FEBS Letters 578 (2004) 236–238 FEBS 29028

Identification and characterization of D8C, a novel domain present in liver-specific LZP, uromodulin and glycoprotein 2, mutated in familial juvenile hyperuricaemic nephropathy

Huirong Yang, Chaoqun Wu, Shouyuan Zhao, Jinhu Guo*

State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Handan Rd 220, Shanghai 200433, PR China

Received 15 August 2004; revised 28 October 2004; accepted 29 October 2004

Available online 16 November 2004

Edited by Gerrit van Meer

Abstract Present work reported a novel domain – D8C (domain with conserved eight cysteines in liver-specific ZP domain-containing protein, glycoprotein 2 (GP-2) and uromodulin (UMOD)), present in liver-specific LZP, UMOD, GP-2 and some uncharacterized proteins, most of which are membrane proteins, extracellular proteins or nuclear membrane proteins. D8C contains eight well-conserved cysteine residues, which were predicted to form four pairs of disulfide bridges. D8C is composed mainly of β-strands. Mutation in the D8C at Cys217 in human UMOD is associated with familial juvenile hyperuricaemic nephropathy, which might be due to the disruption of the disulfide bridge. Identification of D8C would further the understandings of related proteins.

Keywords: D8C; Domain; LZP; Uromodulin; Glycoprotein 2; Familial juvenile hyperuricaemic nephropathy

1. Introduction

Liver-specific ZP domain-containing protein (LZP) is expressed specifically in liver, although it can be secreted into blood. Expression of LZP is down-regulated in hepatocellular carcinoma (HCC) and HCC cell lines; meanwhile, which might be caused, at least in part, by the decreased level of *LZP* mRNA. LZP might be implicated in hepatocellular function and development [1].

The major protein present in isolated zymogen granule membranes of the exocrine pancreas is GP-2 (glycoprotein 2), accounting for up to 40% of the protein [2]. GP-2 is attached to the granule membrane via a glycosyl phosphatidylinositol (GPI) linkage and can be released from the membrane by phosphatidylinositol-specific phospholipase C [3]. A soluble form of GP-2 is also present in the content of the granules [4]. And in acinar cells, it localizes specifically to the secretory granules [4]. GP-2 in pancreatic secretions is further modified in that it is found as a sedimentable protein aggregate in the pancreatic juice [4]. The protein is heavily glycosylated with *N*-linked carbohydrates accounting for about 15–20% of its molecular weight as determined by gel electrophoresis [5].

Uromodulin (UMOD)/Tamm–Horsfall protein, an 85-kDa glycoprotein synthesized by the kidney, shares several characteristics with GP-2 in addition to its sequence similarity. Both are attached to the membrane by a GPI anchor, but are also released from the apical surface of their respective cells and subsequently form large aggregates. Together they may define a new gene family [6–9].

Here we reported a novel domain present in LZP, UMOD, GP-2 and some other uncharacterized proteins. Identification of this domain would further the understandings of the proteins with D8C domain.

2. Materials and methods

LZP is a recently identified liver-specific protein, possibly involved in hepatocellular function and development [1]. Interestingly, while comparing the domain architectures of LZP and UMOD, we found that these two proteins have similar domain composition, but that the location of the three EGF repeats varies. Furthermore, we found that the human LZP (gi|22749297) comprises a cysteine-rich region that, according to searches against the Pfam [10] and SMART [11] databases, does not match any previously identified domain. Therefore we carried out further investigation to determine whether this region represents a novel domain. PSI-BLAST [12] searches, using the sequence of gi|22749297 (21–144 aa), against the non-redundant protein database at GenBank (http://www.ncbi.nlm.nih.gov/blast/) were performed, using an inclusion threshold of 0.005. These searches revealed homology to several proteins. After four iterations, the results of the PSI-BLAST searches converged and retrieved 30 proteins in total. Examples of these sequences include: gi|22749297, *Homo sapiens*, oncoprotein-induced transcript 3, LZP (*E*-value 6×10^{-60}); gi|4507833. *H. sapiens*, UMOD (*E*-value 5×10^{-53}); gi|34534645, *H. sapiens*, unnamed protein product (*E*-value 3×10^{-49}); gi|34858746, *Rattus norvegicus* hypothetical protein XP_346554 (*E*-value 3×10^{-49}); gi|47210737, *Tetraodon nigroviridis*, unnamed protein product (*E*-value 5×10^{-25}) as well as gi|22761725, *H. sapiens*, unnamed protein product (*E*-value 2×10^{-34}).

In total, 21 distinct proteins, representing different molecules, were obtained from the 30 retrieved proteins. This region overlapped with known EGF domains (Smart: SM00181) in some of these sequences, e.g., gi|4507833, gi|137116 and gi|23243412; therefore we trimmed the N′ terminus (for instance, 21–54 aa in sequence gi|22749297 was removed) before generating a multiple sequence alignment using Clustal X [13] and manual editing (Fig. 1). The alignment was colored using Chroma [14]. The sequence alignment was deposited at the EBI database (ftp://ftp.ebi.ac.uk/pub/databases/embl/align/) with alignment number ALIGN_000740.

We named this region the D8C domain (domain with conserved eight cysteines in LZP, GP-2 and UMOD), after the eight well-conserved cysteine residues. To predict the secondary structure of D8C, the profile of the alignment was submitted to Jpred (http://www.compbio.dundee.ac.uk/~www-jpred/submit.html) [15]. To predict possible

*Corresponding author. Fax: +86 21 65642428.
E-mail address: jhguo@fudan.edu.cn (J. Guo).

0014-5793/$22.00 © 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2004.10.092

disulfide bridges in D8C domain, the sequence of gi|4507833 (199–287 aa) was analyzed via the DCON web server (http://gpcr.biocomp.unibo.it/oldpredictors/prototypes.html).

3. Results and discussion

3.1. Features of D8C domain

As a result of PSI-BLAST searches and manual verification, we identified the D8C domain, which contains eight well-conserved cysteine residues. The D8C domain is about 130 amino acid residues in length, and is present in mainly in two classes of proteins: class 1 includes LZP, UMOD and GP-2; class 2 includes some uncharacterized proteins, e.g., gi|49901397, gi|22761725, gi|50732549 and gi|47218596 (Fig. 1). Most of these proteins are membrane or extracellular proteins with signal peptides, and of these proteins human LZP is a nuclear membrane protein that can also be secreted into blood (Fig. 2) [1]. According to their annotation, most protein sequences in class two were derived from in silico prediction and need further verification.

Domains and motifs related to D8C include the signal peptide motif, the EGF domain (SMART: SM00181), the ZP domain [SMART: SM00241, zona pellucida (ZP) domain] and the VWD domain [SMART: SM00216, von Willebrand factor (vWF) type D domain] [16]. One of the functions of vWF is to serve as a carrier of clotting factor VIII (FVIII). The native conformation of the VWD domain of vWF is not only required for FVIII binding but also for normal multimerization and optimal secretion [17]. The domain common to all these proteins is located in the C-terminal portion of the extracellular region and contains 8 conserved Cys residues, which are probably involved in disulfide bond formation. The GPI-anchor contributes to

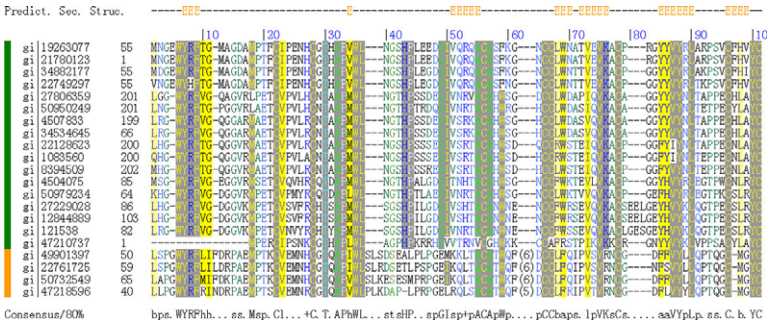

Fig. 1. Alignment of D8C domain, obtained by Clustal X and colored by Chroma software. The secondary structure was predicted by JPRED tool. The sequences are: gi|19263077, *Mus musculus*, LZP variant I; gi|21780123, *M. musculus*, LZP variant II; gi|34882177, *R. norvegicus*, similar to oncoprotein-induced transcript 3; gi|22749297, *H. sapiens*, LZP; gi|27806359, *Bos taurus*, UMOD; gi|50950249, *Canis familiaris*, UMOD precursor; gi|4507833, *H. sapiens*, UMOD; gi|34534645, *H. sapiens*, unnamed protein product; gi|22128623, *M. musculus*, UMOD; gi|1083560, *M. musculus*, UMOD precursor, gi|8394509, *R. norvegicus*, UMOD; gi|4504075, *H. sapiens*, GP-2; gi|50979234, *C. familiaris*, zymogen granule membrane protein GP-2; gi|27229028, *M. musculus*, GP-2; gi|12844889, *M. musculus*, unnamed protein product; gi|121538, *R. norvegicus*, secretory (zymogen) granule membrane glycoprotein GP2; gi|47210737, *T. nigroviridis*, unnamed protein product; gi|49901397, *M. musculus*, unknown protein; gi|22761725, *H. sapiens*, unnamed protein product; gi|50732549, *Gallus gallus*, similar to Wnt inhibitory factor 1 precursor (WIF-1); gi|47218596, *T. nigroviridis*, unnamed protein. These sequences could be classified into two classes, which were denoted in green and orange on the left of the alignments, respectively.

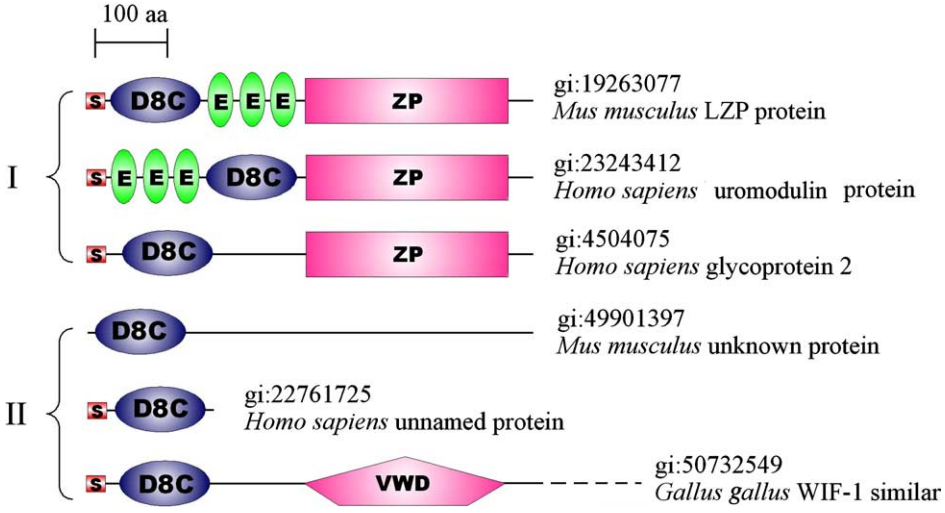

Fig. 2. Domain architecture of representative proteins with D8C domain. There are two classes of proteins containing D8C domain (I and II). S indicates signal peptide, E indicated EGF domain (Smart: SM00181), ZP indicates ZP domain (Smart: SM00241, ZP domain) and VWD indicates VWD domain (Smart: SM00216, vWF type D domain).

the localization at membrane in some of the D8C containing proteins, for instance, GP-2 [18].

The secondary structure of D8C domain contains seven β-strands (Fig. 1). Prediction of disulfide bond results demonstrated that in human UMOD, the cysteine residue pairs of Cys217 and Cys 223 (possibility value: 0.533), Cys 248 and Cys 256 (possibility value: 0.242), Cys 255 and Cys 267 (possibility value: 0.067) as well as Cys 282 and Cys 287 (possibility value: 0.677) have the potential to form disulfide bridges.

The D8C domain shows typical arrangement with other domains in the contexts of LZP, UMOD and GP-2 proteins, especially in LZP and UMOD proteins. In LZP proteins, D8C is upstream of three EGF domains and ZP domain at the N terminus, while in UMOD proteins D8C is between three tandem arranged EGF domains and the ZP domain.

3.2. Functional postulation of D8C domain

The D8C contains four pair of disulfide bridges, disruption of which might lead to functional defect. It has been reported that missense mutation in the D8C at Cys217 (g.2086T > C: p.C217R) in human UMOD (the first conserved cysteine residue) is associated with familial juvenile hyperuricaemic nephropathy but not medullary cystic kidney disease 2 [9]. The heterozygous missense mutation at Cys217 might disrupt the disulfide bond, leading to the depletion of the normal function D8C.

Discovery of D8C domain would be of importance for the research into structure and roles of related proteins.

Acknowledgments: We thank Dr. David J. Studholme (Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, UK) for his kindly revision of present paper, and thank Dr. Zeguang Han (National South Research Center of Human Genome, Shanghai, PR China) for providing related reference. This work was supported by the National Natural Scientific Foundation (No. 30170866, No. 90408000 and No. 30130100), PR China, and the Graduate Innovation Foundation of Fudan University (No. CQH1322014), PR China.

References

[1] Xu, Z.G., Du, J.J., Zhang, X., Cheng, Z.H., Ma, Z.Z., Xiao, H.S., Yu, L., Wang, Z.Q., Li, Y.Y., Huo, K.K. and Han, Z.G. (2003) A novel liver-specific zona pellucida domain containing protein that is expressed rarely in hepatocellular carcinoma. Hepatology 38, 735–744.

[2] Ronzio, R.A., Kronquist, K.E., Lewis, D.S., MacDonald, R.J., Mohrlok, S.H. and O'Donnell, J.J. (1978) Glycoprotein synthesis in the adult rat pancreas. IV. Subcellular distribution of membrane glycoproteins. Biochim. Biophys. Acta 508, 65–84.

[3] Fukuoka, S. and Scheele, G. (1990) Nucleotide sequence encoding the major glycoprotein (GP2) of rat pancreatic secretory (zymogen) granule membranes. Nucleic Acids Res. 18, 5900.

[4] Hoops, T.C. and Rindler, M.J. (1991) Isolation of the cDNA encoding glycoprotein-2 (GP-2), the major zymogen granule membrane protein. Homology to uromodulin/Tamm–Horsfall protein. J. Biol. Chem. 266, 4257–4263.

[5] Havinga, J.R., Strous, G.J. and Poort, C. (1983) Biosynthesis of the major glycoprotein associated with zymogen-granule membranes in the pancreas. Eur. J. Biochem. 133, 449–454.

[6] Pennica, D., Kohr, W.J., Kuang, W.J., Glaister, D., Aggarwal, B.B., Chen, E.Y. and Goeddel, D.V. (1987) Identification of human uromodulin as the Tamm–Horsfall urinary glycoprotein. Science 236, 83–88.

[7] Hession, C., Decker, J.M., Sherblom, A.P., Kumar, S., Yue, C.C., Mattaliano, R.J., Tizard, R., Kawashima, E., Schmeissner, U. and Heletky, S. (1987) Uromodulin (Tamm–Horsfall glycoprotein): a renal ligand for lymphokines. Science 237, 1479–1484.

[8] Rindler, M.J., Naik, S.S., Li, N., Hoops, T.C. and Peraldi, M.N. (1990) Uromodulin (Tamm–Horsfall glycoprotein/uromucoid) is a phosphatidylinositol-linked membrane protein. J. Biol. Chem. 265, 20784–20789.

[9] Hart, T.C., Gorry, M.C., Hart, P.S., Woodard, A.S., Shihabi, Z., Sandhu, J., Shirts, B., Xu, L., Zhu, H., Barmada, M.M. and Bleyer, A.J. (2002) Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. J. Med. Genet. 39, 882–892.

[10] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The Pfam protein families database. Nucleic Acids Res. 30, 276–280.

[11] Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P. and Bork, P.R. (2004) SMART 4.0: towards genomic data integration. Nucleic Acids Res. (Database issue), D142–144.

[12] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

[13] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

[14] Goodstadt, L. and Ponting, C.P. (2001) CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics 17, 845–846.

[15] Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14, 892–893.

[16] Bork, P. and Sander, C. (1992) A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor. FEBS Lett. 300, 237–240.

[17] Voorberg, J., Fontijn, R., van Mourik, J.A. and Pannekoek, H. (1990) Domains involved in multimer assembly of von willebrand factor (vWF): multimerization is independent of dimerization. EMBO J. 9, 797–803.

[18] Kinoshita, T. and Takeda, J. (1994) GPI-anchor synthesis. Parasitol. Today 10, 139–143.