

Structure of Two Crystal Forms of Sheep β -Lactoglobulin with EF-Loop in Closed Conformation

Joanna I. Loch, Marta Molenda, Magdalena Kopeć, Sylwia Świątek, Krzysztof Lewiński,
Department of Crystal Chemistry and Crystal Physics, Faculty of Chemistry Jagiellonian University in Kraków, Ingardena 3,
Kraków 30-060, Poland

Received 20 November 2013; revised 31 January 2014; accepted 7 February 2014

Published online 15 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22471

ABSTRACT:

Ovine β -lactoglobulin has been isolated from whey fraction of sheep milk and crystallized. The high-resolution structures of two crystal forms (triclinic and trigonal) obtained at pH 7.0 have been determined revealing that ovine protein, similarly to its bovine analog, is dimeric. Access to the binding site located in the eight-stranded antiparallel β -barrel in both structures is blocked by the EF loop that has been found in closed conformation. Similarly to bovine lactoglobulin (BLG), conformation of the EF loop is stabilized by hydrogen bond between Glu89 and Ser116 indicating that Tanford transition might occur with the same mechanism. The substitution at six positions in relation to the most abundant isoform B of BLG also affects the distribution of electrostatic potentials and the total charge. © 2014 Wiley Periodicals, Inc. *Biopolymers* 101: 886–894, 2014.

Keywords: β -lactoglobulin; sheep milk; Tanford transition; allergenicity; amino acid substitutions

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of the preprint by emailing the *Biopolymers* editorial office at biopolymers@wiley.com

Additional Supporting Information may be found in the online version of this article.

Correspondence to: K. Lewiński; e-mail: lewinski@chemia.uj.edu.pl

Contract grant sponsor: Polish National Science Centre

Contract grant number: DEC-2012/05/B/ST5/00278

Contract grant sponsor: European Regional Development Fund in the framework of the Polish Innovation Economy Operational Program

Contract grant number: POIG.02.01.00–12-023/08

© 2014 Wiley Periodicals, Inc.

INTRODUCTION

β -lactoglobulin (LG) is a small globular protein that has been identified in the whey fraction of the milk of many mammals, for example, cow, goat, sheep, horse, cat, dog, reindeer, dolphin, donkey, and pig, but not in human milk.¹ Bovine lactoglobulin (BLG), due to its accessibility, low price, and its wide utilization in the food and dairy industry, is the most thoroughly investigated lactoglobulin, in contrast to lactoglobulins from the milk of other mammals. This protein has been recognized as a molecular transporter of hydrophobic compounds such as retinol and fatty acids.¹ Several crystal structures of BLG in complexes with saturated and unsaturated fatty acids are available in the protein data bank (PDB).^{2–7} To date, besides bovine protein, the structures of lactoglobulin have been determined only for reindeer⁸ and pig⁹ protein.

The milk of goat and sheep contain proteins that share the same structural, functional, and biological properties as bovine proteins.¹⁰ The differences and similarities between cow, goat, and sheep milk have been well summarized by Park et al.¹¹ Both in cow and sheep milk, the major whey proteins are LG and α -lactalbumin, other proteins such as serum albumin, immunoglobulins, and lactoferrin are present in lower concentration.¹¹

There is not much information about the physicochemical properties of sheep lactoglobulin (SLG). Reports describing the isolation of SLG from milk appeared in 1965.¹² At that time, it was found that ovine lactoglobulin has two isoforms and that the molecular weight of a polypeptide chain is about 18,000 Da. In 1967, Bell and McKenzie performed starch–gel electrophoresis of SLG at pH 7.6 and designated the faster-moving band of LG as variant A and the slower one as isoform B. The ultraviolet spectrum for variant A and B showed differences that were attributed to different contents of tyrosine.¹³ They also observed that variant A had crystallized readily at pH 5.2, while it was impossible to crystallize variant B. Currently,

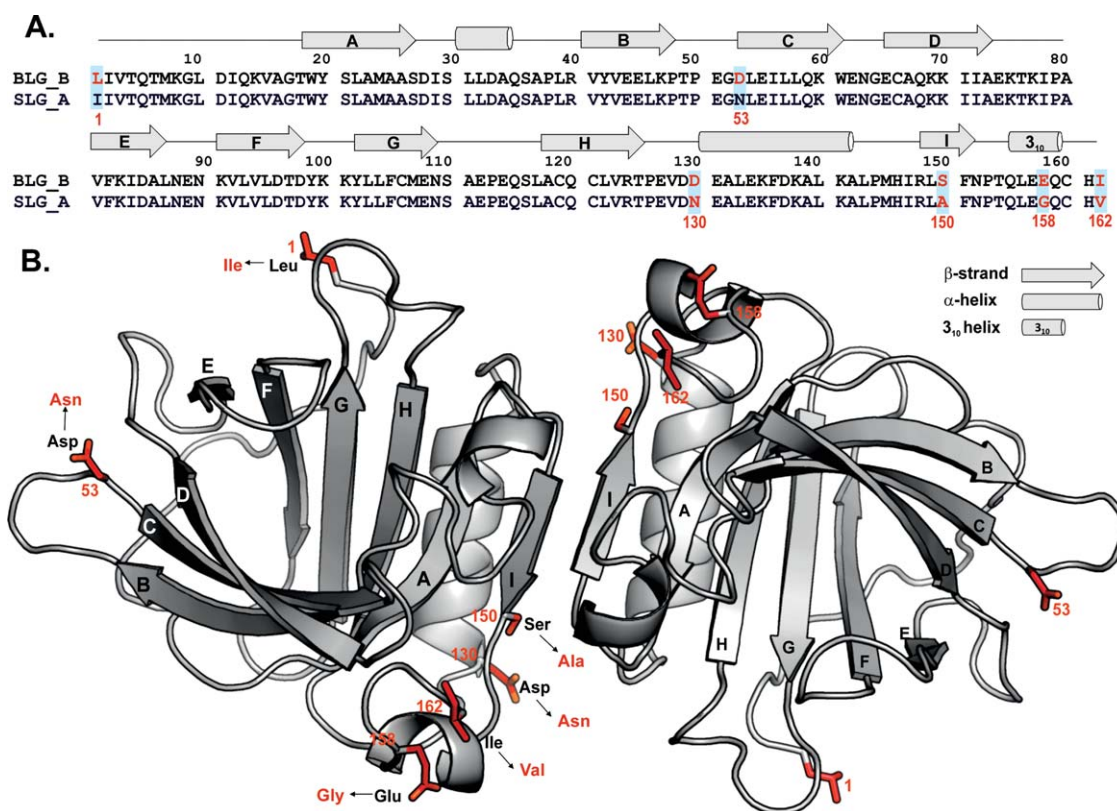


FIGURE 1 (A) Alignment of two sequences: BLG (isoform B) and SLG (isoform A). (B) Positions of amino acid substitutions in three-dimensional structure of BLG dimer.

three isoforms (A, B, and C) of SLG have been identified (UniProt ID: P02756). All of them, similar to the bovine protein, are 162 residues long. In comparison to the most abundant variant B of BLG, the isoform A of sheep protein (SLG-A) that is used in this study contains substitutions at six positions (Figure 1). On the basis of sequence homology, it was concluded that ovine protein, similar to bovine, belongs to the lipocalin family.¹⁴ All lipocalins share the same structural motif in which an eight- or nine-stranded antiparallel β -barrel surrounded by flexible loops forms the core of the protein. The β -strands that in the bovine protein form the β -barrel are named A–H, while loops surrounding the entrance to the ligand binding site are AB, CD, EF, and GH (Figure 1). Access to the β -barrel interior is regulated by the reversible conformational change of the EF loop at pH about 7.0–7.5 in a process called the Tanford transition.^{15–17}

Studies of ligand binding by BLG isoforms A and B have revealed that even small differences in a protein sequence, such as the substitution of two residues, affect the thermodynamic parameters of this process.⁵ Therefore, we decided to investigate also SLG to follow the effect of amino acid substitutions on structural properties and their potential impact on protein allergenicity.

RESULTS AND DISCUSSION

Overall Structure of Ovine Lactoglobulin

Mass spectrometry showed that the LG isolated from ovine milk and used for crystallization is only isoform A. High-resolution structures of two crystal forms of SLG, triclinic SLG-1 (P1) and trigonal (P3121) have been determined. In sheep protein, similar to BLG, the core of the molecule is an eight-stranded antiparallel β -barrel. The entrance to the barrel is surrounded by flexible loops: AB, CD, EF, and GH. The EF loop involved in Tanford transition is in the closed conformation in both SLG-1 and SLG-2 structures. The structures of SLG revealed that similarly to BLG, the ovine protein is dimeric. In triclinic structure SLG-1, the asymmetric unit contains an entire SLG dimer, while in trigonal structure SLG-2, the dimer is located on the crystallographic twofold axis and the asymmetric part of the unit cell contains only one dimer subunit (Figure 2). In both structures, the subunits are linked by four hydrogen bonds between the main chain atoms of His146 to Ala150 and their symmetrical counterparts. Additional contacts are made by side chains of Arg40 and Asp33' on the β -strand B and AB loop, respectively. The crystal

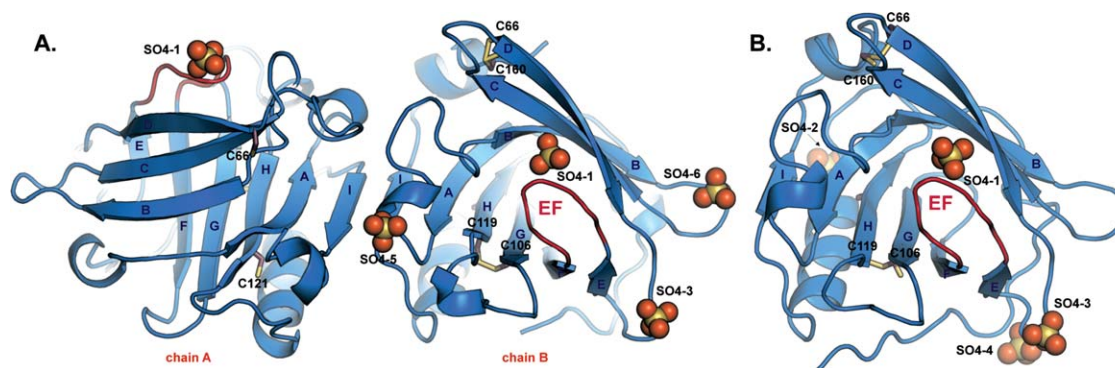


FIGURE 2 (A) Asymmetric unit of SLG-1 structure and (B) SLG-2 structure.

structure of SLG-1 shares many similarities with the structure of the unliganded BLG 1BEB deposited in PDB in 1997.¹⁸ In both structures, the molecule has an EF loop in closed conformation, the unit cell has a similar volume, the same triclinic symmetry P1 and both structures contain a dimer in the asymmetric unit. The second crystal form of SLG, SLG-2, despite its crystallization conditions very similar to the SLG-1, has trigonal symmetry P3₁21. These crystals resemble the most frequently observed trigonal form of BLG; however, the crystal packing, space group, and unit cell parameters are different than those observed in any crystal form of LG deposited in PDB.

In the SLG-2 structure determined at 1.90 Å resolution, the well-defined electron density along the entire protein chain allowed the positions of all 162 residues to be found unambiguously. In contrast, in the triclinic SLG-1 structure, despite its higher resolution (1.40 Å), poor electron density was observed for some fragments of the protein chain indicating a structural disorder different for each subunit; in particular, in the Gly158-Val162 region. The N-terminal residues, Ile1-Thr4, in

chain B were omitted due to poor electron density, while in chain A these residues are very well defined. A similar disorder at the C- and N-terminal part of the protein chain was also observed in the triclinic structure of BLG (PDBid:1BEB).¹⁸

The structures of SLG-1A, SLG-1B, and SLG-2 were compared with the trigonal crystal structure of BLG 3BLG¹⁶ containing an EF loop in closed conformation. The superposition of C α gave RMS values of about 0.7 Å for residues from 5 to 157 indicating high similarities between sheep and bovine protein (Figure 3A). However, analysis of the difference distance matrices (DDM) revealed systematic distortion of the β -barrel in all SLG structures in comparison to 3BLG. The β -strands B, C, and D in sheep protein are up to about 1 Å closer to β -strands F, G, and H located on the opposite side of the β -barrel. The distortion region and shifts vary slightly from chain to chain, but in all cases affect the shape of the barrel in the same way.

Both structures of SLG have been obtained at high concentration of ammonium sulphate; thus, several sulphate anions

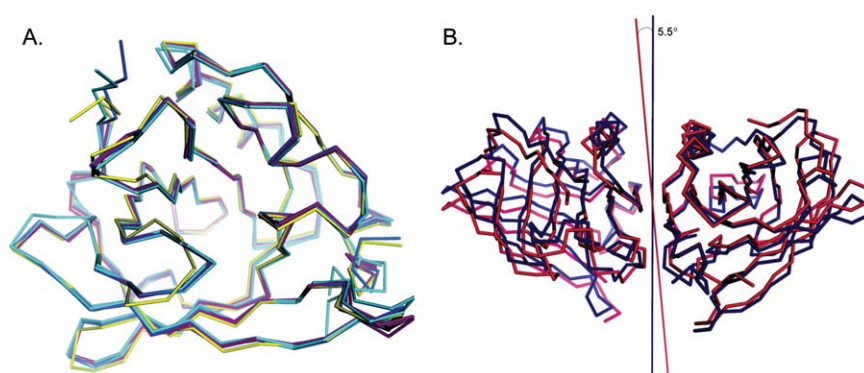


FIGURE 3 (A) C α superposition of BLG structure 3BLG (yellow), trigonal SLG-2 structure (magenta), and two chains of triclinic SLG-1 structure (light blue). (B) C α superposition of residues 146–150 (chain A) of trigonal SLG-1 (blue) and triclinic SLG-2 (magenta) dimers. The two-fold symmetry axis found in SLG-1 dimer is rotated about 5.5° in comparison to crystallographic axis found in SLG-2.

were found attached to the surface of molecule (Figure 2). Four sulphate ions were located in trigonal SLG-2 structure. SO4-1 located at the β -barrel entrance interacts with Lys69 on β -strand D and Leu87 on the EF loop stabilizing the closed conformation of this loop. The next, SO4-2, has been identified at the amino end of the α -helix near residues Asp129 and Asn130, where both the dipole moment of the helix and the availability of NH groups favor binding of negatively charged ion. Other sulphate ions, SO4-3 and SO4-4, have been found at protein surface interacting with residues of loops closing the β -barrel bottom. In the triclinic structure SLG-1, the sulphate ions have been identified at the barrel entrance in both subunits at position corresponding to SO4-1 in SLG-2 structure. Another sulphate ion has been found at a position equivalent to SO4-3 but only in subunit B. Also, and only in this subunit, additional sulphate ions have been located at protein surface in the region of the AB loop (SO4-5) and close to residues Thr49 and Asn53 (SO4-6).

Despite the different symmetry of the crystals and the different molecular packing, the superposition of the SLG-2 molecule and chain A of SLG-1 revealed that they have an almost identical structure with the RMS for C α atoms equal 0.363 Å for residues 5–157 (Figure 3A). The highest shifts of about 1 Å are observed for C-terminal residues that have been omitted in superposition due to their different positions in the molecules.

More subtle differences in the structure of SLG-1 subunits A and B are reflected by RMS. for C α superposition that is equal to about 0.6 Å. These differences comprise, among others, distortion of the β -barrel, conformation of flexible loops and disulphide bridges. The reorientation of the LG dimer in the triclinic lattice was reported first by Brownlow et al.¹⁸ They noticed that the orientation of the second subunit is altered by a rotation of about 5° with respect to that observed in the presence of a twofold axis. A similar effect is observed in the SLG-1 structure. The superposition of the SLG-1 dimer that is present in the asymmetric part of the unit cell and dimer of SLG-2 generated by crystallographic symmetry also showed an altered orientation of the subunits. The distortion also affected the geometry of the dimer interface but retained the hydrogen bond pattern for β -strand I. Analysis of the internal symmetry of the SLG-1 dimer using the Symmetry Detection (SymD) tool^{19,20} revealed that both subunits are related by noncrystallographic twofold symmetry. The orientation of this axis differs by 5.5° in comparison to the crystallographic axis present in trigonal structure SLG-2 (Figure 3B).

LG contains five cysteine residues that are conserved in bovine and sheep protein. They form two disulphide bridges: Cys106-Cys119, that connects the β -strand G and H, and Cys66-Cys160 that connects β -strand D and the C-terminal part of polypeptide chain (Figure 2). In both chains of SLG-1 structure, the Cys66-Cys160 bridge is partially disordered as a

result of the alternative orientations of the Gln159 peptide group. Interestingly, in the SLG-2 structure, this S—S bond seems to be partially reduced. The difference Fourier map indicates the presence of alternative positions of sulphate atoms being more than 4 Å apart. A similar reduction of the disulphide bridge is also observed for Cys106-Cys119. This unexpected result is probably an effect of the radiation damage during experiment on synchrotron.²¹

The Impact of Substitutions on Electrostatic Properties

Isoform A of ovine lactoglobulin differs from the most abundant isoform B of bovine protein only at six positions in the amino acid chain (Figure 1). It is interesting that, except for residue 53, all the other substitutions are located on the rim of the cleft surrounding the dimer interface (Figure 1). Two of those residues, Ile1 and Val162, are N- and C-terminal residues of the SLG chain replaced by leucine and isoleucine in the BLG molecule. In both cases, equivalent amino acids have the same chemical character and similar sizes, so substitution seems to have no effect on crystal structure and should not affect the physicochemical properties of the molecule.

More significant seem to be the differences at positions 53 and 130. They are occupied by Asp in BLG and Asn in SLG, which have similar sizes but different charges and might form different patterns of hydrogen bonds. The substitution of Asp in bovine protein by Asn in SLG at position 53 might be responsible for differences in molecular packing in the trigonal crystals of SLG and BLG. In SLG-2 structure, the packing is stabilized by strong intermolecular hydrogen bond between the Asn53 amide group and Ile2 carbonyl oxygen. Such interaction is impossible in BLG containing aspartate at position 53.

It should be expected that substitution of the Glu158 that is present in BLG by glycine in the sheep protein would influence the structure and properties of the molecule. However, in all crystal forms of bovine protein, Glu158 is exposed to solvent and is not involved in intermolecular interactions. Comparison of bovine and SLG crystal structures shows that substitution does not affect the structure of the protein around residue 158.

Another substitution site is located at the dimer interface. The polar residue Ser150 of bovine protein is replaced in ovine lactoglobulin by Ala. This substitution does not affect the formation of the LG dimer. This observation is consistent with recent results of studies on BLG.²² It has been observed that mutations introduced to the 145–153 region in general do not affect the formation of BLG dimers but only shift the dimerization equilibrium.

Among the described bovine to sheep substitutions, the most important are those that influence the electrostatic

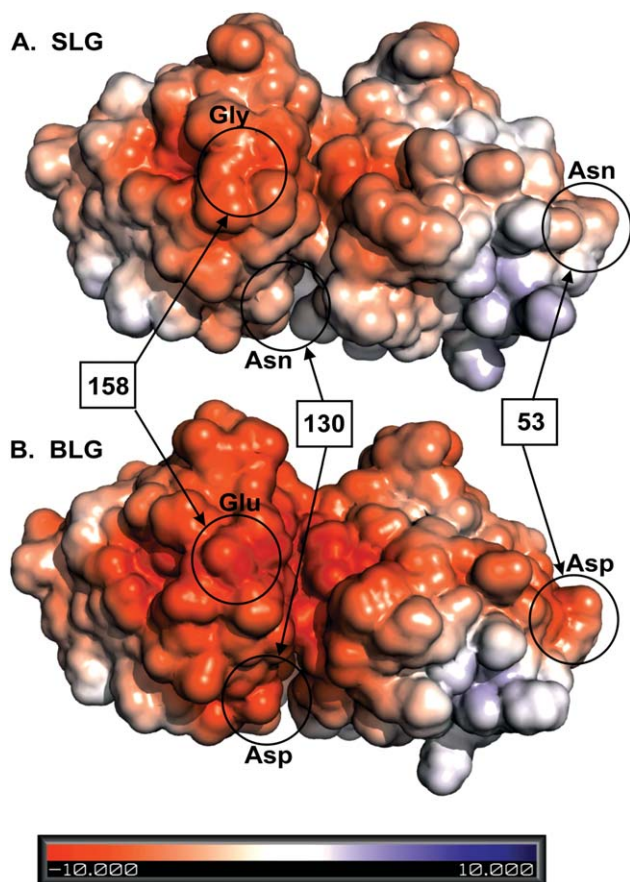


FIGURE 4 The distribution of electrostatic potentials on molecular surface of (A) SLG-2 dimer and (B) BLG dimer (PDB: 3BLG).

properties of the molecule: Asp53Asn, Asp130Asn, and Glu158Gly. The calculated total charge of the BLG dimer at neutral pH is -14.44 , while for SLG the charge is -7.93 . The substitution also affected the theoretical pI that is equal to 4.96 and 5.29 for bovine and sheep protein, respectively. The calculated pI for SLG-A is close to the experimental value 5.54 determined by isoelectric focusing.²³ The differences in sequence are also reflected in the distribution of electrostatic potentials on the molecular surface. As can be seen in Figure 4, for SLG the electrostatic potentials are significantly lower not only at the dimer interface but also at the entrance to the binding site. Different electrostatics around flexible loop EF might affect the pH of the Tanford transition and ligand binding properties. These differences are also reflected in the different conditions of chromatographic separation (see Materials and Methods) and crystallization of bovine and sheep protein.

Conformation of Flexible Loops

In BLG structures, special attention is focused on the EF loop due to its involvement in the Tanford transition that regulates

access to the binding site. The closed conformation of the EF loop observed at pH below 7.0 is maintained by the hydrogen bond between the Glu89 carboxyl group and the carbonyl oxygen of Ser116 and two hydrogen bonds between Asn90 and Ser110 (Figure 5). It has been postulated that an increase in pH to about 7.0–7.5 deprotonates the Glu89 carboxyl group. This results in a break of the hydrogen bond and induces conformational change of the EF loop. To date, the Tanford transition has been investigated only for BLG^{15–17} and there is no experimental evidence that this phenomenon also occurs in LG from other species. As the Glu89 is conserved, it is expected that this mechanism is also valid for SLG.

In contrast to the structure of the BLG isoform A (PDBid:1BSY) obtained at pH 7.1¹⁶ in the crystals of SLG-1 and SLG-2 that have been obtained at pH about 7.0, the EF loop in both structures is observed in closed conformation. The position of loop residues is similar to that observed in a trigonal 3BLG structure¹⁶ at pH 6.2 and the triclinic 1BEB structure of bovine isoform B at pH 6.5.¹⁸ The close conformation of the EF loop observed in sheep protein can be the result of two factors. The relatively high concentration of ammonium

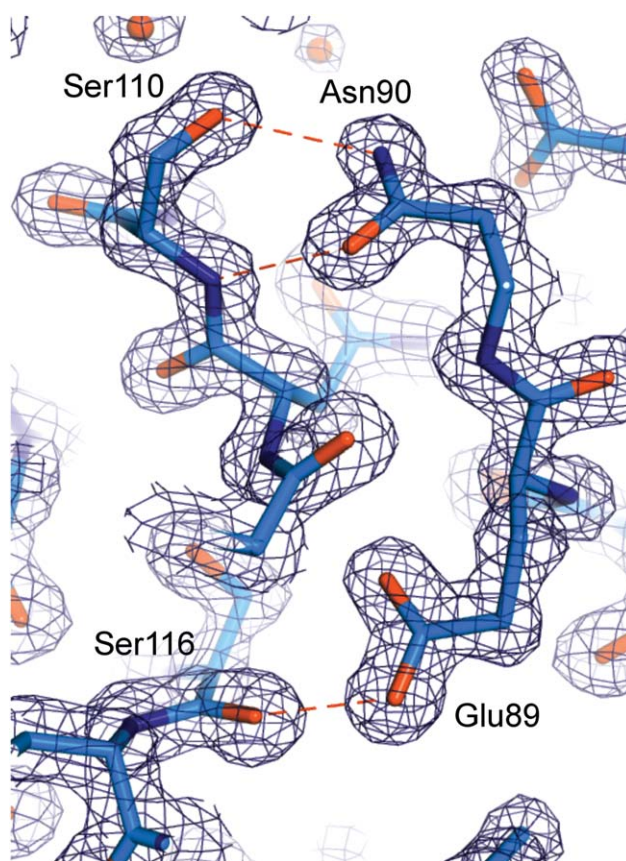


FIGURE 5 Interactions stabilizing closed conformation of EF loop in SLG-1 structure (chain A). Electron density map 2Fo–Fc is shown at 1.5 σ level.

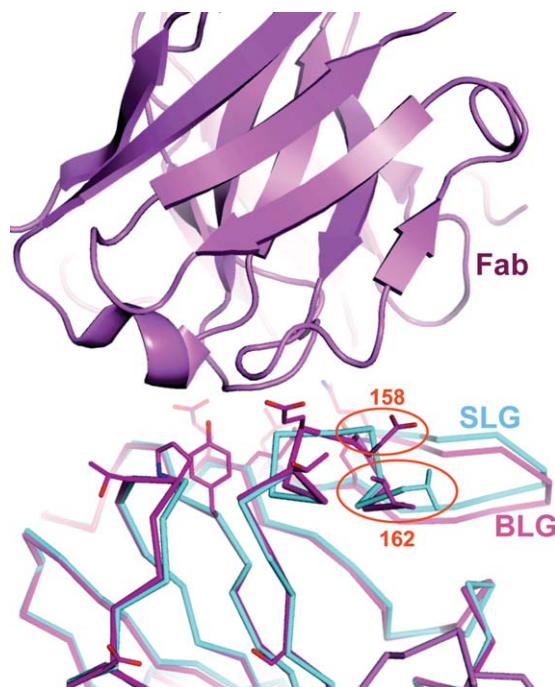


FIGURE 6 Superposition of BLG-IgE(Fab) complex (PDB: 2R56, magenta) and SLG-2 structure. Most of the amino acid substitutions are located outside the BLG epitope. The only exceptions are residues 158 and 162 located at the rim of the contact area.

sulphate lowers the final pH of crystallization drop by approximately 0.5 units. Another explanation is that the change in the global charge of the molecule, being the result of differences in the protein sequence, has shifted the Tanford transition for SLG to a pH higher than that for bovine protein. This hypothesis is supported by calculations of anomalous pK_a for Glu89, the residue that plays the major role in the Tanford transition. According to results provided by the *PROPKA* server, the calculated pK_a for Glu89 in sheep protein is 8.11, while in the bovine is 7.80 for both isoforms. As the *PROPKA* results are sensitive to small structural differences, for calculations we selected molecules from crystals possessing similar trigonal symmetry.

From a comparison of many structures of BLG, it can be concluded that the position of the EF loop also influences the conformation of the GH loop. When the EF loop is in the open position, the GH loop has an increased flexibility that often prevents reliable determination of residues positions. In both SLG structures, the electron density along the protein chain in the region of the EF and GH loops is very well defined. The structures show that the positions of residues in both loops are stabilized by extensive pattern of hydrogen bonds both within loops and between loops. These interactions are not retained in the structures of BLG with the EF loop in open conformation.

The Structural Basis of Sheep Protein Allergenicity

Currently, increasing attention is focused on milk from mammals such as horse²⁴ and camel²⁵ because these milks seem to be less allergenic for people, especially young children, than cow milk.²⁶ The main allergens in milk are caseins (*Bos d 8*) and whey proteins such as α -lactalbumin (*Bos d 4*) and LG (*Bos d 5*).²⁷

The structural basis of BLG allergenicity was studied by Niemi et al.²⁸ They found that dimeric BLG interacts with immunoglobulin E (IgE) Fab fragment. The lactoglobulin epitope involves residues located mainly at β -strands A, B, C, and D.²⁸ The BLG-IgE complex is formed not only by specific interactions (e.g., hydrogen bonds) but also by the three dimensional fit of both molecules. We have superimposed the SLG-2 dimer and BLG dimer in the complex of BLG and IgE (PDB ID: 2R56). This superposition revealed that most of the substitution sites in SLG are distant from the IgE with the exception of Gly158 and Val162 side chains (Figure 6). These two residues are located on the rim of the contact area and the

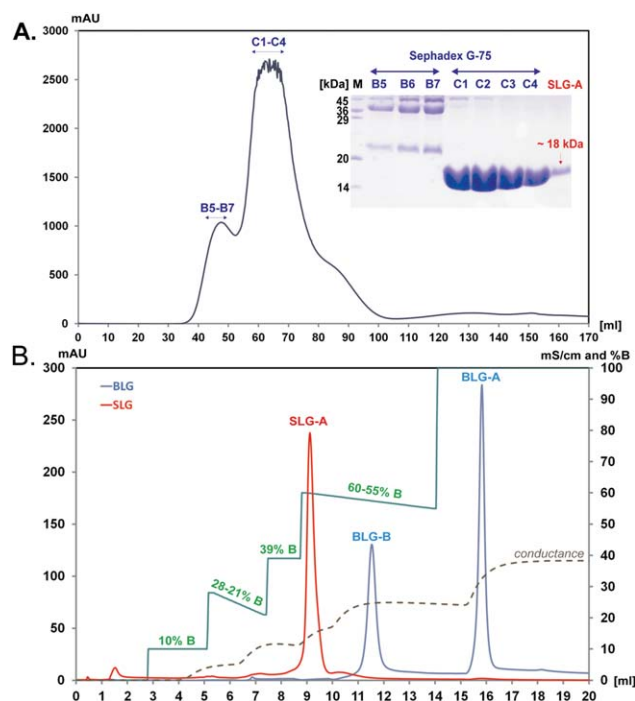


FIGURE 7 (A) Size exclusion chromatography performed on Sephadex G-75. SLG was eluted in a single large peak (fractions C1–C4). SDS-PAGE (15% acrylamide gel): lane 1: molecular weight marker, lanes 2–4: fractions B5–B7 (Sephadex G-75), lanes 5–8: fractions C1–C4 (Sephadex G-75), and lane 9: purified SLG-A eluted from MonoQ column (red peak on Figure 7B). (B) Anion-exchange chromatography on MonoQ 5/50 GL column. Superposition of chromatograms obtained from final purification step of SLG isoform A (SLG-A) and separation of mixture of BLG isoforms A (BLG-A) and B (BLG-B).

Table I Summary of Data Collection and Structure Refinement

	SLG-1	SLG-2
Data collection		
Temperature (K)	100	100
Space group	P1	P3 ₁ 21
Unit cell		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<i>a</i> = 36.18, <i>b</i> = 47.54, <i>c</i> = 50.24	<i>a</i> = <i>b</i> = 97.44, <i>c</i> = 68.24
α , β , γ (°)	α = 70.40, β = 76.56, γ = 72.10	α = β = 90 γ = 120
Resolution range (Å)	14.54–1.40 (1.48–1.40)	35.89–1.90 (2.00–1.89)
Total number of reflections	114,281 (11,292)	326,031 (47,135)
Total number of unique	56,652 (8053)	29,826 (4298)
Completeness (%)	96.6 (94.5)	100 (100)
<i>I</i> / σ <i>I</i>	13.6 (2.1)	22.2 (9.8)
<i>R</i> _{merge}	0.035 (0.401)	0.080 (0.272)
<i>R</i> _{rim} (<i>R</i> _{meas})	0.044 (0.562)	0.082 (0.285)
<i>R</i> _{pim}	0.026 (0.393)	0.025 (0.086)
Refinement		
No. of reflections (test set)	53,791 (2852)	28,117 (1480)
<i>R</i> (%)	18.78	19.44
<i>R</i> _{free} (%)	21.79	20.57
RMSD bonds (Å)	0.009	0.010
RMSD angles (°)	1.542	1.350
No. of water molecules	385	137
No. of sulphate ions	5	4
Average B-factor all atoms	12.84 (chain A) 13.71 (chain B)	31.79
Average B-factor main chain	11.69 (chain A) 12.64 (chain B)	29.13
Ramachandran statistics		
Allowed regions (%)	98.3	98.6
Disallowed regions (%)	1.70	1.40

substitution should not significantly affect interaction between LG and IgE. This result is in agreement with the observation that, due to structural similarities between SLG and BLG, there are often coexistences of cow and sheep milk allergies.²⁹ There are also reports that patients who are allergic to ovine milk tolerate cow milk well.^{30,31} Because other substitution sites are also located on the molecule surface, it may be possible that they are part of other epitopes responsible for different immunological responses to cow and sheep milk. However, the possibility cannot be excluded that proteins other than LG are responsible for differences in allergenicity.

METHODS

Isolation and Purification of Ovine Lactoglobulin

Ovine lactoglobulin was isolated from fresh sheep milk according to a modified procedure previously described for bovine protein.³² Briefly, in the first step milk was defatted by centrifugation at 5000 rpm at 4°C. The top fat layer was removed by filtration. The caseins were precipitated from the defatted milk via the addition of 1.0 M HCl drop-

wise until the pH was 4.6, that is, the isoelectric point of caseins. Precipitated caseins were removed by centrifugation at 5000 rpm and the supernatant (whey) was collected. The pH was adjusted to 2.0 by 6 M HCl and 7% NaCl (w/v) was added to precipitate the first fraction of proteins. Precipitated proteins were removed by centrifugation at 8000 rpm and the supernatant was used in the final salting-out step, in which 30% (w/v) NaCl was added to precipitate SLG. The precipitated SLG was collected by centrifugation at 8000 rpm. The supernatant was discarded, the precipitate was resuspended in ultrapure water, and dialyzed overnight against 0.1 M phosphate buffer (pH 7.1). The crude solution of SLG was loaded on a Sephadex G-75 column (XK 16/100; Pharmacia) equilibrated with 0.1 M phosphate buffer (pH 7.1). In this step, SLG was separated from traces of proteins that have higher molecular masses, such as caseins (Figure 7A). Fractions containing SLG were pooled, dialyzed against ultrapure water, and subjected to a final purification step. The final purification step was performed using ion-exchange chromatography on a MonoQ 5/50 Global column (GE Healthcare) using a stepwise sodium acetate gradient (buffer A: water, buffer B: 0.7 M CH₃COONa, pH 6.3) to separate traces of α -lactalbumin from SLG.³³ The SLG was eluted from the column in single peak (Figure 7B). The collected fractions containing SLG were pooled and dialyzed against ultrapure water for 48 h. The protein was lyophilized and stored at 4°C. The purity and homogeneity of SLG was checked using SDS-PAGE (Figure 7A).

Mass Spectrometry

A protein solution (25 pmol μL^{-1}) was prepared in 0.1% aqueous solution of trifluoroacetic acid (TFA), and 1,5-diaminonaphthalene (1,5-DAN) was used as a matrix. A saturated solution of 1,5-DAN in acetonitrile:0.1% TFA (1:1, v/v) was mixed with the sample solution at a ratio of 2:1, spotted onto a ground steel target plate and left to dry at ambient temperature. An ultrafleXtreme MALDI-TOF/TOF mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a smartbeam IITM laser operating at a 1 kHz repetition rate was used to acquire in-source decay spectra in positive reflectron mode. Ions were accelerated at 25 kV with a PIE time of 130 ns. Spectra were recorded and postprocessed with the use of a Compass for flexSeries 1.3 (Bruker). External calibration was performed with the use of ISD c-fragment ions of intact BSA. BioTools software (Bruker) was used to analyze the data by database searching using a Mascot 2.4 in-house server (Matrix Science, Boston, MA). The recorded ISD-spectrum was also matched with three possible sequences of SLG, including two mutations (see Supplementary Information). The best fit was obtained for isoform A of SLG.

Crystallization

SLG was crystallized using the vapor diffusion method in hanging drop setup. Different crystallization conditions were screened; however, crystals only appeared from ammonium sulphate solutions in drops formed by mixing 1 μL of protein (40 mg/ml in ultrapure water) with 2 μL of 2.1 M $(\text{NH}_4)_2\text{SO}_4$ in 0.05 M Tris-HCl (pH 7.0) equilibrated against 500 μL of well solutions. Typically, small (<0.2 mm) irregular, elongated crystals appeared from heavy precipitate within 2–4 weeks. Despite their small size and poor morphology, these crystals diffracted X-rays well (SLG-1, space group P1). In other drops formed by mixing 2 μL of protein (40 mg/ml in ultrapure water) and 2 μL of 2.2 M $(\text{NH}_4)_2\text{SO}_4$ in 0.05 M Tris-HCl pH 7.0 equilibrated against 500 μL of well solution, large, well-formed crystals (SLG-2, space group P3₁21) appeared after 2–7 days.

Data Collection, Structure Solution, Refinement, and Analysis

Diffraction data for SLG-1 were collected at 100 K on a SuperNova diffractometer (Agilent Technologies) with a CCD Atlas detector and a microfocus CuK α (1.54 Å) radiation source operating at 0.8 mA and 50 kV. Data were processed using *CrysAlis^{Pro}* (Agilent Technologies) and the *Scala*³⁴ from *CCP4* package.³⁵ The diffraction data (100 K) for SLG-2 were collected on a beamline I911-2 (1.00 Å) at a MAX-LAB synchrotron facility with a MAR 225 detector. Data were processed using *Mosflm*³⁶ and *Scala*³⁴. Prior to data collection, crystals were placed for few seconds in cryoprotectant containing 20% (v/v) glycerol in crystallization solution and immediately transferred to a gas nitrogen cryostream (100 K). Data collection statistics are summarized in Table I.

The structures of SLG-1 and SLG-2 were solved by molecular replacement using as a starting model 1QG5 and 3BLG structure, respectively. Molecular replacement was performed in *Phaser*.³⁷ Structures were refined using maximum likelihood refinement in *Refmac5*.³⁸ Electron density maps were investigated in *Coot*.³⁹ Refinement statistics are summarized in Table I. Structures were deposited in PDB as entries 4NLI and 4NLJ.

Calculations of electrostatic potentials were performed using a PDB2PQR server.⁴⁰ The pK_a values and pI were determined using

PROPKA.⁴¹ The surface potentials were determined by adaptive Poisson–Boltzmann solver (APBS) using an on-line server.⁴² All calculations were performed for pH 7.0 and analyzed using an ABPS plug-in to PyMOL.⁴³ The electrostatic properties were calculated for the dimer of sheep protein obtained from the trigonal structure of SLG-2. Calculations for isoform B of the bovine protein were made using a dimer generated from the trigonal structure of isoform A (PDB ID: 3BLG), in which residues Asp64 and Val118 were replaced by Gly64 and Ala118. This structure was selected because it possesses a high similarity of EF loop in closed conformation to the SLG-2 structure.

DDM plots were produced using the DDMP program from the center for structural biology at Yale University, New Haven, CT. The internal symmetry of the LG dimer was analyzed with the SymD program²⁰ from the Galaxy server.⁴⁴

ACKNOWLEDGMENTS

The authors would like to thank Dr. Marta Gawin for performing the mass spectrometry experiments.

REFERENCES

- Kontopidis, G.; Holt, C.; Sawyer, L. *J Dairy Sci* 2004, 87, 785–796.
- Loch, J.; Polit, A.; Górecki, A.; Bonarek, P.; Kurpiewska, K.; Dziedzicka-Wasylewska, M.; Lewiński, K. *J Mol Recognit* 2011, 24, 341–349.
- Loch, J. I.; Polit, A.; Bonarek, P.; Olszewska, D.; Kurpiewska, K.; Dziedzicka-Wasylewska, M.; Lewiński, K. *Int J Biol Macromol* 2012, 50, 1095–1102.
- Loch, J. I.; Bonarek, P.; Polit, A.; Riès, D.; Dziedzicka-Wasylewska, M.; Lewiński, K. *Int J Biol Macromol* 2013, 57, 226–231.
- Loch, J. I.; Bonarek, P.; Polit, A.; Świątek, Ś.; Dziedzicka-Wasylewska, M.; Lewiński, K. *J Mol Recognit* 2013, 8, 357–367.
- Wu, S.Y.; Pérez, M. D.; Puyol, P.; Sawyer, L. *J Biol Chem* 1999, 274, 170–174.
- Gutiérrez-Magdaleno, G.; Bello, M.; Portillo-Téllez, M. C.; Rodríguez-Romero, A.; García-Hernández, E. *J Mol Recognit* 2013, 2, 67–75.
- Oksanen, E.; Jaakola, V.-P.; Tolonen, T.; Valkonen, K.; Akerstrom, B.; Kalkkinen, N.; Virtanen, V.; Goldman, A. *Acta Cryst* 2006, D62, 1369–1374.
- Hoedemaeker, F. J.; Visschers, R. W.; Alting, A. C.; de Kruif, K. G.; Kuil, M. E.; Abrahams, J. P. *Acta Cryst* 2002, D58, 480–486.
- Monaci, L.; Tregoat, V.; van Hengel, A. J.; Anklam, E. *Eur Food Res Technol* 2006, 223, 149–179.
- Park, Y. W.; Juárez, M.; Ramosc, M.; Haenlein, G. F. W. *Small Ruminant Res* 2007, 68, 88–113.
- Maubois, J. L.; Pion, R.; Ribadeau-Dumas, B.; *Biochim Biophys Acta* 1965, 107, 501–510.
- Bell, K.; McKenzie, H. A. *Biochim Biophys Acta* 1967, 147, 123–134.
- Sawyer, L.; Kontopidis, G. *Biochim Biophys Acta* 2000, 1482, 136–148.
- Tanford, C.; Bunville, L. G.; Nozaki, Y. *J Am Chem Soc* 1959, 81, 4032–4036.

16. Qin, B. Y.; Bewley, M. C.; Creamer, L. K.; Baker, H. M.; Baker, E. N.; Jameson, G. B. *Biochemistry* 1998, 37, 14014–14023.
17. Sakurai, K.; Goto, Y. *J Mol Biol* 2006, 356, 483–496.
18. Brownlow, S.; Morais Cabral, J. H.; Cooper, R.; Flower, D. R.; Yewdall, S. J.; Polikarpov, I.; North, A. C.; Sawyer, L. *Structure* 1997, 5, 481–495.
19. Olsson, M. H. M.; Søndergard, C. R.; Rostkowski, M.; Jensen J. H.; *J Chem Theory Comput* 2011, 7, 525–537.
20. Kim, C.; Basner, J.; Lee, B. *BMC Bioinformatics* 2010, 11, 303.
21. Weik, M.; Ravelli, R. B.; Kryger, G.; McSweeney, S.; Raves, M. L.; Harel, M.; Gros, P.; Silman, I.; Kroon, J.; Sussman, J. L. *Proc Natl Acad Sci USA* 2000, 97, 623–628.
22. Ohtomo, H.; Fujiwara, K.; Ikeguchi, M. *J Biochem* 2012, 151, 329–334.
23. Moatsou, G.; Hatzinaki, A.; Samolada, M.; Anifantakis, E. *Int Dairy J* 2005, 15, 123–131.
24. Uniacke-Lowe, T.; Huppertz, T.; Fox, P. F. *Int Dairy J* 2010, 20, 609–629.
25. Ehlayel M. S.; Hazeima, K. A.; Al-Mesaifri, F.; Bener, A. *Allergy Asthma Proc* 2011, 32, 255–258.
26. Cases, B.; García-Ara, C.; Boyano, M. T.; Pérez-Gordo, M.; Pedrosa, M.; Vivanco, F.; Quirce, S.; Pastor-Vargas, C. *J Investig Allergol Clin Immunol* 2011, 21, 398–400.
27. Restani, P.; Beretta, B.; Fiocchi, A.; Ballabio, C.; Galli C. L. *Ann Allergy Asthma Immunol* 2002, 89, 11–15.
28. Niemi, M.; Jänis, J.; Jylhä, S.; Kallio, J. M.; Hakulinen, N.; Laukkanen, M. L.; Takkinen, K.; Rouvinen, J. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2008, 64, 25–28.
29. Bernard, H.; Ah-Leung, S.; Tilleul, S.; Drumare, M. F.; Paty, E.; Bidat, E.; Wal, J. M.; Hazebrouck, S. *Mol Nutr Food Res* 2012, 10, 1532–1540.
30. Bonaguro, R.; Lazzarotto, F.; Toniolo, A.; Zorzin, C.; Muraro, M. A. *Clin Transl Allergy* 2013, 3, P157.
31. Viñas, M.; Carnés, J.; López-Matas, M. A.; Hernández, N.; Castillo, M. J.; Ibero, M. *Allergol Immunopathol (Madr)*, in press.
32. Blanc, F.; Bernard, H.; Alessandri, S.; Bublin, M.; Paty, E.; Leung, S. A.; Patient, K. A.; Wal, J. M. *Mol Nutr Food Res* 2008, 52, S166–S175.
33. Neyestani, T. R.; Djalali, M.; Pezeshki, M. *Protein Expr Purif* 2003, 29, 202–208.
34. Evans, P. *Acta Cryst* 2006, D62, 72–82.
35. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Cryst* 2011, D67, 235–242.
36. Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W. *Acta Cryst* 2011, D67, 271–281.
37. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* 2007, 40, 658–674.
38. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Cryst* 2011, D67, 355–367.
39. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Cryst* 2010, D66, 486–501.
40. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res* 2007, 35, W522–W525.
41. Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* 2005, 61, 704–721.
42. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc Natl Acad Sci USA* 2001, 98, 10037–10041.
43. DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific, San Carlos, CA, 2002. Available at: <http://www.pymol.org>.
44. Goecks, J.; Nekrutenko, A.; Taylor, J. *Genome Biol* 2010, 11, R86.

Reviewing Editor: David E. Wemmer

wwPDB Validation Report

PDB ID: 4NLI
RCSB ID: RCSB083353
TITLE: Crystal structure of sheep beta-lactoglobulin (space group P3121)
AUTHOR(S): J.I.Loch, M.Molenda, M.Kopec, S.Swiatek, K.Lewinski

Summary of PDB Entry Validation

The results of the validation of this PDB entry are shown below. No major issues were raised during data processing.

Geometry Validation

1. Atomic Clashes

No issues found.

2. Peptide Linkage

No issues found.

3. Covalent Geometry

No issues found.

4. Chirality Error

No issues found.

5. Phi/Psi Torsion Angles

No issues found.

Individual Residue Outliers on Real Space R-value

No issues found.

Sequence Validation

The reported biological sequence shows no discrepancy with UniProt sequence (code P67976).

The reported biological sequence and the sequence given in the coordinates show no discrepancy.

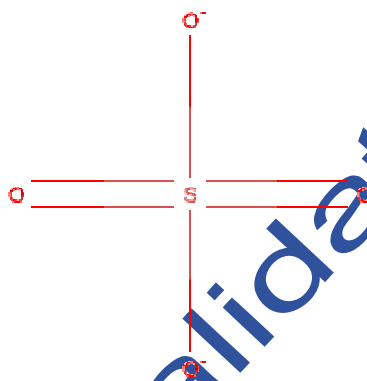
Biological Assembly

The biological assembly predicted by PISA is a monomer while author provided biological assembly is a dimer.

Ligand Chemistry

Ligand chemistry has been checked against the Chemical Component Dictionary. The following is a summary.

Identifier: SO4
Name: SULFATE ION
Formula: O4 S



Type	Program	Version	Descriptor
SMILES	ACDLabs	10.04	[O-]S([O-])(=O)=O
SMILES CANONICAL	CACTVS	3.341	[O-][S]([O-])(=O)=O
SMILES	CACTVS	3.341	[O-][S]([O-])(=O)=O
SMILES CANONICAL	OpenEye OEToolkits	1.5.0	[O-]S(=O)(=O)[O-]
SMILES	OpenEye OEToolkits	1.5.0	[O-]S(=O)(=O)[O-]
InChI	InChI	1.03	InChI=1S/H2O4S/c1-5(2,3)4/h(H2,1,2,3,4)/p-2
InChIKey	InChI	1.03	QAOWNCQODCNURD-UHFFFAOYSA-L

Summary of Structure Factor Validation

Structure quality	
Average Real space R-factor (Calculated by SFCHECK, V7.02.4)	0.0735
Average Real-space correlation coefficient (Calculated by SFCHECK, V7.02.4)	0.9850
Average Occupancy-weighted avg temperature factor	33.217

Resolution	
High Resolution (Author reported)	1.90
High Resolution (Calculated by SFCHECK, V7.02.4)	1.90
High Resolution (Calculated by REFMAC, V5.7.0029)	1.900
Low Resolution (Author reported)	35.89
Low Resolution (Calculated by SFCHECK, V7.02.4)	35.89
Low Resolution (Calculated by REFMAC, V5.7.0029)	35.887

Crystal data	
Space group	P 31 2 1
Total number of reflections	28117
Number of reflections used	26637
Completeness of data	99.3

R-factors	
R-factor (Author reported)	0.194
R-factor (Calculated by SFCHECK, V7.02.4)	0.219
R-factor (Calculated by REFMAC, V5.7.0029)	0.2026
Free R-factor (Author reported)	0.206
Free R-factor (Calculated by SFCHECK, V7.02.4)	0.234
Free R-factor (Calculated by REFMAC, V5.7.0029)	0.2145

Wilson statistics (PHENIX, V1.8-1069)	
Wilson B-factor	28.93

Wilson Scale	0.13
--------------	------

Padilla-Yeates statistics for twin detection (PHENIX, V1.8-1069)	
Padilla-Yeates < L >	0.469
Padilla-Yeates <L*L>	0.299

wwPDB Validation Report

PDB ID: 4NLJ
RCSB ID: RCSB083354
TITLE: Crystal structure of sheep beta-lactoglobulin (space group P1)
AUTHOR(S): J.I.Loch, M.Molenda, M.Kopeck, S.Swiadek, K.Lewinski

Summary of PDB Entry Validation

The results of the validation of this PDB entry are shown below. No major issues were raised during data processing.

Geometry Validation

1. Atomic Clashes

No issues found.

2. Peptide Linkage

No issues found.

3. Covalent Geometry

No issues found.

4. Chirality Error

No issues found.

5. Phi/Psi Torsion Angles

No issues found.

Individual Residue Outliers on Real Space R-value

No issues found.

Sequence Validation

The reported biological sequence shows no discrepancy with UniProt sequence (code P67976).

The reported biological sequence and the sequence given in the coordinates show no discrepancy.

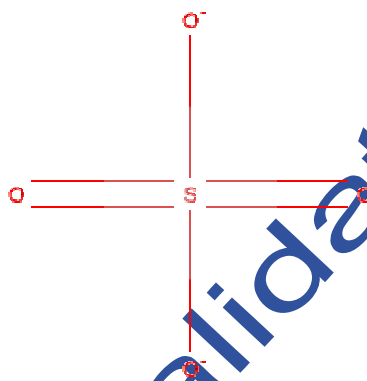
Biological Assembly

The biological assembly predicted by PISA is a dimer. This agrees with author's annotation.

Ligand Chemistry

Ligand chemistry has been checked against the Chemical Component Dictionary. The following is a summary.

Identifier: SO4
Name: SULFATE ION
Formula: O4 S



Type	Program	Version	Descriptor
SMILES	ACDLabs	10.04	<chem>[O-]S([O-])(=O)=O</chem>
SMILES CANONICAL	CACTVS	3.341	<chem>[O-][S]([O-])(=O)=O</chem>
SMILES	CACTVS	3.341	<chem>[O-][S]([O-])(=O)=O</chem>
SMILES CANONICAL	OpenEye OEToolkits	1.5.0	<chem>[O-]S(=O)(=O)[O-]</chem>
SMILES	OpenEye OEToolkits	1.5.0	<chem>[O-]S(=O)(=O)[O-]</chem>
InChI	InChI	1.03	<chem>InChI=1S/H2O4S/c1-5(2,3)4/h(H2,1,2,3,4)/p-2</chem>
InChIKey	InChI	1.03	QAOWNCQODCNURD-UHFFFAOYSA-L

Summary of Structure Factor Validation

Structure quality	
Average Real space R-factor (Calculated by SFCHECK, V7.02.4)	0.0586
Average Real-space correlation coefficient (Calculated by SFCHECK, V7.02.4)	0.9912
Average Occupancy-weighted avg temperature factor	14.827

Resolution	
High Resolution (Author reported)	1.40
High Resolution (Calculated by SFCHECK, V7.02.4)	1.40
High Resolution (Calculated by REFMAC, V5.7.0029)	1.400
Low Resolution (Author reported)	14.54
Low Resolution (Calculated by SFCHECK, V7.02.4)	14.54
Low Resolution (Calculated by REFMAC, V5.7.0029)	14.542

Crystal data	
Space group	P 1
Total number of reflections	53791
Number of reflections used	50939
Completeness of data	96.9

R-factors	
R-factor (Author reported)	0.188
R-factor (Calculated by SFCHECK, V7.02.4)	0.194
R-factor (Calculated by REFMAC, V5.7.0029)	0.1876
Free R-factor (Author reported)	0.218
Free R-factor (Calculated by SFCHECK, V7.02.4)	0.228
Free R-factor (Calculated by REFMAC, V5.7.0029)	0.2176

Wilson statistics (PHENIX, V1.8-1069)	
Wilson B-factor	10.56

Wilson Scale	0.51
--------------	------

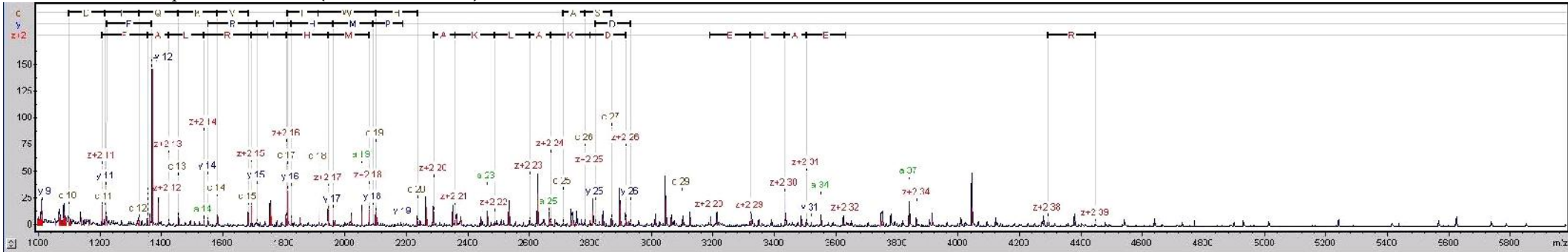
Padilla-Yeates statistics for twin detection (PHENIX, V1.8-1069)	
Padilla-Yeates < L >	0.500
Padilla-Yeates <L*L>	0.333

Supplementary Data

Mass spectrometry results. Fit of three SLG sequences (isoform A, B and C) to experimental data. The best score was obtained for isoform A.

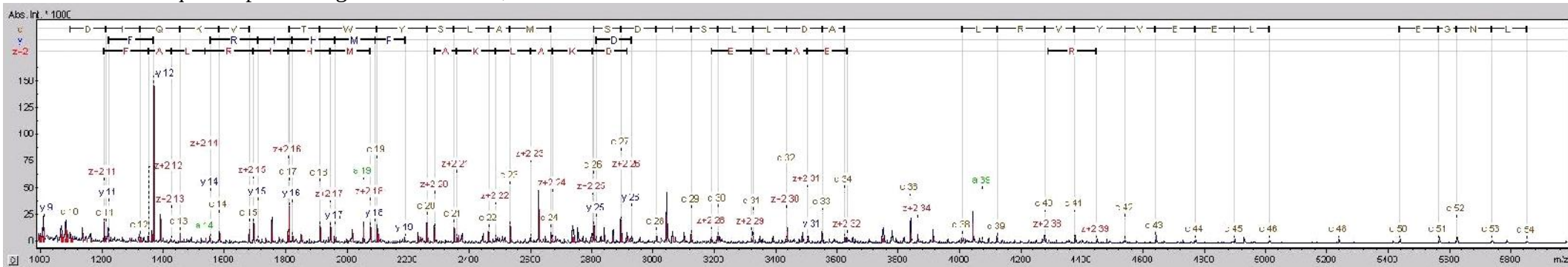
1. Fit of SLG-B sequence P67976 (LACB_SHEEP);

isoform B: score 576



2. Fit of SLG-A sequence possessing mutation H20Y;

isoform A: score 1860



3. Fit of SLG-C sequence possessing mutations H20Y and R148Q;

isoform C: score 112

