VERSITA

Cent. Eur. J. Biol. • 5(4) • 2010 • 407–420
DOI: 10.2478/s11535-010-0039-8

Central European Journal of **Biology**

Homology modeling of 3D structure of human chitinase-like protein CHI3L2

Research Article

Anton Iershov[1,*], Konstantin Odynets[2], Alexander Kornelyuk[2], Vadim Kavsan[1]

[1]*Department of Biosynthesis of Nucleic Acids, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kyiv 03680, Ukraine*

[2]*Department of Protein Engineering and Bioinformatics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kyiv 03680, Ukraine*

Received 7 September 2009; Accepted 6 April 2010

Abstract: The human genome encodes six proteins of family 18 glycosyl hydrolases, two active chitinases and four chitinase-like lectins (chi-lectins) lacking catalytic activity. The present article is dedicated to homology modeling of 3D structure of human chitinase 3-like 2 protein (CHI3L2), which is overexpressed in glial brain tumors, and its structural comparison with homologous chi-lectin CHI3L1. Two crystal structures of CHI3L1 in free state (Protein Data Bank codes 1HJX and 1NWR) were used as structural templates for the homology modeling by Modeller 9.7 program, and the best quality model structure was selected from the obtained model ensemble. Analysis of potential oligosaccharide-binding groove structures of CHI3L1 and CHI3L2 revealed significant differences between these two homologous proteins. 8 of 19 amino acid residues important for ligand binding are substituted in CHI3L2: Tyr^{34}/Asp^{39}, Trp^{69}/Lys^{74}, Trp^{71}/Lys^{76}, Trp^{99}/Tyr^{104}, Asn^{100}/Leu^{105}, Met^{204}/Leu^{210}, Tyr^{206}/Phe^{212} and Arg^{263}/His^{271}. The differences between these residues could influence the structure of the ligand-binding groove and substantially change the ability of CHI3L2 to bind oligosaccharide ligands.

Keywords: *Chitinase 3-like 2 protein (CHI3L2, YKL-39) • Chitinase 3-like 1 protein (CHI3L1, YKL-40) • Chi-lectins • Homology modeling • Protein 3D structure • Oligosaccharide binding*

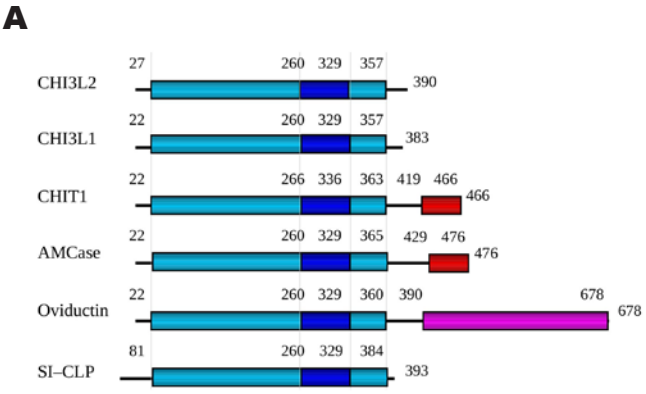
Abbreviations

AMCase – acidic mammalian chitinase;
CHI3L1 (YKL-40) protein – chitinase 3-like 1 protein;
CHI3L2 (YKL-39) protein – chitinase 3-like 2 protein;
CLP – chitinase-like protein;
NAG – β(1,4)-N-acetyl-D-glucosamine;
TIM – triosophosphateisomerase.

1. Introduction

Mammalian genomes, despite the lack of chitin (poly-N-acetyl-D-glucosamine) synthesis in mammals, code for a set of homologous chitinase-like proteins. Based on amino acid sequences and protein structure relationships, they have been grouped in the glycosyl hydrolase family 18, which also includes bacterial and plant chitinases [1]. In humans, there are six proteins of this family [2], and two of them, macrophage chitotriosidase-1 (chitinase-1, CHIT1) [3] and acidic mammalian chitinase (AMCase) [4] are catalytically active enzymes (EC 3.2.1.14).

* *E-mail: a.yershov@yahoo.com*



B

```
CHI3L1
MGVKASQTGF VVLVLLQCCS AYKLVCYYTS WSQYREGDGS CFPDALDRFL CTHIIYSFAN ISNDHIDTWE WNDVTLYGML  60
NTLKNRNPNL KTLLSVGGWN FGSQRFSKIA SNTQSRRTFI KSVPPFLRTH GFDGLDLAWL YPGRRDKQHF TTLIKEMKAE 160
FIKEAQPGKK QLLLSAALSA GKVTIDSSYD IAKISQHLDF ISIMTYDFHG AWRGTTGHHS PLFRGQEDAS PDRFSNTDYA 240
VGYMLRLGAP ASKLVMGIPT FGRSFTLASS ETGVGAPISG PGIPGRFTKE AGTLAYYEIC DFLRGATVHR ILGQQVPYAT 320
KGNQWVGYDD QESVKSKVQY LKDRQLAGAM VWALDLDDFQ GSFCGQDLRF PLTNAIKDAL AAT                   383

CHI3L2
MGATTMDQKS LWAGVVVLLL LQGGSAYKLV CYFTNWSQDR QEPGKFTPEN IDPFLCSHLI YSFASIENNK VIIKDKSEVM  80
LYQTINSLKT KNPKLKILLS IGGYLFGSKG FHPMVDSSTS RLEFINSIIL FLRNHNFDGL DVSWIYPDQK ENTHFTVLIH 160
ELAEAFQKDF TKSTKERLLL TAGVSAGRQM IDNSYQVEKL AKDLDFINLL SFDFHGSWEK PLITGHNSPL SKGWQDRGPS 240
SYYNVEYAVG YWIHKGMPSE KVVMGIPTYG HSFTLASAET TVGAPASGPG AAGPITESSG FLAYYEICQF LKGAKITRLQ 320
DQQVPYAVKG NQWVGYDDVK SMETKVQFLK NLNLGGAMIW SIDMDDFTGK SCNQGPYPLV QAVKRSLGSL             390
```

Figure 1. Domain structure of human chitinase-like proteins and amino acid sequences of CHI3L1 and CHI3L2 precursors. (A). The scheme of domain structure of six human proteins from family 18 glycosyl hydrolases. Main triosophosphateisomerase (TIM) β/α-barrel domains (CHI3L2 residues Tyr^{27}–Thr^{268}+Asp^{337}–Leu^{390}) and FK506 binding protein-like (FKBP-like) chitinase insertion domains (CHI3L2 residues Tyr^{269}–Tyr^{336}) are represented by blue and dark-blue colors, respectively. (B). Amino acid sequences of CHI3L1 and CHI3L2 precursors. Important amino acid residues and motifs are marked on both sequences: signal peptide sequences (orange), YKL – three N-terminal residues of mature proteins (red), chitinase insertion domain (dark blue), two pairs of cysteines, which form disulfide bonds (magenta) and heparin-binding sites (green).

Four other chitinase-like lectins (chi-lectins) are chitinase 3-like 1 protein (CHI3L1, YKL-40, HC gp-39) [5], chitinase 3-like 2 protein (CHI3L2, YKL-39) [6], oviductin (oviductal glycoprotein, oviduct-specific glycoprotein precursor, estrogen-dependent oviduct protein, mucin-9, OVGP1) [7] and less homologous chitinase-like protein, which interacts with endocytic/sorting receptor stabilin-1 (stabilin-1 interacting chitinase-like protein, SI-CLP) [8]. Several other mammalian proteins, encoded by chitinase-like genes, were also isolated and characterized [9-12]. Imaginal disk growth factor-2 of *Drosophila melanogaster* is more distant homologue of chi-lectins [13].

CHI3L1 is the most investigated protein among human chi-lectins. It has a molecular mass of about 40 kDa and is N-glycosylated at Asn^{60} (two residues of β(1,4)-N-acetyl-D-glucosamine, NAG). N-terminal amino acid residues (aa) of the mature form of CHI3L1 are tyrosine, lysine and leucine (Y, K, L), giving rise to its alternative name "YKL-40" [5]. CHI3L1 precursor consists of 383 aa and contains the signal peptide Met^{1}–Ala^{21} for its secretion and two structural domains, representing fold typical for mammalian and bacterial chitinases (Figure 1A). This chi-lectin acts as a proliferative [14], anti-apoptotic [15,16], migration and adhesion factor [17,18]. *CHI3L1* is expressed in synovial cells, articular cartilage chondrocytes, liver and macrophages during the last stage of differentiation. Increased levels of YKL-40 protein and/or mRNA are described for wide range of inflammatory conditions such as rheumatoid arthritis, osteoarthritis, sarcoidosis, inflammatory bowel disease, as well as for solid malignancies including primary breast cancer, small cell lung cancer, colorectal and ovarian carcinomas [19]. *CHI3L1* is overexpressed in glioblastomas in comparison with low-grade gliomas and normal brain tissue, and a correlation between high expression of *CHI3L1* and poor prognosis for patients is observed [20].

Despite detailed information about the structure of many chi-lectins, understanding of their physiological functions is still limited. Human CHI3L2/YKL-39 is the most closely related to CHI3L1 and has high sequence homology with other mammalian chi-lectins [6]. Its alternative name "YKL-39" refers to the same as in CHI3L1 three N-terminal residues of the mature protein after signal peptide (Met^{1}–Ala^{26}) cleavage. CHI3L2 is identified in conditioned medium from human articular cartilage chondrocyte primary culture as a protein that is copurified with CHI3L1. Its amount is approximately 4% of total proteins in chondrocyte-

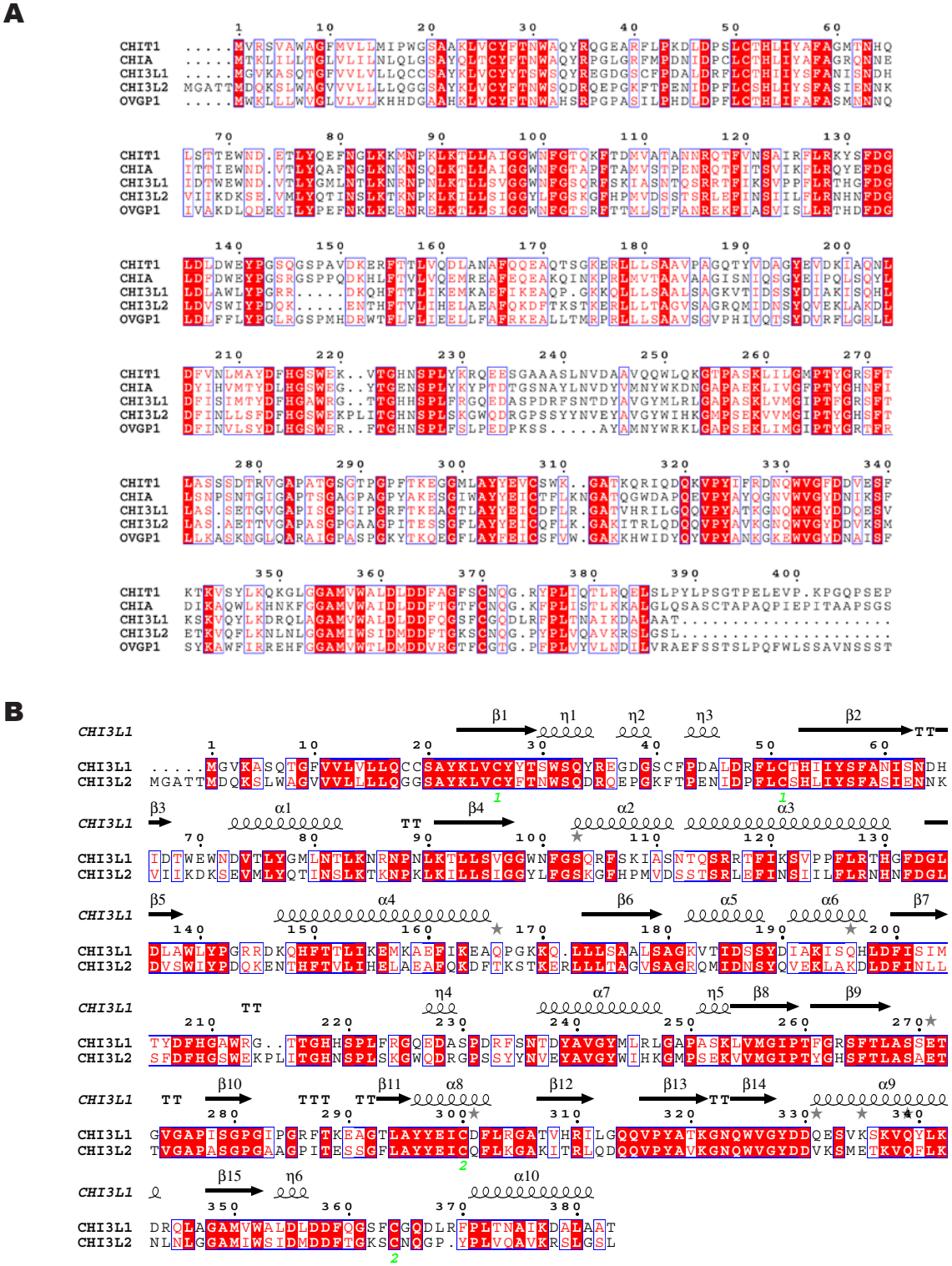


Figure 2. Amino acid sequence alignments of human chitinase-like proteins and CHI3L1/CHI3L2 precursors. (A). Multiple sequence alignment of five human chitinase-like proteins from ClustalW2 and ESPript servers. Identical and similar residues are marked by frames and inverted fonts; and insertions/deletions are marked by points. Protein RefSeq codes: CHI3L2 (NP_003991.2), CHI3L1 (NP_001267.2), CHIT1 (NP_003456.1), AMCase (NP_970615.2), and OVGP1 (NP_002548.3). (B). Pairwise sequence alignment of human CHI3L1 and CHI3L2 precursors. Identical and similar residues are marked by inverted fonts and frames, and insertions/deletions are marked by points. Secondary structure elements are for CHI3L1 free-state crystal structure (PDB code 1NWR:A). The alignment was used for CHI3L2 structure homology modeling by Modeller 9.7 program.

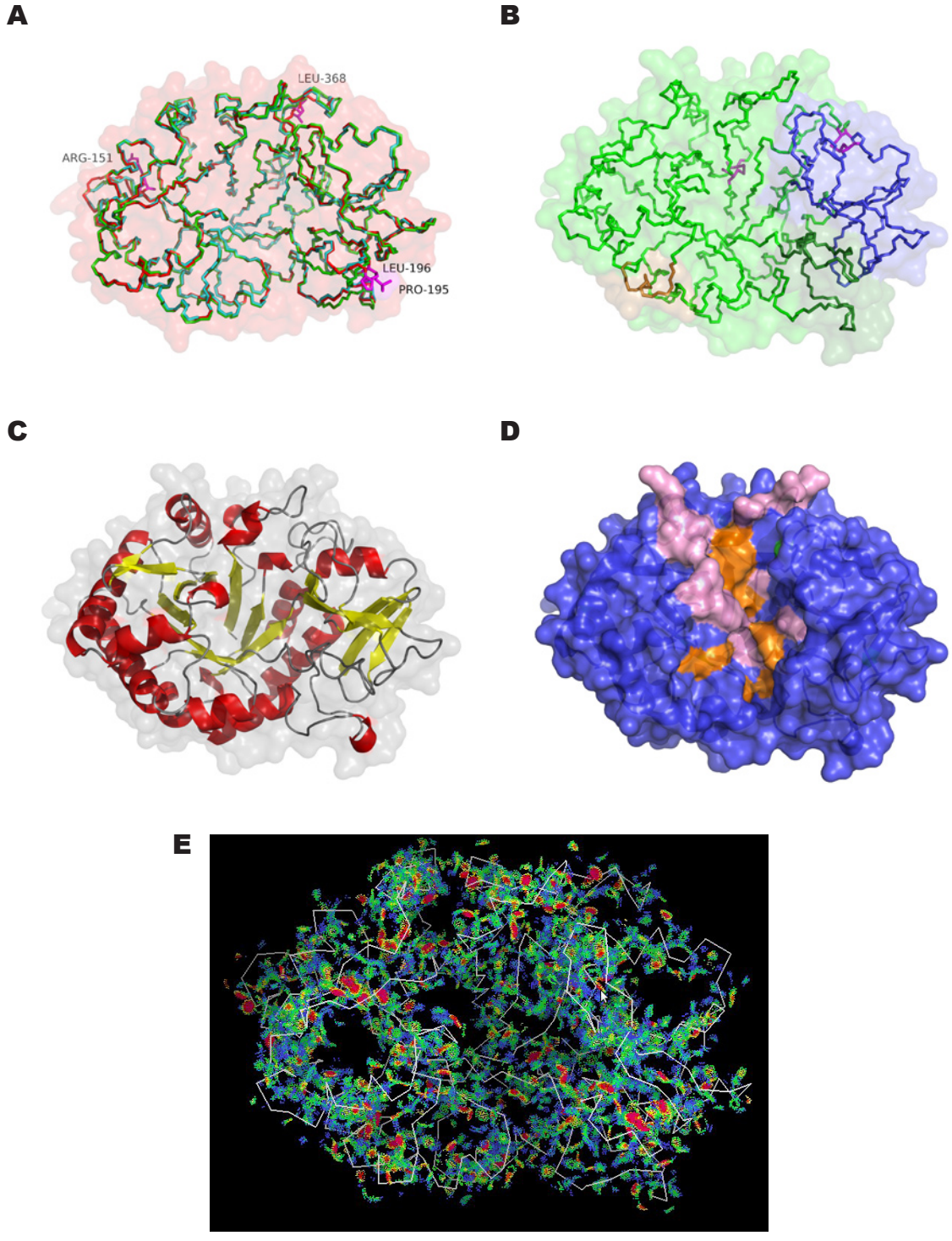


Figure 3. Model 1 of CHI3L2 mature form (region Tyr^{27}–Leu^{390}) built using two CHI3L1 structure templates (PDB codes 1HJX:A and 1NWR:A) by Modeller 9.7 program. (A). Superposition of secondary structures of model 1 (red) and both templates, 1HJX (green) and 1NWR (cyan), shown as backbone models. Two insertions (Arg^{177}, Pro^{221}–Leu^{222} in CHI3L2) and one deletion (Arg^{369} in CHI3L1) are shown as sticks in magenta. (B). Domain structure of CHI3L2 with main TIM β/α-barrel domain (green) and subdomains within it: heparin-binding site (copper), FKBP-like domain (blue) and putative sugar transporter domain (dark green). Two disulfide bridges are shown as sticks in magenta. (C). Secondary structure of CHI3L2 represented as cartoon model. Regions with α-helices are shown in red, β-strands are shown in yellow. (D). Molecular surface representation of CHI3L2 ligand-binding groove for 19 residues, which potentially form contacts with chitooligosaccharides. Residues, identical with CHI3L1 are shown in orange, and different residues are shown in pink. Surface exposed nsSNP (Asn^{126}-Asp, Arg^{318}-Trp) are shown in green. (E). MolProbity image representing clashes near putative ligand-binding site.

	CHI3L1 X-Ray structures				CHI3L2 homology model	
	1NWR:A		1HJX:A		Model 1 27–390 aa	
Resolution, nm	0.270		0.185		–	
R-value (obs.)	0.212		0.197		–	
R-Free	0.239		0.225		–	
Quality assessment scores						
	Mean score	Z-score	Mean score	Z-score	Mean score	Z-score
Procheck scores						
- Most favored regions	89.7	–	92.1	–	91.8	–
- Additionally allowed regions	9.7	–	7.8	–	7.0	–
- Generously allowed regions	0.6	–	0.1	–	1.2	–
- Disallowed regions	0	–	0	–	0	–
- G-factor (psi-psi only)	-0.31	-0.9	-0.1	-0.08	0.06	0.55
- G-factor (all dihedral angles)	-0.26	-1.54	-0.08	-0.47	0.01	0.06
ProsaII	0.55	-0.41	0.52	-0.54	0.62	-0.12
MolProbity clash score	22.89	-2.4	31.55	-3.89	56.99	-10.84

Table 1. Quality assessment scores for CHI3L2 homology model and two CHI3L1 template crystal structures obtained by PSVS meta-server [35] and MolProbity [36].

conditioned medium while the proportion of CHI3L1 is 33% [6]. The highest level of *CHI3L2* mRNA is detected in chondrocytes, synoviocytes, lungs and heart. No *CHI3L2* mRNA is detected in the brain, spleen, kidney, pancreas or liver. The *CHI3L2* gene, but not *CHI3L1*, is upregulated in chondrocytes of patients with osteoarthritis [21].

Our recent studies revealed that both *CHI3L1* and *CHI3L2* genes are upregulated in glioblastoma [22,23]. However, their expression levels differ and often are not simultaneous, at the level of both mRNA and protein.

Further, it is interesting to compare functional features of both homologous proteins encoded by these chitinase-like genes in terms of their 3D structures and oligosaccharide-binding properties. Now, seven crystal structures of CHI3L1 in free state and in complexes with different ligands have been determined (Protein Data Bank (PDB) codes 1HJV, 1HJW, 1HJX, 1NWR, 1NWS, 1NWT and 1NWU) [24,25]. Structures of human chitotriosidase and AMCase are also resolved as well as structures of significant number of signaling chi-lectins of other mammals in free state and in complexes with oligosaccharide ligands.

The main aim of this investigation is sequence-based prediction of 3D structure of mature human CHI3L2, closely related to CHI3L1–the most investigated protein in chitinase-like signal lectin family. Here we present homology modeling of 3D structure of CHI3L2 main secreted isoform A (region Tyr^{27}–Leu^{390}) based on two homologous free-state CHI3L1 crystal structures as templates and comparison of oligosaccharide-binding grooves of both chi-lectins.

2. Experimental Procedures

Homology search of human chitinase-like proteins was performed by Protein BLAST 2.2.18 service [26] available at the National Center for Biotechnology Information (NCBI). Amino acid sequences of conservative domains and domain structure of proteins were analyzed using CDD [27] and CD-Search [28] databases. The pairwise

Properties	CHI3L1		CHI3L2	
	Precursor 1–383 aa	1HJX:A 22–383 aa	Precursor 1–390 aa	Model 1 27–390 aa
Number of aa	383	362	390	364
ProtParam				
Molecular mass, Da	42613.4	40476.8	43500.7	40871.6
Theoretical isoelectric point (pI)	8.69	8.65	7.11	7.24
(Asp+Glu) / (Lys+Arg)	35 / 40	35 / 39	37 / 37	36 / 36
Number of atoms / non-hydrogen atoms	5946 / 3007	5641 / 2860	6106 / 3069	5729 / 2886
Surface area, nm^2	–	130.08	–	131.97
Volume, nm^3	–	51.950	–	51.460

Table 2. Properties of human CHI3L1 and CHI3L2 (isoform A) precursors and mature forms predicted by 3V server [37].

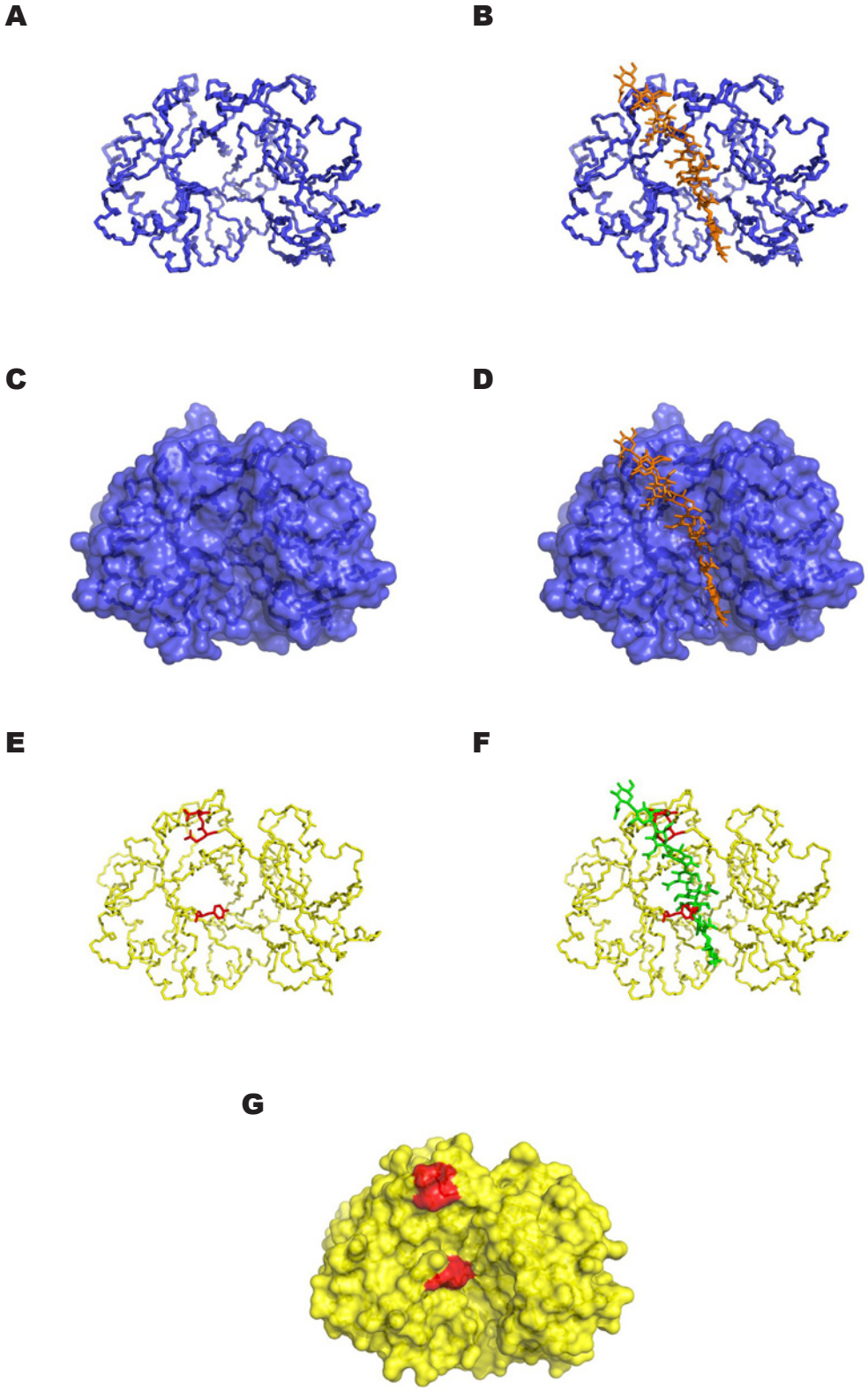


Figure 4. Comparison of CHI3L1 and a model CHI3L2 structures with and without ligands. (A,B). Stick representation of 7 superimposed CHI3L1 X-ray structures without and with chitooligosaccharide ligands. (C,D). Surface representation of CHI3L2 model without and with ligands (E,F). Stick representation of CHI3L1 without and with ligand. (G). Surface representation of CHI3L2 model. Amino acid residues which cause structural clashes are shown in red. See also Supplementary Figure 1 at the end of the text.

and multiple amino acid sequence alignments were created using ClustalW2 [29] and ESPript 2.2 [30] servers. Homologous structures search was performed by BLAST service against PDB database [31]. Coordinate files of protein crystal structures were obtained from PDB. Protein SCOP classification (v.1.75, June 2009) was used to determine definitions of structural domain families [32]. The CHI3L2 model ensemble was constructed using Modeller 9.7 program [33], and the best model was selected by the smallest value of normalized Discrete Optimized Molecule Energy (DOPE) score [34]. Structures were assessed by means of PSVS meta-server v.1.4 [35] and MolProbity [36]. Physico-chemical properties of the model and templates were calculated by 3V server [37]. Structural alignment of proteins was performed using FATCAT server [38], and their multiple superposition with CE-MC server [39]. The amino acid residues of human CHI3L1 in its complexes with chitin tetramers and chitin octamer, which are in contact with the ligands within distance threshold

Subsite	CHI3L1				CHI3L2 Model
	1HJV:A		1HJW:A		
	Residue	Interaction	Residue	Interaction	
-7	–		–		–
-6	Tyr^{34}	VdW	–		Asp^{39}
	Trp^{71}	**VdW**	**–**		**Lys^{76}**
-5	Tyr^{34}	VdW	–		Asp^{39}
	Glu^{70}	OE1–O7	–		Asp^{75}
-4	–		Glu^{70}	OE2–O6	Asp^{75}
-3	–		Trp^{31}	VdW	Trp^{36}
	–		**Trp^{69}**	**VdW**	**Lys^{74}**
	–		Asn^{100}	ND2–O7	Leu^{105}
-2	**Trp^{99}**	**N–O6**	**Trp^{99}**	**VdW**	**Tyr^{104}**
	Asn^{100}	N–O6 ND2–O4	Asn^{100}	N–O6	Leu^{105}
	Trp^{352}	NE1–O7	Trp^{352}	NE1–O7	Trp^{360}
	Leu^{356}	VdW	Leu^{356}	VdW	Met^{364}
-1	Tyr^{27}	VdW	Tyr^{27}	VdW	Tyr^{32}
	Trp^{99}	**N–O3**	**Trp^{99}**	**N–O3**	**Tyr^{104}**
	Met^{204}	VdW	Met^{204}	VdW	Leu^{210}
	Tyr^{206}	OH–O5	Tyr^{206}	VdW	Phe^{212}
	Asp^{207}	OD2–O6	Asp^{207}	OD2–O6	Asp^{213}
	Trp^{352}	VdW	Trp^{352}	VdW	Trp^{360}
1	**Trp^{99}**	**VdW**	**Trp^{99}**	**VdW**	**Tyr^{104}**
	Tyr^{141}	OH–O6	Tyr^{141}	OH–O6	Tyr^{146}
	Met^{204}	VdW	Met^{204}	VdW	Leu^{210}
	Asp^{207}	VdW	Asp^{207}	VdW	Asp^{213}
	Arg^{263}	VdW	Arg^{263}	VdW	His^{271}
2	Tyr^{141}	OH–O3	Tyr^{141}	OH–O3	Tyr^{146}
	Gly^{181}		Gly^{181}		Gly^{187}
	Phe^{208}	VdW	Phe^{208}	VdW	Phe^{214}

Table 3. The ligand-binding residues of two CHI3L1–chitooligosaccharide complexes, selected at distance threshold 0.34 nm by iMolTalk 3.2 server, and equivalent residues of CHI3L2. NAG(-4) and NAG(-3) rings are absent in crystal structure 1HJV:A, as well as NAG(-6) and NAG(-5) rings are absent in 1HJW:A. NAG(-7) does not interact with any CHI3L1 residue. In comparison to CHI3L1, CHI3L2 has three Trp residues replaced by Lys^{74}, Lys^{76} and Tyr^{104} marked in bold.

0.34 nm and form surface of ligand-binding subsites, were selected by iMolTalk server v.3.2 [40]. Structures were visualized in PyMOL v.1.2 [41]. Quantitative data were analyzed in Microsoft Office Excel.

3. Results

Search of homologous to CHI3L2 human chitinase-like proteins by BLASTP service against non-redundant database revealed five homologous proteins in human, CHI3L1 protein being the closest human homologue with 51.7% of sequence identity. To create the CHI3L2 model structure, we performed a BLAST search against the PDB database and selected two CHI3L1 structures in free state (PDB codes 1HJX:A and 1NWR:A) as templates for modeling.

The CHI3L2 precursor consists of 390 aa and contains the signal peptide Met^{1}–Ala^{26} for its secretion and two structural domains, which spread over the regions of Tyr^{27}–Thr^{268}+Asp^{337}–Leu^{390} and Tyr^{269}–Tyr^{336}, respectively (Figures 1A-B). Multiple sequence alignment of five homologous human chitinase-like proteins demonstrates their significant homology and presence of only short insertions/deletions (Figure 2A).

Building of the CHI3L2 3D structure in free state by homology modeling approach was performed by means of Modeller program. The mature CHI3L2 (Tyr^{27}–Leu^{390}) all-hydrogen model structure was constructed using two CHI3L1 crystal structures in free state (PDB codes 1HJX:A and 1NWR:A) as structural templates on the basis of their pairwise sequence alignment (Figure 2B). Conservative disulfide bridges (Cys^{31}-Cys^{56} and Cys^{308}-Cys^{372}) were described in the modeling script. The model 1 was selected from the ensemble of 100 alternative model structures as the best 3D structure according to the smallest value of its normalized DOPE score (-1.89) calculated by Modeller. Regions Trp^{36}-Ile^{51} and Thr^{171}-Leu^{180} of CHI3L2 forming surface loops show relatively low sequence homology to the corresponding regions of CHI3L1. After the model evaluation these protein regions appeared to have excess energy and were optimized by loop refinement procedure using molecular dynamics by Modeller; this resulted in lower normalized DOPE score (-1.91). Comparison of model 1 and 1HJX:A structures gives a root-mean-square deviation (RMSD) value of 0.022 nm for 360 pairs of equivalent C_α atoms (Figure 3A–D). Physico-chemical characteristics of the model and both templates show that CHI3L2 is more acidic; it has higher molecular mass while comparable protein volume and solvent accessible surface indicating that it may have more compact structure (Table 1).

Results of final geometry "quality" estimation of obtained CHI3L2 model structures and both CHI3L1 templates performed by means of PSVS server revealed that these structures corresponded to average statistical values for the most of globular proteins (Table 2). Procheck analysis revealed 91.8% of aa residues in the most favored regions and 7.0% of aa residues in additionally allowed regions. MolProbity analysis shows that there are no significant clashes near ligand-binding groove (Figure 3E).

Same as CHI3L1, CHI3L2 contains two disulfide bonds, Cys^{31}–Cys^{56} and Cys^{308}–Cys^{372}, which are evolutionally conservative in 33 mammalian chitinase and chi-lectin members of family 18 glycosyl hydrolases [2], particularly in 5 of 6 human proteins (except SI-CLP). CHI3L1 is glycosylated at Asn^{60} and contains two β(1,4)-linked residues of NAG. In contrast to CHI3L1, CHI3L2 has serine at this position (other chi-lectins contain serine, glycine or aspartic acid) and is not glycosylated [6].

Two antigenic epitopes described for CHI3L1 (Pro^{259}–Glu^{271} and Arg^{263}–Gly^{275}) [42,43] are created by β-strands β8 and β9 and adjacent surface loop in FKBP-like domain. The model region Pro^{259}–Phe^{265} is situated near oligosaccharide-binding groove, and the region Thr^{266}–Gly^{275} corresponded to exposed segment of protein surface [24]. Comparison of these regions for CHI3L2 showed that both segments (267–279 and 271–283) occupy similar positions within the structure of FKBP-like domain.

Reference non-synonymous single nucleotide polymorphisms (nsSNP) of human CHI3L2 were found in dbSNP (http://www.ncbi.nlm.nih.gov/snp) for CHI3L2 isoform A (NP_003391), namely: rs11556869 (Pro^{48}-Ser), rs35041930 (Asn^{126}-Asp), rs11556868 (Ala^{182}-Val), rs34049547 (Val^{184}-Ile), rs13721 (Arg^{318}-Trp). Only Arg^{318} and Asn^{126} are exposed on the protein surface, but the rest three residues are localized inside of protein structure (Figure 3D).

Houston *et al.* [24] showed that 19 aa of CHI3L1 ligand-binding groove take part in interactions with ligands by forming hydrogen bonds and hydrophobic interactions. However, according to our analysis using iMolTalk with default 0.34 nm threshold, only 14 aa of CHI3L1 form van-der-Waals contacts and nine correct hydrogen bonds with two chitin tetramers in protein-ligand complex (PDB code 1HJV:A), and 9 aa form eight hydrogen bonds with chitin octamer in other complex (1HJW:A) (Table 3). Notably, 11 aa (58%) among these 19 residues are different between CHI3L1 and CHI3L2; the most conservative residues lie on the bottom of the groove (Figure 3D).

Both complexes, 1HJV:A and 1HJW:A, contain different chitin oligomers bound in separate subsites of the CHI3L1 ligand-binding groove. To reconstruct the complete oligosaccharide structure in the ligand-binding site, both protein–ligand complexes were aligned with PyMOL. To examine similarities between CHI3L1 and CHI3L2 ligand-binding grooves we performed multiple structural alignment of two CHI3L1 structures complexed with chitooligosaccharides and CHI3L2 model 1 structure. Then the full reconstructed ligand was placed into CHI3L2 potential oligosaccharide-binding groove and interatomic distances were analyzed with iMolTalk. Several structure clashes were observed, namely side chain of Tyr^{104} (Trp^{99} in CHI3L1) makes sterical difficulties to arrange NAG(-1) and NAG(+1) rings (Figure 4 and Table 3), Gln^{41} and Glu^{42} (Glu^{36} and Gly^{37} in CHI3L1) disrupts NAG(-6) and NAG(-5) binding subsites. The comparison of ligand-binding residues positions in 1HJV:A complex and of the CHI3L2 model structure in free state shows that 8 of 19 (42%) residues are identical in both structures with exception for Tyr^{34}/Asp^{39}, Trp^{69}/Lys^{74}, Glu^{70}/Asp^{75}, Trp^{71}/Lys^{76}, Trp^{99}/Tyr^{104}, Asn^{100}/Leu^{105}, Met^{204}/Leu^{210}, Tyr^{206}/Phe^{212}, Arg^{263}/His^{271}, and Leu^{356}/Met^{364} substitutions (Figure 4). These substitutions could considerably change hydrophobicity and oligosaccharide-binding properties of ligand-binding groove in CHI3L2 as compared to CHI3L1.

4. Discussion

General organization of chi-lectins' structures is evolutionally conservative as evidenced by lack of significant changes in size of main and insertion domains, surface loops and corresponding insertions/deletions. Fold recognition analysis revealed that CHI3L2 has a typical fold of mammalian chitinase-like lectins with the main TIM barrel domain and inserted FKBP-like domain (containing one α-helix and six β-strands). The TIM barrel domain is highly conserved and associated with oligosaccharide-binding function in all mammalian chi-lectins. The FKBP-like domain is shown to bind the immunosuppressive drug FK506 and is found in proteins with prolyl isomerase activity [44]. The FKBP-like fold is characteristic for three protein superfamilies, namely chitinases, *E. coli* immune proteins and FKBP-like family, including immunophilins.

Comparison of physical properties of CHI3L1 and CHI3L2 (Table 2) shows that theoretical isoelectric point (pI) for CHI3L2 mature form is lower than for CHI3L1, and overall charges of both proteins are different due to the greater part of positively charged amino acid residues in CHI3L1. However, there is an important structural difference which has a great impact on physical and biological properties of the protein: CHI3L1 is glycosylated while CHI3L2 is not [6,25]. CHI3L1 is a secreted glycoprotein with pI of 7.6 [45]; in the same time, non-modified CHI3L1 has pI of 8.69. The protein is N-glycosylated at Asn^{60} [24,25], but glycosylation does not give such pI shift. High throughput mass-spectrometry analysis of protein phosphorylation allowed identification of phosphorylation sites in many proteins, including CHI3L1 [46]. PhosphoSite server allows predicting pI for three phosphorylation sites at Tyr^{189}, Tyr^{239} and Tyr^{243} with predicted pIs of 8.36, 7.81 and 7.27 for mono-, bi- and three-phosphorylated forms, respectively. Secreted protein CHI3L2 homologous to CHI3L1 (51.7% identity) has only one predicted phosphorylation site at Tyr^{82}, which is not equal to CHI3L1 phosphorylation sites.

CHI3L1 tends to aggregate and precipitate under low salt conditions [47] and forms dimers that are resolved in SDS-PAGE gel-electrophoresis under non-reducing conditions and disappear in the presence of dithiothreitol [48]. CHI3L1 modulates activity of basic fibroblast growth factor (bFGF) [49], binds to collagen types I, II, and III, and modulates the rate of collagen fibril formation [48]. No one of these interactions is shown for CHI3L2, and no protein-protein interaction partners were found up to now (particularly, there are no reports about participation of CHI3L2 in any protein-protein interaction in BioGRID, HPID, MIPS, MINT, IntAct, DIP, and BRITE databases).

The comparison of ligand-binding residues in CHI3L1 crystal structures and CHI3L2 model structure revealed that only some of ligand-binding residues are identical and occupy similar positions with the exception for Tyr^{34}, Trp^{69}, Trp^{71}, Trp^{99}, Asn^{100}, Met^{204}, Tyr^{206} and Arg^{263} in CHI3L1, which are substituted by Asp^{39}, Lys^{74}, Lys^{76}, Tyr^{104}, Leu^{105}, Leu^{210}, Phe^{212} and His^{271} (Figure 4 and Table 3). The significant structure differences between these CHI3L1/CHI3L2 residues reorganize tiny structure of ligand-binding grooves of both homologous proteins and may change their properties to bind different ligands.

It should be stated that the oligosaccharide-binding groove of CHI3L1 is lined with aromatic residues, which have hydrophobic stacking interactions with the hydrophobic sides of the bounded sugar rings. Among them, Trp^{99} and Trp^{352} additionally participate in ligand recognition by forming hydrogen bonds with the C3–OH of the NAG(-1) and the C7–O of the NAG(-2), respectively [24]. Trp^{352} is conservative in the CHI3L2 ligand-binding groove, but Trp^{99} is replaced by Tyr^{104}. It is important to note that Trp^{69} and Trp^{71} hydrophobic

residues of CHI3L1 are replaced by positively charged Lys^{74} and Lys^{76}, and both substitutions could strongly change the polarity and ligand-binding specificity of CHI3L2 ligand-binding groove. These changes are concerned mainly with (-3) and (-5) NAG-binding subsites, respectively.

CHI3L2, like *CHI3L1*, is overexpressed in astrocytomas grade II-IV, and its expression is absent in normal samples [23,50]. However, expression levels of *CHI3L1* and *CHI3L2* in glioma do not correlate (it was also shown at the protein level by Western blot analysis), thus proposing different functional roles of these chi-lectins. CHI3L2 is a more specific autoantigen than CHI3L1 in rheumatoid arthritis [51]. In addition, CHI3L2 instead of CHI3L1 is upregulated in osteoarthritic chondrocytes [21]. Interestingly, *CHI3L1*, *CHI3L2* and *LUM* were suggested as expansion-related genes in cultured human articular chondrocytes [52]. These genes were upregulated in response to treatment with growth factors (PDGF-BB, but not BMP4), while their levels remained unchanged during matrix depletion.

As previously shown, CHI3L2, like CHI3L1, may play a role in inflammatory and neoplastic processes. CHI3L1 acts synergistically with IGF-I in human fibroblasts and chondrocytes [14], but such experiments for CHI3L2 were not performed.

It is supposed that human CHI3L1 binds heparin [24], while there is no conclusive evidence of this. CHI3L1 presents a cluster of basic residues (GRRDKQH, Gly^{143}–Arg–Arg–Asp–Lys–Gln–His^{149}) similar to other heparin-binding proteins [24]. The putative heparin-binding motif of CHI3L1 lies in the surface loop after strand β5 and the surface-exposed part of α4 helix (Figures 1B, 2B and 3B). Most growth factors, such as EGF, FGF, VEGF, CTGF and HGF, contain heparin-binding sites; however, the sites are rather different with non-overlapping features. Commonly, the heparin-binding site serves as a connector between two proteins, for instance, in the case of FGF and FGFR [53]. Interestingly, other mammalian chitinase-like proteins including chitotriosidase, on the contrast to CHI3L1 and CHI3L2, have insertion of five hydrophobic amino acids into this site, that disrupt it (Figure 2A). It could be concluded that this feature is a unique, evolutionary gain-of-function characteristic of CHI3L1. However, the corresponding region of CHI3L2 (DQKENTH, Asp^{148}–Gln–Lys–Glu–Asn–Thr–His^{154}) has significant differences that could explain a lack of its involvement in heparin binding [6]. Alternatively, the oligosaccharide-binding groove of CHI3L2 could potentially bind heparan sulfates, especially due to the introduction of positively-charged Lys^{74} and Lys^{76} into its structure. Nevertheless, the putative involvement of CHI3L2 into heparan sulfate binding awaits further experimental verification.

Crystallization studies failed to reveal the CHI3L1 chi-lectin binding of sulfated oligosaccharides units within its ligand-binding groove [24]. Nevertheless, there are recent findings proposing that CHI3L1 acts through cell membrane receptor syndecan-1 [54]. Syndecans bear the most of heparan sulfate residues on the cell surface and are involved in angiogenesis and invasion [55]. CHI3L1 stimulates the association of syndecan with a_Vb_3 integrin, and treatment with heparinase or chondroitinase removes this effect.

Differential interactions of heparin and heparan sulfate proteoglycans were observed for selectins [56] and calcium-dependent C-type lectins – a family of glycoproteins mediating the adhesive events and involved in the metastasis of certain cancers. The similar selectivity related to heparin and heparan sulfate proteoglycans may be proposed as a working hypothesis explaining the difference of structures of oligosaccharide-binding grooves of CHI3L1 and CHI3L2 chi-lectins.

Acknowledgements

This work was supported by Ministry of Education and Science of Ukraine (Grant No M/119–2007 “Molecular markers of human glial tumors”).

References

[1] Henrissat B., Davies G., Structural and sequence-based classification of glycoside hydrolases, Curr. Opin. Struct. Biol., 1997, 7, 637-644

[2] Bussink A.P., Speijer D., Aerts J.M.F.G., Boot R.G., Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases, Genetics, 2007, 177, 959-970

[3] Boot R.G., Renkema G.H., Verhoek M., Strijland A., Bliek J., de Meulemeester T. M., et al., The human chitotriosidase gene. Nature of inherited enzyme deficiency, J. Biol. Chem., 1998, 273, 25680-25685

[4] Boot R.G., Blommaart E.F., Swart E., Ghauharali-van der Vlugt K., Bijl N., Moe C., et al., Identification of a novel acidic mammalian chitinase distinct from chitotriosidase, J. Biol. Chem.,2001,276, 6770-6778

[5] Millis A.J., Hoyle M., Kent L., In vitro expression of a 38,000 dalton heparin-binding glycoprotein by morphologically differentiated smooth muscle cells, J. Cell. Physiol., 1986, 127, 366-372
[6] Hu B., Trinh K., Figueira W.F., Price P.A., Isolation and sequence of a novel human chondrocyte protein related to mammalian members of the chitinase protein family, J. Biol. Chem., 1996, 271, 19415-19420
[7] Arias E.B., Verhage H.G., Jaffe R.C., Complementary deoxyribonucleic acid cloning and molecular characterization of an estrogen-dependent human oviductal glycoprotein, Biol. Reprod., 1994, 51, 685-694
[8] Kzhyshkowska J., Mamidi S., Gratchev A., Kremmer E., Schmuttermaier C., Krusell L., et al., Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in altenatively activated macrophages and secreted via lysosomal pathway, Blood, 2006, 107, 3221-3228
[9] Mohanty A.K., Singh G., Paramasivam M., Saravanan K., Jabeen T., Sharma S., et al., Crystal structure of a novel regulatory 40-kDa mammary gland protein (MGP-40) secreted during involution, J. Biol. Chem., 2003, 278, 14451-14460
[10] Srivastava D.B., Ethayathulla A.S., Kumar J., Singh N., Sharma S., Das U., et al., Crystal structure of a secretory signalling glycoprotein from sheep at 2.0A resolution, J. Struct. Biol., 2006, 156, 505-516
[11] Ethayathulla A.S., Srivastava D.B., Kumar J., Saravanan K., Bilgrami S., Sharma S., et al., Structure of the buffalo secretory signalling glycoprotein at 2.8 A resolution, Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun., 2007, 63, 258-265
[12] Sun Y.J., Chang N.C., Hung S.I., Chang A.C., Chou C.C., Hsiao C.D., The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site, J. Biol. Chem., 2001, 276, 17507-17514
[13] Varela P.F., Llera A.S., Mariuzza R.A., Tormo J., Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycoproteins from Drosophila melanogaster, J. Biol. Chem., 2002, 277, 13229-13236
[14] Recklies A.D., White C., Ling H., The chitinase 3-like protein human cartilage glycoprotein 39 (HC-gp39) stimulates proliferation of human connective-tissue cells and activates both extracellular signal-regulated kinase- and protein kinase B-mediated signalling pathways, Biochem. J., 2002, 365, 119-126
[15] Ling H., Recklies A.D., The chitinase 3-like protein human cartilage glycoprotein 39 inhibits cellular responses to the inflammatory cytokines interleukin-1 and tumour necrosis factor alpha, Biochem. J., 2004, 380, 651-659
[16] Junker N., Johansen J.S., Hansen L.T., Lund E.L., Kristjansen P.E.G., Regulation of YKL-40 expression during genotoxic or microenvironmental stress in human glioblastoma cells, Cancer Sci., 2005, 96, 183-190
[17] Nishikawa K.C., Millis A.J.T., gp38k (CHI3L1) is a novel adhesion and migration factor for vascular cells, Exp. Cell. Res., 2003, 287, 79-87
[18] Kawada M., Chen C.-C., Arihiro A., Nagatani K., Watanabe T., Mizoguchi E., Chitinase 3-like-1 enhances bacterial adhesion to colonic epithelial cells through the interaction with bacterial chitin-binding protein, Lab. Invest., 2008, 88, 883-895
[19] Johansen J.S., Studies on serum YKL-40 as a biomarker in diseases with inflammation, tissue remodelling, fibroses and cancer, Dan. Med. Bull., 2006, 53, 172-209
[20] Tanwar M.K., Gilbert M.R., Holland E.C., Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma, Cancer Res., 2002, 62, 4364-4368
[21] Steck E., Breit S., Breusch S.J., Axt M., Richter W., Enhanced expression of the human chitinase 3-like 2 gene (YKL-39) but not chitinase 3-like 1 gene (YKL-40) in osteoarthritic cartilage, Biochem. Biophys. Res. Commun., 2002, 299, 109-115
[22] Shostak K., Labunskyy V., Dmitrenko V., Malisheva T., Shamayev M., Rozumenko V., et al., HC gp-39 gene is upregulated in glioblastomas, Cancer Lett., 2003, 198, 203-210
[23] Kavsan V., Dmitrenko V., Boyko O, Filonenko V., Areshkov P., Avdeev A., et al., Overexpression of YKL-39 Gene in Glial Brain Tumors, Scholarly Research Exchange, 2008, http://www.syrexe.com/biology/2008/814849.html
[24] Fusetti F., Pijning T., Kalk K.H., Bos E., Dijkstra B.W., Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39, J. Biol. Chem., 2003, 278, 37753-37760
[25] Houston D.R., Recklies A.D., Krupa J.C., van Aalten D.M.F., Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes, J. Biol. Chem., 2003, 278, 30206-30212
[26] Altschul S.F., Madden T.L., Schffer A.A., Zhang J., Zhang Z., Miller W., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database

search programs, Nucl. Acids Res., 1997, 25, 3389-3402
[27] Marchler-Bauer A., Anderson J.B., Cherukuri P.F., DeWeese-Scott C., Geer L.Y., Gwadz M., et al., CDD: a Conserved Domain Database for protein classification, Nucl. Acids Res., 2005, 33, D192-D196
[28] Marchler-Bauer A., Anderson J.B., Derbyshire M.K., DeWeese-Scott, C., Gonzales N.R., Gwadz M., et al., CDD: a conserved domain database for interactive domain family analysis, Nucl. Acids Res., 2007, 35, D237-D240
[29] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., et al., Clustal W and Clustal X version 2.0, Bioinformatics, 2007, 23, 2947-2948
[30] Gouet P., Robert X., Courcelle E., ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins, Nucl. Acids Res., 2003, 31, 3320-3323
[31] Deshpande N., Addess K.J., Bluhm W.F., Merino-Ott J.C., Townsend-Merino W., Zhang Q., et al., The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema, Nucl. Acids Res., 2005, 33, D233-D237
[32] Murzin A.G., Brenner S.E., Hubbard T., Chothia, C., SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol., 1995, 247, 536-540
[33] Mart-Renom M.A., Stuart A.C., Fiser A., Snchez R., Melo F., Sali A., Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct., 2000, 29, 291-325
[34] Shen M.-Y., Sali A., Statistical potential for assessment and prediction of protein structures, Protein Sci., 2006, 15, 2507-2524
[35] Bhattacharya A., Tejero R., Montelione G.T., Evaluating protein structures determined by structural genomics consortia, Proteins, 2007, 66, 778-795
[36] Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J., Wang X., et al., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, Nucl. Acids Res., 2007, W375-W383
[37] Voss N.R., Gerstein M., Steitz T.A., Moore P.B., The geometry of the ribosomal polypeptide exit tunnel, J. Mol. Biol., 2006, 360, 893-906
[38] Ye Y., Godzik A., FATCAT: a web server for flexible structure comparison and structure similarity searching, Nucl. Acids Res., 2004, 32, W582-W585
[39] Guda C., Lu S., Scheeff E.D., Bourne P.E., Shindyalov I.N., CE-MC: a multiple protein structure alignment server, Nucl. Acids Res., 2004, 32, W100-W103
[40] Diemand A.V., Scheib H., iMolTalk: an interactive, internet-based protein structure analysis server, Nucl. Acids Res., 2004, 32, W512-W516
[41] DeLano W.L., The PyMOL User's Manual, DeLano Scientific, Palo Alto, CA, USA, 2001
[42] Boots A.M.H., Hubers H., Kouwijzer M., den Hoed-van Zandbrink L., Westrek-Esselink B.M., van Doorn C., et al., Identification of an altered peptide ligand based on the endogenously presented, rheumatoid arthritis-associated, human cartilage glycoprotein-39(263-275) epitope: an MHC anchor variant peptide for immune modulation, Arthritis Res. Ther., 2007, 9, R71
[43] Tsark E.C., Wang W., Teng Y.-C., Arkfeld D., Dodge G.R., Kovats S., Differential MHC class II-mediated presentation of rheumatoid arthritis autoantigens by human dendritic cells and macrophages, J. Immunol., 2002, 169, 6625-6633
[44] Kang C.B., Hong Y., Dhe-Paganon S., Yoon H.S., FKBP family proteins: immunophilins with versatile biological functions, Neurosignals, 2008, 16, 318-325
[45] Renkema G.H., Boot R.G., Au F.L., Donker-Koopman W.E., Strijland A., Muijsers A. O., et al., Chitotriosidase, a chitinase, and the 39-kDa human cartilage glycoprotein, a chitin-binding lectin, are homologues of family 18 glycosyl hydrolases secreted by human macrophages, Eur. J. Biochem., 1998, 251, 504-509
[46] Rikova K., Guo A., Zeng Q., Possemato A., Yu J., Haack H., et al., Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer., Cell, 2007, 131, 1190-1203
[47] Hakala B.E., White C., Recklies A.D., Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family, J. Biol. Chem., 1993, 268, 25803-25810
[48] Bigg H.F., Wait R., Rowan A.D., Cawston T.E., The mammalian chitinase-like lectin, YKL-40, binds specifically to type I collagen and modulates the rate of type I collagen fibril formation, J. Biol. Chem., 2006, 281, 21082-21095
[49] Bonneh-Barkay D., Bissel S.J., Wang G., Fish K.N., Nicholl G.C., Darko S.W., et al., YKL-40, a marker of simian immunodeficiency virus encephalitis, modulates the biological activity of basic fibroblast growth factor, Am. J. Pathol., 2008, 173, 130-143

[50] Boon K., Edwards J.B., Eberhart C.G., Riggins G.J., Identification of astrocytoma associated genes including cell surface markers, BMC Cancer, 2004, 4, 39

[51] Sekine T., Masuko-Hongo K., Matsui T., Asahara H., Takigawa M., Nishioka K., et al., Recognition of YKL-39, a human cartilage related protein, as a target antigen in patients with rheumatoid arthritis, Ann. Rheum. Dis., 2001, 60, 49-54

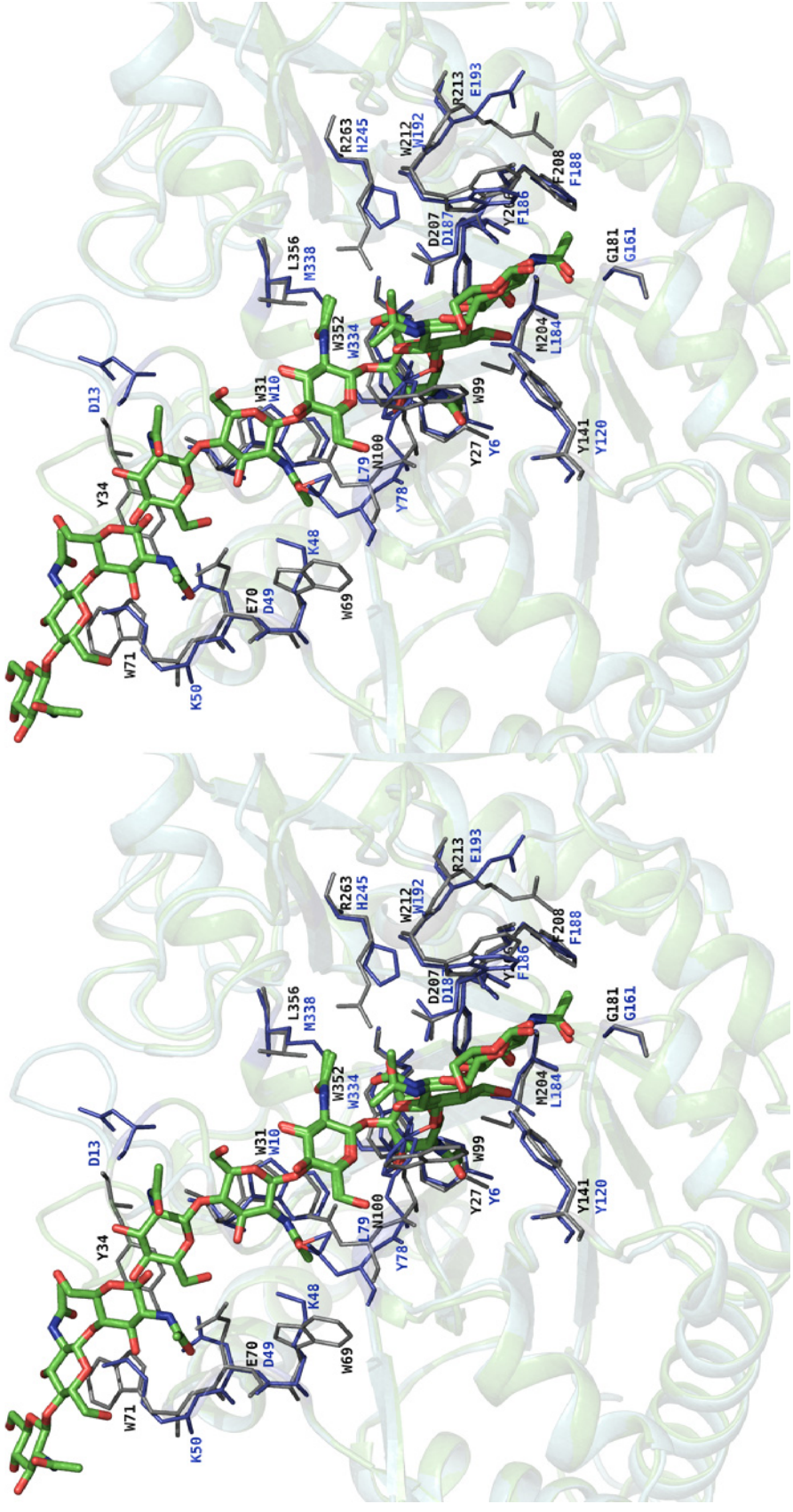
[52] Benz K., Breit S., Lukoschek M., Mau H., Richter W., Molecular analysis of expansion, differentiation, and growth factor treatment of human chondrocytes identifies differentiation markers and growth-related genes, Biochem. Biophys. Res. Commun., 2002, 293, 284-92

[53] Schlessinger J., Plotnikov A.N., Ibrahimi O.A., Eliseenkova A.V., Yeh B.K., Yayon A., et al., Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization, Mol. Cell, 2000, 3, 743-750

[54] Shao R., Hamel K., Petersen L., Cao Q.J., Arenas R.B., Bigelow C., et al, YKL-40, a secreted glycoprotein, promotes tumor angiogenesis, Oncogene, 2009, 28, 4456-4468

[55] Alexopoulou A.N., Multhaupt H.A., Couchman J.R., Syndecans in wound healing, inflammation and vascular biology, Int. J. Biochem. Cell Biol., 2007, 39, 505-528

[56] Koenig A., Norgard-Sumnicht K., Linhardt R., Varki A., Differential Interactions of Heparin and Heparan Sulfate Glycosaminoglycans with the Selectins. Implications for the Use of Unfractionated and Low Molecular Weight Heparins as Therapeutic Agents, J. Clin. Invest., 1998, 101, 877-889



Supplementary Figure 1. Ligand-binding groove residues of CHI3L1–chitin tetramer complex (PDB ID: 1HJV:A, gray color) and equivalent residues of CHI3L2 model 3D structure (blue color) from superposition of both structures. Ligands from 1HJV:A and 1HJW:A are colored by atoms. The residues in CHI3L1 are numbered according to precursor protein, in CHI3L2 the numbers correspond to mature peptide.