The binding of small volatile molecules by bovine β-lactoglobulin

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Chemistry at Massey University

Yu-Ting Hsu

2008

Abstract

Bovine β-lactoglobulin (β-Lg) has been studied extensively but there is no clear identification of its biological function. Hydrophobic molecules have been observed binding into the hydrophobic calyx of β-Lg. By comparison with other members of lipocalin family, it is probable that β-Lg plays a role of transport of ligands, as ligands also bind into the central cavity of lipocalins. The structurally similar MUP is a pheromone-binding protein; therefore, it is possible that β-Lg may also fulfil a similar role. This study has begun to test this hypothesis by investigating the interactions between bovine β-Lg and several small volatile molecules (2-*sec*-4,5-dihydrothiazole, 3-methyl-2-butenal, 3-methyl-2-buten-1-ol and phenylacetic acid). The interactions between the volatile molecules and β-Lg were studied by both two-dimensional NMR spectroscopy and X-ray crystallographic methods. TOCSY spectra were recorded for β-Lg and the complex between β-Lg and the ligands. The observed chemical shifts in the H_N-H_α region are sensitive to the proximity of ligands, and hence chemical shift changes on ligand binding provide information on possible binding sites. It appears that several amino acids with hydrophobic sidechains are affected by interaction with volatile molecules at pH 2.0. The X-ray crystallographic study at pH 8.5 showed that the potential ligand, 2-*sec*-4,5-dihydrothiazole, may have decomposed into a linear 2-methyl-butanol. The refined structure (R=0.281, R_{free}=0.354 for reflections to 2.6 Å resolution) reveals that the potential ligand may bind to the central cavity in a manner similar to the binding of 12-bromodecanoic acid to β-Lg.

Acknowledgements

Thanks to:

My supervisors, Professor Geoff Jameson and Dr. Patrick Edwards for giving me the opportunity to perform this research and guiding me through this project, Dr. Gill Norris for giving assistance on circular dichroism experiments and Mr. Simon Oakley for advice on data X-ray data processing and structural refinement.

Abbreviations

BBP	Bilin-binding protein
δ	Chemical shift
CD	Circular dichroism
CCP4(i)	Collaborative Computational Project Numer 4 (interface)
COSY	Correlation spectroscopy
COOT	Crystallographic Object Orientated Toolkit
kD	KiloDalton
GUI	Graphical User Interface
HSQC	Heteronuclear multiple quantum coherence
Hz	Hertz
β-Lg	β-Lactoglobulin
α-LA	α-Lactalbumin
MUP	Mouse major urinary protein
3MB	3-Methyl-2-butenal
3MBOH	3-Methyl-2-buten-1-ol
mmol	Millimole
min(s)	Minute(s)
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Enhancement
OBP	Odorant-binding protein
PAA	Phenylacetic acid
RBP	Retinol-binding protein
J	Scalar coupling constant

SDS	Sodium dodecyl sulfate
SBT	2-*sec*-Butyl-4,5-dihydrothiazole
TOSCY	Total correlation spectroscopy

Table of contents

List of tables

List of figures

Chapter 1 Introduction

1.1 β-Lactoglobulin and the lipocalins

β-Lactoglobulin (β-Lg) is the most abundant protein present in whey milk [1] with a concentration of 0.3 g/100 mL. [2] It is found in milk of most of the mammalian species but is absent in the milk of rodents and humans. β-Lg is a globular protein, consisting of 162 amino acids. This study has utilized the bovine form of β-Lg for which there are three genetic variants A, B and C commonly occurring in bovine milk. Bovine β-Lg exists as a dimer at pH > 6, but becomes monomeric at pH < 4. [3]

β-Lg was first isolated from bovine milk in 1934, [4] and since then it has been studied extensively using almost every conceivable spectroscopic and analytical tool in order to probe the molecular structure and biological properties and functions. Up to 1975, β-Lg was one of the most thoroughly researched proteins because of the abundance and easy preparation of the protein, [5] and β-Lg was considered to be typical of most globular proteins. β-Lg belongs to the lipocalin family, whose members typically have a length of 150-170 amino acids and a molecular weight of around 20 kDa. This large and diverse family of proteins has various functions, which mostly involve some form of ligand-binding function. The structures of several lipocalins have been determined by X-ray diffraction techniques, including plasma retinol-binding protein (RBP, Fig. 1.1), [6, 7] bilin-binding protein (BBP), [8] odorant-binding protein (OBP), [9] mouse major urinary protein (MUP, Fig. 1.2) [10] and β-Lg (Fig. 1.3). [11-13]

As exemplified in figures 1.1, 1.2 and 1.3, all proteins in the lipocalin family share a very similar three-dimensional structure, characterized by a common fold made up of the calyx-(cup-)like central 8-stranded β-barrel and one three-turn α-helix. The internal cavity or pocket inside the calyx, which is lined with hydrophobic residues, accommodates a range of hydrophobic molecules.

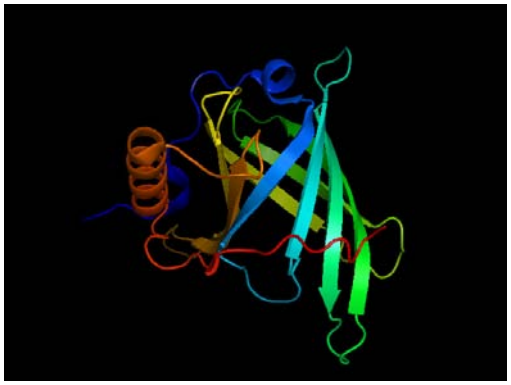


Fig. 1.1 Structure of plasma retinol-binding protein (RBP, pdb: 1aqb).

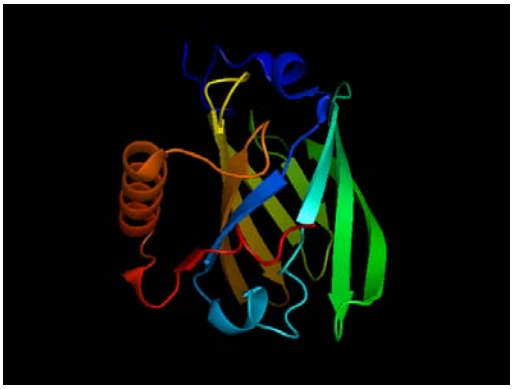


Fig. 1.2 Structure of major mouse urinary protein (MUP, pdb: 1i06).

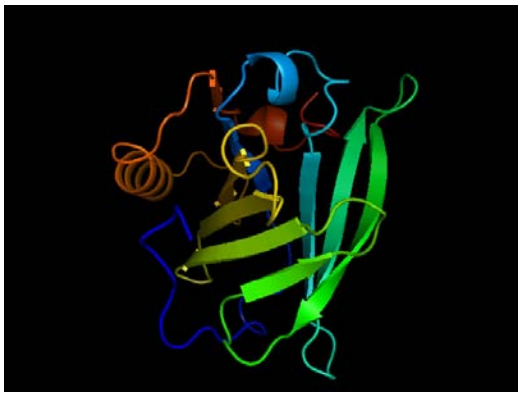


Fig. 1.3 Structure of β-Lg (pdb: 1bso).

1.1.1 Structure and properties of β-Lg

β-Lg (Fig. 1.4) is a relatively small protein and is soluble in dilute salt solutions. As a whey protein it is also soluble in acidic solutions. It exists as a dimer at neutral pH, but dissociates into monomers at pH < 4. [14] The 162 amino acid residues fold up into an 8-stranded β-barrel, which creates a hydrophobic pocket that is commonly referred to as the calyx. A ninth strand, the I strand, forms a significant part of the dimer interface in the bovine protein.

The β-strands A to D form one β-sheet and strands E to H form the second sheet that together fall into the conical-shaped calyx. Strand A has a right-angled kink and forms also an antiparallel strand with strand H and with strand I. Strands D and E, connecting the two sheets, do not exploit their full potential to form hydrogen bonds, hence form a less significant interaction, which nevertheless closes the calyx.

The structure is stabilized by two disulfide bonds, which are located between residues 66 and 160 and residues 106 and 119; β-Lg also has one free thiol (Fig. 1.4, Cys 121). The calyx is able to provide binding sites for ligands. The calyx is involved in the binding of a wide range of hydrophobic molecules including 12-bromodecanoic acid, which has been determined by using X-ray crystallographic techniques. [15]

The 3-turn α-helix is between strands H and I and is located on the outer surface of the calyx. The BC, DE and FG loops connect the β-strands at the closed end of the calyx. These loops are quite short compared to the loops, AB, CD, EF, and GH that are located at the open end of the calyx and are flexible.

The conformation of these surface loops is sensitive to pH. In particular, the EF loop acts as a gate over the binding site inside the calyx. [16] At basic pH (pH greater than 7.3), the EF loop is in its open conformation (articulated away from the entrance to the calyx), which allows the ligand access, whereas at low pH (pH less than 7.3) the EF loop is in its closed conformation, which then inhibits ligand binding within the cavity. Therefore, the conformation of the EF loop is crucial to β-Lg binding to hydrophobic molecules. [17] However, the dynamic properties of EF loop at low pH are not fully understood, it is possible that at pH 2, β-Lg is undergoing conformational changes between open and closed conformations. [18] Hence ligand binding to β-Lg at low pH may be possible.

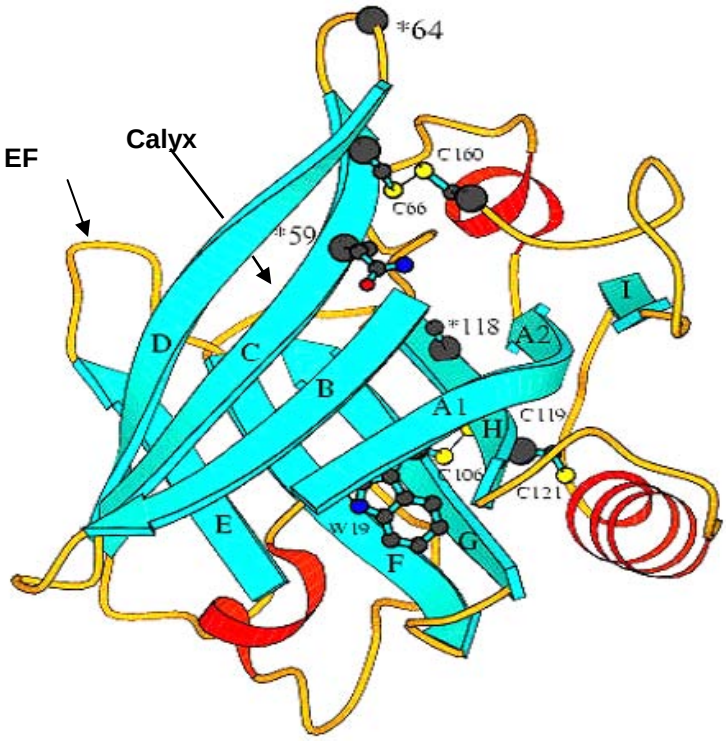


Fig. 1.4 The structure of β-Lg showing location of disulfide bridges (Jameson). C121 is the free Cys121, and C66---C160 and C106---C119 represent the disulfide bridges. The conformation of the highly flexible EF loop is important to ligand binding within the calyx. Residues 64 and 118 are the amino acid substitution sites for the A and B variants.

The three commonly occurring variants of β-Lg in bovine milk are variant A, B and C, which were denoted for convenience by Bell [19] according to the order of appearance on the starch gel electrophoresis from the positive electrode to the negative electrode. The structural difference between β-Lg A, B and C can be distinguished by the variations at positions 59, 64 and 118 in the protein sequence (Table 1.1).

Variant	59	64	118
A	Gln	Asp	Val
B	Gln	Gly	Ala
C	His	Gly	Ala

Table 1.1 Amino acid substitutions for β-Lg A, B and C. [20]

Amino acid 59 is located on the strand C and amino acid 64 is located just before the highly flexible CD loop. It is believed that the conformational changes that are possibly caused by this variation could be masked since the CD loop is highly flexible. The variation on amino acid 118 does not lead to any structural changes at the surface of β-Lg because it is located on strand H and also inside the calyx.

1.1.2 Ligand binding to β-Lg

No definitive biological function has been ascribed to β-Lg, although several proposals have been made due to the structural similarity of β-Lg and retinol-binding protein. It was thought β-Lg might be involved in retinol transport from mother to neonate. [21] Reports on ligand-binding properties of β-Lg can be dated back to 1948 when SDS (Sodium Dodecyl Sulfate) was identified as bound to β-Lg by McMeekin et al. [22] There is evidence that bovine β-Lg binds to a wide range of ligands including fatty acids, [15, 23-27] retinol and retinol derivatives, [28, 29] porphyrin species, [30, 31] and assorted aromatic [32-34] and alkanone species. [35, 36]

β-Lg binds to retinol tightly with one retinol per monomer, and many studies of ligand binding show that the primary binding site is in the hydrophobic pocket or calyx. Some studies suggest the presence of secondary binding sites, especially for smaller hydrophobic molecules (less than ~10 carbon atoms). A possible binding site on the outer surface of β-Lg has been reported for both crystallographic [37] and solution studies. [27, 38]

The retinol-β-Lg complex was then found during the study of retinol-binding protein (RBP) in 1972, [39] and these studies showed that β-Lg appears to bind retinol with an affinity similar to RBP. The interactions of β-Lg and hydrophobic ligands have been determined using various methods, including equilibrium dialysis, affinity chromatography, electron-spin resonance spectroscopy, and the most popular method, the intrinsic Trp fluorescence. From all these studies, β-Lg seems to prefer binding to hydrophobic molecules such as fatty acids (see Fig. 1.8 in next section), [40] cholesterol (Fig. 1.5) [41a] and vitamin D_2 (Fig.

1.6). [41b]

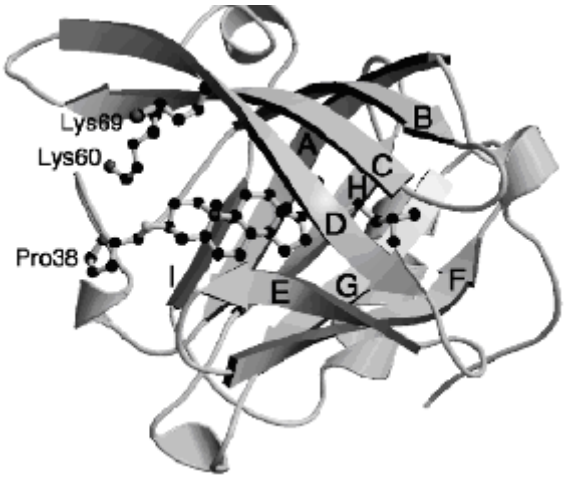


Fig. 1.5 β-Lg with cholesterol complexed within the calyx. [16]

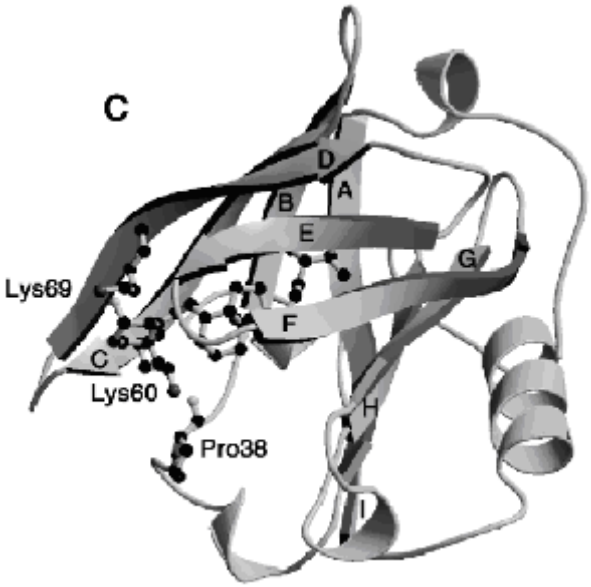


Fig. 1.6 β-Lg with vitamin D_2 complexed within the calyx. [16]

Most of the lipocalins are able to bind to small hydrophobic molecules within the central calyx. Despite the true function of β-Lg being unknown, it has been suggested that β-Lg may play a role in the transportation of retinol or fatty acids. However, it seems that not all β-Lgs bind to fatty acids, at least at pH ~7. For example, porcine β-Lg binds fatty acids only at pH > 8.6. [42]

There is another abundant whey protein present in milk, α-lactalbumin (α-LA). It is a modulator that turns β-galactosyl transferase into lactose synthetase. [43] It is also believed that β-Lg might bind to a cell-surface receptor in such a way as to enhance the receptor's interaction with retinol. It has been postulated that a function similar to α-LA for some metabolic process might be the physiological reason for the significant quantities of β-Lg found in bovine milk. [2] Other studies of β-Lg have led to suggestions that β-Lg may be involved in the transportation of retinol or may be involved in metabolism of lipids [31] where it has also been proposed that β-Lg plays a role in stimulation of pre-gastric lipase. [40]

At pH < 4, β-Lg is a monomer and available to study by NMR methods (the dimer gives poorly resolved spectra). However, ligand-binding and ligand-binding sites, kinetics and thermodynamics may be strongly perturbed by the closed conformation of the EF loop at pH < 7.3 and protonation of external carboxylate residues.

X-ray crystallographic methods are used to determine the structures of ligand-β-Lg complexes. X-ray structures have been characterized only in the pH range of 6.5 to 8.5. At these pH, it is clear that the main binding site, inside the calyx, is able to accommodate both linear molecules such as fatty acids and cyclic species such as the cyclohexenyl ring of retinol.

1.1.3 Mouse major urinary protein – possible implications for bovine β-Lg function

The major urinary protein (MUP) is an important component of the urine from the sexually mature male mouse. It is known that the urine of male mice contains high concentrations of protein [44] and much of the protein has been identified as MUP1. This protein is expressed in the liver under hormonal control and excreted in the urine. [45]

The major mouse urinary proteins are pheromone-binding proteins, which play a role as effectors of mouse physiology and behavior. [46] MUP and β-Lg are structurally similar, and the molecular weight of both proteins is also similar. MUP's molecular weight is about 19kD and that of β-Lg is about 18.2 kD. Like β-Lg, MUPs are lipocalins. [47] It has been proven from crystallographic studies that MUPI binds to SBT [47] and β-Lg binds to a range of molecules, including cholesterol, vitamin D_2, [16] and 12-bromodecanoic acid. [15] The small molecules all bind within the calyx.

MUPs consist of β-barrels shaped into a hydrophobic cavity, which allows ligand-binding interactions. MUPs have eight-stranded β-barrels like β-Lg. It is proven that MUPs bind to male pheromones that play a role in chemical communication. The ligands include 2-*sec*-butyl-4,5-dihydrothiazole (Fig. 1.7) and 3,4-dehydro-*exo*-brevicomin. Both 2-*sec*-butyl-4,5-dihydrothiazole and 3,4-dehydro-*exo*-brevicomin were isolated in 1984 by Novotny and co-workers from the urine of the male mouse. [48] It is believed that the two volatile compounds are the components of the pheromone. It was proposed that MUPs should be capable of binding to other pheromones such as the most

abundant volatile substance, 6-hydroxy-6-methyl-3-heptanone, which induces puberty acceleration in female mice. Therefore, the function of MUPs has been determined as a carrier of small hydrophobic pheromones through an aqueous environment, such that it protects the pheromones from decomposition. MUPs control the pheromone excretion as well as the release of the volatile pheromones from urine. [49] The ligand-bound MUP denatures and releases the volatile ligand when the urine evaporates.

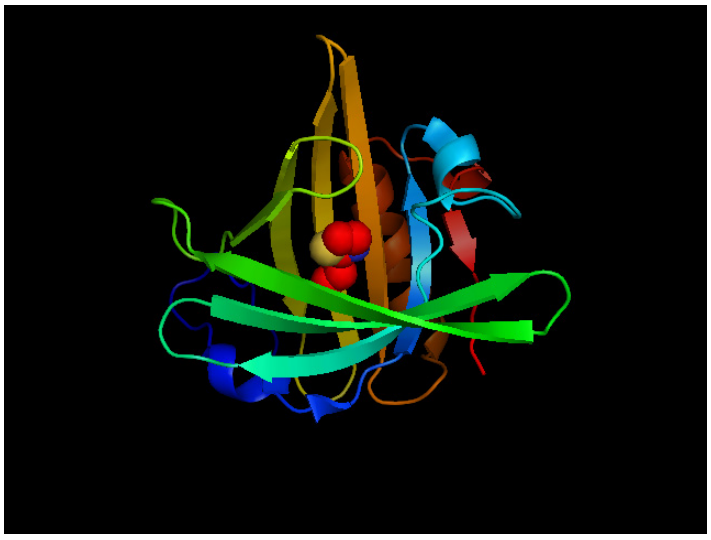


Fig. 1.7 Top view of SBT bound in the central cavity of MUP1.

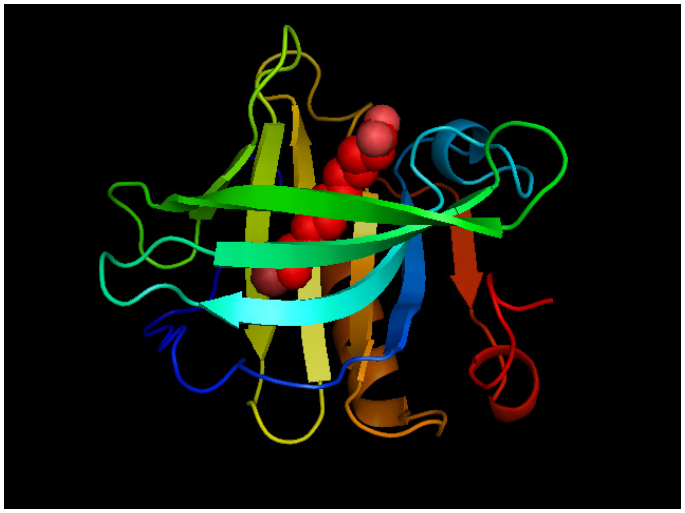


Fig. 1.8 Side view of β-Lg with 12-bromodecanoic acid bound in the central cavity.

Since β-Lg can bind ligands such as 12-bromodecanoic acid (Fig. 1.8) within the calyx, like MUP binds to SBT, our proposed idea is that, rather than being a transport protein, β-Lg physiologically functions as a pheromone-binding protein to help mother cow and calf recognition. The study of new-born rabbit behavior has shown that a pheromone controls the suckling response. [50] Therefore, the pheromone may be carried by the β-Lg, but upon denaturation, as milk dries on the cow's teat, the pheromone is released, which may trigger a similar response to the new-born rabbit and enable cow-calf recognition. In this study we have begun to test the hypothesis by attempting to observe the binding of some small volatile ligands to *bovine β-Lg* using both NMR and X-ray crystallographic methods.

1.2 Selected volatile ligands for binding to bovine β-Lg

1.2.1 Volatile compounds present in bovine milk

Many volatile compounds have been detected in bovine milk that has undergone Ultrahigh Temperature pasteurization, [51] including decalactone and ionone, which have been studied previously at low pH by NMR spectroscopy. [52-54]

It has been suggested that the volatile compounds such as decalactone and ionone are binding not only to the central cavity but also to a hydrophobic surface pocket. However, there is no crystallographic evidence provided so far. Potential ligands, including butyric acid (Rectapur™, Fig. 1.9), phenylacetic acid (Merck-Schuchardt, Fig. 1.10), hexanoic acid (J. T. Baker®, Fig. 1.11) and furaldehyde (Koch-Light Laboratories, Fig. 1.12), all of which have been detected in bovine milk, [51] were used to carry out crystallization trials. The compounds were chosen because of structural similarity to either SBT or 3MB.

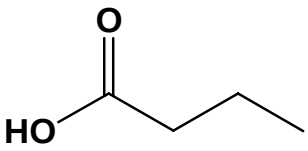


Fig. 1.9 Butyric acid.

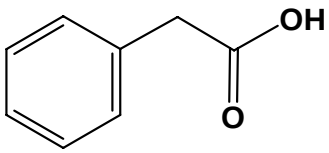


Fig. 1.10 Phenylacetic acid.

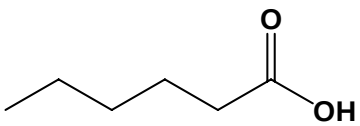


Fig. 1.11 Hexanoic acid.

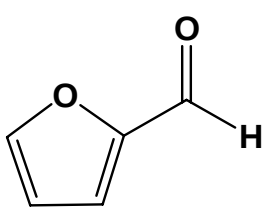


Fig. 1.12 Furaldehyde.

1.2.2 2-*sec*-Butyl-4,5-dihydrothiazole (SBT)

2-*sec*-Butyl-4,5-dihydrothiazole (SBT, Fig. 1.13) was first isolated by Novotny and co-workers in 1984 from the male mouse urine. [48] SBT is a male mouse pheromone, which promotes intermale aggression, and sex attraction and estrus synchronization in female mice and it binds to the mouse urinary protein selectively. [10] Extensive studies of SBT have been carried out; Novotny et al. also determined the absolute configuration of the naturally occurring and the synthetic enantiomer. [55]

SBT may bind with β-Lg due to the structural similarity of β-Lg and MUP. SBT has been chosen as a ligand for X-ray crystallographic study. SBT was synthesized following the scheme from Novotny et al. [55]

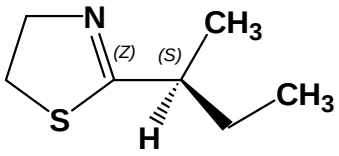


Fig. 1.13 Structure of 2-*sec*-butyl-4,5-dihydrothiazole.

1.2.3 3-Methyl-2-butenal (3MB)

3-Methyl-2-butenal (3MB, Fig. 1.14), is a volatile compound that has been identified from recent studies of rabbit milk. Rabbit newborns show a strong attraction towards to it and this pheromone is responsible for nipple attachment behavior. [50] A number of volatile compounds has been detected from rabbit milk by gas chromatography, although it is still unclear whether all of the compounds play a role as a suckling pheromone.

Therefore, experiments conducted to determine the pheromone that triggers the searching/grasping behaviors identified 3MB as the most effective compound. It has also been stated that the concentration of 3MB is important; it is effective only at concentrations between 10^{-8} and 10^{-2} g/mL. [50]

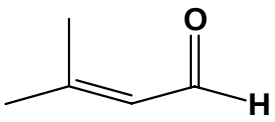


Fig. 1.14 Structure of 3-methyl-2-butenal (3MB).

3MB was purchased from Sigma and used without further purification.

1.2.4 3-Methyl-2-buten-1-ol (3MBOH)

At high concentration (100-fold excess) 3MB was found to react with the lysine side chains of β-Lg. Therefore, simple derivatization of 3MB was carried out whereby the aldehyde functional group of 3MB was reduced to the alcohol. The reduced ligand was also used in protein crystallization trials.

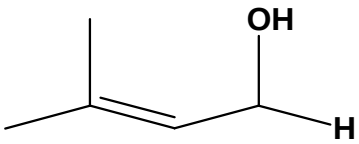


Fig. 1.15 3-Methyl-2-buten-1-ol (3MBOH).

1.3 Protein structure determination methods

There are currently two main methods for determining the three-dimensional structure of a protein: Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography. The NMR spectroscopy method allows the secondary and tertiary structure of a protein to be determined by using two-dimensional or three-dimensional techniques. Two and three-dimensional techniques allow chemical shifts to be dispersed along two and three axes respectively.

The X-ray crystallography method is more suitable for determining the structure of larger proteins. However, the crystallization process can be challenging as not all proteins can be readily crystallized, and sometimes the phase problem cannot be solved from the crystal, or the crystal simply just does not diffract well. Also the structure determined in the crystalline state does not necessarily represent the structure of the protein in the solution state.

NMR spectroscopy and X-ray diffraction methods are complementary to each other. Using both NMR spectroscopy and X-ray crystallographic methods allows us to study the behavior of β-Lg-ligand complexes with the aim of gaining some insight into the function of β-Lg.

1.3.1 Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has developed into the most common technique to determine the structures of organic compounds over the past fifty years. [56, 57] NMR spectroscopy has also become a powerful tool to study and analyze biological macromolecules (especially proteins) in the past two decades, because it can be used to obtain information about the structure and dynamics of the analyte in the solution state, which enables studies of biological macromolecules to be carried out under pseudo-physiological conditions.

The general procedure for NMR-based protein-structure determination is to first perform a sequence-specific peak assignment and then model the structure based on NMR-derived structure restraints such as NOE (Nuclear Overhauser Enhancement) values, residual dipolar couplings and backbone H_N-H_α scalar couplings.

If both the three-dimensional structure (derived from either X-ray or NMR) and the NMR peak assignment of a protein are available then they may be used to obtain residue specific information about ligand binding. Specifically, the presence of a ligand in the proximity of a particular amino acid is likely to perturb the electron density around its constituent atoms. This can lead to small chemical shift changes for nuclei (such as H_N and/or H_α) of the binding site.

For studies using unlabelled protein the experiment of choice to detect these changes is ^{1}H Total Correlation Spectroscopy (TOSCY). Briefly, transverse

magnetization is created by a 90˚ pulse. This is followed by an incremented evolution period to frequency-label this magnetization. Magnetization along a particular axis is then “spin locked” by application of a long (~40-100 ms), but relatively low power, train of pulses. During the spin lock, magnetization is transferred between spins of the same spin system (e.g. ^{1}H nuclei in the same amino acid residue). The spin lock is switched off and the free induction decay recorded. Two-dimensional Fourier transformation yields a two-dimensional spectrum, which in principle contains a grid of peaks for each spin system. For our purposes we are only interested in the peaks at the chemical shift coordinates of the H_N and their associated H_α i.e. of the same amino acid residues, (the so called “finger print region”), making much of the information content of the TOCSY spectrum redundant. A major advantage of the TOCSY experiment is that the peaks all have the same phase and are not prone to partial cancellation as can happen in the simpler Correlation Spectroscopy (COSY) experiment.

Techniques such as ^{15}N-^{1}H, HSQC (for isotopically labelled protein) and COSY or TOCSY (for unlabelled protein) are particularly useful for detecting ligand binding, since binding can often be detected via small changes in crosspeak positions for amino acids involved in the binding site.

Correlation Spectroscopy (COSY) gives the direct correlation of homonuclear spins that share a scalar coupling (e.g. between H_N and H_α). The “finger print” region of the two-dimensional spectrum shows intra-residue correlations between δH_N and δH_α. Therefore this region should show a single crosspeak for each of the protein’s amino acids (except for any prolines since

they do not contain any amide hydrogen atom).

For molecules with broad lines, such as proteins it can be advantageous to use Total Correlation Spectroscopy (TOCSY). The finger print region of a 2D TOCSY spectrum (Fig. 1.16) is similar to a COSY spectrum.

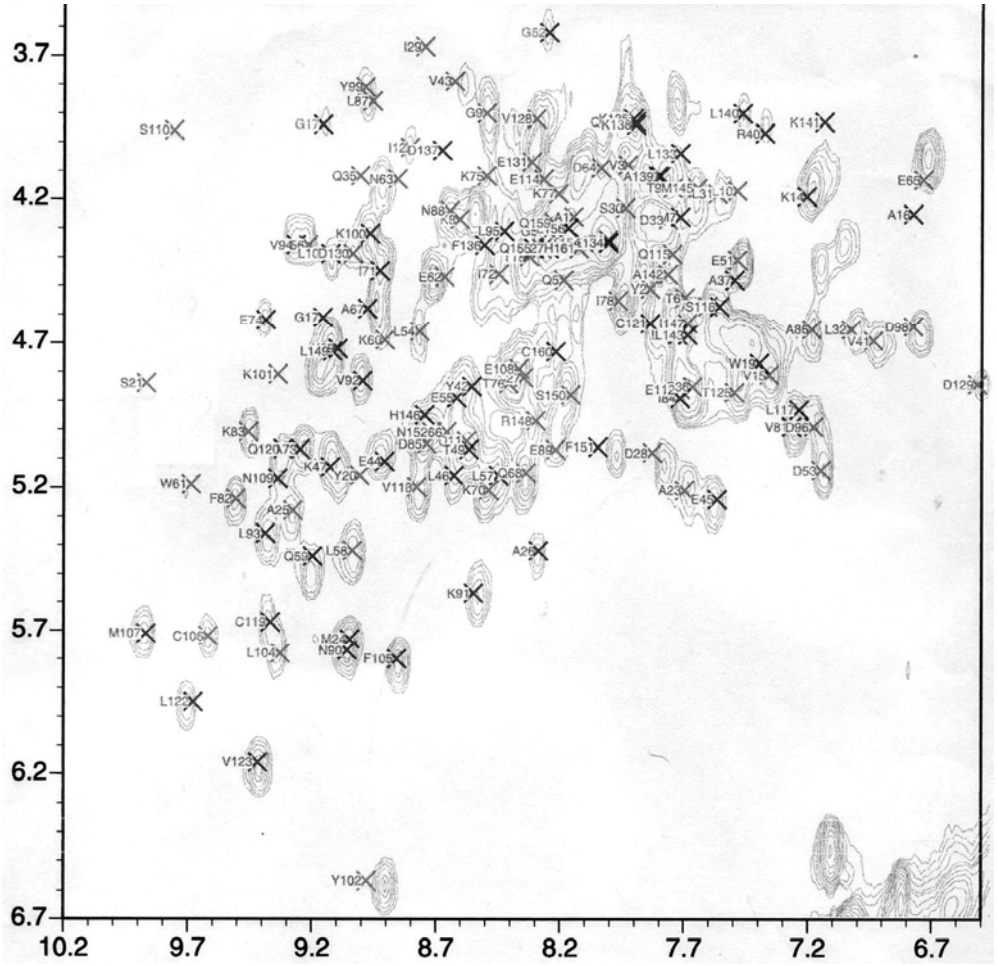


Fig. 1.16 TOCSY spectrum of β-Lg at pH 2.0 and 35 ºC with published peak assignment superimposed. [58]

However, crosspeaks may appear between spins that lack direct scalar couplings so long as they are linked as part of same spin system. [59] The finger print region of a TOCSY spectrum is similar to the COSY spectrum, but unlike COSY the signals are in phase and therefore tend to be more intense. For full peak assignment and structure determination NMR studies require isotopically

labelled protein for molecular weights > ~8000 Da. The molecular weight is an issue even with isotopic labelling; there is an upper limit in molecular weight of approximately 25000 Da, which is approximately 240 amino acids, above which spectral lines become unacceptably broad.

1.3.2 X-ray crystallography

X-ray crystallography is a technique that allows the structure determination of not only small molecules but also macromolecules such as proteins. X-ray crystallography allows us to visualize the protein at the atomic level. Therefore, the biological function and how the protein interacts with other molecules can be studied and possibly understood.

X-ray crystallography is similar to very high-resolution microscopy, so the resolution is limited by the wavelength of the electro-magnetic radiation applied, although for proteins this is rarely the case. In order to observe protein structure at near-atomic resolution, X-rays are used. A typical medium (2 Å) resolution protein crystal data set may require ~3 days of collection time. For a higher resolution, a synchrotron can be used where the X-ray intensity is greater and therefore the collection time is shorter.

The first step of determining an X-ray crystal structure of a macromolecule is usually the most difficult step. This is to grow protein crystals that are sufficiently large to collect X-ray diffraction data. Since the first protein structures determined by X-ray crystallography in the 1950s and 1960s, X-ray crystallography has become one of the most important methods of structure determination for macromolecules. [60]

Once the protein and desired ligand have crystallized, the crystals need to be tested with X-rays and if the crystals diffract to a desired resolution and with signals that are strong enough, the diffraction data then will be collected. The structure determination can be done in several ways: the method used in this

project is Molecular Replacement. The known coordinates for a similar protein are used. The Molecular Replacement method involves rotation and translation of the known model into the new crystal system to best match the calculated diffraction data to the experimental data. The calculated phases from the solution can then be combined with the data to produce an electron density map.

The electron density map is interpreted in terms of a set of atomic coordinates in the model-building process and once the preliminary model is generated, the atomic positions and atomic displacement (thermal) parameters can be refined to improve the phases to get a clearer map and a better model. It normally requires several rounds of refinement and model building to get some improvements. The R factor is a measurement of the average agreement between the structure factor amplitudes calculated from the model and those from the observed data. The lower the R-factor the better the model it is. At the final stage of the refinement the R-factor should fall below 25%. The final model needs to make chemical sense and also that there are no large regions of electron density unaccounted for. [61]

1.4 Summary of goals

The functions of the lipocalin protein family include binding of small hydrophobic molecules, which achieve the physiological functions such as retinol transport and pheromone transport. The physiological function of β-Lg is still speculative. The proposed functions include: transport protein for retinol, stimulation of a pre-gastric lipase, transport protein for unknown acid-sensitive hydrophobic molecules. One of the proposed functions of β-Lg is binding to flavor compounds, which then affects the flavor perception. There are many aroma/volatile compounds separated from bovine milk, including 3-methyl-2-butenal, which is a pheromone to trigger suckling response in new-born rabbits. Therefore, our hypothesis is that β-Lg may bind to volatile compounds such as 3-methyl-2-butenal, which is released through denaturation of β-Lg to trigger cow-calf recognition.

Chapter 2 Experimental Methods

2.1 Introduction

The experimental methods are divided into two parts, firstly ligand synthesis and secondly characterization of ligand binding to β-Lg.

2-*sec*-Butyl-4,5-dihydrothiazole (SBT) was synthesized from 2-methylbutanoic acid and 1,2-ethanolamine. [55] SBT can also be synthesized from 4-methyl-2,3-hexadienenitrile. [62] The method using ethanolamine was chosen because ethanolamine was readily available and the synthesis was straight forward. In the first step ethanolamine was silylated to give 2-(trimethylsilyloxy)ethylamine, which was characterized by ^{1}H NMR.

3-Methyl-2-buten-1-ol (3MBOH) was prepared from the reduction of 3-methyl-2-butenal (3MB). As 3MB had reacted promiscuously with lysine groups of β-Lg at neutral pH, a structurally similar, but less reactive, derivative 3MBOH was sought.

The binding of the purified ligands and the chosen volatile compounds was studied by circular dichroism (CD), NMR, and X-ray diffraction techniques.

2.2 Synthesis of 2-*sec*-butyl-4,5-dihydrothiazole (SBT)

SBT was synthesized starting from 2-methylbutanoic acid. [55] *N*-(2-trimethylsilyloxyethyl)-2-methylbutanamide was formed from the reaction of 2-methylbutanoic acid, which was activated as the acylchloride by $SOCl_2$, with 2-(trimethylsilyloxy)ethylamine. In turn, 2-(trimethylsilyloxy)ethylamine was prepared from ethanolamine by reaction with hexamethyldisilazane to protect the hydroxyl group. Finally, *N*-(2-trimethylsilyloxyethyl)-2-methyl -butanamide was reacted with Lawesson's reagent (Fig. 2.1) to yield SBT. Details of the reactions are provided in the next subsections.

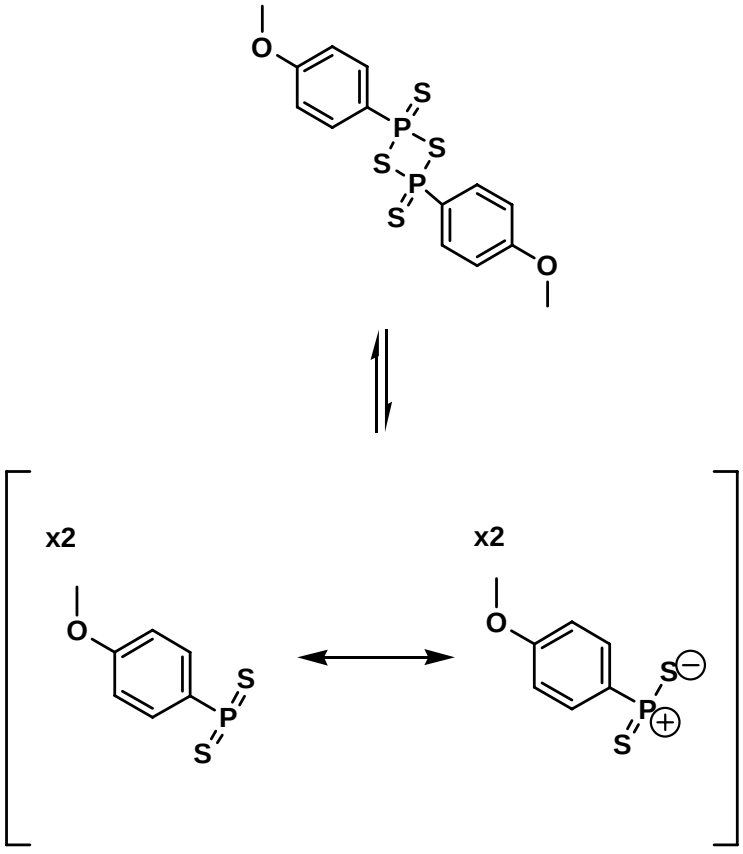


Fig. 2.1

Mechanism of Lawesson's reagent in equilibrium with the dithiophosphine ylide.

Lawesson's reagent is a thionating reagent that allows the conversion of ketones, esters and amides to thioketones, thioesters and thioamides in a reasonable yield at lower temperature when compared with other thionating reagents such as P_4S_{10}. Lawesson's reagent in solution is in equilibrium with the more reactive dithiophosphine ylide, which reacts with the carbonyl functional group on *N*-(2-trimethylsilyloxyethyl)-2-methylbutanamide. This reaction forms the thiaoxaphosphetane intermediate, which has an unstable four-membered ring. The formation of a P=O bond in a cycloreversion reaction is similar to the Wittig reaction (Fig. 2.2).

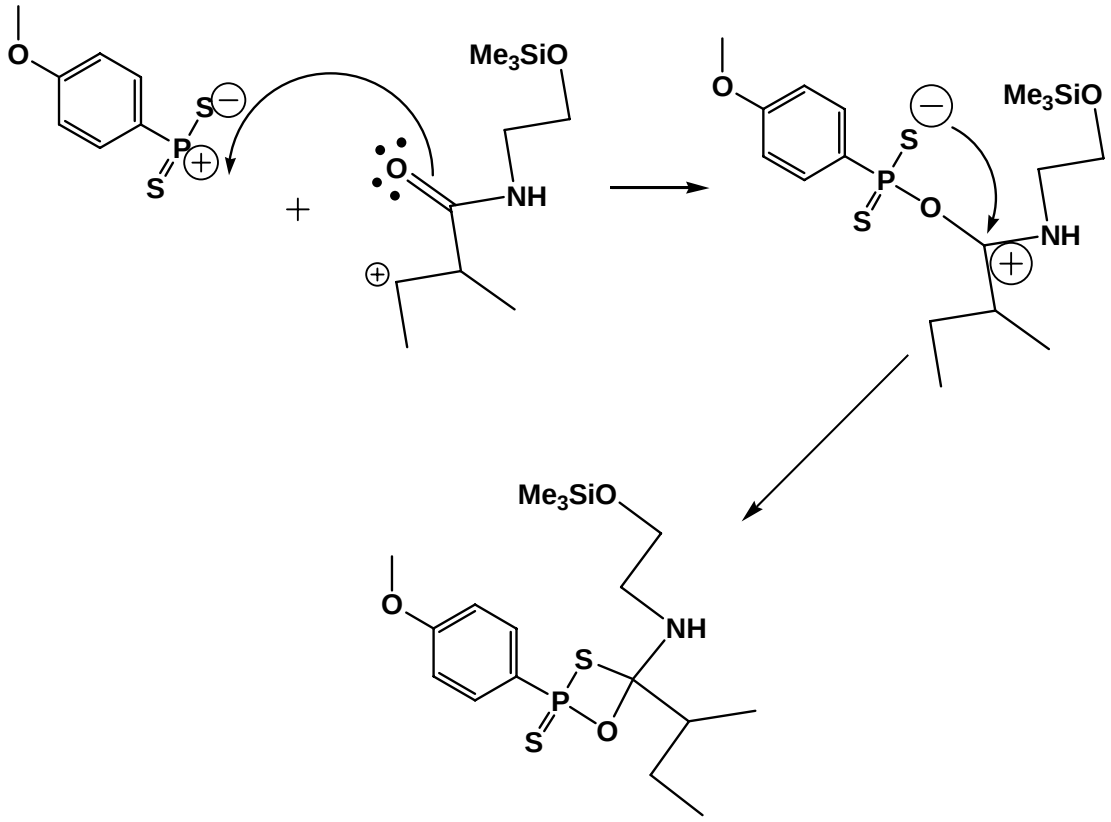


Fig. 2.2 Mechanism of thiaoxaphosphetane intermediate formation.

The cycloreversion is the driving force for the formation of the P=O bond, which results then in the formation of an S=C bond. Intramolecular reaction occurs with loss of trimethylsilanol leading to SBT formation (Fig. 2.3).

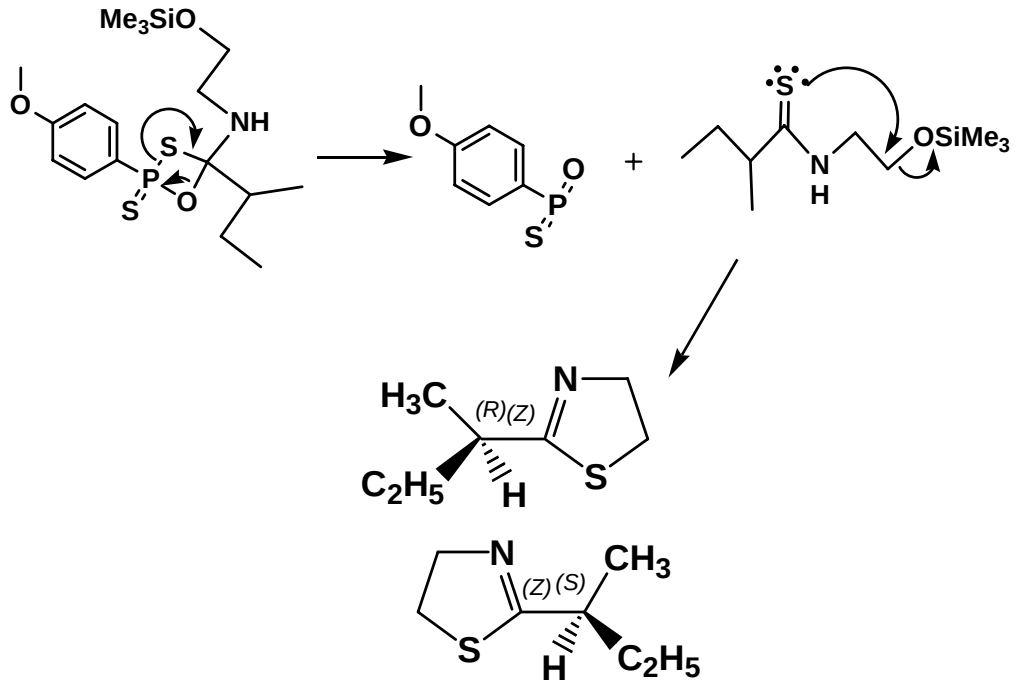


Fig. 2.3 Mechanism of formation of 2-*sec*-butyl-4,5-dihydrothiazole.

Lawesson's reagent not only acts as a thionation reagent but also facilitates the formation of SBT by desilylation; prior silylation had converted the hydroxyl group into a better leaving group. Both *R* and *S* forms of SBT are expected to form in the reaction, since no materials are chiral.

SBT was characterized by ^{1}H NMR (Fig. 2.4) and mass spectrometry (Fig. 2.6).

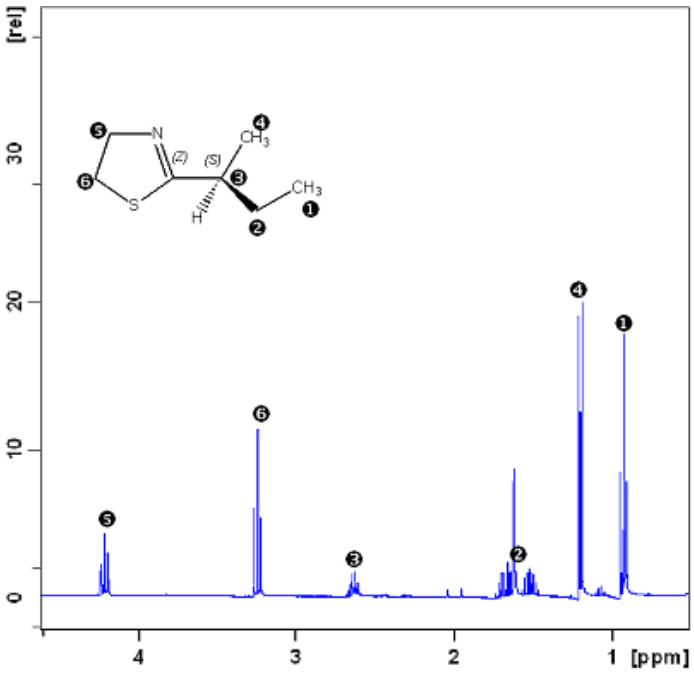


Fig. 2.4 ^{1}H NMR (500 MHz) spectrum of pure 2-*sec*-butyl-4,5-dihydrothiazole.

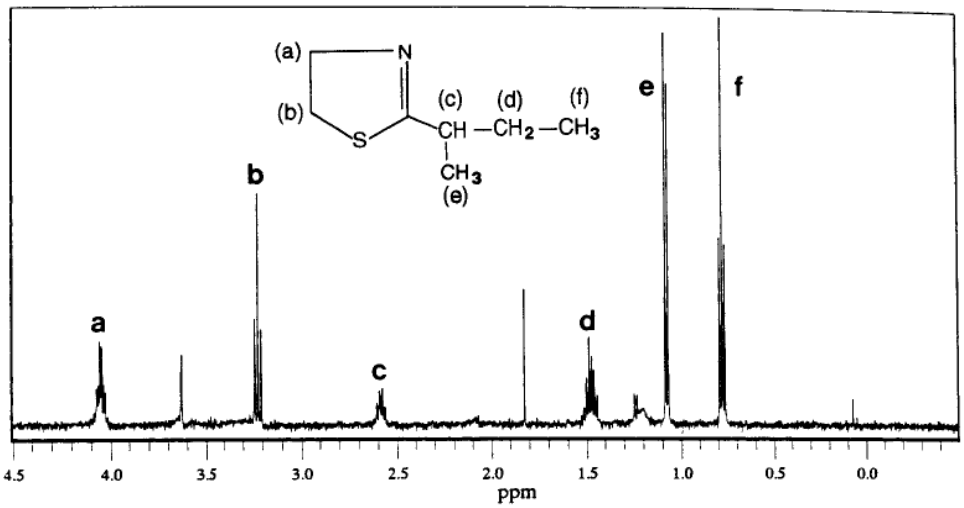


Fig. 2.5 Literature ^{1}H NMR spectrum of 2-*sec*-butyl-4,5-dihydrothiazole. [63]
Note: numerous impurities.

The comparison of the ^{1}H NMR spectrum of 2-*sec*-butyl-4,5-dihydrothiazole (Fig. 2.4) with that from the literature (Fig. 2.5) showed great similarity. However, the product required multiple fractional distillations to separate out the impurities. The mass spectrum (ES, Micromass ZMS) also showed the exact molecular weight of 2-*sec*-butyl-4,5-dihydrothiazole (Calculated molecular weight for $(M+H)^{+}$: 144.25 g/mol, observed molecular weight: 143.92 g/mol). Therefore the formation of 2-*sec*-butyl-4,5-dihydrothiazole was confirmed. The product was stored at 4℃ to avoid decomposition.

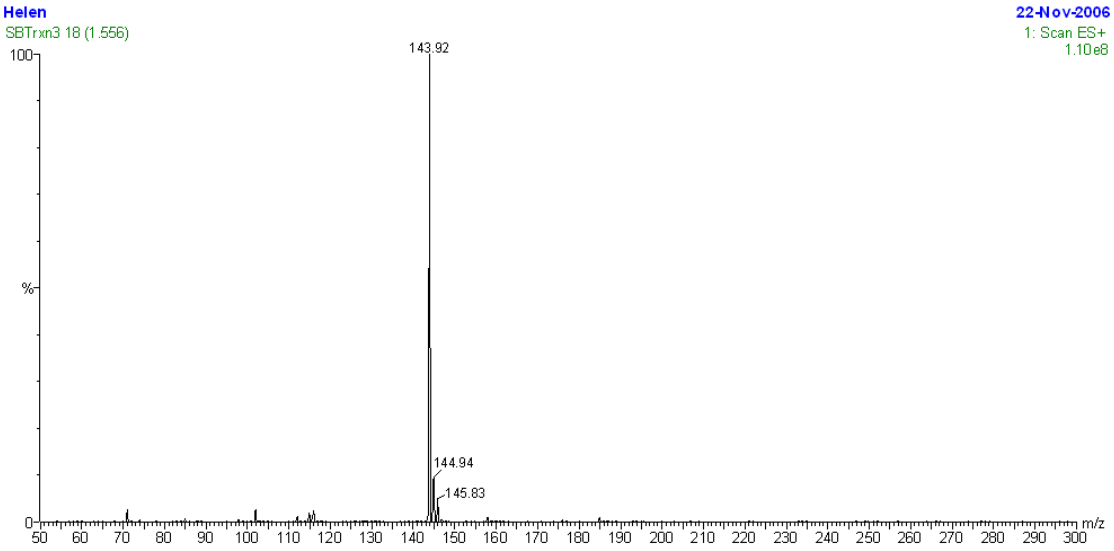
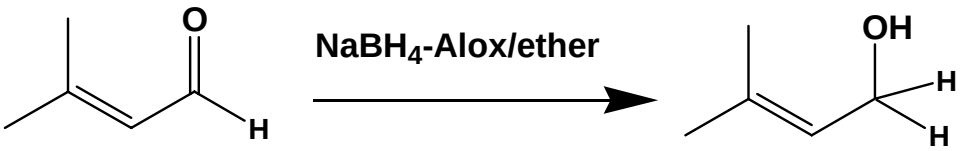


Fig. 2.6 Mass spectrum of 2-*sec*-butyl-4,5-dihydrothiazole.

2.3 Reduction of 3-methyl-2-butenal

The NMR binding experiment (Section 3.2) showed that when 3-methyl-2-butenal is in excess (1:100 protein ligand ratio) the aldehyde functional group of 3MB reacts with the lysine side chains of β-Lg in a reaction that is similar to the common nonenzymatic browning of food. [64]

The derivatization of 3MB would prevent this reaction with the β-Lg side chains. The derivatization of 3MB (Fluka, ≥97%) was carried out by following the scheme below to obtain 3MBOH. [65]



The aldehyde functional group of 3MB can be reduced using $NaBH_4$-Alox, a solid mixture of sodium borohydride adsorbed on alumina.

The reaction was monitored by NMR spectroscopy to observe the disappearance of a doublet at 10 ppm for the aldehyde proton. The yield of product 3MBOH was 48.8 mg and 3MBOH was characterized by ^1H NMR (Table 2.1). The spectrum showed that no trace of 3MB remained. Therefore the reduction of the aldehyde functional group has been accomplished.

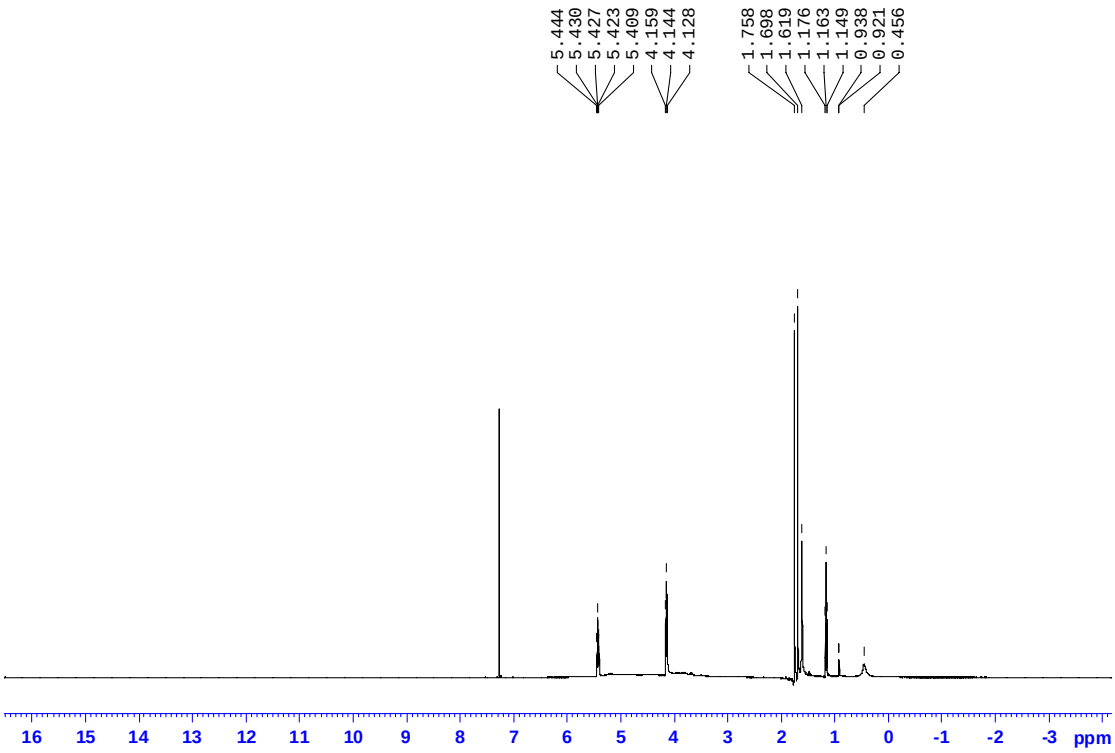


Fig. 2.7 ^{1}H NMR of 3-methyl-2-buten-1-ol in $CDCl_3$.

	^{1}H NMR ($CDCl_3$)
3-methyl-2-butenal	1.99 (s, 3H); 2.18 (s, 3H); 5.90 (d, 1H); 9.97 (d, 1H)
3-methyl-2-buten-1-ol	0.45 (bs, 1H); 1.20 (t, trace of ether); 1.69 (s, 3H); 1.77 (s, 3H); 4.15 (d, 2H); 5.44 (t, 1H)

Table 2.1 ^{1}H NMR chemical shifts for 3-methyl-2-butenal and 3-methyl-2-buten-1-ol.

2.4 Protein purification

Purification of β-Lg variant A (Sigma, Batch 026K7000) was carried out due to presence of impurity; the ^{1}H NMR intense signal of the impurity interfered with the region where H_{α} and H_{N} cross peaks appear in the 2D TOCSY spectrum. The ^{1}H NMR spectrum (Fig. 2.8) showed a distinct triplet at ~7 ppm with J coupling of 50 Hz, indicating that the impurity could very possibly be the ammonium ion.

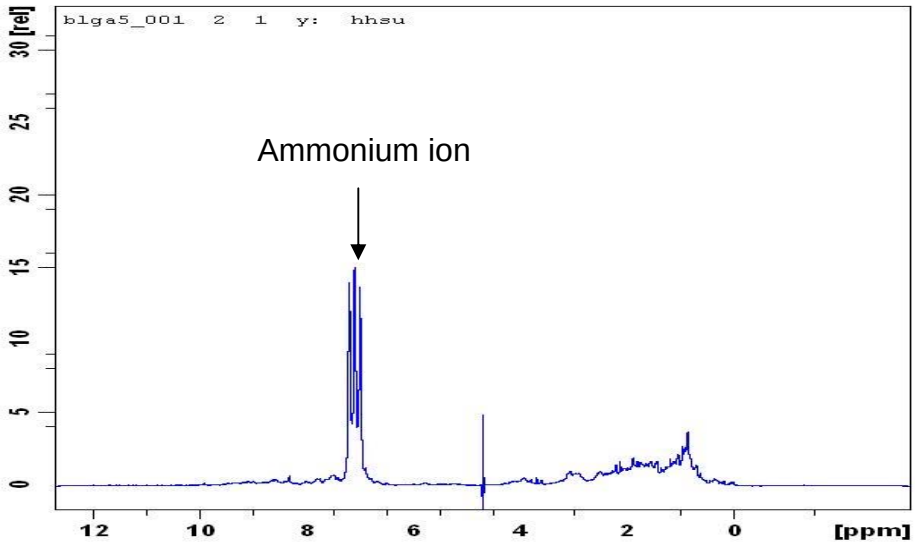


Fig. 2.8 ^{1}H NMR spectrum of contaminated β-Lg A.

2.5 Circular dichroism experiments

Circular dichroism (CD) spectroscopy is a technique that measures the difference in absorption of left-handed polarized light and right-handed polarized light. The absence of chiral structure in the chromophore will result in zero CD intensity, but ordered chiral structures, such as β-sheets, α-helices and aromatic residues in chiral environments, will result in a spectrum that contains both positive and negative signals. [66]

.

The technique allows the structures of a protein obtained from different sources, for example mutants of the same protein, to be compared for structure similarity (or difference). It also enables the study of the stability of protein conformation including thermal stability, pH stability, and stability to denaturants, and how tolerant the protein is to altering the buffer composition.

In the far-UV region (190-230 nm), the backbone amide CD signals are sensitive to protein secondary structure, allowing the percentage of helix and the percentage of β-sheet to be roughly determined and major conformational changes to be identified. Far-UV CD spectra require 20 to 200 μL of solution that contains 1 mg/mL to 50 μg/mL protein in buffer such as phosphate that does not have a high absorbance in the region.

In the near-UV region, from 250 to 350 nm, the CD signals can be sensitive to the tertiary environment of aromatic amino acids and disulfide bonds. Signals that are attributable to phenylalanine, tyrosine and tryptophan are at wavelengths 250-270 nm, 270-290 nm, and 280-300 nm, respectively. The disulfide bonds give broad weak signals throughout the near-UV spectrum.

Near-UV CD spectra can be sensitive to small changes in the tertiary structure due to protein-protein interactions, protein-ligand interactions or changes in solvent conditions. CD signals in the near-UV are much weaker than those in the far-UV CD region. Therefore, the protein solution needs to be more concentrated. It requires about 1 mL of protein solution at a concentration of 0.25 to 2 mg/mL. CD can also be used to detect ligand binding: one of the potential ligands used in this project, racemic SBT, has a chromophore, which is an -N=C-S- moiety, which has an absorbance at ~325 nm the UV-visible spectrum. Although racemic SBT itself lacks any CD signal because of a lack of chirality, when it binds to β-Lg, chirality is induced allowing it to become detectable by CD. [67]

The CD spectrometer used for the experiment was a JASCO J720 belonging to the Fonterra Co-operative Group, Palmerston North. Restrictions on access to this instrument prevented high-quality data being obtained.

The UV-visible spectra of SBT (Fig. 2.9) and 3MB (Fig. 2.10) were also recorded in order to determine the cause of the shifts in the CD spectrum.

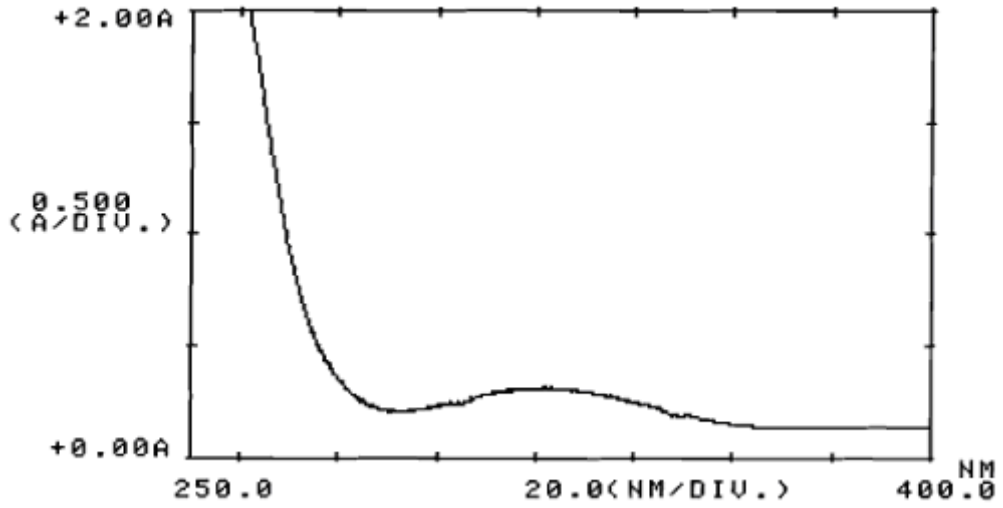


Fig. 2.9 UV-vis spectrum of 16 mg/mL SBT in phosphate buffer.

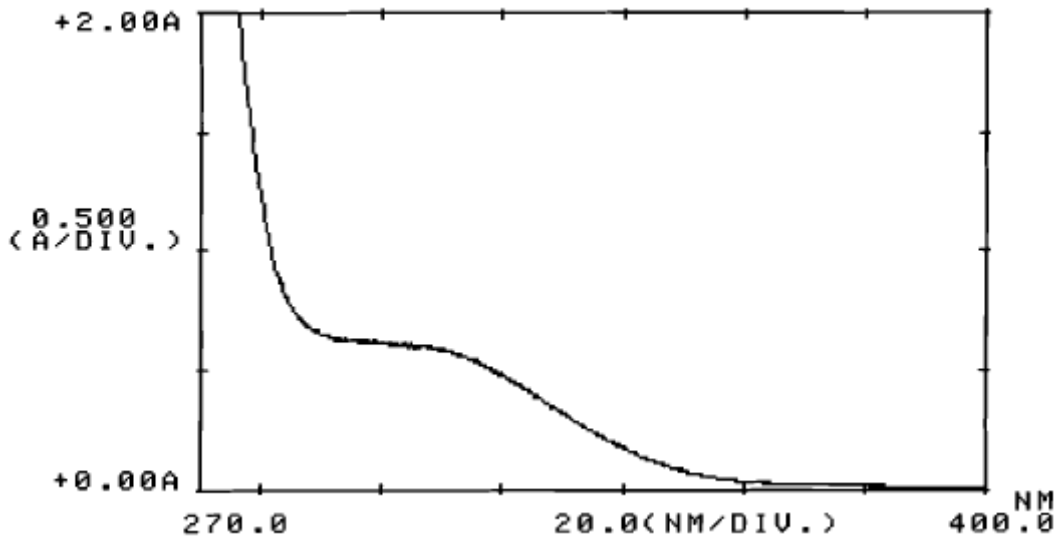


Fig. 2.10 UV-vis spectrum of 10 mg/mL 3MB in phosphate buffer.

SBT has a broad absorption peak centred at ~325 nm, and 3MB has also has a broad absorption peak at ~310 nm. The UV-vis spectrum of each ligand may help distinguish features in the CD spectrum that are induced in the ligand as a result of protein-ligand interactions.

2.5.1 Results of circular dichroism studies

The CD spectrum of β-Lg with ligand SBT (Fig. 2.11) showed a slight shift in wavelength at around 310 to 290 nm. The two tryptophan absorption peaks from residues 19 and 61 (two peaks from 310 nm to 270 nm) showed significant changes in intensity. However, it is not certain whether these changes in intensity are due to a large positive ellipticity being induced in the ligand or are due to increased light scattering by colloidal SBT. The former is more likely as the SBT solution was clear and the UV-visible spectrum was clean. The CD signal intensity changed from -20.6 to -8.3 at 292.6 nm and at 284.6 nm the intensity changed from -11.6 to -6.2. The wavelength of tryptophan absorption peak (maximum) also shifted 2.4 nm. The spectral changes suggest that SBT is possibly interacting or reacting with β-Lg.

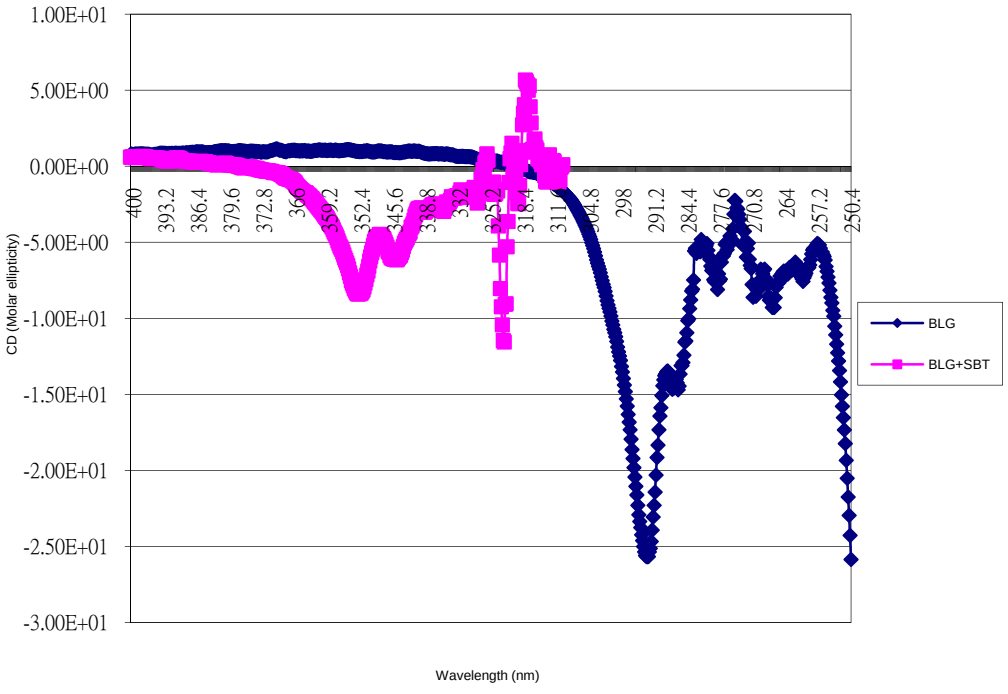


Fig. 2.11 CD spectra of β-Lg solution and β-Lg-SBT solution.

The ellipticity of the β-Lg and 3MB CD spectrum (Fig. 2.12) changed from -6.3 to -6.0 at 292.8 nm, and at 284.8 nm it changed from -4.9 to -4.8 when 2 µL of 3MB was added into β-Lg solution. When a further 2 µL of 3MB was added, the spectrum shifted from -4.8 to -4.5 at 284.8 nm. The small changes in ellipticity of the spectra and the absence of changes in the vicinity of the ligand absorption peak at ~310 nm indicated that β-Lg is possibly not interacting or reacting with the ligand in a chiral manner (the NMR data in section 3,2 indicated that when 3MB is in excess (1:100 molar ratio), it reacts with the lysines' terminal amino groups of β-Lg).

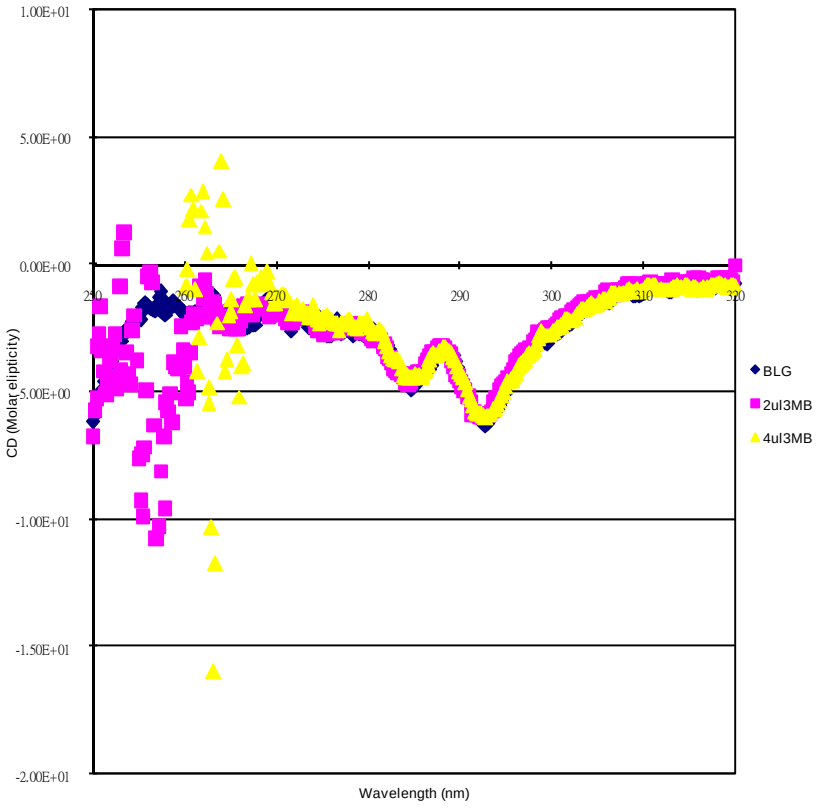


Fig. 2.12 β-Lg and 3MB CD titration spectrum.

Below 270 nm, the spectra of solutions with added ligand are very noisy, possibly due to very strong UV-visible absorptions in this region (see Fig. 2.9 and 2.10). The cause of the significant intensity change in tryptophan peak is unknown but likely to be the interaction between the β-Lg and SBT. Unfortunately experiments were not able to be repeated due to the inaccessibility of the CD spectrometer.

2.6 Cocrystallization of β-Lg and ligands

2.6.1 Protein crystallization

Macromolecules, such as proteins, can be crystallized under the appropriate conditions. In order to crystallize a protein, it is crucial to purify the protein sample before the crystallization. As a result of precipitation from an aqueous solution, individual protein molecules align themselves in a repeating series of unit cells by adopting a consistent orientation. The crystalline lattice is held together by salt bridges and van der Waals interactions.

2.6.2 Vapor diffusion

There are two commonly used methods for protein crystallization that fall under the category of vapor diffusion, the hanging drop and sitting drop methods. Both methods involve droplets that contain purified protein, buffer and precipitant. The droplet equilibrates with a larger reservoir that contains similar buffers and precipitant but at higher concentrations. Both methods require a closed system; therefore, the system must be sealed off from the outside using grease between the cover slip and the well.

All the crystallizations for β-Lg used the hanging drop method. The protein solution is located within a drop hanging on the underside of a siliconized cover slip (Fig. 2.13). 2 µL of protein solution is mixed with 2 µL of the reservoir well solution. The well solution contains a high concentration of precipitant and thus has a lower vapor pressure than the protein solution. The water from the drop diffuses water into the well solution, which raises the concentration of the protein and precipitant in the drop, until the system reaches equilibrium. The optimum conditions are maintained until the protein crystallization process is

complete. In principle, crystallization of the protein will occur over the course of days or weeks.

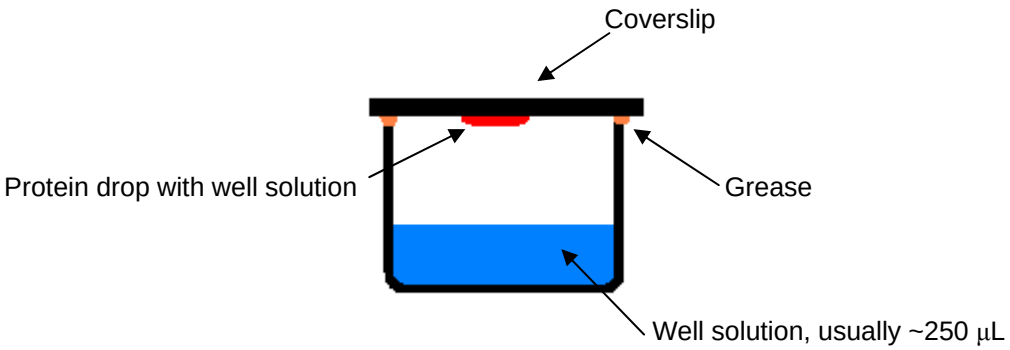


Fig. 2.13 Diagram of the hanging drop method.

2.6.3 Results

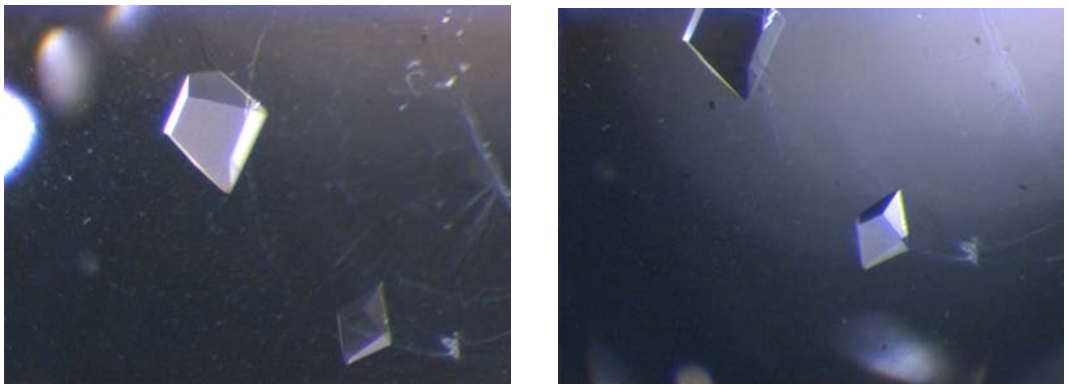


Fig. 2.14 Pictures of β-Lg A and 3MBOH trigonal crystals at pH 7.7.

The crystallizations were initially monitored daily to observe any sign of precipitation and crystallization. The ammonium sulfate screening showed that, with β-Lg variant A, it took over 2 months to generate crystals. However, with β-Lg variant A and B mixture, the crystallization process took 7-10 days. β-Lg crystallized with hexanoic acid, 3MB derivative (3MBOH, Fig. 2.14), SBT (Fig. 2,15), and phenylacetic acid (PAA, Fig. 2.16). All formed crystals within 1 month. All the crystals were harvested and screened for viable X-ray diffraction, and then followed by diffraction data collection.

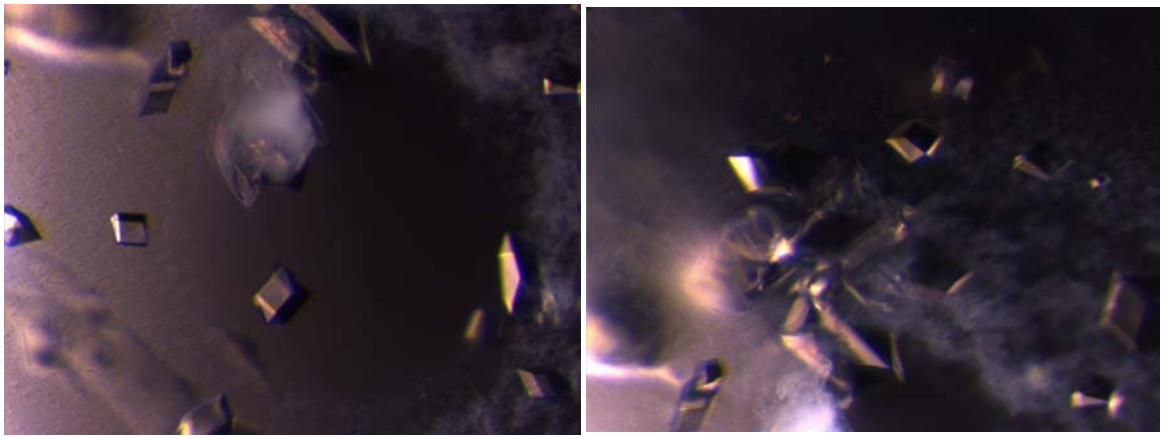


Fig. 2.15 Pictures of β-Lg A+B mixture and SBT crystals at pH 8.5, the sizes of the crystals are relatively smaller than β-Lg-3MBOH crystals and in both trigonal and orthorhombic form.

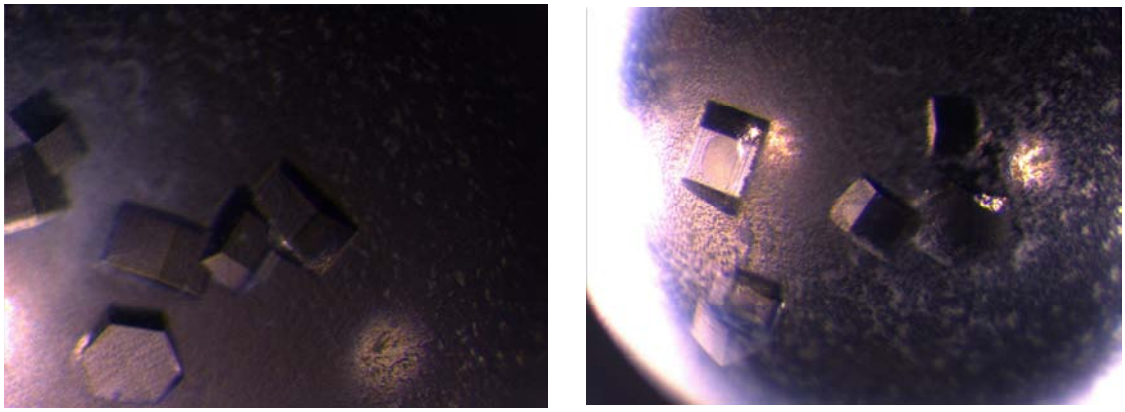


Fig. 2.16 Pictures of β-Lg A+B mixture and PAA crystals at pH 6.1. Most of the crystals are in hexagonal form with a few trigonal-shaped crystals. The sizes of the crystals are relatively larger than other crystals. The X-ray data were collected but the data could not be successfully integrated.

Chapter 3 NMR studies of ligand binding to β-Lg

3.1 Introduction

Ligand-binding studies of β-Lg have been carried out extensively ever since the fortuitous observation that β-Lg bound to retinol (Fig. 3.1). [39] Both X-ray crystallography and nuclear magnetic resonance (NMR) methods have been used to determine the binding site of the ligands. However, it is slightly more difficult to examine ligand binding with X-ray crystallography when compared to NMR techniques since the former requires protein crystals. ^{1}H NMR spectroscopy allows the protein to be studied at atomic resolution but without the requirement of crystals. It provides simultaneous information of different parts of the protein, and the spectra are sensitive to conformational changes and changes in electron density caused by the proximity of a ligand molecule, and thus gives valuable information on ligand-protein interactions.

Several flavor compounds have been used in studies of protein-ligand interactions including β-ionone and γ-decalactone (Fig. 3.2). The interactions with bovine β-Lg were studied by 2D-NMR spectroscopy. [53, 54] TOCSY and NOESY spectra are usually recorded in NMR studies on proteins in order to generate a plot of the H_N-H_α region that shows cross-signals due to the coupling between these protons in the former and spatial proximity in the latter. When the protein reference spectrum and the protein-ligand spectrum are overlayed, any variations in chemical shifts of particular peaks will show up, which can then be analyzed to see whether the shift is significant. The NMR studies provide precise (i.e. residue specific) information on the binding site

and the conformational changes that may be caused by the ligands. For the NMR studies in this project two ligands where chosen, 3-methyl-2-butenal, the pheromone present in rabbit milk, and phenylacetic acid, a volatile compound present in bovine milk. The studies were focused on the possible interactions between the chosen compound and β-Lg.

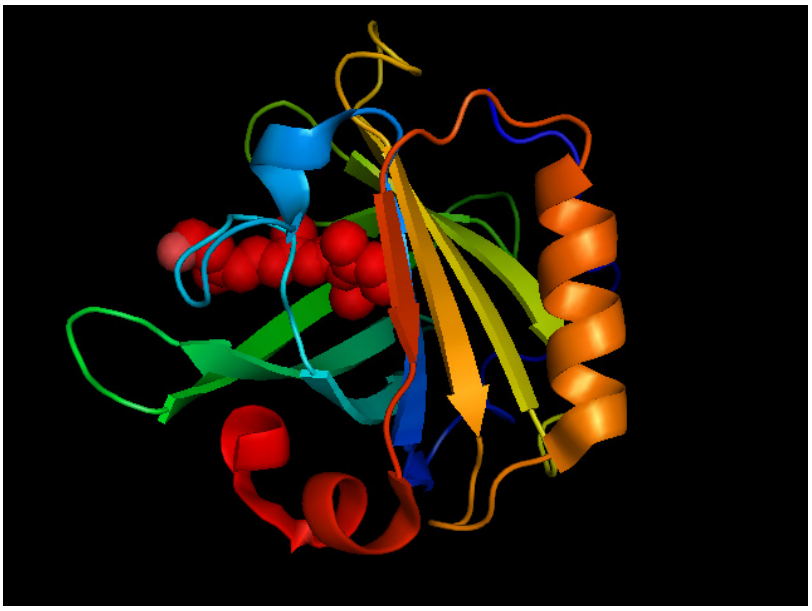


Fig. 3.1 Structure of β-Lg complexed with retinol. [68] The protein is rainbow coloured, starting with blue at the N-terminal end and ending in red at the C-terminal end. The retinol is shown in space-filling representation.

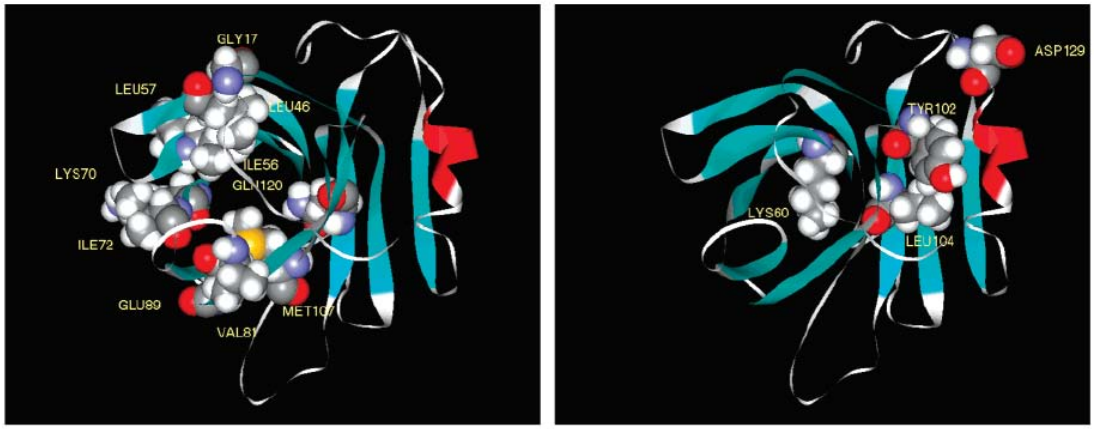


Fig. 3.2 Amino acid side chains (shown in space-filling representation) affected by complex formation with γ-decalactone (left) and β-ionone. [52]

3.2 3-Methyl-2-butenal ligand binding trial at pH 7.4

3-Methyl-2-butenal (3MB, Sigma, 97%) was chosen to carry out the first binding trial at pH 7.4. The chosen pH was 7.4 since the calyx is in its open conformation. The binding experiment was monitored by 700 MHz ^{1}H NMR spectroscopy. A 1 mL aliquot of 80 mg/mL (0.04 mol) solution of β-Lg in phosphate buffer at pH 7.4 was saturated with $1x10^{-2}$ mol/mL ($4x10^{-4}$ mol, 38.59 µL) of 3MB so the protein:ligand ratio was 1:100, to observe if the binding occurred at pH 7.4 and 310 K.

3.2.1 Results

The color of the sample solution that contained ligand 3MB changed from clear to bright yellow then gradually into dark orange with precipitation over 2 hours (Fig. 3.3). The process that caused such color change in the protein solution appears to be the aldehyde functional group of the ligand reacting with the lysine side chains of the protein. The reaction is similar to the common nonenzymatic browning of food (Maillard reaction). The process would be unlikely occur under normal conditions because any potential volatile aldehydes would not be in such excess when present in milk. The binding trial indicated that the ligand:protein ratio needs to be reduced in order to prevent this reaction. Alternatively the aldehyde functional group can also be reduced to an alcohol functional group to avoid this reaction.

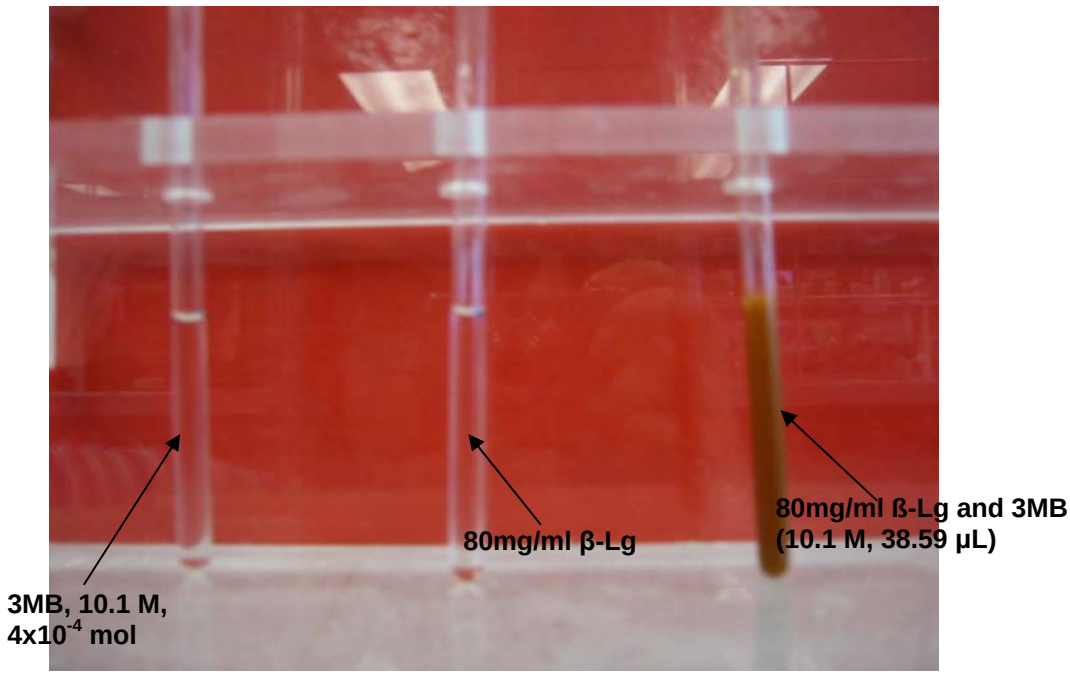


Fig. 3.3 NMR samples for 3MB binding trial.

3.3 3-Methyl-2-butenal ligand-binding experiment at pH 2

β-Lg variant A (purity ≥ 90%; Sigma) was dissolved at a concentration of 1 mM in 10 mM phosphate buffer and with 10% of D_2O, and the pH of the solution was adjusted to pH 2.0. The β-Lg dimer present at neutral pH gives NMR spectra with relatively broad lines. Reduction of the pH causes the dimer to dissociate, but still maintain a native-like tertiary structure. [18] The β-Lg monomer gives good quality spectra. At pH 2 the reaction between lysine and 3MB observed at pH 7.4 seems to be impeded.

The ligand binding experiment was carried out by adding 100 µL aliquots of 10 mM of 3MB (Sigma) to the β-Lg sample to give at ligand:β-Lg ratios of 2:1, and a TOCSY spectrum was recorded at 310 K both before and after addition of ligand in the States-TPPI mode with mixing time of 60 ms, spectral width of 13.0 ppm, 96 scans per FID with a data size of 2048 x 256 points on a Bruker

500 MHz spectrometer.

3.3.1 Results

TOCSY spectra were used to generate maps of the H_N-H_α fingerprint region. The spectra were imported into CARA [69] and overlaid, the peaks were then linked to their respective amino acids using the previously published assignment [58] and peak Val123 was used to calibrate both spectra. The residues that had undergone detectable changes in their H_N chemical shifts were then identified. The fingerprint region of the β-Lg reference spectrum and the β-Lg-3MB sample is shown in figure 3.4.

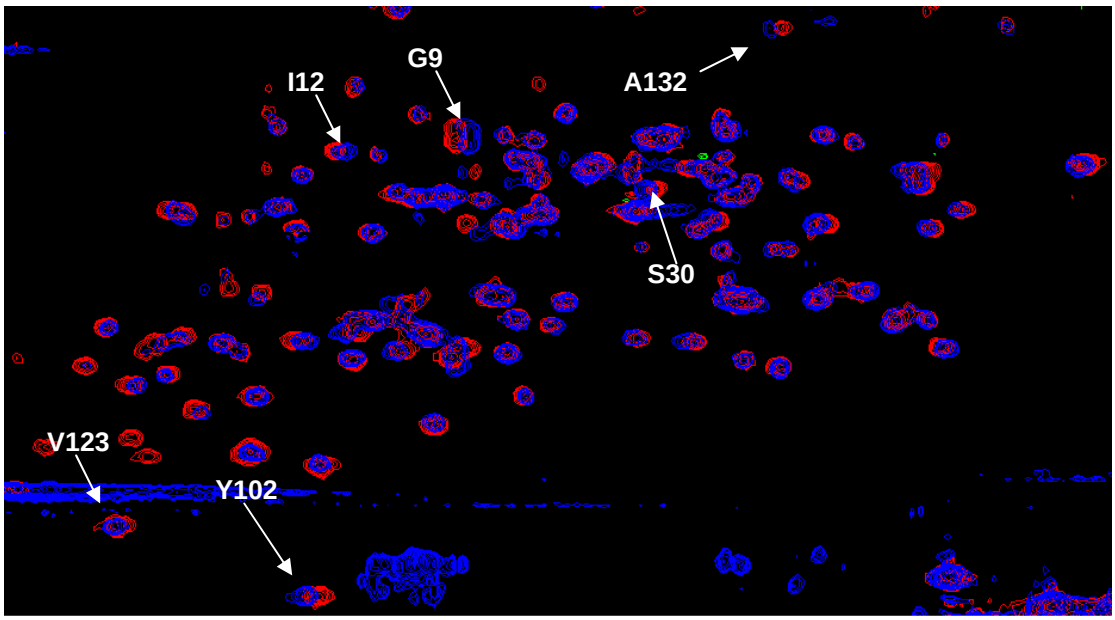


Fig. 3.4 Overlap spectrum of β-Lg reference solution (red) and β-Lg-3MB solution (blue) recorded on the 500 MHz spectrometer.

The residues that appeared to be most affected by the addition of 3MB, based on the size of the change of their H_N chemical shifts, were then related to the tertiary structure of β-Lg.

The residues that seem to have significant chemical shift changes (chemical shift change > 0.01 ppm) were tabulated and compared (Table 3.1).

Residue	δH_N β-Lg	δH_N β-Lg-3MB	$\Delta\delta$
Gly 9	8.544	8.501	-0.043
Ile12	8.880	8.846	-0.034
Ser30	7.986	8.015	0.029
Glu44	8.985	8.963	-0.022
Gln59	9.205	9.185	-0.020
Lys69	9.112	9.123	0.011
Ala86	7.160	7.181	0.021
Val92	9.078	9.091	0.013
Leu93	9.443	9.426	-0.017
Val94	9.290	9.302	-0.012
Tyr102	8.860	8.897	0.037
Ala132	7.637	7.673	0.033

Table 3.1 Residues of β-Lg having significant changes in chemical shift in the 3MB binding experiment on the 500 MHz spectrometer.

The residue that had most significant chemical shift was residue Gly9 (Fig. 3.5), which is located at the bottom of the calyx and is part of the N-terminal coil. The significant chemical shift change is suggesting that 3MB was interacting with Gly9. Residue Ile2 (Fig. 3.6) is also located at the coil between of β-0 and β-A; it also had a relatively large change in chemical shift of 0.034 ppm. Residue Ser30 (Fig. 3.7) also had a 0.029 ppm chemical shift change and it is located in the α-helical loop AB (linking strands A and B).

Residue Tyr102 (Fig. 3.8) on β-strand G and residue Ala132 (Fig. 3.9) on the three-turn α-helix H also had relatively large chemical shift changes of 0.037 and 0.033 ppm.

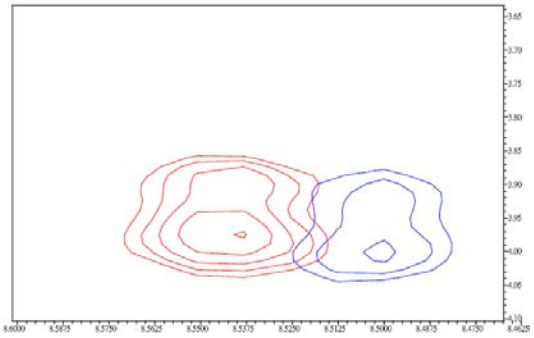


Fig. 3.5 Zoom in of residue Gly9 of overlapped spectra (β-Lg blue, β-Lg-3MB, red).

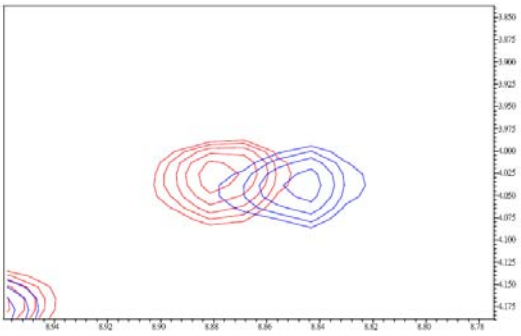


Fig. 3.7 Zoom in of residue Ile12 of overlapped spectra (β-Lg blue, β-Lg-3MB, red).

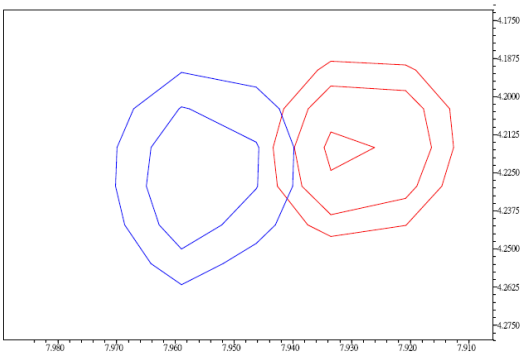


Fig. 3.6 Zoom in of residue Ser30 of overlapped spectra (β-Lg blue, β-Lg-3MB, red).

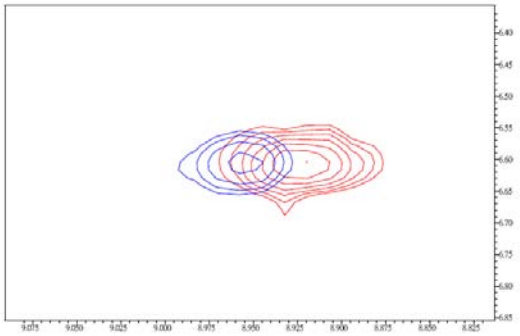


Fig. 3.8 Zoom in of residue Tyr102 of overlapped spectra (β-Lg blue, β-Lg-3MB, red).

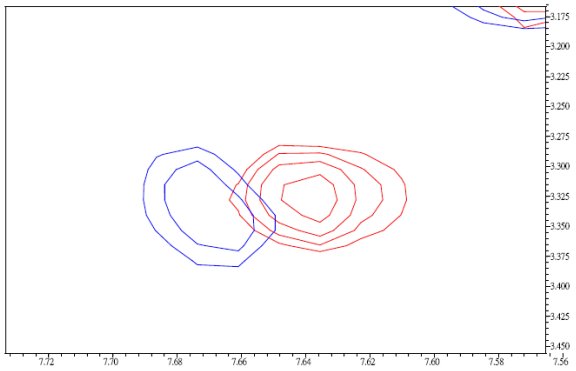


Fig. 3.9 Zoom in of residue Ala132 of overlapped spectra (β-Lg blue, β-Lg-3MB, red).

7The residues with the more significant chemical shift changes (> 0.03 ppm) seem to be at the bottom of the calyx (Fig. 3.10) and could well be caused by interactions between 3MB and β-Lg.

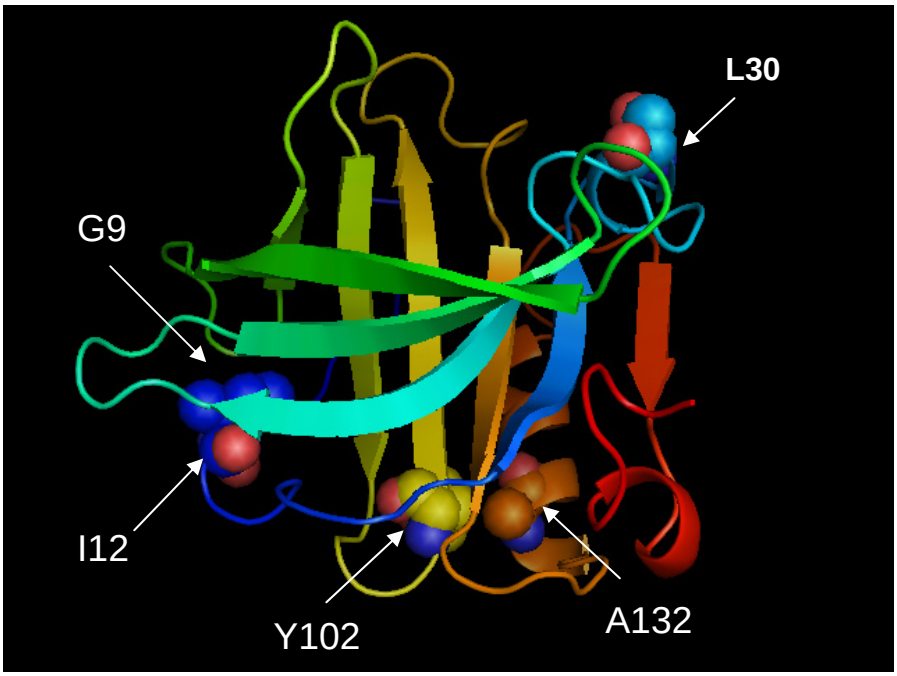


Fig. 3.10 Residues that had significant chemical shift changes caused by addition of 3MB. The figure is colored in the order of the sequence (beginning with blue at the N terminus and ending in red at the C terminus).

Smaller chemical shift changes are also seen on residue Gln44 (on β-strand B), residue Val59 (on β-strand C), residue Lys69 (on β-strand D), residue Ala86 (on β-strand e), and residues Val92, Leu93, Val94 (on β-strand F). The minor changes in chemical shifts are likely triggered by other residues interacting with the ligand to a lesser extent indirectly by perturbation by ligands interacting directly with β-Lg. In order to verify these results a titration experiment was performed, the details of which are given in the following section.

3.4 3-Methyl-2-butenal titration experiment at pH 2 and at 700 MHz

3.4.1 Sample preparation

β-Lg variant A (purity ≥ 90%; Sigma) was dissolved at a concentration of 1 mM in 10 mM phosphate buffer and with 10% of D_2O; the pH of the solution was adjusted to pH 2.0. A titration experiment was carried out by adding 50 µL of 10 mM of 3MB (Sigma) to the β-Lg sample, and after each addition TOCSY spectra were recorded using same parameters as for the preliminary experiment with 3MB at 310 K on a Bruker 700 MHz spectrometer.

3.4.2 Results

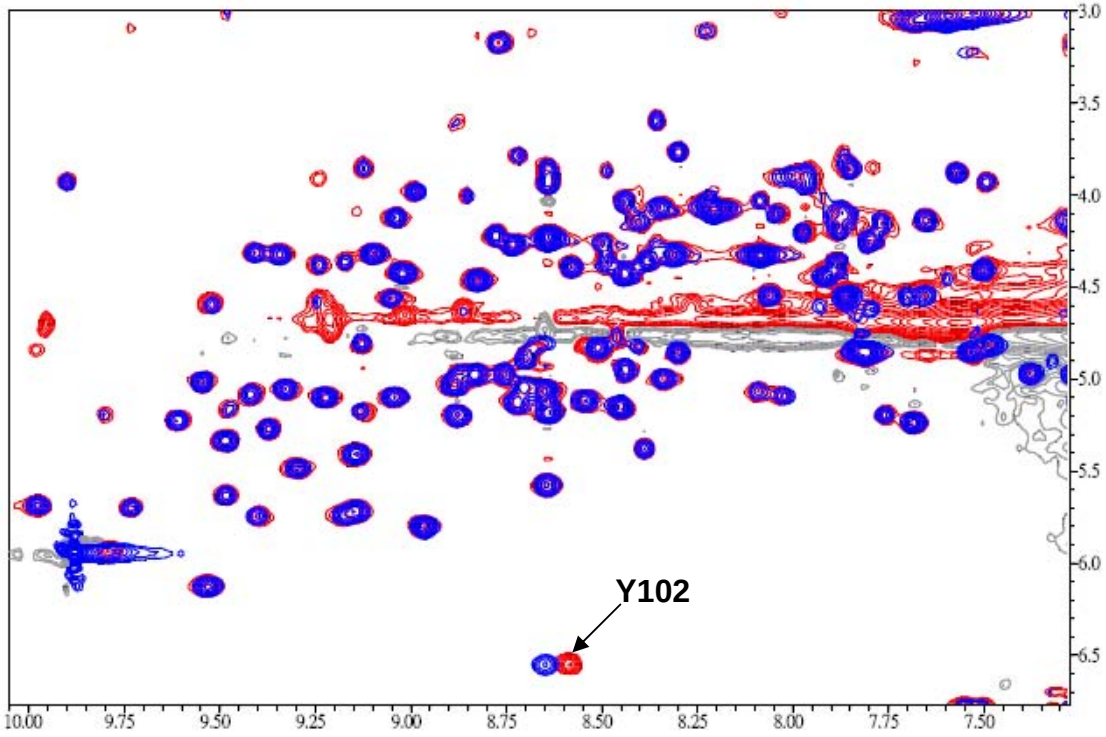


Fig. 3.11 Overlap of reference and titration (1 β-Lg:2 3MB) spectra.

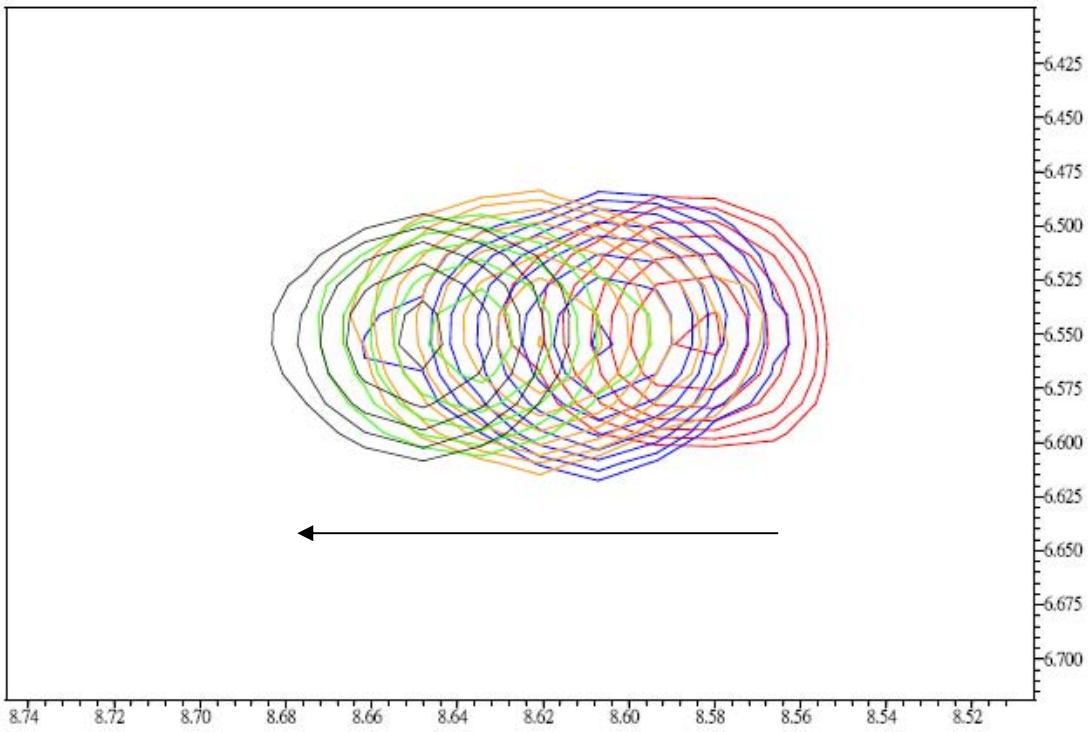


Fig. 3.12 Overlap spectrum of reference and titration spectra for residue Tyr102, Where the red spectrum is the reference β-Lg spectrum, blue spectrum is 1:0.5 protein-ligand ratio, orange is 1:1, green spectrum is 1:1.5 and black spectrum is 1:2. The arrow notes that the consistent downfield shift on addition of 3MB.

The overlap spectrum of the reference β-Lg and the titration spectra (Fig. 3.12) showed that as the ratio of 3MB increased, the chemical shift changes also increased. The chemical shift of residue Tyr102 changed from 8.584 ppm to 8.650 ppm, which is considered a significant change (0.066 ppm). The titration experiment showed that residue Tyr102 which is located at the bottom of the calyx (Fig. 3.10), had a result that is consistent with the preliminary titration experiment.

However, residue Gly9, Ile12 and Ser30 only had very small chemical shift changes (< 0.01 ppm) unlike the apparently significant changes shown in the preliminary NMR experiment on the 500 MHz spectrometer. Hence it is confirmed that the chemical shift change in residue Tyr102 observed from the

preliminary titration experiments was more likely to be caused by ligand-protein interaction rather than by the environment change. 3MB could potentially be bound to residue Tyr102 at the hydrophobic surface area rather than within the calyx.

3.5 Phenylacetic acid titration experiment

The titration experiment for phenylacetic acid (0.52 M, with 5% ethanol to dissolve the acid completely) with β-Lg was carried out using same method as the previous experiment with the 500 MHz NMR spectrometer. The protein:ligand ratio was 1:2 and the experiment was carried out overnight (15 hours). TOCSY spectra were recorded using same parameters as for the preliminary experiment with 3MB at 310 K on a Bruker 500 MHz spectrometer.

3.5.1 Results

The comparison of the 2D spectra showed that only residues Lys101, Tyr102 and Ala132 have significant changes in chemical shifts (Fig. 3.13). The reference spectrum showed that the H_N chemical shift of residue Tyr102 (Fig. 3.14) is at 9.027 ppm, and changed to 9.052 ppm after phenylacetic acid was added (0.025 ppm). The H_N chemical shift for residue Lys101 (Fig. 3.15) from the reference spectrum was at 9.361 ppm and the titration spectrum showed the chemical shift became 9.401 ppm with a change in chemical shift of 0.040 ppm. The H_N chemical shift for residue Ala132 (Fig. 3.16) from the reference spectrum was at 7.671 ppm and the titration spectrum showed the chemical shift become 7.692 ppm with a change in chemical shift of 0.021 ppm.

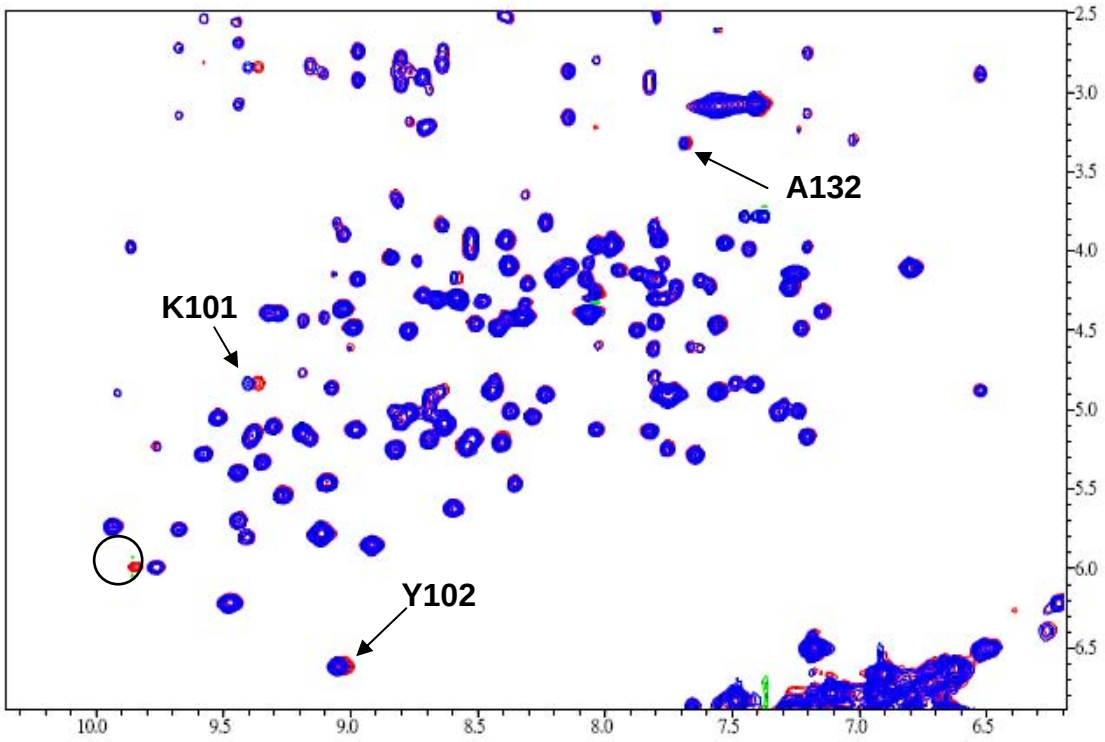


Fig. 3.13 Overlap of β-Lg reference and β-Lg-phenylacetic acid TOCSY spectra, recorded at 500 MHz. The circled red peak is not due to protein/ligand interaction.

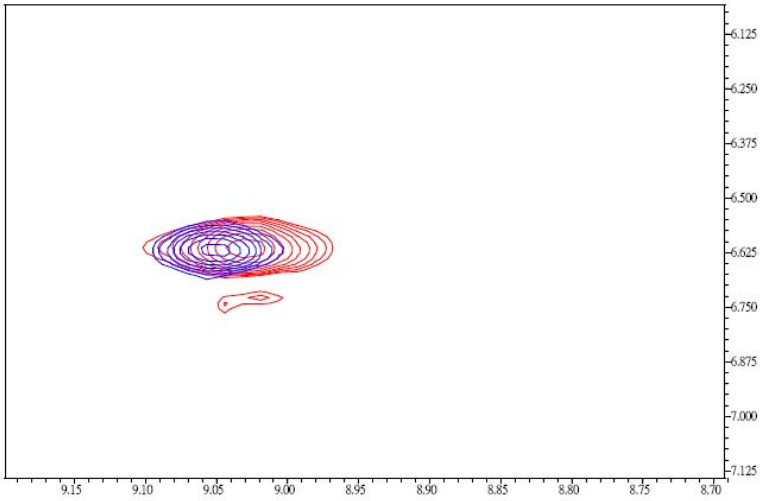


Fig. 3.14 Overlap of reference and titration spectra for residue Tyr102.

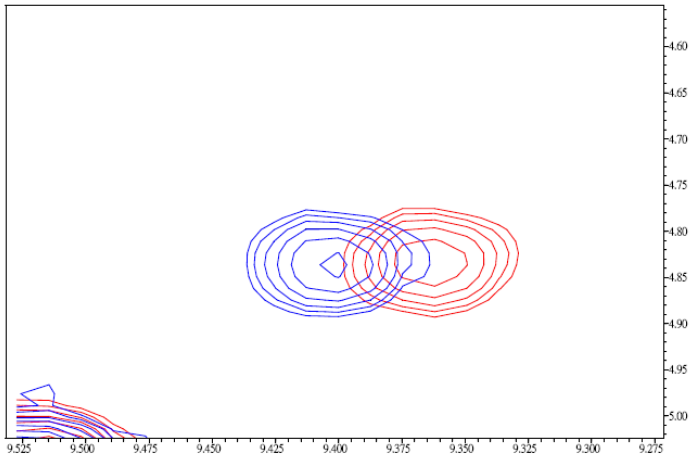


Fig. 3.15 Overlap of reference and titration spectra for residue Lys101.

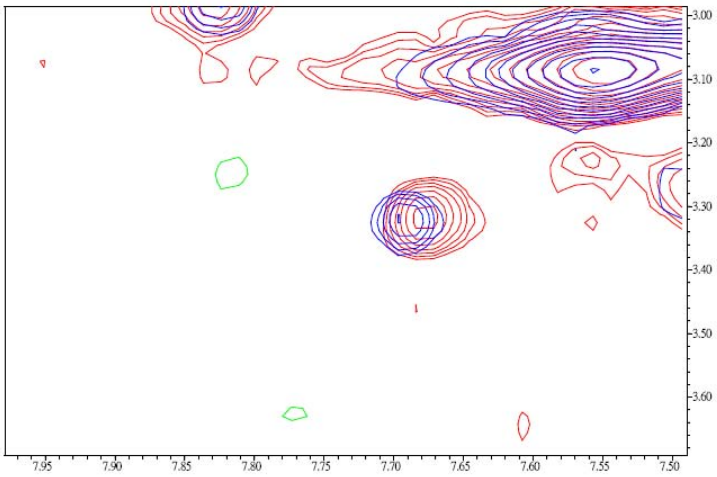


Fig. 3.16 Overlap of reference and titration spectra of residue Ala132.

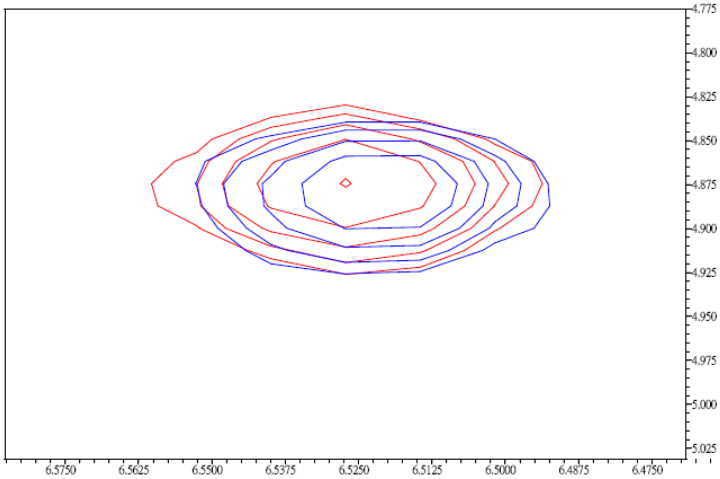


Fig. 3.17 Overlap of reference and titration spectra of residue Asp129.

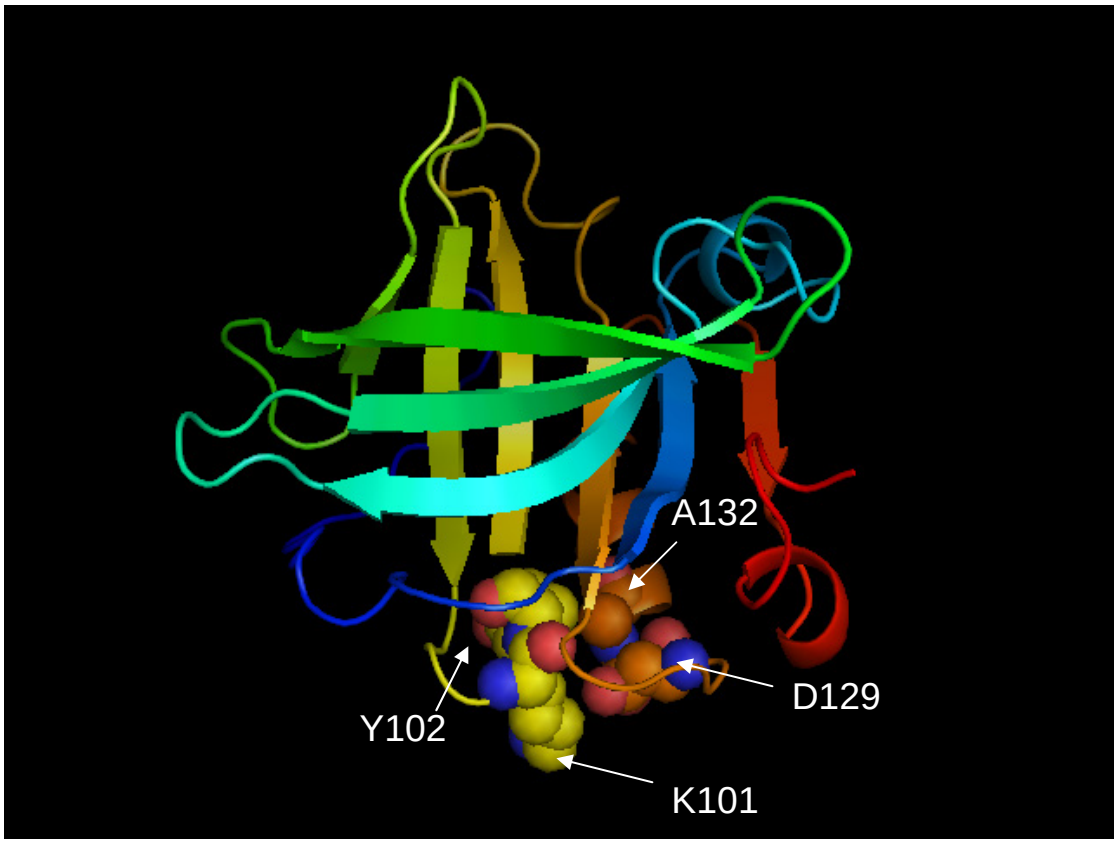


Fig. 3.18 Residues having significant chemical shift changes upon binding of phenylacetic acid. The figure is colored in order of sequence (beginning with blue at the N terminus and ending in red at the C terminus).

Residue Lys101 and Tyr102 are close to the bottom of the calyx and belong to β-strand G. The results showed that phenylacetic acid was also interacting with β-Lg Tyr102. By looking at residue Asp129 closely, the chemical shift changed from 6.592 ppm to 6.516 ppm (Δ=-0.013 ppm), indicating that phenylacetic acid was interacting with β-Lg at the hydrophobic surface area. Residue Ala132 belongs to the three-turn α-helix and projects into the solvent accessible surface area, which also indicates that phenylacetic acid may be interacting with β-Lg at an alternative binding site to the central calyx (Fig. 3.18).

3.6 Discussion and conclusions

The results from the β-Lg NMR binding experiments revealed information about the preferential binding site for small volatile compounds. It is possible that the EF loop of β-Lg at pH 2 is undergoing conformational change between open and closed conformations; hence potential ligands 3MB and PAA may bind within the calyx. The interactions between β-Lg and flavor compounds also have been studied at pH 2.0 by 2D NMR technique. [52, 53, 56] At 500 MHz the binding of γ-decalactone took place in the calyx but β-ionone binds to a hydrophobic surface area. [52] The chemical shift changes from the 3MB ligand binding experiment at 500 MHz showed that it is possible that 3MB may have multiple binding sites similar to γ-decalactone, inside the calyx and at an external hydrophobic surface area. However, titration experiments at 700 MHz for both potential ligands 3MB and PAA indicated they interacted only with the residues that are located at the bottom of the calyx, similar to β-ionone. Binding of 3MB may have taken place at an external hydrophobic surface area due to a significant chemical shift change of Tyr102 (observed also in the β-ionone study [52]). This Tyr is located at the bottom of the calyx. However, the lack of other chemical shift changes that should appear when the ligand interacts with/binds to β-Lg at the hydrophobic surface area suggests that 3MB only binds to β-Lg on residue Tyr102, whereas residues Tyr102 and Lys101, Asp129 and Ala132 appear all to interact with PAA at a hydrophobic surface area.

Chapter 4 The X-ray crystallographic studies

4.1 Introduction

Crystallographic methods have become one of the most important tools in structural determination. To carry out protein crystallographic studies, there are several necessary steps: firstly, protein needs to be purified and concentrated for crystallization. Secondly, conditions must be found to get crystals to grow and then to determine if crystals diffract well enough for collection of an X-ray diffraction data set. Thirdly, the structure of the protein must be solved by an appropriate method and the structure refined. Fig. 4.1 describes the steps for X-ray analysis of a macromolecule.

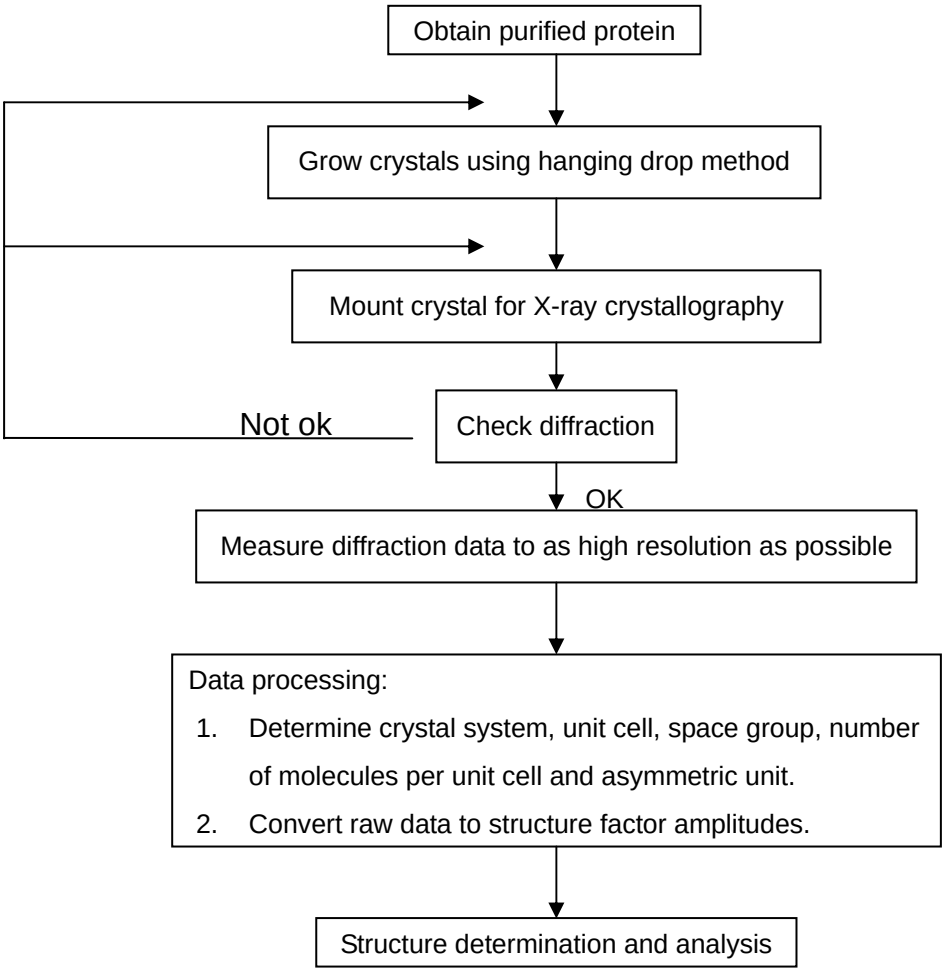


Fig. 4.1 Flow diagram for X-ray analysis of a macromolecule.

The Collaborative Computational Project Number 4 in Protein Crystallography (CCP4) [70] was set up in 1979 in the UK to support collaboration between researchers and resulted in a suite of programs spanning data processing to structure determination and analysis. CCP4i [71] was then developed in 1997 by Liz Potterton, to provide a graphical user interface (GUI) for the CCP4 suite. The CCP4 suite is a set of separate but integrated programs that perform individual tasks via data files. Each task requires an input file, and the output file is usually the file required for the next step. The CCP4i suite is used because it is convenient for crystallographic calculations. Many programs have standard default parameters but the GUI for CCP4i allows calculations to be easily changed.

In addition to CCP4i, there are several other programs used for this project in order to solve and display the protein crystal structure including: Crystallographic Object Orientated Toolkit (COOT) and PyMOL. COOT is a program for model building, model completion and validation. [72] COOT displays maps and allows model manipulation such as real space refinement, manual rotation/translation, and many other functions that are useful for solving the crystal structure. It also has many structure validation tools. PyMOL is a molecular visualization system that provides a useful tool to generating high quality three-dimensional images for both small and biological macromolecules. [73]

4.2 X-ray data collection system

The X-ray data collection system (Fig. 4.2) included the Rigaku MicroMax-007 microfocus X-ray generator with an Axco P70 Capillary optic to focus and monochromate the X-ray beam. This optic is designed for poorly diffracting crystals and it provides flux intensity of ~8 x 10^{10} photons/mm^2/sec.

The Rigaku R-AxisIV++ detector is the most popular X-ray area detector for macromolecular crystallography. The detector combines two large active area (300mm x 300mm) imaging plates with low background noise and wide dynamic range, which makes the device ideal for collecting the diffraction data accurately. The system is supported by CrystalClear™, a unified graphical user interface for data collection and processing. CrystalClear allows data collection be automated almost completely. It allows data to be collected at optimized speed, redundancy or coverage and then the data can be processed after the collection.

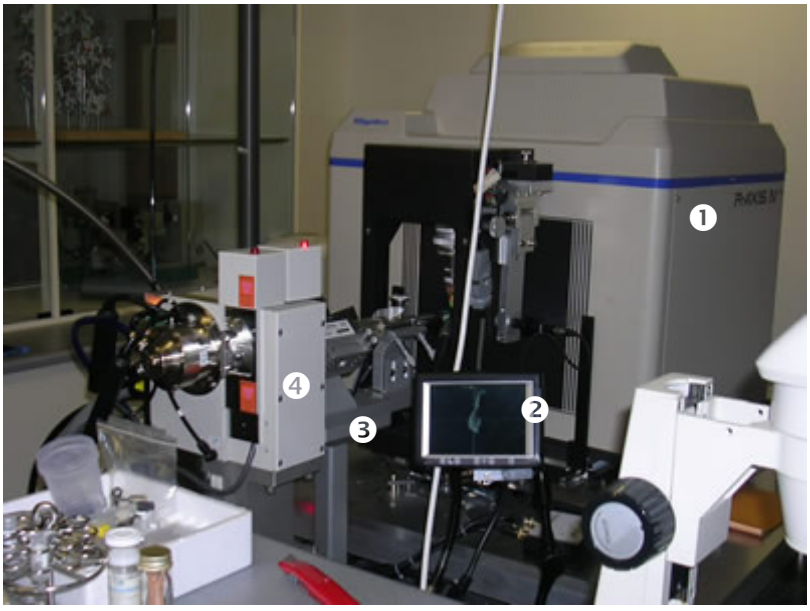


Fig. 4.2 Protein X-ray diffraction system. ❶R-AxisIV++ image-plate detector ❷Video microscope showing loop with frozen droplet containing a protein crystal ❸Monochromator (Osmic confocal blue)* ❹MicroMax007 X-ray generator.
*** The Osmic Confucal blue optic was replaced by the AxcoPX 70 capillary optic after this photo was taken.**

4.3 Data collection

β-Lg-ligand crystals were mounted onto the appropriate size of loop with cryoprotectant to protect the crystal from freezing damage. The crystal was quickly dipped into the cryoprotectant (20% glycerol in mother liquor).
The crystal symmetry, unit cell parameters, the crystal orientation and the resolution limit were determined before the data collection with the aid of CrystalClear. The data collection strategy was then derived to maximize the resolution as well as the completeness and redundancy of the data set. The collected data are used for calculating the electron density or Patterson maps via Fourier transforms. The intensity of a single reflection is a factor to a single term in the Fourier summation; hence, missing too many strong data can cause loss of significant features in the electron density or Patterson maps. Therefore, to achieve a high accuracy in reconstructing electron density map by Fourier transforms, a high degree of completeness and redundancy in diffraction data is necessary.

The ideal completeness of a data set is 100% with most reflections measured several times. However, it is not always possible to achieve 100% completeness. The target for completeness of data set is usually >99%, and if the completeness of the data set is less than 90%, the noise level of the Fourier transform appears to be high. As the redundancy increases, the quality of X-ray data also increases because the systematic and random error will be reduced. Assuming the data do not contain significant anomalous scattering, applying Friedel's Law (Fig. 4.3), the redundancy of space group P1 will be 2-fold. As the crystal is rotated during data collection redundancy is increased as a given reflection $I(h, k, l)$ is recorded in several different orientations. The

redundancy aimed for was 4.

$$\mathrm{I}(h,k,l) = \mathrm{I}(-h,-k,-l)$$

Fig. 4.3 Friedel's Law

The resolution of the X-ray diffraction data is usually measured in Ångstroms (Å), and the higher the resolution the more reliable is the information on a biological macromolecule that can be extracted. High-resolution structures diffract to resolution better than 1.4 Å. However, most biological structures do not diffract to such high resolution. Data to a resolution of 3.5 Å can still be useful, but beyond that at 4.0 Å, the side-chains will be questionable. Therefore, a very similar structure would be needed for comparison. The resolution aimed for was 2.5 Å.

The space group of the orthorhombic form of β-Lg crystals is $C222_1$ for both variants A and B, [74] β-Lg crystal structures in the orthorhombic form have been solved previously and there appears to be no ligand binding within the cavity or on the surface.

4.4 Data processing

X-ray data processing is a key stage for protein crystallography. The data processing procedures involving integration of the raw diffraction data and scaling are closely related to the final X-ray data quality. X-ray data can be processed, while the data set is still collecting and any errors in data collection strategy or machine problems can be corrected. Once the X-ray data are processed, phases can be obtained to generate electron density maps that can be solve the protein structure (Fig. 4.4). The procedure and programs used to process the data are illustrated in figure 4.5 and described in the next sections

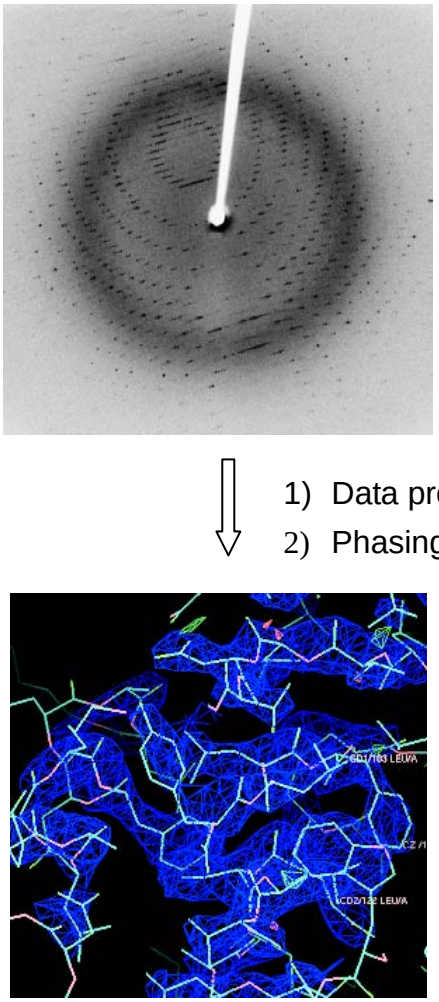


Fig. 4.4 Electron density map can be transformed from raw diffraction data. The stick diagram is an interpretation of the electron density map.

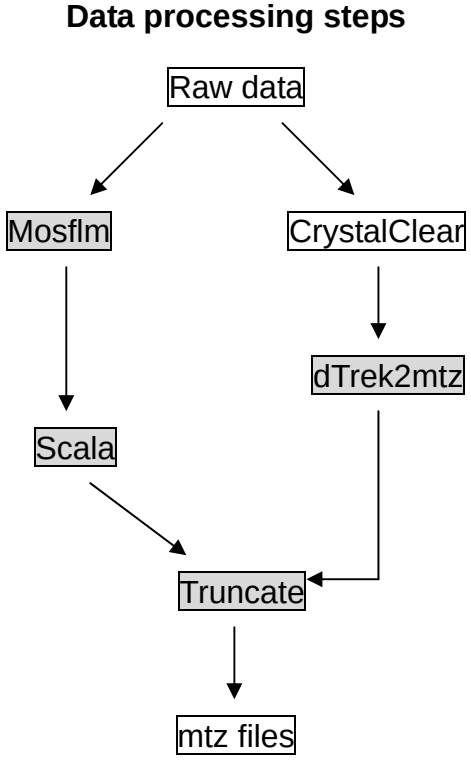


Fig. 4.5 Steps and programs used for X-ray diffraction data processing. Shaded boxes are programs from the CCP4 suite.

4.4.1. MOSFLM

MOSFLM, a program that is part of the CCP4i suite of protein crystallography programs, provides an alternative to CrystalClear for data processing. The data processing can be divided into three sections:

1) Determining the crystal orientation, unit cell parameters, possible crystal system and possible space group.
2) Generating the reflection list and integrating the images, which involves indexing the diffraction spots using the reflection list and removing the background X-ray scatter. As each reflection may be spread over several frames, integration involves creating a three-dimensional profile for each reflection.

Parameters such as crystal-to-detector distance, wavelength and direct beam

coordinates are needed in order to determine unit cell parameters and crystal orientation and to index reflections.

4.4.2 SCALA/ TRUNCATE

SCALA is one of the data harvesting program in CCP4 suite, and, after the integration of reflections, SCALA scales and merges the intensities. In addition to scaling and averaging the data, it converts the intensities to structure factors. The output data are in a standard MTZ format, which is then used to carry out molecular replacement or experimental phasing.

The overall R_{merge} from the scaling job should be very low for an excellent data set and is typically ~0.05. If the overall value of R_{merge} is greater than 0.10, this is cause for concern.

4.4.3 Molecular replacement

There are several methods for macromolecular phase determination, which includes molecular replacement (MR), multiple isomorphous replacement (MIR) and multi-wavelength anomalous diffraction (MAD). There are several advantages to using the molecular replacement method: the method only requires the measurement of one data set, it is a good method to use for closely related proteins such as mutants, and it is widely applicable as more protein structures become known. However, the initial structure is biased toward the model.

Molecular replacement can be used to derive phase information for a partially known structure. It also provides an initial starting model for structure refinement. The method focuses on the rotation function, which determines the

rotation matrix, and the translation function, which determines the translation vector that with the rotation matrix gives the position and orientation of the molecule. The rotation function allows the determination of search molecule orientation in absence of phase information. It is achieved by comparing the Patterson self-vectors (the interatomic vectors which can be formed within a molecule) of the known and unknown structures at different orientations of the search model in the experimental unit cell. The concept of the translation function is similar to rotation function. Once the orientation of the protein molecule in the unknown structure is determined, the correct position of the molecule can be found by translating the correctly orientated model within experimental unit cell to give the best agreement between calculated and observed structure factor amplitudes.

The initial phases are calculated based on the positions of all the atoms of the known model. As long as there is more than about 30% identity in sequence identity between the known model and the unknown structure, the phases are often sufficient to obtain a usable electron density map, which can be used to complete the structure.

MOLREP is an automatic molecular replacement program; it allows the number of molecules expected in the asymmetric unit to be determined. The best solution is generated as a PDB file. MOLREP is a powerful program, not only allowing processing for a standard molecular replacement but also being able to perform several other procedures including: multi-copy search for molecular replacement, fitting two models, using an electron density map as the search model, and determining search model orientation in an electron

density map for particular positions generated by a phased rotation function.

AMoRe is another program used for automatic molecular replacement and it is designed to locate model electron densities within the asymmetric unit of a crystal cell. The AMoRe procedure is divided into several steps: SORTING, TABLING, ROTING, TRAING, and FITTING. The X-ray data and a search probe first undergo preparation steps to format the data into suitable representation. The SORTING step transforms the X-ray data into AMoRe internal format. The TABLING step calculates the continuous Fourier coefficients to be used in the structure factor calculation. The rotation function is then calculated by ROTING using a search model. The ROTING step results in a list of possible solutions for the orientation of the molecules in the unit cell These solutions are the starting points to determine the position(s) of the molecule(s) in the unit cell using the translation function. The correct rotation function solution should also give a highest peak in the translation function generated in the TRAING step and the correlation coefficient should be as high as possible where the R-factor, the residual or agreement factor, should be as low as possible. In the last step, FITTING uses the obtained solutions to carry out rigid-body refinement. As the solutions to the rotation and translation functions are only approximate, the rigid-body refinement procedure is able to improve the correctness of the obtained solutions, and if the solutions are correct, the correlation coefficient should increase and R-factor should decrease. Validation procedures will be carried out to verify the correctness of the obtained molecular replacement solution, the final results should be consistent with the initial solutions. The best solution should give an uninterrupted electron density map that also shows clearly the differences

between the model structure and unknown structure, and there should be no overlaps of symmetry-related molecules.

4.4.4 Electron density map

The result of a crystallographic experiment is a map that shows the distribution of electrons in molecules, which is called the electron density map.

The Structure Factor $F(\boldsymbol{h})$ is proportional to the square root of the measured intensity $\mathrm{I}(\boldsymbol{h})$, and the structure factors can be deduced from the electron density $\rho(\boldsymbol{x})$ of the unit cell by the following integral expression:

$$F(\boldsymbol{h}) = \int_{cell} \rho(\boldsymbol{x}) e^{2\pi i \boldsymbol{h}\boldsymbol{x}} d\boldsymbol{x} = \sum_{\mathrm{j}=1}^{n} f_{\mathrm{j}}(\boldsymbol{h}) e^{2\pi i \boldsymbol{h}\boldsymbol{x}_{\mathrm{j}}}$$

The integral over continuous electron density can be replaced by a sum over discrete atoms, where $f_{\mathrm{j}}(\boldsymbol{h})$ is the scattering factor of atom j, ***h*** is a vector in reciprocal or diffraction space, often prescribed as (h,k,l) and **x** is a vector in real or crystal space, often presented by the components (x,y,z).

The expression is also known as a Fourier transform, and the structure factor $F(\boldsymbol{h})$ is the Fourier transform of the electron density $\rho(\boldsymbol{x})$ expressed in electrons per unit volume (e/Å^3). An electron density map can be calculated from the following expression, the inverse Fourier transform:

$$\rho(\boldsymbol{x}) = \frac{1}{V} \sum F(\boldsymbol{h}) e^{-2\pi i \boldsymbol{h}\boldsymbol{x}}$$

4.5 Structure solution and refinement

The goal of crystallographic refinement is to generate a model that has the best match to the experimental data and is also consistent with known stereochemistry. The accuracy of three-dimensional protein models can be improved by crystallographic refinement. The refinement process will minimize the difference between the experimentally observed diffraction amplitudes and those calculated from the hypothetical crystal model. The average difference is called the $\boldsymbol{R}$ factor (Fig. 4.6).

$$R = \frac{\sum_{\boldsymbol{h}} \left\| F_{obs}\left(\boldsymbol{h}\right) \right| - \left| F_{calc}\left(\boldsymbol{h}\right) \right\|}{\sum_{\boldsymbol{h}} F_{obs}\left(\boldsymbol{h}\right)}$$

Fig. 4.6 Crystallographic *R*-factor; it can vary between 0.00 and 0.59.

The $\boldsymbol{R}$ factor can vary between 0.00, which is exact agreement, and 0.59, which is total disagreement for a random arrangement of atoms. A well-determined or well-refined crystal structure for proteins will have an R-factor between 0.16 and 0.25.

There are several methods that can be used for phase generation of a protein structure and the molecular replacement method was used for phase generation for β-Lg-ligand complex. The electron density map then can be generated and displayed by COOT; [73] the structure is then refined by Refmac5. Refmac5 [75] is a program distributed as part of CCP4i and used for macromolecular structure refinement. Refmac5 is highly optimized and allows various tasks including: model idealization, rigid-body refinement, phased and

non-phased, restrained and unrestrained refinement. At the end of refinement Refmac5 then produces an improved or refined set of coordinates in the PDB output file and a new set of $F_{calc}(\boldsymbol{h})$ and phases in the MTZ output file. Refmac5 calculates a special weighted difference $W(\boldsymbol{h})|F_{obs}(\boldsymbol{h})|-|F_{calc}(\boldsymbol{h})|$ and $2W(\boldsymbol{h})|F_{obs}(\boldsymbol{h})|-|F_{calc}(\boldsymbol{h})|$ that are used as coefficients for electron density maps. The weighting $W(\boldsymbol{h})$ depends upon the estimated accuracy of the phases $\varphi_{calc}(\boldsymbol{h})$. The structure determination process is illustrated in figure 4.7.

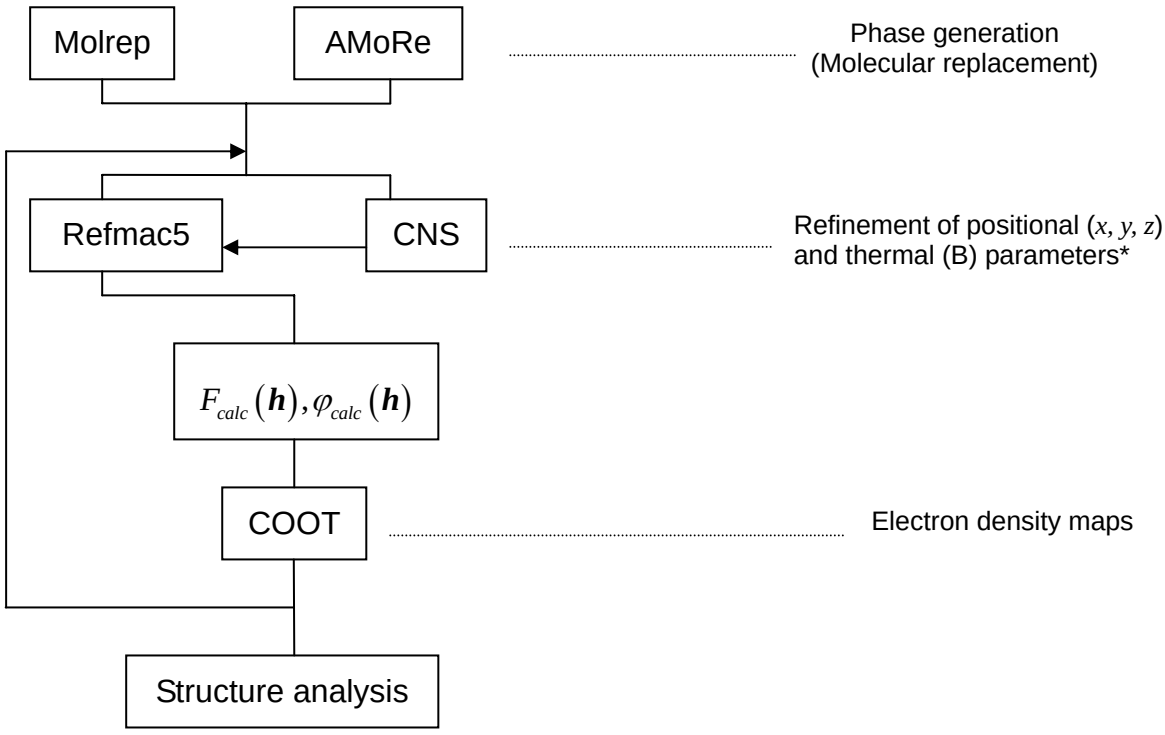


Fig. 4.7 Programs used to generate of $F_{calc}(\boldsymbol{h}), \varphi_{calc}(\boldsymbol{h})$ and to refine structural parameters. *Initially a rigid-body refinement is performed where only four parameters (*x, y, z* and B factor for the entire molecule) are refined. Program CNS offers different refinement protocols to Refmac5.

4.6 Data processing of β-Lg-SBT complex at pH 8.5

The first four images were collected for screening the crystal and to determine the orientation as well as optimized start rotation angle for data collection. The primitive hexagonal lattice was the suggested lattice type for β-Lg-SBT complex. 426 images were collected and automatically indexed by dTREK, and the images were processed with space group $P3_121$. The overall completeness was 94.2 % to 2.6 Å resolution with redundancy of 4, as detailed in Table 4.1.

Spacegroup	$P3_121$		
Unit cell dimensions (Á)	53.62	53.62	110.30
(deg)	90.00	90.00	120.00
Resolution range	35.52	2.60	(2.69 - 2.60)
Total number of reflections	23058		
Number of unique reflections	5695		
Average redundancy	4.05		(4.23)
% completeness	94.2		(97.5)
Rmerge	0.112		(0.477)
Reduced ChiSquared	0.76		(0.71)
Output <I/sigI>	7.3		(1.8)

Note: Values in ( ) are for the highest resolution shell.

Table 4.1 Summary of data collection statistics for the β-Lg-SBT complex.

4.7 Structural refinement of β-Lg-SBT complex at pH 8.5

The diffraction data set was reprocessed using Mosflm with space group $P3_221$, and molecular replacement refinement was performed using Molrep with the 12-bromodecanoic acid-β-Lg complex as the search model (1bso DOI:10.2210/pdb1bso/pdb); the waters, ligand and external loops were excluded from the search model.

Resolution limits (Á)	46.42	2.60
Number of used reflections	5434	
Percentage observed	94.10	
Percentage of R_{free} [a]	4.600	
Overall R factor	0.3608	
Free R factor	0.3785	
Overall weighted R factor	0.3625	
Free weighted R factor	0.3789	
Overall correlation coefficient	0.8360	
Free correlation coefficient	0.8133	
Cruickshank's DPI[b] for coordinate error	0.4339	
DPI based on free R factor	0.4552	
Overall figure of merit	0.6214	
ML based su[c] of positional parameters	0.4335	
ML based su of thermal parameters	21.1540	

Table 4.2 Data summary after molecular replacement, where R_{free} factor is an R-factor calculated in a partial data set (of ~5% of data set) that is not used in the refinement of a structure. [a] R_{free}, the Free R factor. [b] Cruickshank's dispersion precision indicator. [c] ML based su of positional parameters is the overall standard uncertainties of positional parameters based on the likelihood function.

Resolution limits (Á)	46.42	2.60
Number of used reflections	5434	
Percentage observed	94.1002	
Percentage of R_{free}	4.5662	
Overall R factor	0.3459	
Free R factor	0.3593	
Overall weighted R factor	0.3478	
Free weighted R factor	0.3554	
Overall correlation coefficient	0.8509	
Free correlation coefficient	0.8321	
Cruickshanks DPI for coordinate error	0.4159	
DPI based on free R factor	0.4321	
Overall figure of merit	0.6362	
ML based su of positional parameters	0.4074	
ML based su of thermal parameters	20.1611	

Table 4.3 Data summary following after rigid-body refinement

The initial R factor was 0.3606 with a correlation coefficient of 0.8133 (Table 4.2); the reflections used in the refinement were in the resolution range of 46.42 to 2.60 Å, and a total 5434 reflections were used. The overall R factor then was decreased to 0.3459 by rigid-body refinement using Refmac5 (Table 4.3).

The electron density maps (Fig. 4.8) show the electron density within the calyx, which is about the length of five carbon-carbon bonds. The structure requires further refinement to solve the structure of the molecule bound inside of the calyx.

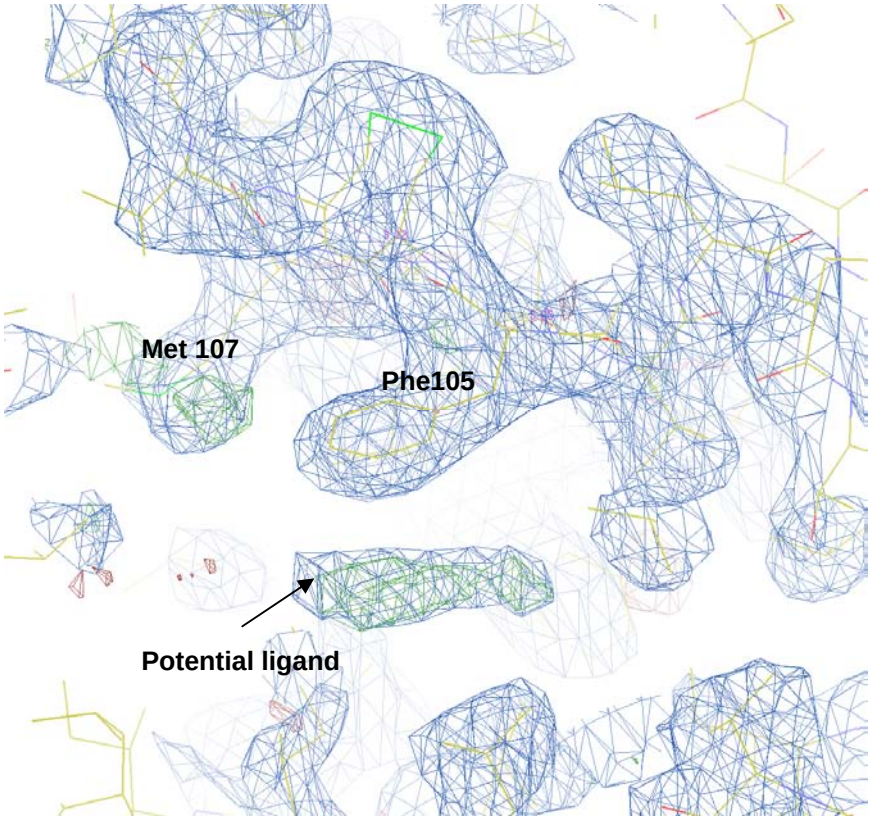


Fig. 4.8 The electron density maps of β-Lg after rigid-body refinement with Refmac5, contoured in blue at 1.1 σ for the 2Fo-Fc density map and at ±3.0 σ for the difference map (green is positive, red is negative electron density).

Restrained refinement was then carried out, resulting in slightly larger density within the calyx next to Phe105 and Met107 (Fig. 4.9) with R factor 0.293 and R_{free} 0.352.

The undefined residues (Leu1, Ile2, Pro50, Glu51, Gln155, Leu156, Glu158, Gln159, His161, Ile162), residues from CD loop (Trp61, Glu62, Asn63, Asp64,

Glu65), EF loop (Asp85, Ala86, Leu87, Asn88, Glu89), GH loop (Asn109, Ser110, Ala111, Glu112, Pro113, Glu114, Gln115) were left out in the initial refinements (rigid body refinement and restrained refinement). The electron density maps were then used to rebuild the absent residues and add them to the coordinate pdb file.

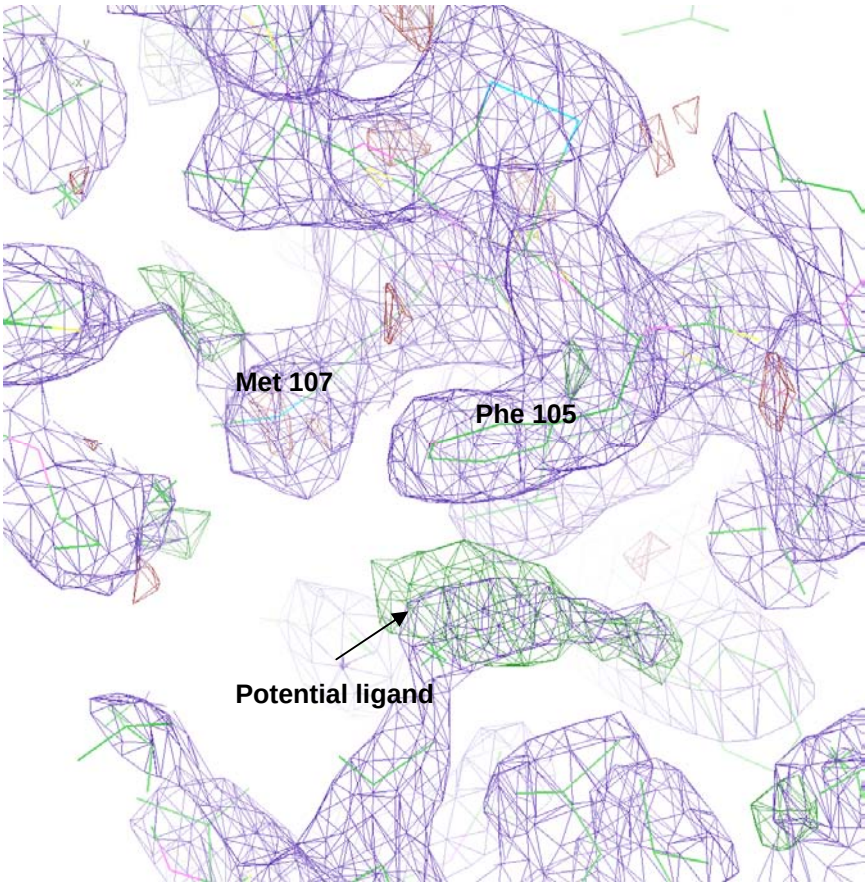


Fig. 4.9 The electron density maps of β-Lg after restrained refinement with Refmac5, contoured at 1.1 (purple) for 2Fo-Fc map and 3.0 σ for the difference electron density map (green is positive and red is negative electron density).

The structure of β-Lg was rebuilt by using COOT, and the side chains were refined by using Rotamers and Real Space Refine Zone functions. The Rotamers function uses a library of observed side-chain conformations from high-resolution structures. [76] The Real Space Refine Zone function allows real space refinement (automated interactive fitting of model to electron density) for

selected zone/residues, taking into account standard peptide geometry. The Regularize Zone function forces the molecule to conform to ideal values for peptide geometry by adjusting the selected zone, which is ≤ 20 residues. Residues Leu1, Ile2, Pro50, Glu51, the CD loop, Asp85, Asn109, Ser110, His161, and Ile162 were built and the first round of restrained refinement carried out. Restrained geometry and individual B-factor refinement was performed by crystallographic conjugate gradient minimization refinement, followed by calculation for $F_{calc}(\boldsymbol{h})$ and $\varphi_{calc}(\boldsymbol{h})$. The R factor was lowered to 0.2978 with R_{free} 0.3471. The generated electron density and difference map and coordinates were used to carry on further model rebuilding.

Residues 86 and 89, were built after the 1st round of refinement and a 2nd round of refinement was performed. Residue 111 was built after the 2nd round refinement which was then followed by the 3rd round of refinement. The 4th refinement was carried out after residue refinements by COOT. The summary of refinements us shown in tables 4.4 and 4.5.

	Initial rigid body refinement	Initial restrained refinement	1st round refinement
Initial R factor	0.3680	0.3392	0.2978
Initial R_{free}	0.3785	0.3383	0.3471
Final R factor	0.3459	0.293	0.2892
Final R_{free} value	0.3593	0.352	0.3488

Table 4.4 Summary of R and R_{free} values for initial and 1st round refinements.

	2nd round refinement	3rd round refinement	4th round refinement
Initial R factor	0.3170	0.3079	0.2978
Initial R_{free}	0.3366	0.3202	0.3471
Final R factor	0.2855	0.2878	0.2892
Final R_{free} value	0.3216	0.3209	0.3488

Table 4.5 Summary of R and R_{free} values for 2nd to 4th round refinements.

The remaining absent residues were added (residues from the CD loop) and a 5th refinement was performed. The R factor increased by 0.8% (from 0.2880 to 0.2956) but with slightly lower R_{free} value (0.3424 to 0.3404). The 3rd refinement resulted in the best R and R_{free} value. However, further refinements were carried out because the structural rebuilding was not completed (residues from CD loop were still excluded in the 3rd refinement). The chiral volumes for Cα (except glycine residues) and the Ramachandran plot were checked throughout the model rebuilding, and used to check the distribution of ϕ and φ angles for the polypeptide backbone. Further refinements (Table 4.6) were performed in order to generate better coordinates of β-Lg followed by addition of seven water molecules.

	5th round refinement	6th round refinement	7th round refinement
Initial R factor	0.2880	0.2856	0.2847
Initial R_{free}	0.3424	0.3430	0.3474
Final R factor	0.2835	0.2805	0.2805
Final R_{free} value	0.3404	0.3418	0.3539

Table 4.6 Summary of R and R_{free} values for refinements after all residues were added.

After the 7th refinement the electron density map still showed a clear linear density close to Phe105 and Met107 within the calyx. The electron density does not resemble the potential ligand SBT. The ^{1}H NMR spectrum confirmed that SBT decomposed within a week at room temperature. However, the decomposed compounds could not be easily identified from ^{1}H NMR. There are two proposed molecules that could potentially be the species bound to β-Lg within the calyx.

The possible molecules are the ring-opened derivative of SBT, *N*-ethyl-3-methyl-butanethioamide, and the hydrolyzed SBT, 2-methyl-1-butanol. However, the mass of the complex could not be checked by mass spectroscopy because the precipitant, ammonium sulfate used for crystallization is non-volatile.

Both small molecules were built using GaussView [77] and refined by Gaussian03; [78] the coordinate pdb file was converted into a CIF coordinate file by Sketcher (from CCP4i suite). The new CIF file contains idealized coordinates for each molecule, which allows them be used in model refinements in COOT (e.g. Real Space Refinement and Structure Regularization).

The difference map revealed a 5-carbon length density within the calyx and it remained clear through out the refinement procedures, which resulted in a poor electron density fit for the SBT and the *N*-ethyl-3-methyl-butanethioamide model (Fig. 4.10, 4.11).

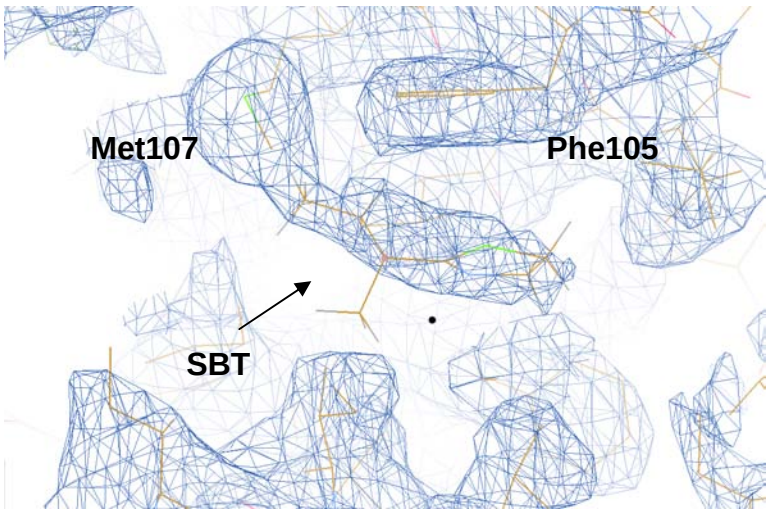


Fig. 4.10 Electron density map contoured at 1.3 σ with modeled SBT close to Met107 and Phe105. The butyl group of SBT does not fit into the more linear-like density.

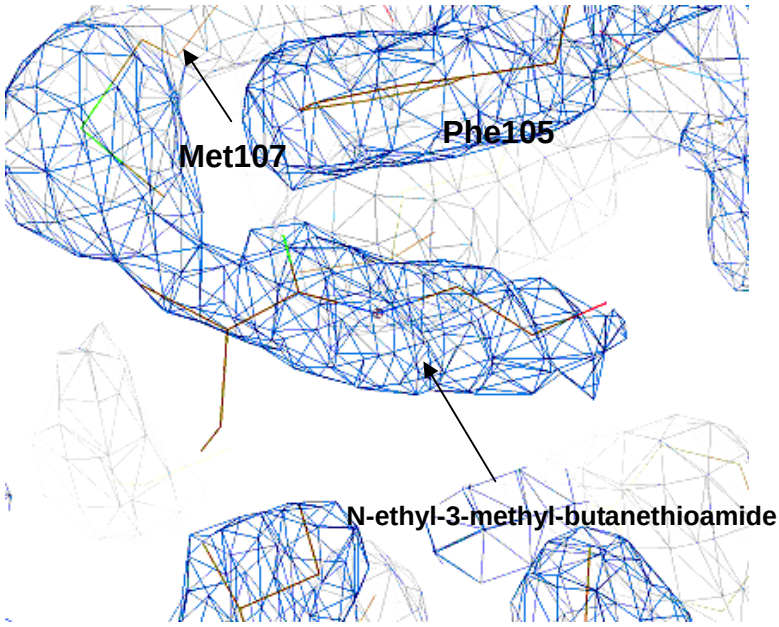


Fig. 4.11 Electron density map contoured at 1.3 σ with modeled *N*-ethyl-3-methyl -butanethioamide close to Met107 and Phe105.

The potential ligand may possibly be the hydrolyzed SBT species, 2-methyl-1-butanol judged by the size and the shape of the density. A model for 2-methyl-1-butanol binding inside the calyx was made. The hydrophobic end of the molecule is orientated towards to Met107, and the OH group is pointing towards the top of the calyx (Fig. 4.12, 4.13). Met107 has been

proposed as a preferential binding site for NMR studies [52] and both Met107 and Phe105 have been observed to be involved in hydrophobic interactions with the hydrocarbon chain of 12-bromodecanoic acid. [15] Therefore, the electron density map showed that the density was valid and not caused by structure bias.

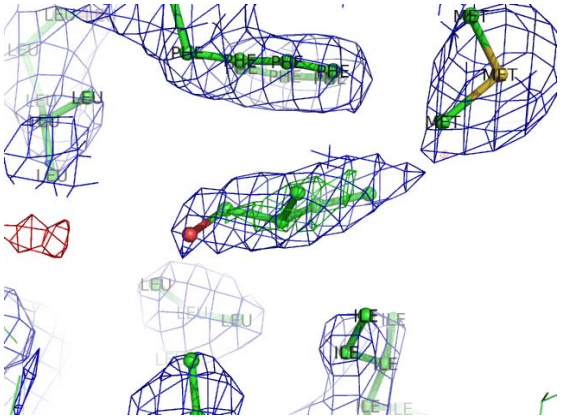


Fig. 4.12 Side-view of β-Lg electron density map (blue) contoured at 1.1 σ and difference electron density map (green is positive and red is negative density) contoured at ±3.0 σ with modeled potential ligand 2-methyl-1-butanol positioned within the calyx.

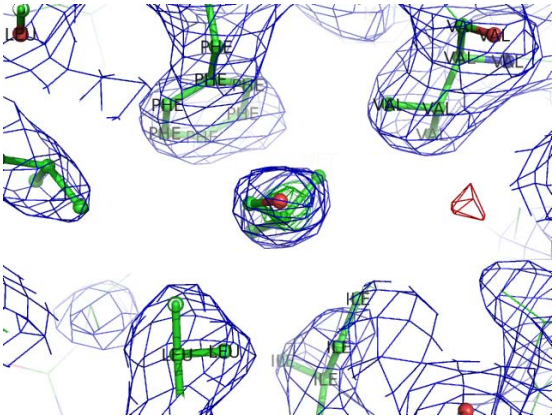


Fig. 4.13 Top-view of β-Lg electron density map (blue) contoured at 1.1 σ and difference electron density map (green is positive and red is negative density) contoured at ±3.0 σ with modeled potential ligand 2-methyl-1-butanol positioned within the calyx.

The composite omit map was calculated after the structural building and refinement to see if omitted features (~5% of the model is omitted for phasing) can be recovered. The recovered densities in the omitted region formed a composite omit map which then reduces model bias (Fig. 4.14, 4.15). The overlap of the electron density map and the composite omit map (Fig. 4.16) showed a very little difference (both contoured at 1.1 σ), which suggests that the structural rebuilding was reasonable and the density remained defined within the calyx. Sample contained SBT in phosphate buffer was monitored

over a time (2 weeks) by NMR, which showed SBT decomposed over time. Unfortunately the potential ligand could not be definitely characterized using technique such as Mass Spectroscopy, hence the model structure is not refined.

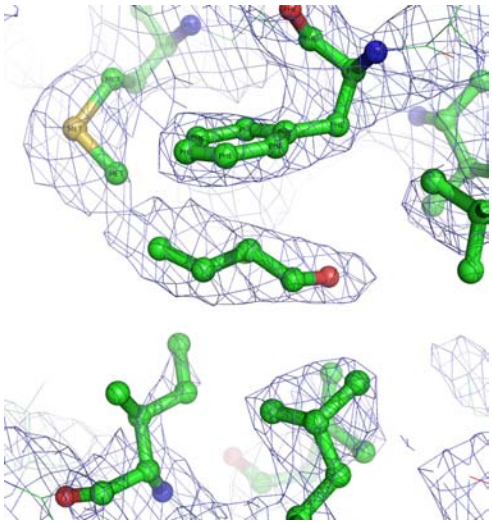


Fig. 4.14 The side view of β-Lg composite omit map countered at 3.5 σ with modeled potential ligand, 2-methyl-1-butanol, positioned within the calyx.

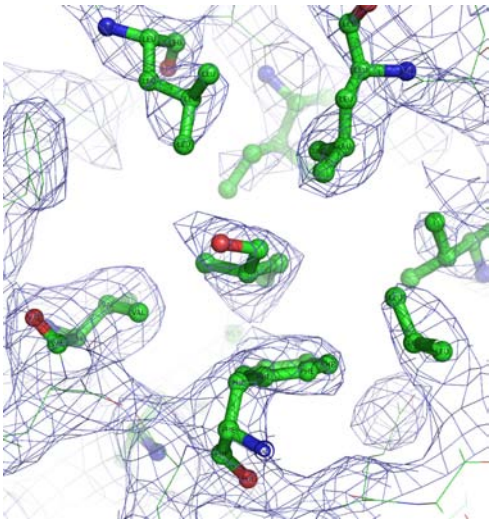


Fig. 4.15 The top view of β-Lg composite omit map countered at 3.5 σ with modeled potential ligand, 2-methyl-1-butanol, positioned within the calyx.

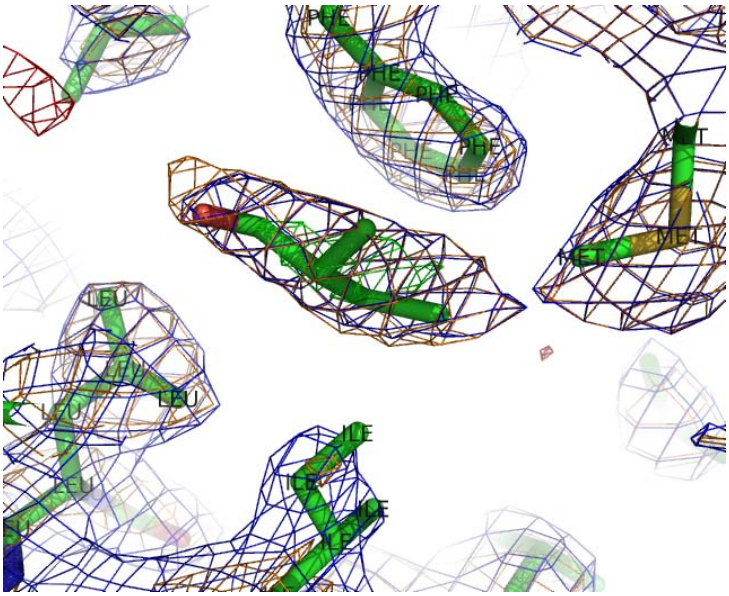


Fig. 4.16 Overlap of electron density map (blue), difference map (green for positive density and red for negative density) and composite omit map (orange) of β-Lg and modeled ligand.

4.8 Discussion and conclusions for X-ray structures

After several model rebuilding and refinement, the final R factor for the β-Lg structure was 0.281, and R_{free} was 0.354 for 5434 unique reflections to 2.6 Å resolution, the R and R_{free} values were higher than the published β-Lg-retinol structure (1gx8) by 8% and 5% with same space group to 2.4 Å resolution.
It is clear that there is a clear density locates close to Met107 and Phe105, which does not belong to either residue sidechain. Therefore, it is possible that β-Lg bound to ligand at Met107 and Phe105, where Met107 has been reported as a binding site for ligands. The final R and R_{free} values indicated that the structure fitted the published structure (1bso) relatively well without the ligand, and the poor electron density fit for SBT model suggested that the molecule bound to β-Lg is not the SBT but a more linear molecule.

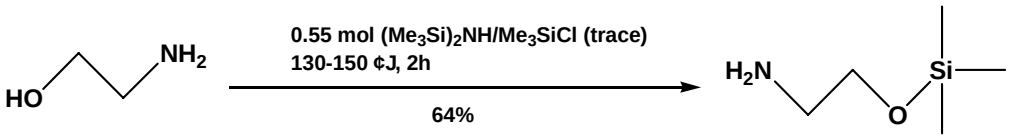
β-Lg was crystallized with all ligands at both neutral pH (pH 6.5 to 8.5). However, the crystallization at the condition at which NMR studies were performed (pH 2.0) was unsuccessful.

The potential ligand SBT decomposed into two possible molecules, the ring-opened derivative of SBT, *N*-ethyl-3-methyl-butanethioamide, and the hydrolyzed SBT, 2-methyl-1-butanol. However, the potential ligand bound to β-Lg within the calyx can not be characterized due to non-volatile precipitant was used for protein-ligand crystallization. The poor electron density fit for *N*-ethyl-3-methyl-butanethioamide indicated that 2-methyl-1-butanol is more likely to be the potential ligand. The electron density map also indicated that there was no ligand found outside of the calyx.

Chapter 5 Experimental details

5.1 Synthesis of 2-(trimethylsilyloxy)ethylamine

Ethanolamine (Unilab, 99.0% min.) was used to synthesize 2-(trimethylsilyloxy)ethylamine. The hydroxyl functional group of ethanolamine was protected by silylation with hexamethyldisilazane following an efficient scheme shown below: [79]



The reflux apparatus was dried and set up. Then the system was evacuated, followed by purging with nitrogen gas. 1.00 mol of ethanolamine (61.08 g), 0.55 mol of hexamethyldisilazane (88.76 g, Sigma), and a trace of chloromethylsilane (5 drops, Merck) were injected into the three-neck flask. The reaction mixture was refluxed at 150℃ under nitrogen for 2 hours. The by-product, ammonia, was then removed by fractional distillation with a 40 cm column to give (85 g, 64% yield based on ethanolamine). The product was characterized by ^{1}H NMR spectroscopy (Fig. 5.1). The ^{1}H NMR of 2-(trimethylsilyloxy)ethylamine showed no trace of ammonia remaining and the chemical shifts are similar to the literature values (Table 5.1).

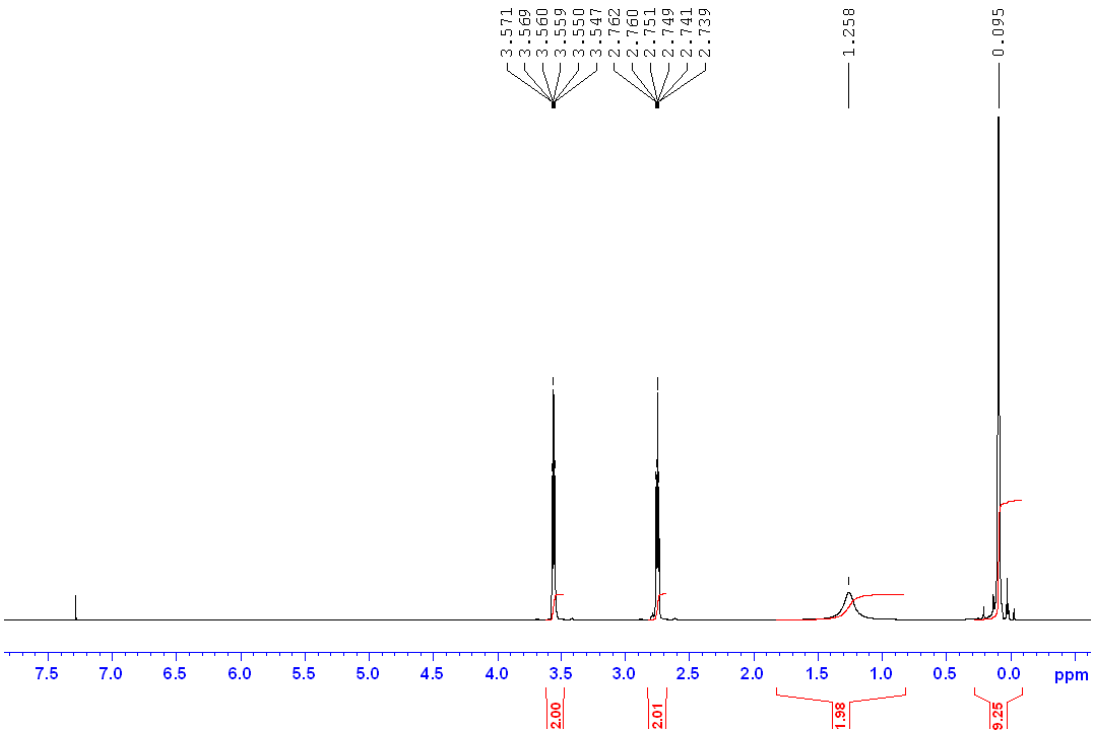


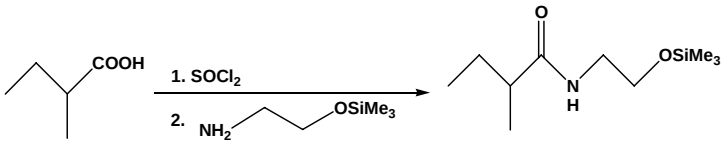
Fig. 5.1 ^{1}H NMR spectrum of 2-(trimethylsilyloxy)ethylamine.

	Bp (℃)	^{1}H NMR ($CDCl_3$)
Lit.	134-134.5	0.0 (s, 9H); 1.07 (s, 2H); 2.71 (t, 2H); 3.57 (t, 2H)
Exp.	134	0.095 (s, 9H, -Si($C\underline{H_3}$)$_3$); 1.258 (s, 2H, -$N\underline{H_2}$); 2.751 (t, 2H, NH_2-$C\underline{H_2}$-CH_2); 3.56 (t, 2H, NH_2-CH_2-$C\underline{H_2}$)

Table 5.1 Literature and experimental values of ^{1}H NMR chemical shifts for 2-(trimethylsilyloxy)ethylamine.

5.2 Synthesis of *N*-(2-trimethylsilyloxyethyl)-2-methylbutanamide

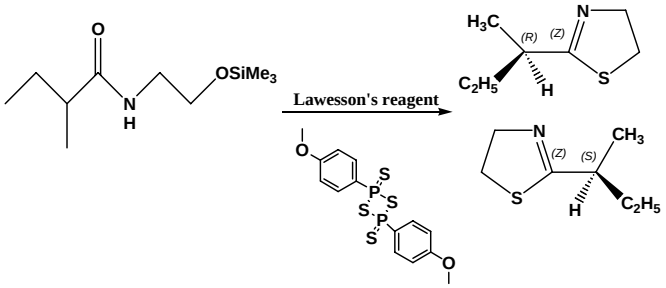
N-(2-trimethylsilyloxyethyl)-2-methylbutanamide is synthesized from 2-methyl-butanoic acid (Aldrich, >98%) following the scheme below: [55]



2-Methylbutanoic acid (4.57 mL, 41.9 mmol) and thionyl chloride (3.06 mL, 42.6 mmol; Merck, ≥ 99%) were reacted under reflux in 17 mL of chloroform for 1 hour. A solution of 2-(trimethylsilyloxy)ethylamine (5.58 g) and trimethylamine (5.83 mL, 41.9 mmol; Merck >99%) in 8 mL chloroform was added to the cooled solution dropwise at 5℃. The chloroform was removed by rotary evaporation and the residue then taken up in diethyl ether followed by filtration. The filtrate was dried with $MgSO_4$, then concentrated by rotary evaporation to yield 7.51 g of *N*-(2-trimethylsilyloxyethyl)-2-methylbutanamide (34.6 mmol, 83%).

5.3 Synthesis of 2-*sec*-butyl-4,5-dihydrothiazole

Synthesis of SBT was accomplished by means of Lawesson's reagent. [55]



5.98 g of Lawesson's reagent (ACROS ORGANIC, 99%) was added to *N*-(2-trimethylsilyloxyethyl)-2-methylbutanamide solution (6.42 g, 29.6 mmol) with 15 mL of toluene, then refluxed under nitrogen at 100℃ for 3.5 hours.

The cooled mixture was then partitioned against diethyl ether and aqueous sodium bicarbonate. The organic phase was dried with $MgSO_4$ and then subjected to rotary evaporation followed by multiple vacuum fractional distillations at 10 torr to yield 0.8 g (5.6 mmol, 18.7%) of product.

5.4 Reduction of 3-methyl-2-butenal

The $NaBH_4$-Alox was prepared from addition of a solution of sodium borohydride (Riedel-de Haën) in water (1 g/mL) to 10 g of alumina (Merck, activity I, 70-230 mesh) with stirring until the white powder was obtained. The powder was kept under vacuum overnight; the solid may then stand at ambient pressure and temperature without any particular precaution.

A solution of 3MB (99.0 µL, 1 mmol) in 1 mL of ether was reduced by adding it to the mixture of $NaBH_4$-Alox (1 g) and diethyl ether (3 mL) with stirring. The solid material was filtered off after the mixture was stirred for 5 minutes and washed with diethyl ether (5 mL x 3). The filtrates were combined then left at room temperature until the diethyl ether evaporated.

5.5 Protein purification

The β-Lg samples for NMR studies were made up in 10 mM phosphate buffer at a concentration of 20 mg/mL. The sample solution was transferred into a concentrator (VivaScience Vivaspin 20 mL), where the salts and contaminants can be removed by centrifugation in a swing bucket rotor at 36 k rpm for 10 minutes followed by re-dilution with clean buffer solution. The process was repeated three times to ensure the ammonium ions were removed completely. The concentration of the desalted sample solution was determined by uv-visible spectrometry (Cary 100 Bio) at 280 nm where the absorbance is 1

when the concentration of β-Lg is 1 mg/mL.

5.6 Circular dichroism study of β-Lg and 2-*sec*-butyl-4,5-dihydrothiazole

The β-Lg sample solution (2 mg/mL) was made up in 10 mM phosphate buffer at pH 7.4 then filtered with 2 μm filter. The ligand -*sec*-butyl-4,5-dihydrothiazole (SBT) was dissolved in phosphate buffer to 16 mg/mL (0.112 M) and filtered. The spectrum (average of 7 scans) of the buffer solution was recorded for base-line correction purposes, and then a reference spectrum for β-Lg solution was recorded (average of 5 scans) from 400-250 nm. The 0.112 M SBT solution (2 μL) was then added to give a 1 β-Lg: 2 SBT molar ratio and the spectrum was recorded over the range 250-340 nm (average of 5 scans). The CD signals of the buffer solution were subtracted from the β-Lg solution data and from the β-Lg-SBT solution data. The spectra of β-Lg reference solution and β-Lg-SBT solution were then plotted after noise subtraction. Time and access problems prevented titration of additional aliquots of SBT.

5.7 Circular dichroism study of β-Lg and 3-methyl-2-butenal

β-Lg sample solution (2 mg/mL) was prepared in 10 mM phosphate buffer at pH 7.4 and filtered. The 3-methyl-2-butenal (3MB) was dissolved to 10.1 mg/mL (0.10 M) with the phosphate buffer then filtered. The spectrum (average of 7 scans) of the buffer solution was recorded for base-line correction. The spectrum (average of 5 scans) of the β-Lg reference solution was first recorded. The spectrum was then recorded following addition of 2 μL of 0.10 M 3MB (a 1 β-Lg: 2 3MB ratio). A further 2 μL of 3MB was added and this spectrum recorded to determine the relationship between ligand concentration and protein-ligand interaction.

5.8 Ammonium sulfate screening for β-Lg and SBT

Bovine β-Lg A (20 mg/mL) was saturated with SBT (16 mg) in the molar ratio of 1:100. The crystals were grown from 2 µL protein solution and 2 µL precipitant solution using the hanging drop method. The concentrations of the precipitant solutions were 2.2 M to 2.7 M ammonium sulfate in the 6x4 screen matrix (Fig. 2.15) with four tris-HCl buffers (at concentration of 0.20 M) at pH 6.1, 6.9, 7.7, and 8.5. All crystallizations for other chosen ligands (Chapter 1) followed the same method with alternation in ligand concentration.

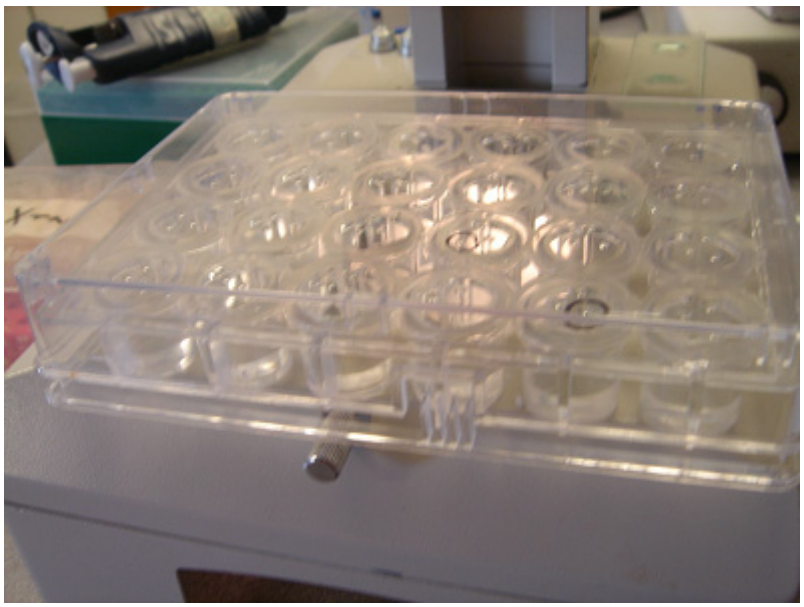


Fig. 5.2 A 6 x 4 crystal screen matrix.

5.9 Ammonium sulfate screening for β-Lg and 3-methyl-2-buten-1-ol

Bovine β-Lg A (20 mg/mL) was saturated with 3-methyl-2-buten-1-ol (8.6 mg) in the molar ratio of 1:100. The crystals were grown from 2 µL protein solution and 2 µL precipitant solution using the hanging drop method. The concentrations of the precipitant solutions were 2.2 M to 2.7 M ammonium sulfate in the 6x4 screen matrix with four tris-HCl buffers (at concentration of 0.20 M) at pH 6.1, 6.9, 7.7, and 8.5.

Chapter 6 Conclusions and perspectives

The study of β-Lg binding pheromones and milk volatile compounds has revealed information about the preferential binding site of β-Lg. Several small volatile molecules including 3-methyl-2-butenal, phenylacetic acid and 2-*sec*-4,5-dihydrothiazole were chosen for NMR investigation. The NMR investigation at pH 2.0 has shown that 3-methyl-2-butenal may bind to the Tyr102, which is located at the bottom of the calyx, and that phenylacetic acid appears to be preferentially bound to Try102 and Asp129 at an external hydrophobic surface area. Therefore, the preferential binding site for small volatile molecules is likely to be the external hydrophobic surface area instead of inside the calyx, consistent with the behavior of the flavor molecule β-ionone at pH 2.0. [52]

The X-ray crystallographic studies showed that the central binding site, the calyx seems to accommodate linear molecules such as fatty acids and retinol. The crystallographic studies showed no success to crystallize β-Lg at low pH, the condition used for NMR studies. However, β-Lg crystallized in trigonal form for most of the chosen ligands at pH 8.5. Several volatile molecules which present in bovine milk were chosen as ligands and, the proposed ligand, 2-*sec*-4,5-dihydrothiazole has decomposed into a linear molecule that has a length of about 5 carbon atoms, which may be 2-methyl-1-butanol. It is clear that the calyx is very hydrophobic, and would be the preferential binding site for larger linear hydrophobic ligands and 2-methyl-1-butanol is the potential ligand. The potential ligand was not characterized, but it is likely to be the

hydrolyzed SBT, 2-methyl-1-butanol. Further crystallization for fully characterized 2-methyl-1-butanol will required confirming the model structure.

Both NMR and crystallographic studies showed the possible interactions between volatile molecules and β-Lg indicated the proposed hypothesis that β-Lg acts as a pheromone binding protein similar to MUP may be true, but interactions between volatile molecules and β-Lg will require further research.

References

1. Hambling, S. G.; McAlpine, A. S.; Sawyer, L. (1992) β-Lactoglobulin. *Advanced dairy chemistry: Volume 1: proteins*, 2nd Ed. Elsevier Applied Science: London, 141.
2. Bell, K.; McKenzie, H. A. (1964). β-Lactoglobulin. *Nature*, 204, 1275.
3. McKenzie, H. A.; Sawyer, W. H. (1967) Effect of pH on β-Lactoglobulins. *Nature*, 214, 1101
4. Palmer, A. H. (1934). The preparation of a crystalline globulin from the albumin fraction of cow milk. *Journal of biological chemistry*, 104, 359.
5. Creamer, L. K.; MacGibbon, A. K. H. (1996) Some recent advances in basic chemistry of milk proteins and lipids. *International Dairy Journal*, 6, 539.
6. Cowan, S. W.; Newcomer, M. E.; Jones, T. A. (1990). Crystallographuc refinement of human serum retinol binding protein at 2 Å. resolution. *Proteins: Structure, Function, and Genetics*, 8, 144.
7. 頁: 91

   Zanotti, G.; Malpeli, G.; Berni, R. (1993). The interaction of *N*-ethylretinamide with plasma retinol-binding protein (RBP) and crystal structure of the retinoid-RBP complext at 1.9 Å resolution. *Journal of Biological Chemistry*, 268, 24873.
8. Huber, R.; Schneider, M.; Epp, O.; Mayr, I.; Messerschmidt, A.; Pflugrath, J.; Kayser, H.; (1987). Crystallization, crystal structure analysis and preliminary molecular model of the bilin binding protein from the insect Pieris brassicae. *Journal of Molecular Biology*, 195, 423.

9. Monaco, H.; Zanotti, G. (1992). Three-dimensional structure and active site of three hydrophonic molecule-binding proteins with significant amino acid sequence similarity. *Biopolymers*, 32, 457.
10. Bocskei, Z.; Groom, C. R.; Flower, D. R.; Wright, C. E.; Phillips, S. E. V.; Cavaggioni, A.; Findlay, J. B. C.; North, A. C. T. (1992). Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature*, 360, 186.
11. Green, D. W.; Aschaffenburg, R.; Camerman, A.; Coppola, J. C.; Dunnill, P.; Simmons, R. M.; Komorowski, E. S.; Sawyer, L.; Turner, E. M. C.; Woods, K. F. (1979). Structure of bovine β-lactoglobulin at 6 Å resolution. *Journal of Molecular Biology*, 131, 375.
12. Brownlow, S.; Cabral, J. H. M; Cooper, R.; Flower, D. R.; Yewdall, S. J.; Polikarpov, I.; North, A. C. T.; Sawyer, L. (1997). Bovine β-lactoglobulin at 1.8 Å resolution – still an enigmatic lipocalin. *Structure*, 5, 481..
13. Qin, B. Y., Bewley, M. C.; Creamer, L. K.; Baker, E. N.; Jameson, J. B. (1998). Structural basis of the Tanford transition of bovine β-lactoglobulin. *Biochemistry*, 37, 14014.
14. Bell, K.; McKenzie, H. A. (1967). Whey proteins of ovine milk. β-Lactoglobulins A and B. *Biochimica et Biophysica Acta, Protein Structure*, 147, 123.
15. Qin, B. Y.; Creamer, L. K.; Baker, E. N.; Jameson, G. B. (1998). 12-Bromododecanoic acid binds inside the calyx of bovine β-lactoglobulin. *FEBS Letters*, 438, 272.
16. Kontopidis, G.; Holt, C.; Sawyer, L (2004). B-lactoglobulin: binding properties, structure, and function. *Journal of Dairy Science*, 87, 785.

17. Fogolari, F.; Ragona, L.; Zetta, L.; Romagnoli, S.; De Kruif, K. G.; Molinari, H. (1998). Monomeric bovine β-lactoglobulin adopts a β-barrel fold at pH 2. *FEBS Letters*, 436, 149.
18. Uhrinova S; Smith M H; Jameson G B; Uhrin D; Sawyer L; Barlow, P. N. (2000) Structural changes accompanying pH-induced dissociation of the beta-lactoglobulin dimer. *Biochemistry*, 39, 3565.
19. Bell, K., McKenzie, H. A.; Shaw, D. C. (1981). Bovine β-lactoglobulin E, F and G of Bali (banteng) cattle, Bos (Bibos) javanicus. *Journal of Biological Science*, 34, 133.
20. Bewley, M. C.; Qin, B. Y.; Jameson, J. B.; Sawyer, L.; Baker, E. N. (1997). Bovine β-lactoglobulin and its variants: a three-dimensional structural prespective. *Milk protein polymorphism, IDF special issue*, 9702, 100.
21. Papiz, M. Z.; Sawyer, L.; Eliopoulos, E. E.; North, A. C; Findlay, J. B.; Sivaprasadarao, R.; Jones, T. A.; Newcomer, M. E.; Kraulis, P. J. (1986). The structure of β-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature*, 324, 383.
22. McMeekin, T. L.; Polis, B. D.; DellaMonica, E. S.; Custer, J. H. (1948). Heterogeneity of crystalline β-lactoglobulin. *Journal of American Chemical Society*, 70, 881.
23. Seibles, T. S. (1969). Interaction of dodecyl sulfate with native and modified β-lactoglobulin. *Biochemistry*, 8, 2949.
24. Spector, A. A.; Fletcher, J. E. (1970). Binding of long chain fatty acid to β-lactoglobulin. *Lipids*, 5, 403.
25. Jones, M. N.; Wilkinson, A. (1976). Interaction between β-lactoglobulin

and sodium n-dodecyl sulfate. *Biochemical Journal*, 153, 713.

26. Pérez, M. D.; Diaz de Villegas, C.; Sanchez, L.; Aranda, P.; Ena, J. M.; Calvo, M. (1989). Interaction between fatty acid with β-lactoglobulin and albumin from ruminant milk. *Journal of Biochemistry* (Tokyo, Japan). 106, 1094.
27. Frapin, D.; Dufour, E.; Haertlé, T. (1993). Probing the fatty acid binding site of β-lactoglobulin. *Journal of Protein Chemistry*, 12. 443.
28. Creamer, L. K. (1995). Effect of sodium dodecul sulfate and palmitic acid on the equilibrium unfolding of bovine β-lactoglobulin. *Biochemistry*, 34, 7170.
29. Puyol, P.; Pérez, M. D.; Ena, J. M.; Calvo, M. (1991). Interaction of bovine β-lactoglobulin and other bovine and human whey proteins with retinol and fatty acid. *Agricultural and Biological Chemistry*, 55, 2515.
30. Marden, M. C.; Dufour, E.; Christova, P.; Huang, Y.; LeclercL'Hostis, E.; Haertlé, T. (1994). Binding of heme-CO to bovine and porcine β-lactoglobulin. *Archives of Biochemistry and Biophysics*, 311, 258.
31. Marden, M. C.; Dufour, E.; Christova, P.; Huang, Y.; LeclercL'Hostis, E.; Haertlé, T. (1994). Binding of heme-CO to bovine and porcine β-lactoglobulin. *Archives of Biochemistry and Biophysics*, 311, 258.
32. Robilliard, K. A. Jr.; Wishnia, A. (1972). Aromatic hydrophobes and β-lactoglobulin A. Thermodynamic of binding. *Biochemistry*, 11, 3835.
33. Robilliard, K. A. Jr.; Wishnia, A. (1972). Aromatic hydrophobes and β-lactoglobulin A. Kinetics of binding by nuclear magnetic resonance. *Biochemistry*, 11, 3841.
34. Dufour, E.; Roger, P.; Haertlé, T. (1992). Binding of benzo(a)pyrene, ellipticine, and cis-parinaric acid to β-lactoglobulin: influence of protein

modifications. *Journal of Protein Chemistry*, 11, 645.

35. O'Neill, T. E.; Kinsella, J. E. (1987). Binding of alkanone flavors to β-lactoglobulin: effect of conformational and chemical modification. *Journal of Agriculture and Food Chemistry*, 35, 770.
36. Dufour, E.; Haertlé, T. (1990). Binding affinities of β-ionone and related flavor compounds to β-lactoglobulin: effects of chemical modifications. *Journal of Agriculture and Food Chemistry*, 38, 1691
37. Monaco, H. L.; Zanotti, G.; Spadon, P.; Bolognesi, M.; Sawyer, L.; Eliopoulos, E. E. (1987). Crystal structure of the trigonal form of bovine β-lactoglobulin and of its complex with retinol at 2.5 Å resolution. *Journal of Molecular Biology*, 197, 695.
38. Lange, D. C.; Kothari, R.; Patel, R. C.; Patel, S. C. (1998). Retinol and retinoic acid bind to a surface cleft in bovine β-lactoglobulin: a method of binding site determination using fluorescence resonance energy transfer. *Biophysical Chemistry*, 74, 45.
39. Futterman, S.; Heller, J. (1972). Enhancement of fluorescence and the decreased susceptibility to enzymic oxidation of retinol complexed with bovine serum albumin, β-lactoglobulin, and the retinol-binding protein of human plasma. *Journal of Biological Chemistry*, 274, 5168.
40. Perez, M. D.; Calvo, M. (1995) Interaction of β-lactoglobulin with retinol and fatty acids and its role as a possible biological function for this protein: a review. *Journal of dairy science*, 78, 978.
41. A) Sawyer, L.; Kontopidis, G. (2000). The core lipocalin, bovine β-lactoglobulin. *Biochimica et Biophysica Acta*, 1482, 136. B) Kontopidis, G.; Holt, C.; Sawyer, L. (2004). β-lactoglobulin: binding

properties, structure, and function. *Journal of Dairy Science*, 87, 785.

42. Ragona, L.; Fogolari, F.; Catalano, M.; Ugolini, R.; Zett, L.; Molinari, H. (2003). EF loop conformational change triggers ligand binding in β-lactoglobulins. *Journal of Biological Chemistry*, 278, 38840.

43. O'Leeffe, E. T.; Mordick, T.; Bell, J. E. (1980). Bovine galactosyltransferase: interaction with α-lactalbumin and the role of α-lactalbumin in lactose synthase. *Biochemistry*, 19, 4962.

44. Zidek, L.; Stone, M. K.; Lato, S. M.; Pagle, M. D.; Miao, Z.; Ellignton, A. D.; Novotny, M. V. (1999). NMR mapping of the recombinant mouse major urinary protein I binding site occupied by the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Biochemistry*, 38, 9850.

45. Flower, D. R. (1996). The lipocalin protein family: structure and function. *Biochemistry*, 318, 1.

46. Timm, D. E.; Baker, L. J.; Mueller, Zidek, L.; H.; Novotny, M. V.; (2001). Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Science*, 10, 997.

47. Böcskei, Z.; Findlay, J. B. C.; North, A. C.; Phillips S. E.; Somers, W. S.; Wright, C. E.; Lonetti, C.; Tirindelli, R.; Cavaggioni, A. (1991). Crystallization of and preliminary X-ray data for the mouse major urinary protein and rat α-2u globulin. *Journal of Molecular Biology*, 218, 699.

48. Schwende, F. J.; Jorgenson, J. W.; Novotny, M. (1984). Possible chemical basis for histocompatibility-related mating preference in mice. *Journal of Chemical Ecology*, 10, 1603.

49. Krizova, H.; Zidek, L.; Stone, M. J.; Novotny, M. V.; Sklenar, V. (2004). Temperature-dependent spectral density analysis applied to monitoring

backbone dynamics of major urinary protein-I complexed with the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Journal of Biolmolecular NMR*, 28, 369.

50. Schaal, B.; Coureaud, G.; Langlois, D.; Ginies, C.; Semon, E.; Perrier, G. (2003). Chemical and behavioral characterization of the rabbit mammary pheromone. *Nature*, 424, 68.
51. Colahan-Sederstrom, P; Peterson, D. (2005). Inhibition of key aroma compound generated during ultrahigh-temperature processing of bovine milk via Epicatechi addition. *Journal of Agricultural and Food Chemistry*, 53, 398.
52. Lübke, M.; Guichard, E.; Tromelin, A.; Quéré, J. L. L. (2002). Nuclear magnetic resonance spectroscopic study of β-lactoglobulin interactions with two flavor compounds, γ-decalactone and β-ionone. *Journal of Agricultural and Food Chemistry*, 50, 7094.
53. Jung, D.; Ebeler, S. E. (2003). Investigation of binding behavior of α- and β-ionone to β-lactoglobulin at different pH values using a diffusion-based NOE pumping technique. *Journal of Agricultural and Food Chemistry*, 51, 1988.
54. Tromelin, A.; Guichard, E. (2006). Interaction between flavor compounds and β-lactoglobulin: approach by NMR and 2D/ 3D-QSAR studies of ligands. *Flavour and Fragrance Journal*, 21, 13.
55. Novotny, M. V.; Xie, T. -M.; Harvey, S.; Wiesler, D.; Jemiolo, B.; Carmack, M. (1995). Stereoselectivity in mammalian chemical communication: male mouse pheromones. *Experientia*, 51, 738
56. Braun, S.; Halinowski, H.-O.; Berger, S. (1996) *100 and more basic NMR experiments*, VCH: New York.

57. Friebolin, H. (1991). *Basic one- and two-dimensional NMR spectroscopy*, VCH: New York.
58. Uhrinova, S.; Uhrin, D.; Denton, H.; Smith, M.; Sawyer, L.; Barlow, P. N. (1998). Complete assignment of ^{1}H ^{13}C and 15N chemical shifts for bovine β-lactoglobulin: seconardy structure and topology of the native state is retained in a partially unfolded form. *Journal of Biomolecular NMR*, 12. 89.
59. Claridge, T. D. W. (1999) *High-Resolution NMR techniques in Organic Chemistry*, Pergamon: Oxford.
60. Müller, P. (EDT);Herbst-Irmer, R.; Spek, A.; Schneider, T. (2006). *Crystal structure refinement A crystallographer's guide to SHELXL*, Oxford University Press: New York.
61. Rhodes, G. (2006). *Crystallpgraphy made crystal clear: a guide for users of macromolecular midels*, Academic Press: Boston.
62. Landor, S. R.; Landor, P. D.;Fomum, Z. T.; Mbafor, J. T,; Mpango, G. W. B. (1984). The addition of thiols to allenyl- and phenylpropynylnitrile and the formation of thiazolines and benzothiazoles. *Tetrahedron*, 40, 2141.
63. Lehman-McKeeman, L. D.; Caudill, D.; Eddy, C.; Rodriguez, P. A. (1998). 2-sec-butyl-4,5-dihydrothiazole is a ligand for mouse urinary protein and rat α-2u0globulin: physiological and toxicological relevance. *Toxicology and Applied Pharmacology*, 149, 32.
64. Van Boekel, M. (1998). Effect of heating on Maillard reactions in milk. *Food Chemistry*, 70, 403.
65. Santaniello, E.; Ponti, F.; Manzocchi, A. (1978). Reduction of carbonyl compound by sodium borohydride adsorbed on alumina. *Synthesis*, 12,

891.

66. Brittain, H. G.; Purdie, N. (1994). *Analytical applications of circular dichroism*, Elsevier: New York

67. Zsila, F. (2003). A new ligand for an old lipocalin: induced circular dichroism spectra reveal binding of bilirubin to bovine β-lactoglobulin. *FEBS Letters*, 539, 58.

68. Zanotti, G.; Panzalorto, M.; Marcato, A.; Malpeli, G.; Folli, C.; Berni, R. (1998). Structure of pig plasma retinol-binding protein at 1.65 Å resolution. *Acta Crystallographica*. D54, 1049.

69. Keller, R. (2004). *The Computer Aided Resonance Assignment Tutorial*, 1st Ed. Cantina Verlag: Switzerland.

70. COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4, (1994) *Acta Crystallographica*, D50, 760.

71. Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallographica*, D59, 1131.

72. Emsley, P.; Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica*, D60, 2126.

73. DeLona, W. L. (2002)*The PyMOL Molecular Graphics System*, DeLano Scientific Palo Alto, CA, USA. http://www.pymol.org

74. Oliveira, K. M. G.; Valente-Mesquita, V. L.; Botelho, M. M.; Sawyer, L.; Ferreira, S. T.; Polikarpov, I. (2001). Crystal structures of bovine β-lactoglobulin in the orthorhombic spece group c2221 structural differences between genetic variants A and B and features of the Tanford transition. *European Journal of Biochemistry*, 268, 477.

75. Murshudov, G. N.; Vagin, A. A.; Dodson , E.,J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica.* D53, 240.
76. Bower, M. J.; Cohen, F. E.; Dunbrack, R. L. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology*, 267, 1268.
77. GaussView, Version 3.09, Dennington II, R.; Roy, K.; Todd, M. et. al. Semichem, Inc., Shawnee Mission, KS, **2003**.
78. Gaussian 03, C.02, Frisch, M. J.; Trucks, G. W; Schlegel, G. E.; et. al. Gaussian, Inc., Wallingford CT, **2003**.
79. Mormann, W.; Leukel, G. (1988). A simple and versatile synthesis of trimethylsiloxy-substituted isocyanates. *Synthesis*, 12, 990.