

BEL β -trefoil: A novel lectin with antineoplastic properties in king bolete (*Boletus edulis*) mushrooms

Michele Bovi², Lucia Cenci², Massimiliano Perduca², Stefano Capaldi², Maria E Carrizo³, Laura Civiero⁴, Laurent R Chiarelli⁵, Monica Galliano⁵, and Hugo L Monaco^{1,2}

²Department of Biotechnology, Biocrystallography Laboratory, University of Verona, Ca Vignal 1, strada Le Grazie 15, 37134 Verona, Italy;

³Departamento de Química Biológica, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, 5016 Córdoba, Argentina; ⁴Department of Biology, University of Padova, via Ugo Bassi 58b, 35121 Padova, Italy; and

⁵Department of Molecular Medicine, University of Pavia, via Taramelli 3b, 27100 Pavia, Italy

Received on September 7, 2012; revised on November 29, 2012; accepted on November 29, 2012

A novel lectin was purified from the fruiting bodies of king bolete mushrooms (*Boletus edulis*, also called porcino, cep or penny bun). The lectin was structurally characterized i.e its amino acid sequence and three-dimensional structure were determined. The new protein is a homodimer and each protomer folds as β -trefoil domain and therefore we propose the name *Boletus edulis* lectin (BEL) β -trefoil to distinguish it from the other lectin that has been described in these mushrooms. The lectin has potent anti-proliferative effects on human cancer cells, which confers to it an interesting therapeutic potential as an antineoplastic agent. Several crystal forms of the apo-protein and of complexes with different carbohydrates were studied by X-ray diffraction. The structure of the apoprotein was solved at 1.12 Å resolution. The interaction of the lectin with lactose, galactose, *N*-acetylgalactosamine and T-antigen disaccharide, Gal β 1-3GalNAc, was examined in detail. All the three potential binding sites present in the β -trefoil fold are occupied in at least one crystal form and are described in detail in this paper. No important conformational changes are observed in the lectin when comparing its co-crystals with carbohydrates with those of the ligand-free protein.

Keywords: *Boletus edulis* / galactose / lactose / lectin mushroom / *N*-acetylgalactosamine / structure / T-antigen / β -trefoil

Introduction

Proteins of nonimmune origin that selectively bind and recognize carbohydrates without modifying them enzymatically are called lectins (Sharon 2007). In general, ligand binding precedes the fulfillment of an important biological function, which in some cases is still not known. Some members of this family are also called agglutinins because of their ability to agglutinate red blood cells, but this term does not necessarily imply the recognition of a carbohydrate and is therefore less precise (Sharon and Lis 2004). Although initially characterized in plants, lectins are ubiquitous in nature and have been identified in most living species, from viruses to humans. Their sugar selectivity can be very useful and they are widely used in both basic and applied science, for example, for the purification of glycoproteins (Naeem et al. 2007), stem cell fractionation (Reisner et al. 1978) and in targeted drug delivery (Bies et al. 2004).

Fungi are the members of a group of living organisms, once considered to be close to plants, but that in fact form a separate kingdom. They rely on symbiosis for their survival and associate with a wide variety of hosts through an interaction that in many cases appears to imply the recognition of a glycoconjugate by a fungal lectin (Imberty et al. 2005). In addition to this function, fungal lectins are also believed to play a role in morphogenesis, development and in defense, but in many cases the function of these proteins in the physiology of the fungus is not clearly established.

The presence of lectins in edible mushrooms was recognized about one century ago. Since then many mushroom lectins have been thoroughly studied and structurally characterized and their ligand-binding specificity has been described (Guillot and Kanska 1997; Goldstein and Winter 2007; Singh et al. 2010; Khan and Khan 2011). The reasons for this interest is that several mushroom lectins have been proven to be useful in glycobiology and that targeting carbohydrates in glycoproteins is emerging as a potential very powerful tool in medicine (González De Mejía and Prisecaru 2005). In this regard, the lectin present in the common and very widely diffused edible mushroom *Agaricus bisporus*, ABL, provides an interesting example because it selectively inhibits the proliferation of human malignant epithelial cell lines without any noticeable toxicity for normal cells (Yu et al. 1993). The selectivity of the lectin for the malignant cells is explained by the presence on the surface of neoplastic cells of the Thomsen–Friedenreich antigen or T-antigen, a disaccharide, Gal β 1-3GalNAc, bound to either serines or threonines in glycoproteins. The

¹To whom correspondence should be addressed: Tel: +39-045-8027-903; Fax: +39-045-8027-929; e-mail: monaco@sci.univr.it

disaccharide is hidden in healthy cells and exposed in most human carcinomas (Springer 1984, 1997; Yu 2007). The X-ray structure of the ABL molecule identified a new fold present in several other fungal lectins of the saline soluble family in which two different carbohydrate-binding sites are present in a single monomer: One which is specific for *N*-acetylgalactosamine (GalNAc) and binds the T-antigen disaccharide and the other which selectively binds *N*-acetylglucosamine (GlcNAc) and most probably chitin (Carrizo et al. 2005).

Another lectin with the same fold and similar antineoplastic properties was subsequently isolated from the popular edible wild mushroom *Boletus edulis* (king bolete, penny bun, porcino or cep). The lectin was named *Boletus edulis* lectin (BEL) and its X-ray structure and ligand-binding properties, including its interaction with the T-antigen disaccharide, were examined in detail (Bovi et al. 2011). It was purified using two different methods: The first exploited its affinity for chitin and the second used a column of human erythrocytic stroma incorporated into a polyacrylamide gel (Betail et al. 1975). While using the second method it was found that a second lectin was present in very significant quantities in the mushroom extracts. Here, we present our results on this second, totally different lectin. We have determined its X-ray structure, examined its ligand-binding and antineoplastic properties and its interaction with cancer cells. The protein is a dimer and each monomer folds as a β -trefoil domain and therefore we propose the name BEL β -trefoil for it to distinguish it from the other BEL which belongs to the saline soluble family of fruiting body-specific lectins. The β -trefoil fold is very versatile and has been observed in proteins with very different functions and therefore it is being very extensively studied (Gosavi et al. 2008; Lee et al. 2011). These facts enhance the potential biotechnological applications of BEL β -trefoil.

Results

Amino acid sequence

The X-ray structure of BEL β -trefoil was determined before its amino acid sequence was known and therefore information on the sequence based on the electron density maps was available before the other methods were used. Automated Edman sequence analysis of the most abundant isoform of BEL β -trefoil established unambiguously the initial 56 N-terminal residues of the protein. An almost complete coverage of the lectin amino acid sequence was obtained by tandem mass spectrometric analysis of the peptides derived from tryptic, chymotryptic and *S. aureus* V8 endopeptidase cleavage. The results confirmed the amino acid sequence deduced from the electron density maps of the BEL β -trefoil crystals and the information was used to obtain the cDNA encoding the sequence of the protein starting from total RNA extracts of the fresh fruiting bodies of the mushroom. There are some positions in the sequence that show variability confirmed in most cases by more than one of the sequencing methods used. The molecules of the crystal forms listed in Table I were modeled using three different sequences: Crystal forms 1, 2, 4, 6, 7, 8 and 9 (first sequence), crystal form 3 (apo, trigonal, second sequence) and crystal form 5 (complex with lactose, third sequence). In order to give an idea of the

variability observed, Figure 1A represents the three sequences aligned identified with the numbers 1 (seven crystal forms), 3 (trigonal form) and 5 (complex with lactose). The sequence of the most abundant isoform will be used for all the following discussions and is indicated with the number 1 in Figure 1A; the amino acids of the other two isoforms that are different are represented red. Note that there are three tryptophans that are conserved in the three isoforms. The sequence QxW, where x can be any residue, has been identified in the three subdomains of many lectins with the β -trefoil fold (Hazes 1996). In the case of BEL β -trefoil, it is present only in the third subdomain, amino acids 139–141, QLW and has been highlighted in Figure 1A.

A sequence similarity search in the ExPASy server identifies the N-terminal module of the hemolytic pore-forming lectin of the parasitic fungus *Laetiporus sulphureus* (Mancheño et al. 2005; Angulo et al. 2011) as the protein with highest sequence similarity with BEL β -trefoil, 43% amino acid identity. Figure 1B represents the two sequences aligned. Note that the sequence QxW is not present in the *L. sulphureus* lectin.

Structure of the protomer

Four different crystal forms of apo BEL β -trefoil were obtained; they belong to space groups P2₁2₁2₁, P3₂21 and P2₁ (see Table I). One of the two monoclinic forms diffracts to the highest resolution, 1.12 Å, and will be used to describe the structure of the protomer. These crystals belong to space group P2₁ with unit cell parameters $a = 34.5$ Å, $b = 66.8$ Å, $c = 63.9$ Å, $\beta = 96.1^\circ$ and contain one dimer in the asymmetric unit. The stereochemical quality of the protein model was assessed with the program PROCHECK (Laskowski et al. 1993). In total, 88.9% of the residues are in the most favorable region of the Ramachandran plot and the remaining 11.1% in the additionally allowed regions. A residual electron density, continuous in the $F_{\text{obs}} - F_{\text{calc}}$ map at a 2.5σ level was not modeled and is present only in this crystal form at the interface between the two monomers in the dimer, apparently in contact with the side chains of Ile 98 and Lys 135 of the two chains A and B. This space is occupied in the other crystal forms by solvent molecules. The fold is that of the β -trefoil family (Murzin et al. 1992), i.e. the structure presents a β barrel with a pseudo-3-fold symmetry axis. The molecule contains three subdomains generally called α , β and γ , starting from the N terminus of the protein each with four β strands. The fold can also be described in terms of three trefoil subunits, segments of continuous polypeptide chain, that form the core of the subdomains with three β strands and contribute with a fourth strand to the formation of an adjacent subdomain.

The 12 β strands span the following residues: A, 13–17, B, 23–26, C, 34–37, D, 47–51, E, 60–64, F, 70–73, G, 81–84, H, 95–99, I, 107–111, J, 116–120, K, 128–131 and L, 141–145. The β sheet of subdomain α contains strands L, A, B and C, all antiparallel to each other, the β sheet of subdomain β contains strands D, E, F and G and that of subdomain γ strands H, I, J and K. Figure 2A is a stereo diagram showing one protomer in approximately the direction of the pseudo-3-fold axis. The three trefoil subunits are colored red, green and blue; strands A, B, C and D are red, strands E, F, G and H green and strands I, J, K and L blue. There are four 3₁₀

Table I. Data collection and refinement statistics

Data set	BEL β -trefoil APO	BEL β -trefoil + lactose	BEL β -trefoil + galactose	BEL β -trefoil + <i>N</i> -acetyl galactosamine	BEL β -trefoil + T-antigen disaccharide	BEL β -trefoil + T-antigen			
Space group	$P2_1$	$P2_12_12_1$	$P3_221$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Crystal form	1	2	3	4	5	6	7	8	9
<i>a</i> (Å)	34.5	41.6	77.7	65.4	72.1	66.4	66.5	66.4	66.4
<i>b</i> (Å)	66.8	66.6	77.7	70.0	39.1	70.0	69.9	70.0	70.1
<i>c</i> (Å)	63.9	105.1	52.8	71.8	119.4	71.6	71.4	71.8	71.5
α	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
β	96.1	90.0	90.0	106.8	93.2	106.6	106.3	106.4	106.7
γ	90.0	90.0	120.0	90.0	90.0	90.0	90.0	90.0	90.0
Protomers in the asymm. unit	2	2	1	4	4	4	4	4	4
Resolution range (Å)	63.6–1.12	30.0–1.28	16.8–1.51	30.0–1.77	20.5–1.40	27.6–1.57	24.5–1.50	30.0–1.72	30.0–1.90
Observed reflections	279,827	457,601	151,341	200,480	534,265	303,728	361,483	240,053	151,328
Independent reflections	103,696	70,848	29,039	59,260	131,858	83,522	99,935	64,980	48,087
Multiplicity	2.7 (2.3)*	6.5 (6.4)	5.2 (4.7)	3.4 (3.4)	4.1 (4.0)	3.6 (3.5)	3.6 (3.6)	3.7 (3.6)	3.1 (3.2)
Rmerge (%) ^a	3.2 (9.2)	6.2 (27.9)	7.5 (26.2)	7.4 (35.5)	7.8 (28.0)	6.7 (27.7)	8.1 (33.8)	9.3 (34.6)	7.0 (30.4)
$\langle I/\sigma(I) \rangle$	18.0 (8.8)	17.8 (6.1)	3.9 (1.3)	12.2 (3.2)	11.6 (4.7)	11.4 (4.1)	9.8 (3.4)	12.4 (5.3)	11.8 (3.5)
Completeness (%)	93.8 (81.6)	93.5 (86.5)	96.8 (89.5)	98.0 (97.0)	99.6 (99.1)	95.8 (87.8)	99.6 (98.6)	97.1 (95.7)	97.1 (96.1)
Wilson B factor	6.9	9.7	16.6	19.8	14.2	16.2	15.4	12.3	19.1
Reflections in refinement	103,652	70,792	28,987	59,224	125,098	79,300	94,920	61,668	45,649
Rcryst (%) ^b	15.1	16.7	18.8	17.6	18.9	20.0	18.7	17.8	19.2
Rfree (%) (test set 5%) ^c	16.0	18.5	21.5	22.0	20.4	22.4	20.7	20.5	22.5
Protein atoms	2378	2384	1185	4802	4768	4775	4813	4803	4783
Specific ligand atoms	–	–	–	–	115	96	90	182	192
Nonspecific ligand atoms	65	32	18	48	54	16	24	–	8
Sites occupied	–	–	–	–	β, γ	α, β, γ	α, β, γ	α, β, γ	α, β, γ
Water molecules	380	412	168	502	747	605	609	714	382
r.m.s.d. on bond lengths (Å) ^d	0.005	0.006	0.006	0.006	0.005	0.006	0.005	0.006	0.007
r.m.s.d. on bond angles (Å) ^d	1.146	1.128	1.127	1.060	1.071	1.151	1.029	1.036	1.148
Planar groups (Å) ^d	0.007	0.006	0.006	0.005	0.005	0.005	0.004	0.005	0.007
Chiral volume dev. (Å ³) ^d	0.083	0.084	0.081	0.081	0.073	0.084	0.072	0.073	0.079
Average B factor (Å ²)	10.1	11.4	14.8	21.1	7.8	12.9	11.7	8.2	15.9
Protein atoms	8.9	10.1	13.5	20.6	6.4	11.9	10.5	6.8	14.8
Specific ligand atoms	–	–	–	–	13.8	20.1	17.9	16.1	38.3
Nonspecific ligand atoms	14.3	14.2	14.1	27.0	11.3	16.6	19.7	–	26.3
Solvent atoms	17.0	18.8	24.1	24.9	15.6	20.0	20.0	15.4	22.7

*The values in parentheses refer to the highest resolution shells. For the data collection of the different crystal forms (identified with a number in the second line) they are the following: 1, 1.18–1.12 Å; 2, 1.35–1.28 Å; 3, 1.56–1.51 Å; 4, 1.87–1.77 Å; 5, 1.47–1.40 Å; 6, 1.66–1.57 Å; 7, 1.58–1.50 Å; 8, 1.81–1.72 Å and 9, 2.00–1.90 Å.

The highest resolution shells used in the refinements are the following: Crystal form 1, 1.13–1.12 Å; 2, 1.30–1.28 Å; 3, 1.56–1.51 Å; 4, 1.80–1.77 Å; 5, 1.43–1.40 Å; 6, 1.61–1.57 Å; 7, 1.54–1.50 Å; 8, 1.76–1.72 Å and 9, 1.95–1.90 Å. The nonspecific ligands are tris and glycerol.

^aRmerge = $\sum_i |I_i - \langle I \rangle| / \sum_i I_i$, where $\langle I \rangle$ is the mean intensity of the i observations of reflection h .

^bRcryst = $\sum (|F_{obs}| - |F_{calc}|) / \sum |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes, respectively. Summation includes all reflections used in the refinement.

^cRfree = $\sum (|F_{obs}| - |F_{calc}|) / \sum |F_{obs}|$, evaluated for a randomly chosen subset of 5% of the diffraction data not included in the refinement.

^dRoot mean square deviation from ideal values.

A		
BEL β -T (1)	VNFPNIPAEGVQFRLRARDTGYVIYSRTENPPLVWQYNGPPYDDQLFTLIYGTGPRKNLYAIKSVPNGRVLFSRTSASPY	80
BEL β -T (3)	VNFPNIPAEGRFRRLRARDSGYVIYSRTENDPLVWHYNGPPYDDQLFTLIHGTGSRNLNLYAIKSVPNGRVLFSRNSASPT	80
BEL β -T (5)	VNFPNIPAEGAQFRLRARDTGYVIYSRTENPPLVWQYNGPPYDDQLFTLIYGTGPHQNLNLYAIKSVPNGRVLFSRTSASPH	80
BEL β -T (1)	VGNIAGDGTYNWNWFQFIQDDNDPNSFRIYNLASDVTLYSRTTADPKFGNFTGAKYDDQLWHFELV	146
BEL β -T (3)	VGNI V GDGTYNWNWFQFIQDDNDANSFRIYSLASDSVLYSRTTGAPQFGNYTGP K FDDQLWHFEIV	146
BEL β -T (5)	VGNIAGDGTYNWNWFQFIQDDNDPNSFRIYSLASDVTLYSRTT P DPQFGNYT G AKYDDQLWHFELV	146
B		
BEL β -T	VNFPNIPAEGVQFRLRARDTGYVIYSRTENPPLV-WQYNGPPYDDQLFTLIYGTGPRKNLYAIKSVPNGRVLFSRTSASPY	80
LSL ₁₅₀	MTDIYIPPEGLYFRLLGFASRQVIFARNSPSPDVGLSPVNDQATDQYFSLIYGTGEHAGLYAIKSKATGKVLFSRRPAEY	81
BEL β -T	VGNIAGDGTYNWNWFQFIQDDNDPNS-FRIYNLASDVTLYSRTTADPKFGNFTGAK-YDDQLWHFELV	146
LSL ₁₅₀	VGQIDGGRYPDNWFKIEFGKTYLSKYFRLVQPSGTALVSRTHLQYFVNHQPQTEVEDDQYFTFLFED	150

Fig. 1. Amino acid sequence. (A) Sequence variability in three isoforms of BEL β -trefoil. The numbers identifying the isoforms correspond to those used for the different crystal forms presented in Table I. The sequence of isoform 1 is taken as a reference and the amino acids that are different in the other two isoforms are represented red. The highlighted sequence is the QLW motif. (B) Sequence comparison with the N-terminal module of the *Laetiporus sulphureus* lectin. The sequence of the β -trefoil domain of the *L. sulphureus* lectin (LSL₁₅₀, protein data bank (PDB) accession code 2Y9F) presents the highest amino acid sequence identity with BEL β -trefoil (43%). The amino acids that are identical in the two sequences are in red. (The colour version of this figure is available on the website.)

helices after strands C, D, G and K that span residues 43–45, 54–56, 91–93 and 137–139. Only three are represented gray in the figure, they are those present before the last β strand of each subunit.

Figure 2B shows the sequences of the three subunits in the fold aligned on the basis of three strictly conserved residues SRT (the second with an important role in ligand binding, see below), numbers 26–28, 73–75 and 120–122. The amino acids that are identical in at least two subunits are red and the 12 β strands are underlined and labeled. Figure 2C is a superposition of the models of the BEL β -trefoil protomer and the N-terminal module of the *L. sulphureus* lectin (Angulo et al. 2011). BEL β -trefoil is represented red while the N-terminal domain of the *L. sulphureus* lectin (LSL₁₅₀, PDB accession code 2Y9F) is blue.

BEL β -trefoil is a dimer

The first line of evidence that indicated that BEL β -trefoil is a dimer under normal conditions was the appearance of a single band corresponding to the molecular weight of the dimer in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) when the sample was not boiled before running the electrophoresis and a single band with the molecular weight of the monomer if the sample was boiled instead (data not shown). This experiment indicated that the dimer was stable even in the presence of SDS and that exposure to high temperature was required to dissociate it into the monomeric units. This observation was confirmed using dynamic light scattering. Figure 3A is a scattered intensity distribution plot of different concentrations of BEL β -trefoil, showing that the hydrodynamic diameter of the major peak is consistent with the presence of a dimer and that the samples contain a relatively low percentage of higher molecular weight aggregates. In a similar plot of the percent volume or number of particles, the second peak becomes totally insignificant.

Examination of the first two crystal forms of the apoprotein listed in Table I (1 and 2) reveals that in both cases the asymmetric unit contains the same dimer and the same contacts between molecules are also found in the other two forms (3 and 4). In addition, the trigonal crystal form contains only one protomer in the asymmetric unit and therefore the two monomers in the dimer are related by a 2-fold axis. In the PDB file of crystal form 4, the two monomers in each dimer are labeled A and B for molecule 1 and C and D for molecule 2.

We have also calculated the solvent accessible area of protomers and dimer in crystal form 1 that diffracts to the highest resolution, 1.12 Å. The A monomer of BEL β -trefoil has a solvent accessible area of 6148 Å² and the B monomer has that of 6263 Å². The contact areas are 1850 and 1885 Å², i.e. ~30% of the total surface of the monomers and a value that justifies the observed stability of this molecule when compared with those reported for other physiologically relevant dimers (Collaborative Computational Project Number 4 1994; Jones and Thornton 1995; Ponstingl et al. 2000).

The contacts between the two protomers in the dimer are established through the first two amino acids, strands H and I and the loops connecting strands D to E, G to H, H to I, I to J and K to L. Most of the interactions are hydrophobic although an interesting specific contact is established between Gln 96 of one protomer and the OH of Tyr 110 of the other and Lys 135 and Asp 101. Table II lists the distances <3.5 Å measured between atoms in the interacting protomers and Figure 3B is a representation of the dimeric molecule with the elements of secondary structure colored as in Figure 2A.

Effect of BEL β -trefoil on neoplastic cells

The anti-proliferative effect of the lectin on neoplastic cells was assayed using two methods based on different principles. The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay measures the enzymatic activity of mitochondrial reductases that can directly be related to the

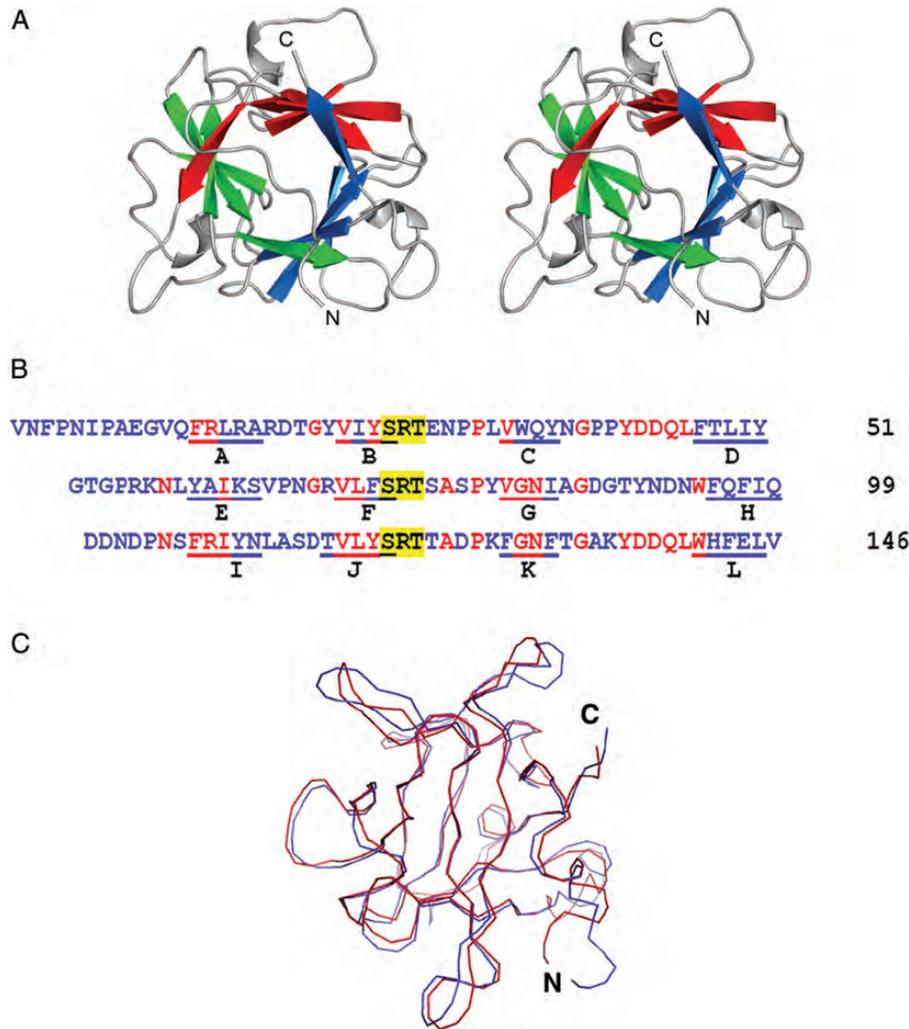


Fig. 2. X-ray structure of BEL β -trefoil. (A) Stereo representation of the protomer. The view is approximately looking down the pseudo-3-fold axis. The first four β strands (A, B, C and D) are in red, the next four (E, F, G and H) are in green and the final four (I, J, K and L) are in blue. (The colour version of this figure is available on the website.) The subdomain α is made up of three red and one blue strands, subdomain β of three green and one red strands and subdomain γ of three blue and one green strands. (B) Sequence alignment of the three polypeptide trefoil subunits. The three partial sequences were aligned on the basis of three strictly conserved residues SRT (the second with an important role in ligand binding) numbers 26–28, 73–75 and 120–122, highlighted in yellow. The amino acids that are identical in at least two subunits are in red and the 12 β strands are underlined and labeled as described in the text. (C) Superposition of the models of the BEL β -trefoil protomer and the N-terminal module of the *L. sulphureus* lectin (Angulo et al. 2011). The models were superimposed using the program LSQKAB (Kabsch 1978). BEL β -trefoil is represented in red, whereas the N-terminal domain of the LS lectin (PDB accession code 2Y9F) is in blue.

number of viable cells. If the amount of the yellow tetrazole MTT reduced to purple formazan by cells treated with the lectin is compared with that produced by untreated control cells, the effectiveness of the protein in inhibiting cell proliferation can be deduced by examination of the dose–response curve. The second method measures the inhibition of tritium-labeled thymidine incorporation into DNA as a result of the presence of the cytotoxic lectin. Both assays revealed that the lectin has a dose-dependent anti-proliferative effect against several human carcinoma cell lines.

Figure 4A sums up the outcome of these experiments. The top panel summarizes the results of the MTT assay. The plot shows the percentage viability relative to the control as a function of the lectin concentration. Different colors are used

for different human neoplastic cells. The inset shows the dose–response effect to doxorubicin of the same cell lines that was used as a positive control of the assay. The bottom panel shows the [3 H]-thymidine incorporation into the DNA of different cell lines as a function of the lectin concentration. Note that the scale of the lectin concentration is similar for both assays and that comparable inhibitory effects are observed with the two different protocols. The MTT assay shows that BEL β -trefoil is particularly effective in inhibiting the proliferation of the HepG-2 hepatocellular carcinoma cells and somewhat less effective in inhibiting the CaCo-2 human colorectal adenocarcinoma and MCF-7 breast cancer cell lines. The [3 H]-thymidine incorporation assay reveals that BEL β -trefoil produced similar cell growth inhibition in HeLa, SK-MEL-28

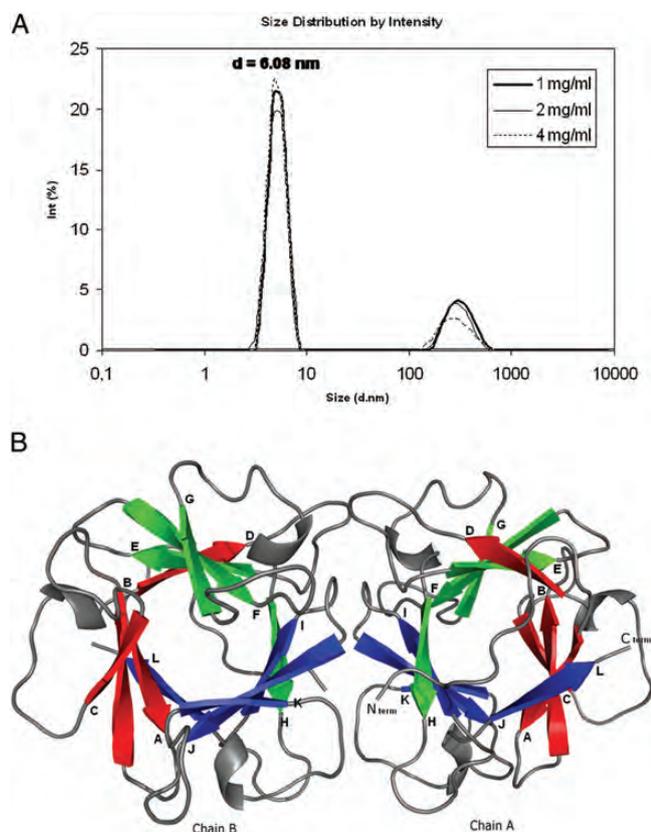


Fig. 3. The BEL β -trefoil dimer. (A) Dynamic light scattering experiment of BEL β -trefoil. The size distribution is by intensity and three curves corresponding to three different concentrations of the protein were recorded. The small high-molecular-weight peak corresponds to the presence of a low percentage of aggregates. The hydrodynamic diameter of the highest peak is consistent with the presence of a dimer in solution. (B) Ribbon representation of the BEL β -trefoil dimer. The three subunits are in red, green and blue. (The colour version of this figure is available on the website.) The 2-fold axis present in the dimer is in the plane of the figure and is approximately vertical to the bottom of the page. Note the position of the N and C termini of the protomers.

and U-87 MG cell lines (89, 87 and 82% inhibition at 40 $\mu\text{g}/\text{mL}$, respectively). Under the same conditions, the inhibitory effect of BEL β -trefoil was less pronounced in HT-29 and A549 cell lines (66 and 57% inhibition, respectively). Two different cell lines that show a very clear response to the lectin, the HepG-2 hepatocellular carcinoma cells and HeLa cells were chosen to test the effect of the addition of carbohydrates to cell cultures in which the lectin was present at a concentration of either 40 or 20 $\mu\text{g}/\text{mL}$. Figure 4B shows that galactose (Gal) as well as GalNAc and lactose have a dose-dependent protective effect on the lethal action of the lectin and that the disaccharide is more protective than the monosaccharides.

Internalization of BEL β -trefoil by CaCo-2, MCF-7 and CFPAC-1 cells was demonstrated by confocal microscopy using fluorescein-conjugated lectin. After incubation of the cells for 2 h, with 50 $\mu\text{g}/\text{mL}$ fluorescein isothiocyanate (FITC)-BEL β -trefoil, strong cell surface and intracellular fluorescence was observed (Figure 4C shows the results for CaCo-2 cells).

Table II. Significant contacts between the two protomers of BEL β -trefoil in the dimer in crystal form I

Amino acid residue monomer A	Atom	Distance (Å)	Amino acid residue monomer B	Atom
Asn 2	N	3.49	Asp 115	CB
Asn 2	ND2	3.29	Asp 115	CG
Asn 2	ND2	3.45	Asp 115	OD2
Asn 2	N	2.81	Asp 115	OD2
Gly 54	C	3.43	Asn 91	OD1
Asn 58	ND2	3.03	Leu 112	O
Asn 58	ND2	3.22	Ala 113	O
Asn 91	CG	3.36	Gly 54	O
Asn 91	ND2	3.24	Gly 54	O
Asn 91	OD1	3.33	Pro 55	CA
Gln 96	NE2	2.94	Tyr 110	OH
Asp 101	OD2	3.43	Lys 135	CD
Asp 101	OD2	3.45	Lys 135	CE
Asp 101	OD2	2.79	Lys 135	NZ
Tyr 110	OH	2.93	Gln 96	NE2
Leu 112	O	2.97	Asn 58	ND2
Ala 113	O	3.17	Asn 58	ND2
Asp 115	OD2	2.82	Asn 2	N
Asp 115	CG	3.29	Asn 2	ND2
Asp 115	OD2	3.45	Asn 2	ND2
Asp 115	OD2	3.46	Val 1	CA
Lys 135	CE	3.45	Asp 101	OD2
Lys 135	NZ	2.72	Asp 101	OD2

The table lists all the contacts <3.5 Å.

Interaction with carbohydrates

The three sugar-binding sites present in BEL β -trefoil are identified by the letters α , β and γ , starting from the N terminus of the polypeptide chain and related by the pseudo-3-fold axis of the protomer. The physiological dimer has thus six potential binding sites. Using the X-ray diffraction of single crystals, we have studied the interaction of the lectin with lactose, Gal, GalNAc, the T-antigen disaccharide (Gal β 1-3GalNAc) and the T-antigen (the disaccharide linked to serine), the most probable mediator of the anti-proliferative effect of the molecule. Examination of five co-crystals with the different ligands reveals that all the three sites present in the protomer bind carbohydrates. Table I summarizes the data collection and structure refinement statistics of the different co-crystal forms and Table III collects our observations on their occupancies. Note that although all the five crystal forms contain two dimers in the asymmetric unit, the crystals with lactose, prepared by co-crystallization and not like all the others by soaking, are different. It is only in this case that the α site is not occupied, the reason being that it is involved in the intermolecular contacts in the lattice. Figure 5A is a stereodiagram showing the electron density of lactose in the β - and γ -binding sites with the $2F_{\text{obs}} - F_{\text{calc}}$ map contoured at a 1.5σ level and Figure 5B represents the main interactions observed between lactose and the binding site β of BEL β -trefoil.

The lectin appears to bind with equal ease Gal and GalNAc and in a very similar fashion the carbohydrates in the three different binding sites. Table IV lists the shortest distances between BEL β -trefoil and Gal. Note the role played by arginines 27, 74 and 121 in the α , β and γ sites, respectively, and bound to the O4 of Gal in every case. Three aspartates: 44,

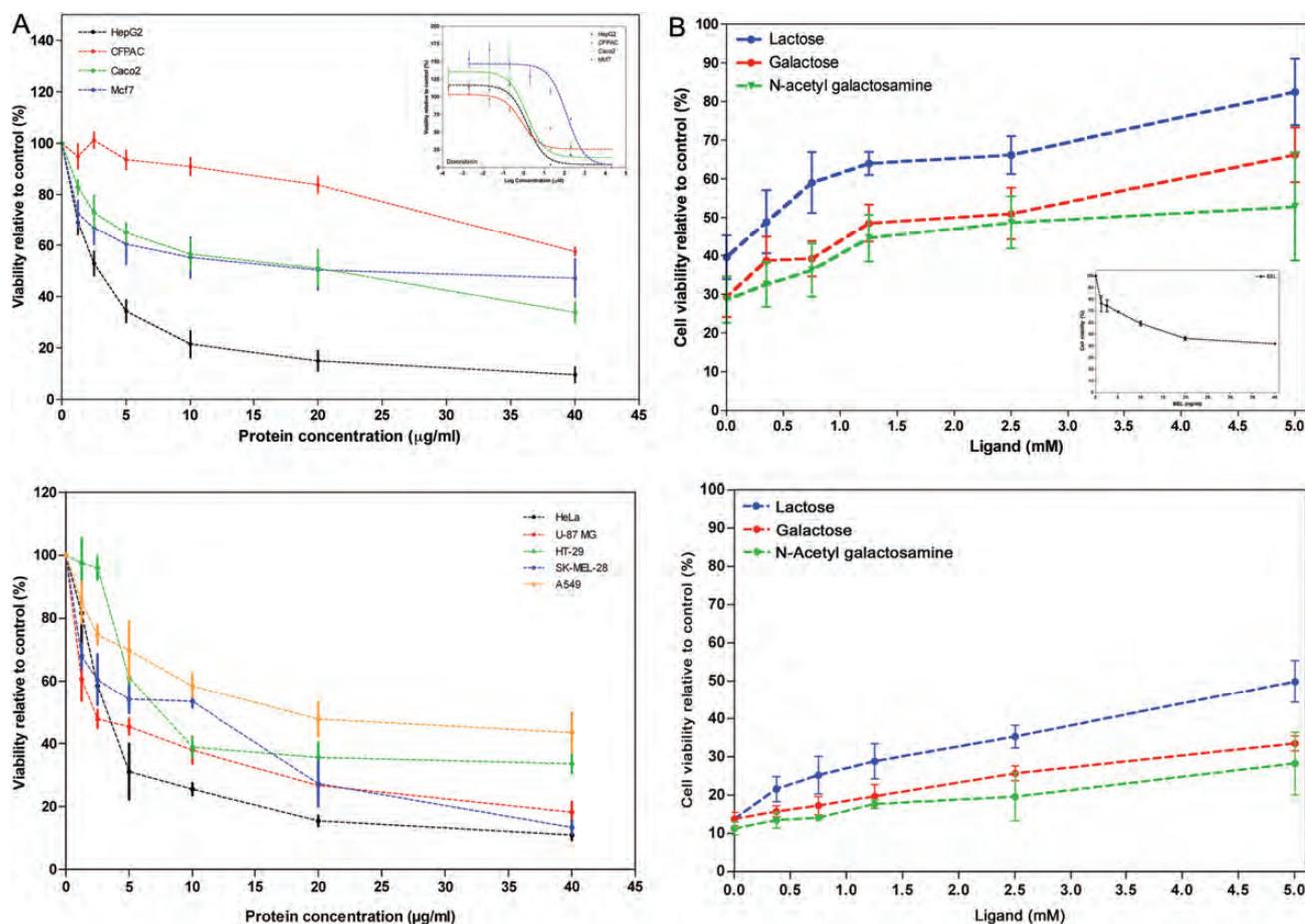


Fig. 4. Antineoplastic properties of BEL β -trefoil. (A) The top panel shows the percent viability of four different cell lines as a function of BEL β -trefoil concentration in the MTT cell proliferation assay (Alley et al. 1988). The inset shows the dose-response effect of doxorubicin (positive control) on the same cell lines. The bottom panel represents the analogous results obtained with the [3 H]thymidine incorporation into the DNA assay (Yu et al. 1993). Nine different cell lines were tested with the two assays and different colors were used to represent the results obtained with the different cells. The values plotted are the means \pm SD of triplicate determinations. (B) The top panel shows the protective effects of carbohydrates on the inhibition of HepG2 cell proliferation in the MTT assay. The cells grew in the presence of a BEL β -trefoil concentration of 40 μ g/mL and variable concentrations of the carbohydrates. The inset shows the dose-response effect of the cells to increasing concentrations of the lectin in the absence of carbohydrates. The bottom panel shows the protective effects of carbohydrates on the inhibition of HeLa cell proliferation in the [3 H]thymidine incorporation into the DNA assay. For these experiments, the BEL β -trefoil concentration was 20 μ g/mL. Note the ligand concentration range and the greater effect of the disaccharide lactose. (C) Interaction of fluorescein-labeled BEL β -trefoil with CaCo-2 cells. The top left panel shows the DAPI fluorescence (nucleus), the top right panel shows the FITC-labeled BEL β -trefoil fluorescence, the left bottom is a white light image while the bottom right panel is a superposition of the other three images. The scale bar indicates 10.0 μ m.

92 and 138 are hydrogen bound to the O4 and O6 of the saccharide and another important contact of O6 is with two different amino acids that play the same role: Gln 45 and 139 and Asn 93. The two contacts with O4 appear to confer selectivity to the lectin that does not bind glucose, which differs from Gal only in the conformation of this epimeric oxygen. The other amino acids that are equivalent in the three sites are tyrosines 42, 90 and 136 and Tyr 25 and 37 (site α), Phe 72 and Ile 84 (site β) and Tyr 119 and Phe 131 (site γ). In these last three cases, the main contacts are hydrophobic. Figure 5C shows the interactions of Gal in the three different binding sites.

The interactions described for Gal are quite similar to those observed for GalNAc and the three disaccharides. It is also worth mentioning that the T-antigen disaccharide,

Gal β 1-3GalNAc, was found to bind in the two possible ways to the lectin, i.e. with both Gal and GalNAc in contact with the protein at the active site. The electron density of these and the other ligands studied in this work can be examined in Figure 5D.

Table V lists the contacts measured in binding site β of protomer B of the five crystal forms. In the table, the prime is used to indicate the atoms of the glucose moiety of lactose. Note that the distances of the O4 of the two monosaccharides and the three disaccharides to Arg 74 and Asp 92 are very similar as is the distance of O6 to Asn 93. Figure 5D is a ribbon representation of a protomer with the electron density of the two monosaccharides, the T-antigen and the T-antigen disaccharide. In the case of the latter, the electron density of the two possible orientations is represented in the figure. The

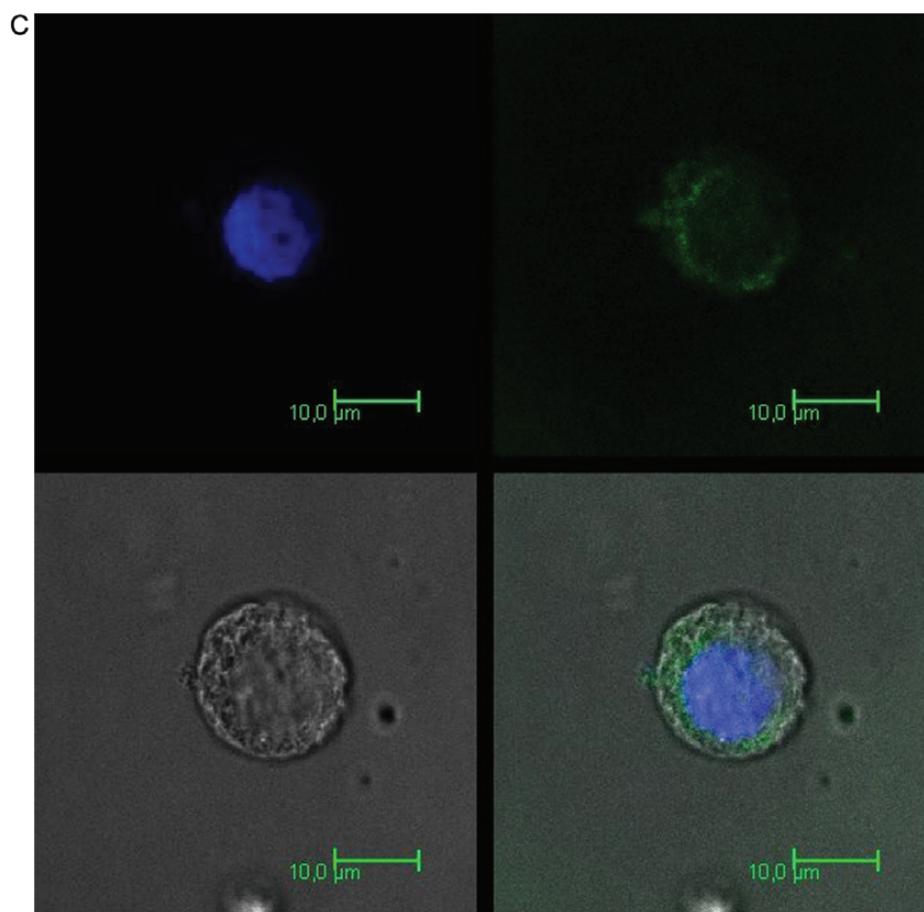


Fig. 4 Continued

Table III. Ligand binding in the different co-crystal forms

Crystal form	Specific ligand	α binding site	β binding site	γ binding site
5	Lactose	–	A–B–D	B–D
6	Galactose	A–D	A–B–D	A–B–D
7	<i>N</i> -acetyl galactosamine	A–D	B–D	A–B
8	T-antigen disaccharide	B–C	B–D	A–B–D
9	T-antigen	B–C	B–D	B–D

The table lists the sites occupied by carbohydrates in the different crystal forms. The letters A, B, C and D identify the four protomers present in the asymmetric unit, A and B form the first and C and D the second dimer. The α binding site of crystal form 5 is involved in the lattice contacts.

ligand-binding site shown is γ and the ball and stick models of the carbohydrates are oriented as when bound to the lectin model represented in the figure. No important conformational changes are observed in the lectin when comparing its co-crystals with carbohydrates with those of the ligand-free protein.

Discussion

Exploiting its affinity for human erythrocytic stroma, we have isolated a new lectin from the widely diffused and very

popular species of wild edible mushrooms *Boletus edulis*. The protein, an effective inhibitor of human cancer cell growth, was also structurally characterized, i.e. its amino acid sequence and three-dimensional structure were determined. The result of the X-ray analysis of single crystals revealed that the lectin belongs to the well-known structural protein family that shares the β -trefoil fold, characterized by 12 β strands folded into three similar trefoil subunits associated to give an overall structure with pseudo-3-fold symmetry (Murzin et al. 1992). Many proteins with relatively low amino acid sequence similarity and very different binding specificities share this fold and perform very diverse functions. Within the group, an increasing number is being found to bind carbohydrates and therefore to belong to the lectin family. In particular, the β -trefoil fold is sometimes called “a galactose-binding fold” to indicate a specificity confirmed here by our results.

BEL β -trefoil is a relatively stable novel dimer in which the two protomers associate very differently from those of the *Sclerotinia sclerotiorum* agglutinin (SSA) that has an identical fold and binds the same carbohydrates but possesses a single binding site per protomer (Sulzenbacher et al. 2010; PDB accession code 2X2S). Figure 6 shows the two dimers with one protomer superimposed, the red model is BEL β -trefoil and the yellow is SSA. It is evident that the intermolecular contacts in the dimers are completely different and that the six

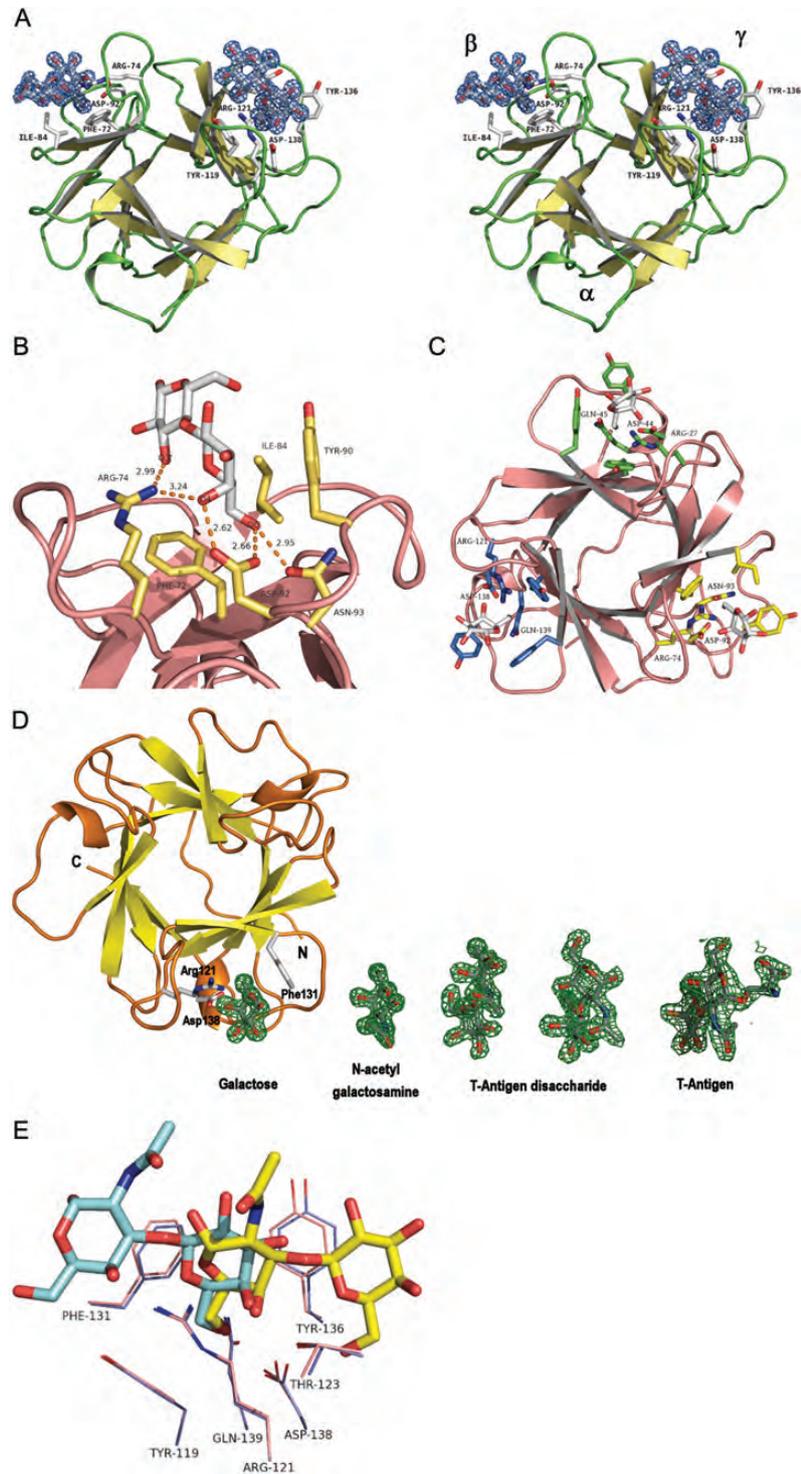


Fig. 5. Binding of carbohydrates to BEL β -trefoil. (A) Stereo diagram of a protomer of BEL β -trefoil with two molecules of lactose bound at the β and γ sites. The electron density of the $2F_{\text{obs}} - F_{\text{calc}}$ map was contoured at the 1.5σ level. The side chains of the main amino acids involved in the interactions are represented. (B) Residues involved in the interaction of BEL β -trefoil with lactose in binding site β . The most relevant distances are indicated in the figure. (C) Main residues involved in the interaction of BEL β -trefoil with carbohydrates. The sugar represented in the figure is Gal bound to the three sites of a protomer (crystal form 6). (D) Ribbon representation of the BEL β -trefoil protomer with the $F_{\text{obs}} - F_{\text{calc}}$ electron density of the different saccharides that can occupy its binding sites. The four carbohydrates represented as ball-and-stick models are Gal (bound in binding site γ), GalNAc, the T-antigen disaccharide and the T-antigen. The electron density of the $F_{\text{obs}} - F_{\text{calc}}$ maps was contoured at a 1.5σ level. The side chains of the main amino acids involved in the interactions are represented in the figure. (E) Alternative binding of the T-antigen disaccharide to BEL β -trefoil. The figure represents two binding sites γ superimposed using the program LSQKAB (Kabsch 1978). The two molecules of the T-antigen disaccharide are represented blue (with the lectin side chains in contact violet) and yellow (with the side chains orange). The figure was prepared using the program PyMOL (<http://www.pymol.org>). (The colour version of this figure is available on the website.)

binding sites of the lectin we describe here are accessible for ligand binding.

The novel lectin was found to effectively inhibit the proliferation of nine different human cancer cell lines by using two alternative methods based on completely different principles. This effect is inhibited by the presence of the carbohydrates that selectively bind to the lectin, in a dose-dependent manner. Confocal microscopy using fluorescein-conjugated lectin was then used to show that the lectin exerts its lethal action by penetrating into the cells.

Table IV. Selected distances between the closest BEL β -trefoil residues and the Gal molecule in the three different binding sites

BEL β -trefoil residue	Atom	Galactose atom	Distance (Å)
Tyr 25	CG	C6	3.73 [α]
Arg 27	NH2	O4	3.02 [α]
Tyr 37	CG	C6	3.81 [α]
Tyr 42	CD2	C5	3.69 [α]
Asp 44	OD1	O6	2.67 [α]
Asp 44	OD2	O4	2.52 [α]
Gln 45	OE1	O6	2.87 [α]
Phe 72	CG	C6	3.59 [β]
Arg 74	NH1	O4	3.10 [β]
Ile 84	CD1	C6	3.86 [β]
Tyr 90	CB	O6	3.52 [β]
Asp 92	OD2	O6	2.60 [β]
Asp 92	OD1	O4	2.65 [β]
Asn 93	OD1	O6	2.78 [β]
Tyr 119	CG	C6	3.79 [γ]
Arg 121	NH2	O4	3.03 [γ]
Phe 131	CG	C6	3.80 [γ]
Tyr 136	CD2	C5	3.76 [γ]
Asp 138	OD1	O6	2.73 [γ]
Asp 138	OD2	O4	2.53 [γ]
Gln 139	OD1	O6	2.78 [γ]

The distances refer to protomer A of the asymmetric unit of crystal form number 6. Only one distance per residue has been included in the table.

The lectin was initially eluted from the human erythrocytic stroma columns by exploiting its high affinity for lactose. Subsequent experiments showed that glucose, GlcNAc, chitobiose and mannose do not bind to the protein whereas Gal as well as GalNAc do. This different affinity is explained by the position of the epimeric O4 in the moiety that binds which together with O6 participates in very strong interactions with the protein that confer selectivity to the binding. Three amino acids participate in hydrogen bonds with the saccharides: Arg 27, Asp 44 and Gln 45 in binding site α ; Arg 74, Asp 92 and Asn 93 in binding site β and Arg 121, Asp 138 and Gln 139 in binding site γ . In the high-resolution model of the apoprotein, these three amino acids are hydrogen bonded to a glycerol molecule present at high concentration in the mother liquor. The overall equivalence of Gal and GalNAc as ligands of this lectin is also supported by the observation that the T-antigen disaccharide, Gal β 1-3GalNAc can bind at the active sites with both moieties (see Figure 5E). Another lectin with a binding site for these carbohydrates has been purified from king bolete mushrooms and has been given the name BEL (Bovi et al. 2011). One of the major differences between BEL and BEL β -trefoil is that the former has two different binding sites per protomer with different binding specificities, one that binds glucose, and is absent in BEL β -trefoil and the other that binds Gal and GalNAc and is therefore comparable with the sites present in BEL β -trefoil. The binding of GalNAc is mediated in BEL by several hydrophobic interactions and the hydrogen bond established by its O7 and the OG of Ser 48. At the other end of the sugar molecule, there are two important interactions between O6 and the NH2 of Arg 106 and O5 and the NE2 of His 71. Other hydrogen bonds are formed with the carbonyl of Gly 49 and the N of Asn 72. These differences are not unexpected because the two lectins belong to very different folds and have totally different quaternary structures. Angulo et al. (2011) have examined the binding of carbohydrates to a series of β -trefoil lectins and have found that

Table V. Selected distances between the closest BEL β -trefoil residues of binding site β and the different carbohydrates

BEL β -trefoil residue	Atom	Distance (Å) [Lactose]	Distance (Å) [Galactose]	Distance (Å) [N-acetylgalactosamine]	Distance (Å) [T-antigen disaccharide]	Distance (Å) [T-antigen]
Phe 72	CB	3.69 [C6]	4.00 [C6]	3.89 [C6]	3.90 [C6]	–
Phe 72	CG	3.58 [C6]	3.69 [C6]	3.59 [C6]	3.68 [C6]	3.66 [C6]
Phe 72	CD2	3.43 [C6]	3.43 [C6]	3.41 [C6]	3.50 [C6]	3.44 [C6]
Arg 74	CD	3.47 [O4]	3.62 [O4]	3.56 [O4]	3.50 [O4]	3.37 [O4]
Arg 74	NH1	3.00 [O3']	–	–	–	–
Arg 74	NH1	2.93 [O5]	3.11 [O5]	3.05 [O5]	3.27 [O5]	3.08 [O5]
Arg 74	NH1	3.23 [O4]	3.35 [O4]	3.15 [O4]	3.13 [O4]	3.02 [O4]
Arg 74	NH2	3.42 [O3']	–	–	–	–
Ile 84	CD1	3.83 [O6']	–	–	–	–
Ile 84	CD1	3.75 [C5]	3.84 [C5]	3.81 [C5]	3.94 [C5]	3.97 [C5]
Ile 84	CD1	3.88 [C6]	3.79 [C6]	3.85 [C6]	3.83 [C6]	3.77 [C6]
Tyr 90	CB	3.50 [O6]	3.70 [O6]	3.80 [O6]	3.56 [O6]	3.64 [O6]
Tyr 90	CG	3.67 [O6]	3.88 [O6]	3.99 [O6]	3.75 [O6]	3.75 [O6]
Tyr 90	CD1	3.70 [C3]	3.94 [C3]	3.73 [C3]	3.85 [C3]	3.73 [C3]
Asp 92	OD1	2.63 [O4]	2.67 [O4]	2.63 [O4]	2.67 [O4]	2.73 [O4]
Asp 92	OD2	2.67 [O6]	2.69 [O6]	2.67 [O6]	2.65 [O6]	2.75 [O6]
Asn 93	CG	3.78 [O6]	3.89 [O6]	3.80 [O6]	3.73 [O6]	3.91 [O6]
Asn 93	OD1	2.96 [O6]	2.88 [O6]	2.78 [O6]	2.76 [O6]	2.92 [O6]

The distances refer to site β of protomer B of the asymmetric unit of crystal forms number 5, 6, 7, 8 and 9. In the model of lactose the prime is used to indicate the oxygens of the glucose moiety. The maximum distance cutoff is 4.0 Å.

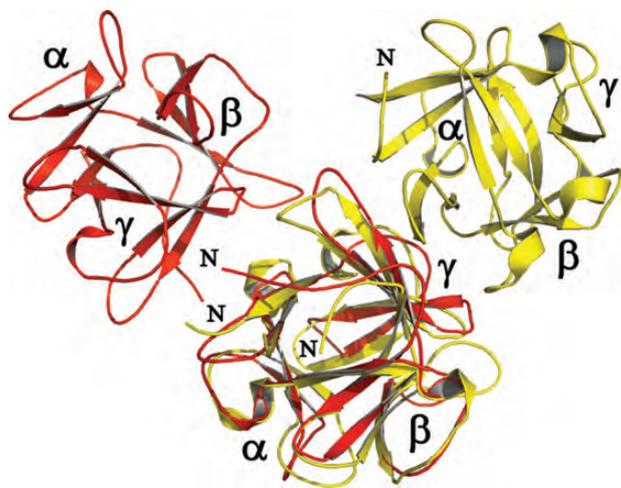


Fig. 6. Comparison of two β -trefoil dimeric lectins. The two models superimposed are those of BEL β -trefoil (red) and the lectin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* (yellow, Sulzenbacher et al. 2010, PDB accession code 2X2S). The superposition of two protomers, one of each of the two lectins, was carried out using the program LSQKAB (Kabsch 1978) and shows clearly that the contacts between the protomers in the two dimers are very different. (To view the colour version of this figure please see the online version of this article at glycob.oxfordjournals.org.)

despite the common ligands and three-dimensional similarity of the scaffold, the sugar-binding sites differ significantly. Since BEL β -trefoil is most similar in both three-dimensional structure and amino acid sequence to the N-terminal module of the *L. sulphureus* lectin (LSL₁₅₀), the binding of lactose, the only carbohydrate for which there is information for both proteins, is worth comparing. Both LSL₁₅₀ and BEL β -trefoil do not bind the saccharide at site α , a fact explained in the case of LSL₁₅₀ by the absence of an aromatic residue at this site that should be preferably a tryptophan. That explanation cannot be used for BEL β -trefoil that binds Gal and other Gal-containing carbohydrates at this site where Tyr 25 and 37 make hydrophobic contacts with the saccharides. Other aspects of ligand binding are very similar in both proteins.

In most cases, the functional roles of fungal lectins are still not clearly established. An exception is the lectin CCL2, from the ink cap mushroom *Coprinopsis cinerea* whose nuclear magnetic resonance (NMR) β -trefoil structure was determined recently and which is known to be involved in the defence of the mushroom against predators and parasites (Schubert et al. 2012). Two other possible functions have been suggested for the other characterized dimeric β -trefoil fungal lectin SSA and they are as a storage protein or as a molecule with a role in development and morphogenesis (Sulzenbacher et al. 2010, PDB accession code 2X2S). Clearly, these same three physiological functions may be attributed to BEL β -trefoil.

It is also worth mentioning that lectins have been proposed as potentially useful candidates to coat nanoparticles in targeted drug delivery toward cancer cells (Lehr 2000; Smart 2004), but a very severe obstacle to this use is the likely generation of antibodies against the foreign molecule. A strategy used for other proteins that are being tested for the selective delivery of the drug loaded particles is to “humanize” them, in other words, to use protein engineering to render them as

similar as possible as a human counterpart and therefore nonrecognizable by the immune system. Many fungal lectins that recognize the T-antigen disaccharide and therefore malignant cells belong to folds that are not found among human proteins but BEL β -trefoil is a potential candidate for this type of application for two reasons. The first is that its cDNA sequence is available and we have found that the protein is expressed readily in a soluble form by *E. coli* (data not shown). The second reason is that when confronting the coordinates of the lectin in the Dali server (Holm and Sander 1999), a human acidic fibroblast growth factor is found to be structurally highly similar to BEL β -trefoil (Blaber et al. 1996). If there exists a human protein very similar to the foreign molecule, it is conceivable to modify the latter to attempt to evade the immune response. These characteristics make BEL β -trefoil an interesting target for potential biotechnological applications.

Materials and methods

Protein purification

BEL β -trefoil was initially purified from *Boletus edulis* fruiting bodies by affinity chromatography in a column of human erythrocytic stroma incorporated into a polyacrylamide gel (Betail et al. 1975) as described elsewhere (Bovi et al. 2011). A second method was subsequently developed that was based on the binding of the lectin to desialylated hog gastric mucin immobilized to Sepharose 4B (Freier et al. 1985) and desorption with lactose. Both methods gave unsatisfactory yields and therefore a third method was developed.

In each preparation, 400 g of king bolete mushrooms were homogenized in a blender using the same volume of phosphate-buffered saline (PBS). After centrifugation at 5000 \times g for 30 min, the supernatant was filtered through glass wool to eliminate solid aggregates and 50% w/v ammonium sulfate was added to the supernatant. The suspension was centrifuged at 10,000 \times g for 15 min and BEL β -trefoil was recovered in the pellet that was dissolved in roughly the starting volume of PBS. The proteins were then dialyzed against Tris-HCl 20 mM, pH 7.5, and loaded onto a DEAE cellulose column, previously equilibrated with the same buffer. The column was first washed with NaCl 30 mM Tris-HCl 20 mM, pH 7.5, until the absorbance of the effluent at 280 nm was negligible and the bound proteins were eluted with 250 mM NaCl Tris-HCl, pH 7.5. The eluted proteins were concentrated and further purified by gel filtration chromatography, using a Superdex G75 column. Elution was carried out with the Tris-HCl 20 mM buffer containing 150 mM NaCl. The fractions containing BEL β -trefoil, identified using SDS-PAGE, were then dialyzed exhaustively against 20 mM Tris-HCl, pH 7.5, to remove salts and, after concentration, were loaded onto a MonoQ column that was used to separate the different isoforms that had been previously identified by IEF. A 0–300 mM NaCl gradient resolved six different peaks, corresponding to six different BEL β -trefoil isoforms. The most abundant was used for the structural studies and the yield was ~40 mg of this isoform and ~20 mg of all the others.

Before use, the pure protein was further submitted to a hydrophobic interaction chromatography step using a Lipidex

1000 column at 37°C to remove potential hydrophobic ligands bound to the lectin.

Amino acid sequence analysis

Automated Edman degradation and N-terminal sequence analysis of the most abundant isoform of BEL β -trefoil (0.5 nmol) was carried out on a Hewlett–Packard model G 1000A sequencer connected on-line to a phenylthiohydantoin analyzer from the same manufacturer.

Proteolytic digestions and LC-ESI-MS/MS analysis of peptides

Aliquots (50 μ g) of BEL β -trefoil were dissolved in 100 μ L of 100 mM ammonium carbonate, pH 8.5, and reduced for 20 min at 56°C with 1 mM DTT. Proteolysis was then performed with either trypsin (Trypsin Gold, Promega), chymotrypsin (sequencing grade, Promega) or endopeptidase V8 from *S. aureus* (Sigma-Aldrich). The digestions with trypsin and chymotrypsin were performed at 37°C for 2 h, adding 1.8 μ g of proteases and 0.01% (v/v) ProteaseMAX™ Surfactant (Promega). The reaction with V8 endopeptidase was carried out with 2 μ g of enzyme at 37°C for 4 h. In every case, the reaction was stopped by the addition of 1% (v/v) formic acid.

Peptides were then resolved in an analytical Jupiter C18 column (4 μ m, 150 \times 2 mm, Phenomenex) at a flow rate of 0.2 mL/min with a gradient 3–70% B in 90 min, 70–97% B in 5 min and 98% B for 10 min. Solvents consisted of water (A) and acetonitrile (B) both containing 0.1% formic acid. The eluent was analyzed online with an LCQ ion trap mass spectrometer (Thermo Scientific) with Electrospray ionization (ESI) ion source controlled by the Xcalibur software 1.4 (Thermo Scientific). Mass spectra were generated in positive ion mode under constant instrumental conditions: Source voltage 5.0 kV, capillary voltage 46 V, sheath gas flow 55 (arbitrary units), auxiliary gas flow 33 (arbitrary units), sweep gas flow 1 (arbitrary units), capillary temperature 200°C, tube lens voltage –105 V. MS/MS. Spectra, obtained by collision-induced dissociation (CID) in the linear ion trap, were recorded with an isolation width of 3 Da (m/z), the activation amplitude was 35% of ejection RF amplitude that corresponds to 1.58 V. Tandem mass spectra were interpreted manually with the assistance of the prediction algorithm for peptide fragmentation ProteinProspector (Clauser et al. 1999) and were automatically analyzed using the Peaks Studio software (Zhang et al. 2012) version 5.2 (Bioinformatic Solution, Inc., Waterloo, Ont., Canada).

Cloning and cDNA sequence

Total RNA was isolated from fresh *Boletus edulis* mushrooms using the TRIZOL (Sigma) protocol. One microgram of RNA was used to synthesize the first cDNA strand by means of the Superscript First Strand Synthesis System (Invitrogen) using an Oligo(dT)₁₈ primer. The BEL β -trefoil sequence was amplified by PCR using a High Fidelity Taq DNA polymerase (Jena Biosciences) with degenerated primers deduced from the amino acid sequence of the N- and C-terminal portions of the protein. The primers used were the following: Forward, ATGGTNAAYTTYCCNAAYATHCC, reverse, YTCRAAYTTCCANAGYTCRTT. The amplicon, separated on a 1% agarose gel, was

eluted and directly TA cloned into the pGEM-T Vector System (Promega). The plasmids from four positive colonies were isolated and sequenced using the standard T7 forward and SP6 reverse primers.

Cell proliferation assays

MTT assay (Alley et al. 1988). One hundred microliters of MCF-7 (breast adenocarcinoma), HepG-2 (hepatocellular carcinoma), CaCo-2 (colorectal adenocarcinoma) and CFPAC-1 (pancreatic duct adenocarcinoma) cells in RPMI-1640 medium without phenol red (Sigma) supplemented with 10% fetal bovine serum (FBS) were seeded onto 96-well plates at a final density of 5×10^3 cells/well in a total volume of 180 μ L and incubated overnight at 37°C in humidified atmosphere with 5% CO₂. BEL β -trefoil (2 mg/mL in sterile PBS) was added to each well to the final concentrations of 40, 20, 10, 5, 2.5, 1.25 and 0 μ g/mL and the plates were incubated at 37°C, 5% CO₂ for 48 h. Doxorubicin at concentrations of 200, 20, 2, 0.2, 0.02 and 0.002 μ M was used as positive control. Twenty microliters of 5 mg/mL MTT (Thiazolyl Blue Tetrazolium Bromide, Sigma) solution in PBS were added to each well and the plates incubated again at 37°C for 3 h. The MTT assay is based on the enzymatic conversion of a yellow tetrazolium salt to an insoluble formazan product by the mitochondria of viable cells. At the end of the incubation period, the medium was removed and the cells were washed twice with PBS. Two hundred microliters of DMSO were added to each well and the plates were shaken until the complete solubilization of the dye was achieved. The absorbance of the converted dye was read in a plate reader at a wavelength of 570 nm. Each experiment was carried out in triplicate.

³Hthymidine incorporation into DNA assay (Yu et al. 1993)

Cells from human cell lines HeLa (cervix adenocarcinoma), U-87 MG (glioblastoma), HT-29 (colon carcinoma), SK-MEL-28 (melanoma) and A549 (lung adenocarcinoma) were seeded at a density of $1.0\text{--}2.0 \times 10^4$ cells/well in 0.5 mL of Dulbecco's modified Eagle's medium (DMEM) containing 5% FBS in 24-well plates. After 48 h incubation at 37°C in 5% CO₂, the growth medium of each well was replaced by 0.5 mL of DMEM containing 250 μ g/mL of bovine serum albumin and BEL β -trefoil at the concentrations used in the MTT assay. The cells were incubated for a further 48 h prior to a 1 h pulse with 0.5 μ Ci/well [methyl-³H]-thymidine. In the subsequent step, the cells were washed twice with PBS and precipitated by addition of 0.5 mL/well of 5% trichloroacetic acid at 4°C. After two washes with 0.5 mL/well of 95% ethanol at 4°C, the precipitate was solubilized with 0.5 mL/well of 0.2 M NaOH. The dissolved precipitate was mixed with scintillation solution, and the cell-associated radioactivity was counted using a liquid-scintillation spectrometer. As before, each experiment was carried out in triplicate.

Confocal microscopy

For confocal microscopy, purified BEL β -trefoil (2 mg) was chemically conjugated to (fluorescein isothiocyanate (FITC), Sigma) following a standard protocol. The fluorescein/protein molar ratio, as determined by absorbance at 280 and 495 nm,

was 0.85. One hundred microliters of MCF-7, CaCo-2 and CFPAC-1 cells ($\sim 1 \times 10^6$ cells/mL) in RPMI-1640 medium without phenol red (Sigma) supplemented with 10% fetal bovine serum were seeded onto 18 mm round cover slips in Petri dishes and left to attach overnight at 37°C in humidified atmosphere with 5% CO₂. The cover slips were washed three times with 2 mL of PBS, incubated for 2 h with 50 µg/mL of FITC-labeled lectin in PBS and washed again. The cells were fixed with 4% paraformaldehyde for 15 min. To visualize the nuclei, the slips were treated with a 4' 6-diamidino-2-phenylindole (DAPI) solution (100 ng/mL) for 10 min at room temperature. The cover slips were fixed on glass slides with one or two drops of aqueous anti-fading mounting medium and sealed with nail polish. Images at different focal planes were collected on a Leica tcs-sp5 confocal microscope. Both 405 and 488 nm lasers were used for the excitation of DAPI and FITC, respectively.

Crystallization and X-ray structure determination

Three different isoforms crystallized in different space groups. Two were obtained using the second purification method based on the binding of the lectin to immobilized desialylated hog gastric mucin. These two isoforms crystallized in two different space groups under the same crystallization conditions, depending also on whether the protein was apo or complexed with lactose. The two crystal forms were grown using the vapor diffusion method by mixing equal volumes of the protein solution at concentration of 18 mg/mL and a solution 20% in polyethylene glycol (PEG) 8000, 0.2 M magnesium acetate and 0.1 M sodium cacodylate, pH 6.5, as the precipitant. The crystals of the apoprotein belong to the trigonal space group $P3_221$ with unit cell parameters $a = 77.7$ Å and $c = 52.8$ Å and contain one protomer in the asymmetric unit. The crystals of the complex of the lectin with lactose belong to the monoclinic space group $P2_1$ with $a = 72.1$ Å, $b = 39.1$ Å, $c = 119.4$ Å and $\beta = 93.2^\circ$.

The third isoform was crystallized as apoprotein using two different precipitants and yielded three different crystal forms. The precipitant solution of the first form contained 1.5 M ammonium sulfate, 10% glycerol and 0.1 M Tris-HCl, pH 8.5. These are the monoclinic crystals, space group $P2_1$, that diffract to the highest resolution, 1.12 Å, and the unit cell parameters are $a = 34.5$ Å, $b = 66.8$ Å, $c = 63.9$ Å and $\beta = 96.1^\circ$. The precipitant solution of the second and third form contained 25% PEG 4000, 0.2 M magnesium chloride, 0.2 M 1-butyl-3-methylimidazolium chloride and 0.1 M Tris-HCl, pH 8.5. The crystals of the second form are orthorhombic, space group $P2_12_12_1$, those of the third form belong instead to the space group $P2_1$ but the unit cell parameters are $a = 65.4$ Å, $b = 70.0$ Å, $c = 71.8$ Å and $\beta = 106.8^\circ$. These crystals contain two dimers in the asymmetric unit and turned out to be very useful because the intermolecular contacts did not block the ligand-binding sites and for that reason they were used to prepare co-crystals by soaking them in solutions that contained the different mono and disaccharides in mother liquor saturated with the carbohydrates.

A summary of the data of nine different crystal forms is presented in Table I.

The final diffraction data were collected from crystals frozen at 100 K after a brief immersion in a mixture of 80% of the mother liquor and 20% PEG 400. The orthorhombic crystal form of the apoprotein was solved by the SIRAS method using a gold derivative and collecting the data with the home source. The native protein and derivative data were obtained using copper K α radiation from a Rigaku RU-300 rotating anode X-ray generator with a Mar345 imaging plate area detector.

The heavy atom derivative was prepared by overnight soaking of a crystal in mother liquor with the addition of NaAuBr₄ at a final concentration of ~ 10 mM. Initial phases to 2.5 Å resolution were determined by single isomorphous replacement with the Bijvoet pairs of the heavy atom derivative of the orthorhombic form. The major gold site was located in a difference Patterson map (Sheldrick 2008) and refined using the program MLPHARE (Collaborative Computational Project Number 4). The site was subsequently used as input for the program autoSHARP (Bricogne et al. 2003) that was used to locate three other sites using also the anomalous data, and for density modification and final phasing. The electron density map thus produced was of excellent quality and could be readily interpreted. The initial model of the apoprotein was built in the high-quality map at 2.5 Å resolution using the program Xtalview (McRee 1999). Refinement was carried out initially using the program REFMAC (Murshudov et al. 1997). During the process of refinement and model building, the quality of the model was controlled with the program PROCHECK (Laskowski et al. 1993). The refined model thus produced was used to solve all the other crystal forms by molecular replacement using the program MOLREP (Vagin and Teplyakov 2000).

The data used for the final refinement of all the forms, apoprotein and co-crystals, were collected at the XRD1 beamline of the Elettra synchrotron in Trieste and various beamlines of the European Synchrotron Radiation Facility in Grenoble. Data were indexed, integrated and reduced using the programs MOSFLM (Leslie 1992) and Scala (Collaborative Computational Project Number 4). A nonstandard protocol was used to collect the high-resolution data of the apoprotein (crystal form 1), i.e. 1600 0.25° oscillation frames were collected and processed. The excellent statistics of the highest resolution data are explained by this choice and the use of the ID29 beamline in Grenoble. The final refinement of the four crystal forms of the apoprotein was carried out initially using the program REFMAC (Murshudov et al. 1997) and, in a second stage, with the program Phenix.refine (Adams et al. 2010). The models were built using the program Coot (Emsley et al. 2010) and finally subjected to a round of TLS refinement. The mono and disaccharides in the co-crystals were modeled into difference Fourier maps phased by the refined, unliganded structure. The models of the complexes were refined with REFMAC using the same criteria followed in the refinement of the apoprotein. Solvent molecules were added to the models in the final stages of refinement according to hydrogen-bond criteria and only if their B factors refined to reasonable values and if they improved the R free.

The diffraction data and refinement statistics of all the models are summarized in Table I. The figures of the models

were prepared using the program PyMOL (<http://www.pymol.org>).

Acknowledgements

We are grateful to Maili Zimmermann and Erika Lorenzetto for assistance with the confocal microscopy, to Michela Sega for the help with cell cultures and to the staff of the Elettra synchrotron and the ESRF in Grenoble (Proposal MX 1309) for assistance during data collection. The coordinates of the models and the structure factors of the apoprotein and the complexes with carbohydrates have been deposited in the protein data bank; accession codes 4I4O, 4I4P, 4I4Q, 4I4R, 4I4S, 4I4U, 4I4V, 4I4X and 4I4Y.

Funding

This work was supported by Fondazione Cassa di Risparmio di Verona, Vicenza, Belluno e Ancona and Rete delle Imprese per la Tutela dei Funghi di Bosco. M.B. is recipient of a scholarship granted by Fondazione Cassa di Risparmio di Verona, Vicenza, Belluno e Ancona.

Conflict of interest

None declared.

Abbreviations

BEL, *Boletus edulis* lectin; BSA, bovine serum albumin; CID, collision-induced dissociation; DAPI, 4'-6-diamidino-2-phenylindole; DMEM, Dulbecco's modified Eagle's medium; FBS, fetal bovine serum; FITC, fluorescein isothiocyanate; Gal, galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; LC, liquid chromatography; MS, mass spectrometry; MTT, 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide; NMR, nuclear magnetic resonance; PBS, phosphate buffered saline; SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis; SSA, *Sclerotinia sclerotiorum* agglutinin.

References

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 66:213–221.
- Alley MC, Scudiero DA, Monks A, Hursey ML, Czerwinski MJ, Fine DL, Abbott BJ, Mayo JG, Shoemaker RH, Boyd MR. 1988. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Res*. 48:589–601.
- Angulo I, Acebrón I, de las Rivas B, Muñoz R, Rodríguez-Crespo I, Menéndez M, García P, Tateno H, Goldstein IJ, Pérez-Agote B, et al. 2011. High-resolution structural insights on the sugar-recognition and fusion tag properties of a versatile β -trefoil lectin domain from the mushroom *Laetiporus sulphureus*. *Glycobiology*. 21:1349–1361.
- Betail G, Coulet M, Genaud L, Guillot J, Scandariato M. 1975. Erythrocyte stroma included in polyacrylamide gel. Applications to affinity chromatography. *C R Séances Soc Biol Fil*. 169:561–566.
- Bies C, Lehr CM, Woodley JF. 2004. Lectin-mediated drug targeting: History and applications. *Adv Drug Deliv Rev*. 56:425–435.
- Blaber M, DiSalvo J, Thomas KA. 1996. X-ray crystal structure of human acidic fibroblast growth factor. *Biochemistry*. 35:2086–2094.
- Bovi M, Carrizo ME, Capaldi S, Perduca M, Chiarelli LR, Galliano M, Monaco HL. 2011. Structure of a lectin with antitumoral properties in king bolete (*Boletus edulis*) mushrooms. *Glycobiology*. 21:1000–1009.
- Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. 2003. Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. *Acta Crystallogr Sect D Biol Crystallogr*. 59:2023–2030.
- Carrizo ME, Capaldi S, Perduca M, Irazoqui FJ, Nores GA, Monaco HL. 2005. The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J Biol Chem*. 280:10614–10623.
- Clauser KR, Baker P, Burlingame AL. 1999. Role of accurate mass measurement (10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem*. 71:2871–2882.
- Collaborative Computational Project Number 4. 1994. *Acta Cryst*. D50:760–767.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 66:486–501.
- Freier T, Fleischmann G, Rüdiger H. 1985. Affinity chromatography on immobilized hog gastric mucin and ovomucoid. A general method for isolation of lectins. *Biol Chem Hoppe Seyler*. 366:1023–1028.
- Goldstein IJ, Winter HC. 2007. Mushroom lectins. In: Kamerling JP, editors. *Comprehensive Glycoscience: From Chemistry to Systems Biology*, Vol. 3. Amsterdam: Elsevier Ltd. p. 601–621.
- González De Mejía E, Prisecaru VI. 2005. Lectins as bioactive plant proteins: A potential in cancer treatment. *Crit Rev Food Sci Nutr*. 45:425–445.
- Gosavi S, Whitford PC, Jennings PA, Onuchic JN. 2008. Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci USA*. 105:10384–10389.
- Guillot J, Kanska G. 1997. Lectins in higher fungi. *Biochem Syst Eco*. 25:203–230.
- Hazes B. 1996. The (QxW)₃ domain: A flexible lectin scaffold. *Protein Sci*. 5:1490–1501.
- Holm L, Sander C. 1999. Protein folds and families: Sequence and structure alignments. *Nucleic Acids Res*. 27:244–247.
- Imberty A, Mitchell EP, Wimmerová M. 2005. Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Curr Opin Struct Biol*. 15:525–534.
- Jones S, Thornton JM. 1995. Protein-protein interactions: A review of protein dimer structures. *Prog Biophys Mol Biol*. 63:31–65.
- Kabsch W. 1978. A solution for the best rotation to relate two sets of vectors. *Acta Cryst*. A32:922–923.
- Khan F, Khan MI. 2011. Fungal lectins: Current molecular and biochemical perspectives. *Int J Biol Chem*. 5:1–20.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 26:283–291.
- Lee J, Blaber SI, Dubey VK, Blaber M. 2011. A polypeptide “building block” for the β -trefoil fold identified by “top-down symmetric deconstruction”. *J Mol Biol*. 407:744–763.
- Lehr CM. 2000. Lectin-mediated drug delivery: The second generation of bioadhesives. *J Control Release*. 65:19–29.
- Leslie AGW. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACMB Newslett Protein Crystallogr*. 26:27–33.
- Mancheño JM, Tateno H, Goldstein IJ, Martínez-Ripoll M, Hermoso JA. 2005. Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J Biol Chem*. 80:17251–17259.
- McRee DE. 1999. XtalView/Xfit A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol*. 125:156–165.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 53:240–255.
- Murzín AG, Lesk AM, Chothia C. 1992. β -Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 β and 1 α and fibroblast growth factors. *J Mol Biol*. 223:531–543.
- Naeem A, Saleemuddin M, Khan RH. 2007. Glycoprotein targeting and other applications of lectins in biotechnology. *Curr Protein Pept Sci*. 8:261–271.
- Ponstingl H, Henrick K, Thornton JM. 2000. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins*. 41:47–57.

- Reisner Y, Itzicovitch L, Meshorer A, Sharon N. 1978. Hemopoietic stem cell transplantation using mouse bone marrow and spleen cells fractionated by lectins. *Proc Natl Acad Sci USA*. 75:2933–2936.
- Schubert M, Bleuler-Martinez S, Butschi A, Wälti MA, Egloff P, Stutz K, Yan S, Wilson IB, Hengartner MO, Aebi M, et al. 2012. Plasticity of the β -trefoil protein fold in the recognition and control of invertebrate predators and parasites by a fungal defence system. *PLoS Pathog*. 8:e1002706.
- Sharon N. 2007. Lectins: Carbohydrate-specific reagents and biological recognition molecules. *J Biol Chem*. 282:2753–2764.
- Sharon N, Lis H. 2004. History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology*. 14:53R–62R.
- Sheldrick GM. 2008. A short history of SHELX. *Acta Cryst*. A64:112–122.
- Singh RS, Bhari R, Kaur HP. 2010. Mushroom lectins: Current status and future perspectives. *Crit Rev Biotechnol*. 30:99–126.
- Smart JD. 2004. Lectin-mediated drug delivery in the oral cavity. *Adv Drug Deliv Rev*. 56:481–489.
- Springer GF. 1984. T and Tn, general carcinoma autoantigens. *Science*. 224:1198–1206.
- Springer GF. 1997. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J Mol Med*. 75:594–602.
- Sulzenbacher G, Roig-Zamboni V, Peumans WJ, Rougé P, Van Damme EJ, Bourne Y. 2010. Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a beta-trefoil domain. *J Mol Biol*. 400:715–723.
- Vagin A, Teplyakov A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*. 56:1622–1624.
- Yu LG. 2007. The oncofetal Thomsen–Friedenreich carbohydrate antigen in cancer progression. *Glycoconj J*. 24:411–420.
- Yu L-G, Fernig DG, Smith JA, Milton JD, Rhodes JM. 1993. Reversible inhibition of proliferation of epithelial cell lines by *Agaricus bisporus* (edible mushroom) lectin. *Cancer Res*. 53:4627–4632.
- Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. 2012. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomic*. 11:1–8.