of h-SH3BGRL with T4 Grx. Stereo diagram of superposition between Cα traces of h-SH3BGRL (blue) and T4 Grx (red). The unique structural elements (α3, α5 and α6) of h-SH3BGRL are indicated by Ribbon diagram. **B**: Topology diagram of h-SH3BGRL in comparison with that of Grx. The secondary structures distinct in h-SH3BGRL are shown in gray.

class of the thioredoxin fold proteins due to its distinct general architecture and abnormal local structures relative to the active sites of thioredoxin fold proteins. Most recently the crystal and NMR solution structures of a mouse SH3BGRL3 (m-SH3BGRL3) protein were determined.[11] The m-SH3BGRL3 consists of 93 residues and shows the sequence identity of 35% to h-SH3BGRL according to the PSI-BLAST. The general structural features of m-SH3BGRL3 are similar to those of h-SH3BGRL, except for deletions of the tiny helix (α3) and about 20 residues in the C-terminal lobe that forms a stretch of helix (α_6) in h-SH3BGRL. The active site (Cys-X-X-Cys) is also completely missing in m-SH3BGRL3. Reasonably, m-SH3BGRL3 should also belong to the structural class

represented by h-SH3BGRL. It appears that the h-SH3BGRL protein may be involved in redox-related bioprocesses in a unique way through a special structural framework. The 3D structure of h-SH3BGRL will provide a sound basis for the further studies to define its biological function and structure-function relationship.
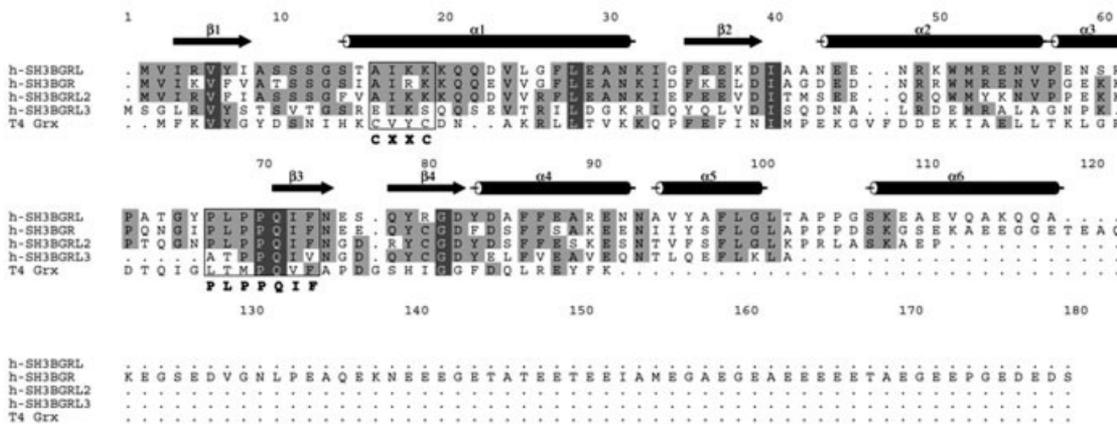
Fig. 3. ClustalW alignment of the amino acid sequences of the h-SH3BGRL family with T4 Grx. Identical and conserved residues are shaded in black and gray, respectively. The conservative proline-rich motif PLPPQIF in the h-SH3BGRL family and the catalytic motif Cys-X-X-Cys of Grx are boxed.

REFERENCES

1. Egeo A, Mazzocco M, Arrigo P, Vidal-Taboada JM, Oliva R, Pirola B, Giglio S, Rasore-Quartino A, Scartezzini P. Identification and characterization of a new human gene encoding a small protein with high homology to the proline-rich region of the SH3BGR gene. Biochem Biophys Res Commun 1998;247:302–306.
2. Mazzocco M, Maffei M, Egeo A, Vergano A, Arrigo P, Lisi RD, Ghiotto F, Scartezzini P. The identification of a novel human homologue of the SH3 binding glutamic acid-rich (SH3BGR) gene establishes a new family of highly conserved small proteins related to thioredoxin superfamily. Gene 2002;291:233–239.
3. Powson T, Scott JD. Signaling through scaffold, anchoring, and adaptor proteins. Science 1997;278:2075–2080.
4. Niebuhr K, Ebel F, Frank R, Reinhard M, Domain E, Carl UD, Walter U, Gertler FB, Wehland J, Chakraborty T. A novel proline-rich motif present in ActA of Listeria monocytogenes and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. EMBO J 1997;16:5433–5444.
5. Sudbrak R, Wieczorek G, Nuber UA, Mann W, Kirchner R, Erdogan F, Brown CJ, Wöhrle D, Sterk P, Kalscheuer VM, et al. X chromosone-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications. Hum Mol Genet 2001;10:77–83.
6. Martin JL. Thioredoxin-a fold for all reasons. Structure 1995;3:245–250.
7. Collaborative Computational Project Number 4, The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
8. De La Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. In: Sweet RM and Carter CW, editors. Methods in enzymology, Volume 276. Macromolecular crystallography. New York: Academic Press; 1997. p 472–494.
9. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and thelocation of errors in these models. Acta Crystallogr A 1991;47:110–119.
10. Eklund H, Ingelman M, Soderberg BO, Uhlin T, Nordlund P, Nikkola M, Sonnerstam U, Joelson T, Petratos K. Structure of oxidized bacteriophage T4 glutaredoxin (thioredoxin). J Mol Biol 1992;228:596–618.
11. Nardini M, Mazzocco M, Massaro A, Maffei M, Vergano A, Donadini A, Scartezzini P, Bolognesi M. Crystal structure of the glutaredoxin-like protein SH3BGRL3 at 1.6 Å resolution. Biochem Biophys Res Commun 2004;247:302–306.