



**UNIVERSITA' DEGLI STUDI DI VERONA**

**DIPARTIMENTO SCIENTIFICO E TECNOLOGICO**

**DOTTORATO DI RICERCA IN:  
BIOTECNOLOGIE AMBIENTALI E INDUSTRIALI**

**XVII CICLO**

**NMR structural studies, dynamics simulations and relaxation  
measurements for functional characterization of proteins: calycins  
and proteins rich in disulphide bridges**

**Coordinatore:** Ch.mo Prof. Hugo Luis Monaco

**Tutore:** Ch.mo Prof.ssa Henriette Molinari

**Dottoranda** Maddalena Catalano

## INDEX

<b>Riassunto</b>	pag	1
<b>Introduction</b>	pag	3
<b>Chapter 1</b> Bowman Birk inhibitor from Snail Medic seeds ( <i>Medicago scutellata</i> )	pag	6
<b>Chapter 2</b> Viscotoxins	pag	15
<b>Chapter 3</b> Beta-lactoglobulins	pag	23
<b>Chapter 4</b>		
4.1 Bile Acid Binding Protein (cl-BABP)	pag	43
4.2 NMR structure of chicken liver bile acid binding protein and dynamics characterisation of its apo and holo forms		49
<b>Chapter 5</b> NMR Methods	pag	103
5.1 Introduction	pag	103
5.2 Heteronuclear NMR experiments		105
5.3 Structure calculation		107
5.4 The relaxation		
5.4.1 Dipole-dipole interaction and rotational motion in liquid	pag	110
5.4.2 Spin-lattice relaxation		111
5.4.3. Spin-spin relaxation		112
5.4.4. Nuclear Overhauser effect		113
5.4.5 Relaxation experiments		115
5.4.6 Relaxation data analysis		118

## Appendix

Paper N° 1

Solution structure of a Bowman Birk Inhibitor isolated from snail medic seeds (*Medicago Scutellata*) showing anticarcinogenic activity and analysis of self-association behaviour

Paper N°2

Effects of the *Nedicago scutellata* Trypsin Inhibitor (MsTI) on Cisplatin-induced Cytotoxicity in human Breast and Cervical Cancer Cells

Paper N°3

NMR solution structure of viscotoxin C1 from *Viscum album* ssp. *coloratum* Ohwi: towards structure-function analysis of viscotoxins

Paper N°4

Competitive binding of fatty Acids and the fluorescent probe 1-8-anilinonaphtalene sulfonate to bovine beta-lactoglobulin

Paper N°5

EF loop conformational change triggers ligand binding in beta-lactoglobulins

Paper N°6

Heterologous expression of bovine and porcine b-lactoglobulins in *Pichia pastoris*: towards a comparative functional characterization

Paper N°7

Determinants of protein stability and folding: comparative analysis of beta-lactoglobulins and liver basic fatty acid binding protein

Paper N°8

Solution structure of chicken liver fatty acid binding protein

## **Introduction**

The research project of my PhD concerns essentially the application of NMR spectroscopy to the structure-function characterisation of proteins.

A fundamental understanding of biological processes requires knowledge of molecular structure at atomic resolution, as well as a detailed description of time-dependent changes in molecular conformation.

NMR spectroscopy is unique among the methods available for three-dimensional structure determination of proteins since the NMR data can be recorded in solution. In the NMR experiments solution condition such as temperature, pH and salt concentration can be adjusted so as to closely mimic a given physiological fluid. Conversely, the solution may also be changed to extreme non-physiological conditions, for example for protein denaturation.

Furthermore, in addition to protein structure determination, NMR applications include investigations of dynamic features of the molecular structures, as well as studies of structural, thermodynamic and kinetic aspects of interactions between proteins and ligands.

Proteins are dynamic molecules that often undergo conformational changes while performing their specific functions, such as, for example, enzymatic catalysis or ligand binding.

NMR spectroscopy can be used to monitor the dynamic behaviour of a protein at a multitude of specific sites. Moreover, protein movements on a broad range of timescales can be monitored using various types of NMR experiments. Nuclear spin relaxation measurements report the internal motions on fast (sub-nanoseconds) and slow (microseconds to milliseconds) timescales as well as overall rotational diffusion of the molecule (5-50 ns), whereas rates of magnetisation transfer among protons with different chemical shifts and proton exchange report movements of protein domains in very slow time scales (milliseconds to days).

I have exploited all of these NMR features to carry out my research work on two groups of proteins:

- a) proteins rich in disulphide bridges displaying anti-tumoral activity
  
- b) proteins belonging to the calycin superfamily

The present thesis has been divided in several parts:

Chapters 1 and 2 describe the work carried out on proteins rich in disulphide bridges displaying anti-tumoral activity, i.e. the Bowman Birk inhibitor from Snail Medic seeds (MSTI) (Chapter 1) and viscotoxin C1 from *Coleratum ohwi* (Chapter 2).

Chapters 3 and 4 describe the work carried out on proteins belonging to the calycin superfamily, i.e. bovine and porcine beta-lactoglobulins (Chapter 3) and chicken liver bile acid binding protein (cl-BABP) (Chapter 4).

The main emphasis in these chapters has been given to the description of the biological problem studied, and the results obtained, without giving any description of the NMR methods employed to obtain the mentioned results.

Indeed all NMR and simulation methods employed are described in Chapter 5

All chapters describe the results obtained during my thesis work: the published papers referred to in each chapter are reported in Appendix.

For each published paper presented the level of my contribution to it is clearly stated.

**PROTEINS RICH IN DISULPHIDE BRIDGES DISPLAYING ANTI-TUMORAL  
ACTIVITY.**



proteases. Most BBIs inhibit trypsin at the first reactive site (N-terminal) and chymotrypsin at the second reactive site (C-terminal). The amino acid residues around the reactive site of the inhibitors are usually designed as P3, P2, P1, P1', P2', P3'. P1 residue is involved in the so-called "primary" contact region and confers inhibitory specificity: arginine and lysine for trypsin, leucine, phenylalanine and tyrosine for chymotrypsin and alanine for elastase (1). However, in contrast to primary specificity, trypsinolysis of soybean BBI (BBI-I) occurs not only in the predicted trypsin inhibitory domain, but also in the predicted chymotryptic inhibitory domain (4).

Despite extensive studies of BBIs, only a few three-dimensional structures have been solved by X-ray or by NMR. They include structures of tracy soybean (5), peanut (6) and pea seeds (2) and soybean inhibitor (7, 8). The X-ray structures of trypsin complexes of inhibitors from azuki bean (9), mung bean (10) and soybean (11) have also been reported. Recently a novel 14 amino acid residue cyclic peptide, cyclo (RCTKSIPPICFPDG), isolated from sunflower seeds, was found to be a potent inhibitor of trypsin ( $K_d = 100$  pM) and the X-ray structure of its complex with bovine trypsin showed both sequence and conformational similarity with the trypsin-reactive loop of the Bowman-Birk family of serine protease inhibitors (12, 26). This inhibitor, however, is unique in being mono-functional, cyclic and far shorter than inhibitors belonging to this family. It has been suggested that the high potency of this peptide is likely to arise from the considerable structural rigidity achieved through its cyclic nature which is further stabilised by a single internal disulphide bond.

The 11-residue peptide NleCTASIPPQCY, designed to reproduce the short beta-sheet segment that forms the reactive site loop of BBIs, cyclised via a disulphide bridge, retains the "canonical conformation" typical of BBIs, indicating that this sequence represents an independent structural  $\beta$ -hairpin motif, and maintains the biological activity of the native protein (13).

It has been suggested that BBIs have anti-tumoral and radio-protective activity, and it is still unknown whether these activities are correlated to the inhibitory activity of trypsin/chymotrypsin or other enzymes or to the capability of interacting with other macromolecules. (11, 14, 15, 16, 17, 18, 19, 20).

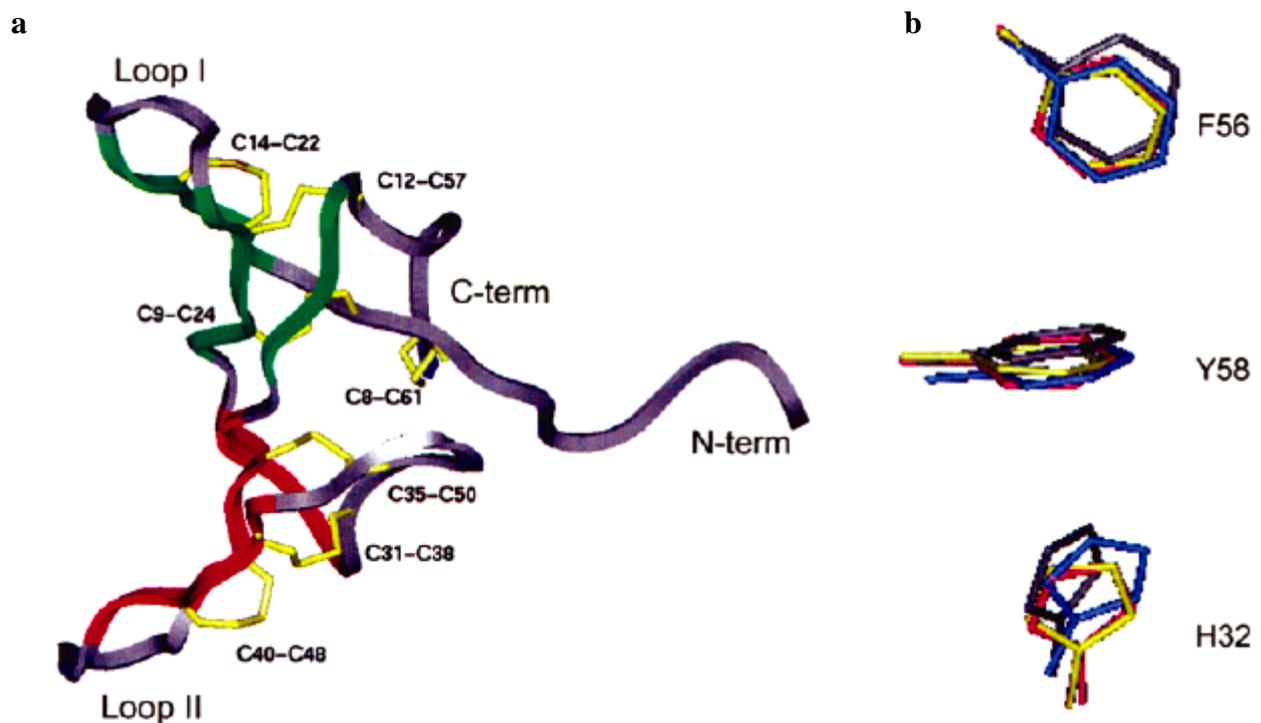
In any case, all parameters related to the proteinase recognition are correlated to the inhibitor conformation, to its charge distribution and residue exposure at the surface, in other words, to the active folding of the inhibitor.

As far as the anti-tumoral activity of this class of compounds is concerned, literature data so far published have been focalised into three main lines, namely: i) cellular proteases involved in carcinogenesis and possibly inhibited by BBIs; ii) the effect of the inhibitors on the nuclear excision repair system, through the stabilisation of TP53 protein (21, 22), or the expression regulation of



Most BBIs inhibit trypsin at the first reactive site (N-terminal) and chymotrypsin at the second reactive site (C-terminal), but MSTI has two antitryptic active sites, as the P1 residues at both active sites are Arginines.

Interestingly three aromatic residues, highly conserved in all BBIs (H32, F56, Y58) form in MSTI an aromatic cluster showing a typical geometry in which pairs of aromatic residues are almost orthogonal to each other and with centroids at distances shorter than 7Å (Figure 1.3b). This geometry is known to confer thermo stability to many proteins (25) in agreement with the observation that MSTI shows high temperature stability.



**Figure 1.3** Lowest-energy MSTI structure drawn in ribbon representation. Residues 11-25 and 53-55, corresponding to the first three-stranded  $\beta$ -sheet, are colored in green; residues 27-29 and 37-51, corresponding to the second three-stranded  $\beta$ -sheet, are colored in red. The active loops, the disulfide bridges and N and C-termini are indicated (a) Superposition of residues belonging to the aromatic cluster present in all BBIs of known structure (MSTI, gray; BBI-I, yellow; PI-II, red; PsTI-IVb, blue) (b)

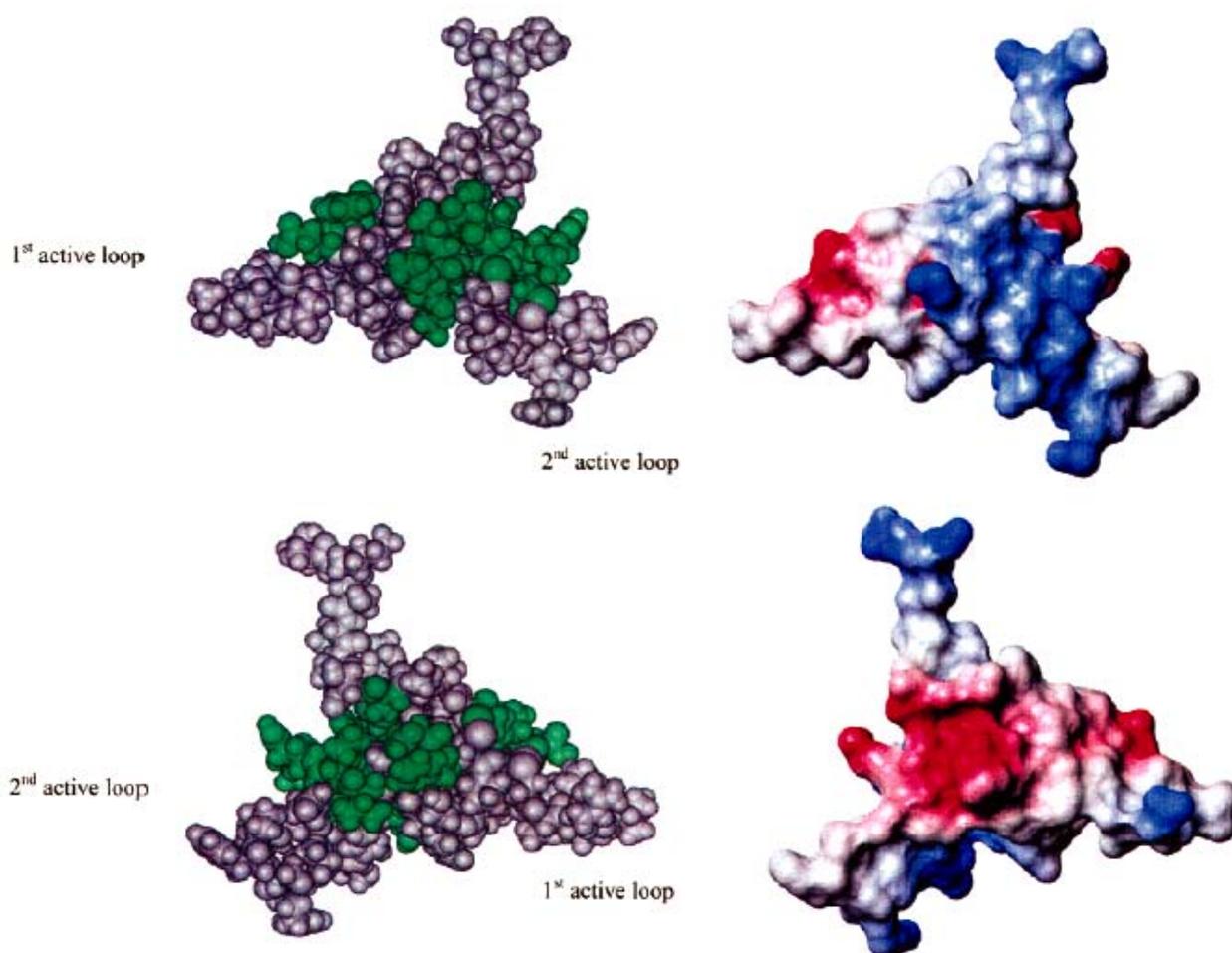
Polar and charged side-chains are predominant in both domains of MSTI as expected for trypsin inhibition specificity, however the second domain presents higher hydrophobicity. This observation suggests a low specificity of the second domain against trypsin, supported by the structural results pointing at its lower conformational rigidity. In agreement with this hypothesis, titration studies have shown a stoichiometric ratio MSTI:trypsin of 1:1.5.

The analysis of the electrostatic properties at the surface shows that there are two faces of the protein with opposite electrostatic potential, thus suggesting that intermolecular electrostatic interactions may play an important role in MSTI association. Furthermore, residues involved in self-association are highly conserved in BBIs from different seeds suggesting a functional relevance of these two superficial patches. We have proposed that MSTI residues involved in self-association may represent those regions of the protein capable of interacting with other enzymes or macromolecules, which could be of highest importance in anti-tumoral activity.

We further analysed self-association behaviour of MSTI collecting NMR data at different protein concentrations ranging from 0.6 to 2.0 mM in order to characterize the self-association behaviour.

The comparison of  $H_\alpha$  conformational shifts at 1.2 mM and 2.0 mM indicates that the secondary structure of the domains is preserved upon association. At 2.0 mM protein concentration a few amide and side-chain protons shifts were observed, clearly due to self-association.

The residues affected by self-association are localized in two well defined regions on opposite faces of the molecule (Figure 1.4, left panel). The calculation of the electrostatic potential surfaces shows that these two faces present an opposite potential (Figure 9, right panel), and we suggest that electrostatic interactions between them may play an important role in MSTI association. Residues involved in MSTI self-association, located at the border of the domains, are highly conserved in BBIs from different seeds suggesting a functional relevance of the two superficial patches. The presence of self-association has also been reported for PsTI-IVb (2), PI-II (5) and BBI-I (7), but the structural data for the dimer have been reported only for PsTI-IVb. It is important to stress that different residues are involved in self-association of MSTI and PsTI-IVb. Residues involved in PsTI-IVb self-association are not conserved in BBI family, at variance with those of MSTI. From these data we infer that residues of MSTI affected by self-association may represent those regions of the protein capable of interacting with other enzymes or macromolecules, which could be of the highest importance in anti-tumoral activity. The identification of association/interaction interfaces will be further investigated to correlate the association properties of MSTI with its anti-tumoral activity.



**Figure 1.4** Self-association interfaces and electrostatic potential surfaces of MSTI. Two views of MSTI are offered in the lower and upper panel (one rotated by 180 degrees with respect to the other about z-axis). Residues affected by self-association are colored in green on CPK representation of the molecule (left panel); electrostatic potential surfaces were calculated by MOLMOL program (27) (right panel). The location of the first and second active loop in the two views are indicated.

The effect of MSTI on cisplatin-induced cell killing in MCF7 human breast carcinoma cells and HeLa human cervical carcinoma cells were evaluated in order to assess i) whether MSTI inhibitor shows anti-tumoral activity and ii) enhances cisplatin induced cytotoxicity.

It was observed that the 24-hour treatment of MSTI in the cell culture medium decreased the clonogenic survival of MCF7 and HeLa cells in dose-dependent manner.

In order to verify if the presence of MSTI enhances the cisplatin induced cytotoxicity, different experiments were performed by using colony formation assay. The presence of 50 mg/ml MSTI in the culture medium before, during and after the treatment with cisplatin caused a reduction of the percentage of cells that were able to form clones. Treatment of MCF7 and HeLa cells with cisplatin

in combination with MSTI decreased the clonogenic survival of these cells to a much greater extent than treatment with cisplatin in the absence of MSTI. Among the two cell lines analysed, the effects of MSTI on clonogenic capacity were relatively higher in MCF7 cells as compared to HeLa cells.

Interestingly our data underline that the cytotoxic properties of MSTI indicate that anti-chymotriptic activity is not a strict requirement for anti-tumoral activity.

These results have suggested that MSTI could be a useful agent for the potentiation of cisplatin-mediated cancer treatment.

This work has been performed in collaboration with the medical team of prof. Robustelli della Cuna at the Fondazione Maugeri of Pavia.

### ***Work in progress and future perspectives***

In order to characterise in detail the structural features of the dimer and to produce mutated proteins for biological tests, it is crucial to set up a protocol for the expression of <sup>13</sup>C and <sup>15</sup>N enriched protein. The group of Prof. Carbonera, at the University of Pavia, has isolated the gene and gave it to us for protein production. This is however a very difficult task, given the high number of disulphide bridges present in the protein. At the moment few collaborative projects are under evaluation in the NMR laboratory.

### **This project led to the publication of the following papers:**

- Catalano, M., Ragona, L., Molinari, H., Tava, A., and Zetta, L., Solution structure of a Bowman Birk Inhibitor isolated from snail medic seeds (*Medicago Scutellata*) showing anticarcinogenic activity and analysis of self-association behaviour, (2003) **Biochemistry**, 42, 2836-2846. (Paper N° 1 in Appendix)
- Lanza, A., Tava, A., Catalano, M., Ragona, L., Singuaroli, I., Robustelli della Cuna, F. S., Robustelli della Cuna, G., Effects of the Nedicago scutellata Trypsin Inhibitor (MsTI) on Cisplatin-induced Cytotoxicity in human Breast and Cervical Cancer Cells (2004) **Anticancer Res**, 24, 227-234. (Paper N°2 in Appendix)

My contributions to these papers were related to the sequence and NMR structure determination and structure function correlation. The anti-tumoral tests were carried out by the group of Prof. Robustelli della Cuna at the Fondazione Maugeri of Pavia.

## References

1. Laskowski, M. Jr, Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593-626
2. Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J., Brunie, S. (1999) *J. Mol. Biol.* **285**, 1195-207
3. Odani, S., Ikenaka, T. (1978) *J. Biochem.* **83**, 747-53
4. Jensen, B., Unger, K.K., Uebe, J., Gey, M., Kim, Y.M., Flecker, P. (1996) *J. Biochem. Biophys. Methods* **33**, 171-85
5. Chen, P., Rose, J., Love, R., Wei, C.H., Wang, B.C. (1992) *J. Biol. Chem.* **267**:1990-1994
6. Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S., Ikenazka, T. (1993) *J. Mol. Biol.* **234**, 722-734
7. Werner, M.H., Wemmer, D.E. (1992) *Biochemistry* **31**, 999-1010
8. Voss, R.H., Ermler, U., Essen, L.O., Wenzl, G., Kim, Y.M., Flecker, P. (1996) *Eur. J. Biochem.* **242**, 122-31
9. Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S., Takahashi, K. (1986) *J. Biochem.* **100**, 1637-1646
10. Lin, G., Bode, W., Huber, R., Chi, C. and Engh, R. A. (1993) *Eur. J. Biochem.* **212**, 549-555
11. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., Flecker, P. (2000) *J. Mol. Biol.* **298**, 477-491
12. Luckett, S., Garcia, R.S., Barker, J.J., Konarev, A.V., Shewry, P.R., Clarke, A.R., Brady, R.L. (1999) *J. Mol. Biol.* **290**, 525-33
13. Brauer, A.B., Kelly, G., McBride, J.D., Cooke, R.M., Matthews, S.J., Leatherbarrow, R.J. (2001) *J. Mol. Biol.* **306**, 799-807
14. Kennedy, A.R. (1998) *Am. J. Clin. Nutr.* **68**, 1406S-1412S
15. Miyata, S., Koshikawa, N., Yasumitsu, H., Miyazaki, K. (2000) *J. Biol. Chem.* **275**, 4592-4598
16. Sorsa, T., Salo, T., Koivunen, E., Tyynelä, J., Konttinen, Y.T., Bergmann, U., Tuuttila, A., Niemi, E., Teronen, O., Heikkilä, P., Tschesche, H., Leinonen, J., Osman, S., Stenman, U.H. (1997) *J. Biol. Chem.* **272**, 21067-21074
17. Miyata, S., Miyagi, Y., Koshikawa, N., Nagashima, Y., Kato, Y., Yasumitsu, H., Hirahara, F., Misugi, K., Miyazaki, K. (1998) *Clin. Exp. Metastasis* **16**, 613-622
18. Ware, J.H., Wan, X.S., Rubin, H., Schechter, N.M., Kennedy, A.R. (1997) *Arch. Biochem. Biophys.* **344**, 133-138
19. Ware, J.H., Wan, X.S., Kennedy, A.R. (1999) *Nutr. Cancer.* **33**, 174-177
20. Billings, P.C., St. Clair, W., Owen, A.J., Kennedy, A.R. (1998) *Cancer. Res.* **48**, 1798-17802
21. Dittmann, K.H., Gueven, N., Mayer, C., Ohneseit, P., Zell, R., Begg, A.C., Rodemann, H.P. (1998) *Radiat. Res.* **150**, 648-655

22. Dittmann, K.H., Gueven, N., Mayer, C., Rodemann, H.P. (1998) *Int. J. Radiat. Biol.* **74**, 225-230
23. St. Clair, W.H., Billings, P.C., Kennedy, A.R. (1990) *Cancer. Lett.* **52**, 145-152
24. Ceciliani, F., Tava, A., Iori, R., Mortarino, M., Odoardi, M., Ronchi, S. (1997) *Phytochemistry* **44**, 393-398
25. Burley, S. K., and Petsko, G. A. (1985) *Science* **229**,23-28.
26. Korsinczky, M. L. J., Clark, R. J., Craik, D. J., (2005) *Biochemistry* **44**, 1145-1153.
27. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.***14**,51-5, 29-32.

## CHAPTER 2

### *Viscotoxins*

Viscotoxins are a group of basic low molecular weight proteins (MW about 5000 Da) found in several European mistletoe species (*Viscum album*). On the basis of sequence homology, they have been classified as belonging to the  $\alpha$ - and  $\beta$ -thionins family (1, 2). Thionins include several plant proteins from cereals (hordothionins, purothionins and avenothionins), Brassicaceae (crambins), *Pyralia pubera* and Viscaceae (viscotoxins from European mistletoes and phoratoxins from American mistletoes). The main characteristic of all thionins, except crambin, is their toxic effect on different biological systems: evidences of bactericidal and fungicidal properties as well as toxicity in insects were reported (3,4). Besides the poisonous effects on the whole organisms, viscotoxins have been reported to be cytotoxic against many tumoral cell lines including Yoshida sarcoma cells (5), K 562 human myelogenous leukaemia (6) and HeLa cells (7). Thionins' toxicity is exerted through cell membranes destabilisation and disruption, but their mechanism of action is not yet fully understood. To account for this generalised toxicity it was proposed that thionins induce ion channels formation in cell membranes, causing the dissipation of ion concentration gradients (8). An alternative theory suggests that thionins lead to cell membranes breakdown through purely electrostatic interactions with the negatively charged phospholipids (10).

Viscotoxins share a high sequence similarity with thionins (10), and their cysteines residues, all involved in disulphide bridges, are kept in conserved positions within the amino acid sequence defining a structural motif, known as "concentric motif". This kind of disulphide pattern is suggested to stabilise a common structure occurring in various small proteins able to interact with cell membranes (1).

A research project was undertaken in the NMR laboratory aimed at the determination of the solution structure of viscotoxins displaying different anti-tumoral activities, with the final aim of correlating the biological activity with structural features.

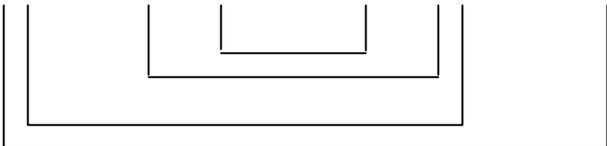
The structure of viscotoxin A3, displaying the highest biological activity, was previously determined in our laboratory (10).

I have been working on the structural determination of a newly isolated viscotoxin, from *Coloratum Ohwi* named viscotoxin C1.

The sequence alignment of the known viscotoxins, including the newly determined sequence of C1 (A1, A2, A3, B, 1-PS, U-PS and C1), together with those of related thionins of known structure is reported in Figure 2.1, which also contains data relative to viscotoxins cytotoxicity. Cytotoxicity was detected by measuring the inhibition of  $^3\text{H}$ -thymidine incorporation into Yoshida Sarcoma

cells (5). The concentration of substance inhibiting 50% of  $^3\text{H}$ -thymidine corresponds to the  $\text{ED}_{50}$  and was defined as 1 inhibitory unit.

PROTEIN	SEQUENCE	IDENTITY %	$\text{ED}_{50}$ $\mu\text{M}$
			
viscotoxin A3	KSCCPNTTGRNIYNACRLTGAPRPTCAKLSGCKIISGSTCPSDYPK	100.0	0.068
viscotoxin 1PS	KSCCPNTTGRNIYNTCRFGGGSREVCAKLSGCKIISASTCPSDYPK	78.3	0.089
viscotoxin A1	KSCCPNTTGRNIYNTCRLTGSSRETCAKLSGCKIISASTCPSNYPK	87.0	0.177
viscotoxin C1	KSCCPNTTGRNIYNTCRFAGGSREKCAKLSGCKIISASTCPSDYPK	82.6	0.198
viscotoxin A2	KSCCPNTTGRNIYNTCRFGGGSROVCAKLSGCKIISASTCPSDYPK	80.4	0.223
viscotoxin B	KSCCPNTTGRNIYNTCRLGGGSREKCAKLSGCKIISASTCPSDYPK	82.6	0.942
phoratoxin-A	KSCCPTTARNIYNTCRFGGGSRPVCAKLSGCKIISGTKCDSNGNH	67.4	
$\alpha$ -purothionin	KSCCRSTLGRNCYNLCRARGAQK-LCAGVCRCKISSGLSCPFGFPK	56.5	
$\beta$ -purothionin	KSCCKSTLGRNCYNLCRARGAQK-LCANVCRCKLTSGLSCPKDFPK	56.5	
$\beta$ -hordothionin	KSCCRSTLGRNCYNLCRVRGAQK-LCANACRCKLTSGLSCPKGFPK	56.5	
$\alpha$ -hordothionin	KSCCRSTLGRNCYNLCRVRGAQK-LCAGVCRCKLTSTGSCPFGFPK	52.2	
crambin	TTCCPSIVARSNFNVCRLPGTPEALCATYTGCIIPGATCPGDYAN	47.8	

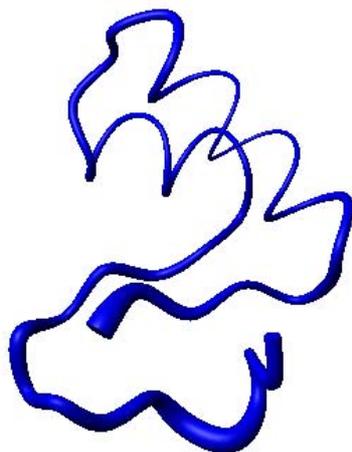
CONCENTRIC MOTIF

**Figure 2.1** Alignment of all reported viscotoxin isoforms, listed in order of decreasing cytotoxicity ( $\text{ED}_{50}$ ) together with related  $\alpha$  and  $\beta$  thionins of known structure. Conserved residues in the viscotoxin sequences are shaded. Sequence identity and biological activity data ( $\text{ED}_{50}$ ) are reported in the last two columns. Secondary structural elements of viscotoxin A3 are indicated. The disulphide bridges pattern, known as “concentric motif”, is shown.

As a part of this research project we have determined the primary sequence of viscotoxin C1, by Edman degradation coupled to HPLC. The obtained sequence presented few ambiguities relative to four residues: residue 38 was identified with the same probability as an alanine or serine and the last three residues were only supposed to be  $\text{Y}_{44}$ ,  $\text{P}_{45}$  and  $\text{K}_{46}$  due to low signal. The sequence was then completely determined through the NMR sequential assignment, confirmed by mass spectrometry and deposited in the Swiss-Prot data bank (Swiss-Prot accession number P83554).

I have then determined the 3D structure in solution of viscotoxin C1.

The viscotoxin C1 fold (deposited in the PDB data bank as 1ORL) resulted in a helix-turn helix motif with a short stretch of an antiparallel beta-sheet, common to that found for the highly similar viscotoxin A2 (10), A3 (10) and for other related thionines (1) (Figure 2.2).



**Figure 2.2** NMR solution structure of viscotoxin C1 (PDB code 1ORL). A sausage representation is reported, where the radius of the sausage is proportional to the RMSD among the ten best structures superimposed

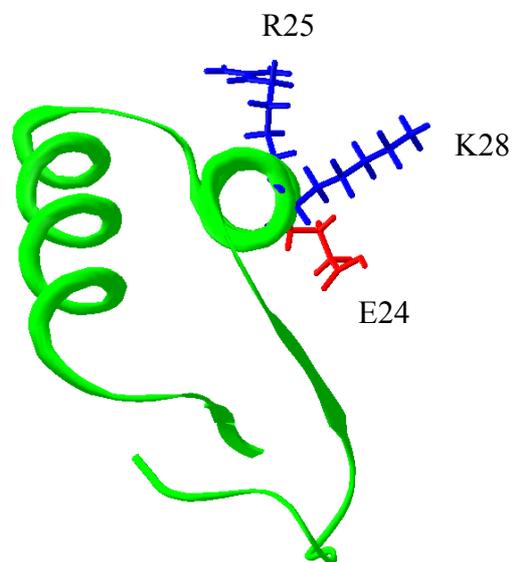
Viscotoxins, belonging to the thionin family, differs from the other members of the family for their high toxicity against tumoral cells.

The structural comparison of the three structurally known viscotoxins, together with the modelled structures of viscotoxins A1, B and 1-PS allowed for the identification of a conserved hydrophobic cluster, involving residues 12 and 13 (helix I), 27 and 29 (helix II) and 34 (C-terminal beta-strand), which could bear some relevance for viscotoxin topology, stabilizing the packing of the two helices on the beta-sheet. Electrostatic potential maps for C1, A1, A2, A3, B and 1-PS viscotoxins were computed using a Poisson-Boltzmann based methodology (12) as implemented in the software package UHBD (13). In the paper reported in the appendix, all the electrostatic surfaces are reported.

The presence of extended high positive potential areas at the molecule surface of all the studied viscotoxins hints at possible interaction sites with molecules bearing negative surface potential, such as the heads of cell membrane phospholipids. From the analysis of the mutations among viscotoxin sequences and the inactive crambin, it was possible to predict that few key residues,

localised on the solvent exposed face of the second helix, could behave as a recognition site for membrane interaction.

Interestingly, when pairs of sequences exhibiting the lowest number of mutations (three) are considered, i.e. the pairs C1/B, 1PS/A2 and A2/B, the relative role of every mutation can be discussed. Only one non conservative mutation (K28S) is observed for the pair C1/B, underlying the relevance of a positively charged residue at position 28. Two non conservative mutations (E24Q, R28S) are observed for the pair 1PS/A2, where only a 2.5-fold decrease in activity is measured, suggesting that the lack of a negatively charged residue at position 24 may enhance the toxicity. The four-fold decrease in activity on going from A2 to B isoform can be ascribed to the two non conservative mutations Q24E and V25R, where a further decrease in activity is possibly introduced by a positively charged residue at position 25. When the comparison is extended to the pair 1PS/B, where a ten-fold decrease in activity was measured, only two (V25R, R28S) out of the observed four mutations (F18L, V25R, R28S, I29L) are non conservative, strengthening our observation on the relevance of a non charged residue at position 25 and of a positively charged residue at position 28. The most relevant residues 24, 25 and 28 are localized on the solvent exposed face of the second helix (Figure 2.3), which we suggest can behave as a recognition site for membrane interaction.



**Figure 2.3** Sausage representation of viscotoxin C1 structure (PDB code 1orl). Three residues exposed to solvent are reported.

Like all other thionins, viscotoxins are synthesized in plant cells like precursor proteins with a common structure: a signal peptide for vacuolar targeting at the N-terminus, a central thionin domain and a long acidic peptide at the C-terminus (4-7) (Scheme 1). The function of this precursor structure may consist in the right targeting of the protein, and the maintenance of the inactive condition of the mature viscotoxin, in order to avoid self-toxicity events in the plant cell. It is therefore important to investigate the mechanism by which the C-terminal acidic domain of viscotoxin precursor neutralises the cytotoxic activity of the mature protein. Viscotoxin precursors are formed by a signal peptide for vacuolar translocation, the segment forming the mature viscotoxin and the C-terminus acidic polypeptide. The leader sequence is cleaved off just after the entry in the vacuolar space, while the acidic domain is post translationally processed. The viscotoxin domain is characterized by the presence of seven positively charged amino acids, while C-terminal domain contains seven negatively charged residues (Scheme 2.1).



Scheme 2.1

The charge difference between the basic viscotoxin domain and the acidic portion of the precursor polypeptide could favour a close interaction of these domains, through electrostatic attraction, thus explaining the non-toxicity of precursor protein in terms of a charge shielding within the mature toxin. The acidic domain may also induce conformational alteration in the three dimensional

viscotoxin structure, thus preventing the correct folding of the active protein. Furthermore it is worth noting that cysteines (six in the viscotoxin domain and six in the acidic one) are highly conserved, not only within different thionins, but also within the corresponding acidic portions of their precursor molecules (14), with the consequence that thionins might be kept inactive through intramolecular disulphide bonds between the thionin and the acidic protein domain. Experimental evidences seem to support both these hypothesis: even though the acidic part of the viscotoxin and other thionin precursors so far analyzed are different in the amino acidic length, neither the negative charge, nor the number and relative position of cysteine residues within the acidic domain have been lost by amino acid deletions (14). The determination of the three dimensional structure of the entire precursor protein could constitute one of the major tools for the comprehension of the acidic domain role. Since it is not possible to isolate viscotoxin precursors from plant extracts, due to the post translation processing by cell proteases, it is necessary to resume to molecular biology techniques for over expression of recombinant proteins in transgenic organisms. This is a quite difficult target for disulfide bonded polypeptides, since one of the major drawbacks of *E. coli*, as an expression system, is the limited ability to facilitate extensive disulphide bridges formation in its reducing cytoplasm (15). Different protocols are now being tested in our laboratory to obtain high level expression, refolding and purification of recombinant viscotoxin A3 precursor in its oxidised form.

**This project led to the publication of the following paper:**

- Romagnoli, S., Fogolari, F., Catalano, M., Zetta, L., Schaller, G., Urech, K, Giannattasio, M., Ragona L., and Molinari, H., NMR solution structure of viscotoxin C1 from *Viscum album* ssp. *coloratum* Ohwi: towards structure-function analysis of viscotoxins (2003) **Biochemistry**, 42, 12503-12510 (Paper N° 3 in Appedix)

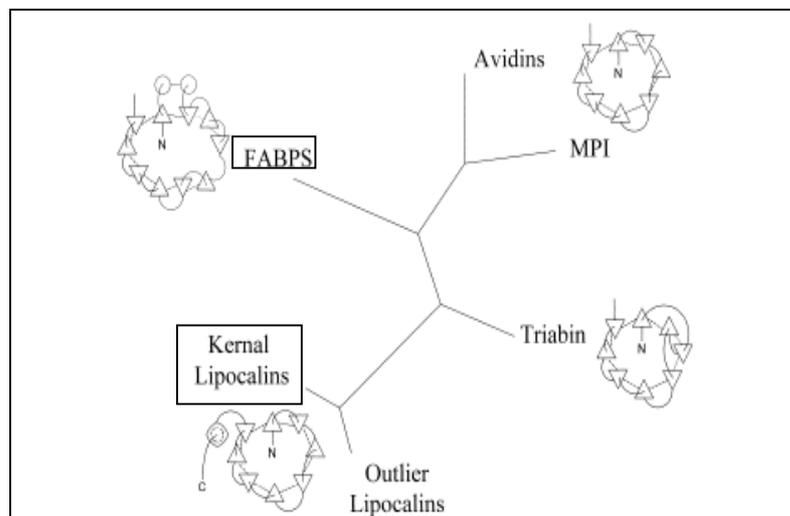
My contribution to this paper consisted in the viscotoxin C1 structure determination, using DYANA MD simulations, and in subsequent comparative analysis of various viscotoxins.

## References

1. Orrù, S., Scaloni, A., Giannattasio, M., Urech, K., Pucci, P. and Schaller, G. (1997) *Biol. Chem.* **378**, 986-996
2. Samuelsson, G. and Jayawardene, A. L. (1974) *Acta Pharm. Suec.* **11**, 175-184
3. Bohlmann, H. and Apel, K. Thionins. *Annu. Rev. Plant Mol. Biol.* (1991) **42**, 227-40
4. Bohlmann, H., Clausen, S., Behnke, S., Giese, H., Hiller, C., Ulrich, R. P., Shrader, G., Vibeke, B. and Apel, K *The EMBO J.* (1998) **7(6)** 1559-1565
5. Schaller, G., Urech, K. and Giannattasio, M. (1996) *L. Phytother. Res.* **10**, 473-477
6. Urech, K., Schaller, G., Ziska, P. and Giannattasio, M.. (1995) *Phytother. Res.* **9**: 49-55.
7. Konopa J., Woynarowski J.M., and Lewandowska-Gumieniak, M. (1980) *Physiol. Chem.*, **361**, 1525-1533
8. Huges, P., Dennis, E., Whitecross, M., Llewellyn, D. and Gage, P. (2000) *J. Biol. Chem.* **275**, 823-827.
9. Thevissen, K., Ghazi, A., De Samblanx, G. W., Brownlee, C., Osborn, R. W. and Broekaert, W. F.,(1996) *J. Biol. Chem.* **271**, 15018-15025
10. Romagnoli, S., Ugolini, R., Fogolari, F., Schaller, G., Urech, K., Giannattasio, M., Ragona, L. and Molinari, H, (2000) *Biochemistry. J.* **350**: 569-577.
11. Piantini, U., Sorensen, O. W. and Ernst, R. R., (1982) *J. Am. Chem. Soc.* **104**, 6800-6801.
12. Fogolari, F., Brigo, A., and Molinari, H. (2002) *Journal of Molecular Recognition* **15**, 377-392.
13. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Llin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., and Mc Cammon, J. A., (1995) *Comp. Phys. Comm.* **91**, 57-95.
14. Gausin, K., Thionin (1987) *Planta*, **171**: 241-246
15. Makrides, S. C., Strategies for achieving high-level expression of genes in *Escherichia coli*. (1996) *Microbiological Reviews*, **60**, 512-538

## PROTEINS BELONGING TO THE CALYCIN SUPERFAMILY

This work is part of a wide research project, going on in our Laboratory, involving the comparative study of structure, folding and interaction properties of members of the Calycin superfamily. The guiding principle of this research has been the idea that the close comparison of similar proteins, sharing the same folding topology and various degree of sequence identity (from 20 to 65%) could help in learning fundamental aspects of folding and interaction mechanisms.



## CHAPTER 3

### *Beta-lactoglobulins*

Bovine and porcine beta-lactoglobulins (BLG, PLG), the most abundant milk whey proteins, belong to the lipocalin family.

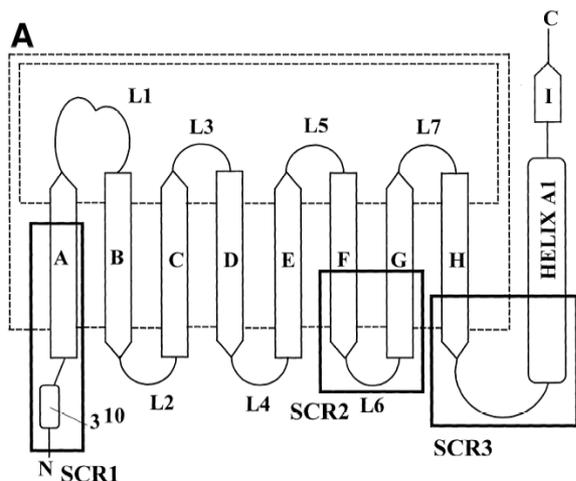
Lipocalins constitute a remarkable protein family, the members of which exhibit extraordinary diversity at the level of sequence and function although they have a highly conserved structure and a shared ancestor that appeared very early in evolution.

This family consists of more than 30 small extracellular proteins found in a wide variety of tissues in vertebrate and invertebrate animals, plants and bacteria (1). Lipocalins (from the greek words lipos and calyx) are characterised by three molecular recognition properties: the binding of small hydrophobic substances, the binding to cell-surface receptors, and the formation of complexes with other soluble macromolecules.

A typical lipocalin consists of a 160–190 amino-acid peptide, with 8–9  $\beta$ -strands folded into a  $\beta$ -barrel with a hydrophobic interior.

Many lipocalins have a strong impact in clinical medicine and great therapeutic potential. The plasma proteins retinol-binding protein, orosomucoid, and  $\alpha$ 1-microglobulin, the nasal odorant-binding proteins, the bilin-binding proteins of butterflies, and some temperature-induced proteins that confer on a plant its resistance to cold, are all examples of lipocalins (21).

The typical lipocalin fold is represented in Figure 3.1



**Figure 3.1** Schematic representation of the typical lipocalin fold: beta strands are presented with arrows and labelled with letters, loops are presented by solid lines connecting beta strands. The C-terminal helix is presented by a rectangle.

Beta-lactoglobulins belong to the group of the so called “kernel” lipocalins, including retinol-binding protein (RBP),  $\beta$ -lactoglobulin (BLG), insecticyanin, bilin-binding protein (BBP), major urinary, protein,  $\alpha_2\mu$ -globulin, epididymal retinoic, acid-binding protein, and neutrophil lipocalin . Their structure-based sequence alignment is reported in Figure 3.2.

```

1RBP   RDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA-----EFSVDETGQMSATAKGRV
1MUP   EEASSTGRNFNVEKINGEWHITIIASDKREKIEDNGNFRLFLE-QIHVLENSL-VLKFHTVR
1BBP   HDGACPEVKPVDNFDWSNYHGKWWEVAKYPNSVEKYGKCGWA-----EYTPGKSV-KVSNYHVI
1EPA   VKDFDISKFLGFWYEIFAFASKMGTPLAHKKEKMGAM-VVELKENLL-ALTTTYYS
2A2G   EEASSTRGNLDVAKLNGDWFISIVVASNKREKIEENGSMRVFMQ-HIDVLENSL-GFKFRIKE
INS    YPGYCPDVKPVNDFLSAFAGAWHEIAKLPLENENQ GKCTIA-----EYKYDGKKA-SVYNSFVS
1BEB   QTMKGLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTPEGDLEILLQKWE
1NGL   PAPPLSKVPLQQNFQDNQFQGGKQWYVVGLAGNAIILREDKDPQKMYAT--IYELKEDKSYNVTSVLF
          *****
          ***          ****

1RBP   RLLNNWDVDCADMVCTFTDTEDP-----AKFKMKYWCVASFLQKGNDDHWIVDTDYDTYAVQYSCRLL
1MUP   DEECSEL-----SMVADKTEKA-----GEYSVTYDGFN-----TFTIPKTDYDNFLMAHLINEK
1BBP   HGKEYFI-----EGTAYPVGDSKI---GKIYHKLTYGGVTKEN---VFNVLSTDNKNYIIGYCYKYD
1EPA   EDHCVLE-----KVTATEGDGP-----AKFQVTRLSGKK-----EVVVEATDYLTyaiIDITSLV
2A2G   NGECREL-----YLVAYKTPED-----GEYFVEYDGGN-----TFTILKTDYDRYVMFHLINEK
INS    NGVKEYM-----EGDLEIAPDAKYTKQGKYVMTFKFGQRVVNL---VPWVLATDYKKNYAINYNCDYH
1BEB   NGECAQK-----KIIAEKTKIP-----AVFKIDALNEN-----KVLVLDTDYKYYLLFCMENSA
1NGL   KKKCDYW-----IRTFVPGCQP-----GEFTLGNIKSYPLTSY--LVRVSTNYNQHAMVFFKKVS
          ****          ***          *****

1RBP   NLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYC
1MUP   DGETFQ---LMGLYGREPDLSSDIKERFAQLCEEHGILRENIIDLSNANRC
1DDP   EDKKGHQD FVWVLSRSKVLGTGEAKTAVENYLIGSPVVDSSQKLVYSDPF3EAACKVN
1EPA   AGAVHR---TMKLYSRSLDDNGEALYNFRKITS DHGPFSETDLYILKHDLTVCVKVLSAA
2A2G   NGETFQ---LMVL YGRTKDLSSDIKEKFAKLCEAHGITRDNIIIDLTKTDRCL
INS    PDKKAHSI-HAWILSKSKVLEGNTEKVVNDVNLKTFSHLIDASKFISNDFSEAACQYSTTYSLTGPDR
1BEB   EPEQSL---VCQCLVRTPEVDDEALEKFDKALKALPMHIRLSFNPTQLEEQC
1NGL   QNREYF---KITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDG
          *****

```

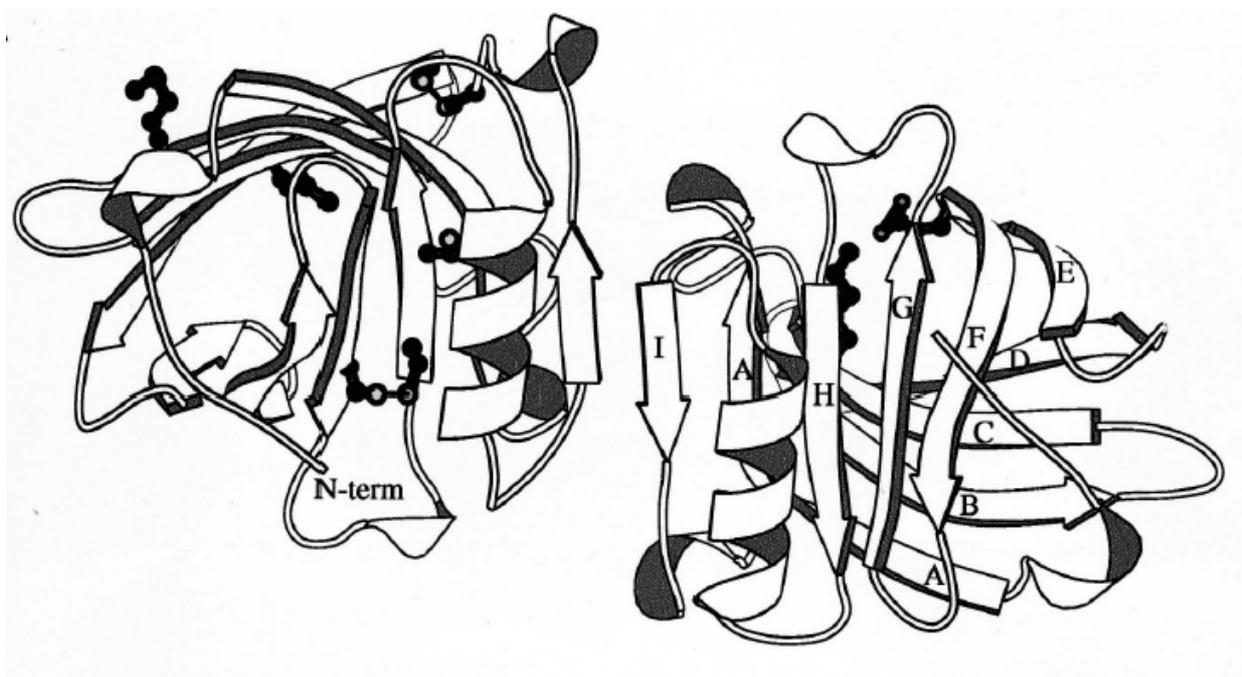
**Figure 3.2.** Structure-based sequence alignments of those kernel lipocalin structures which have been determined experimentally by crystallography and multidimensional NMR. Equivalent structural positions within the lipocalin common core are indicated by an asterisk. 1MUP, mouse major urinary protein; 1NGL, neutrophil lipocalin; INS, insecticyanin; 1BBP, bilin binding protein; 1EPA, epididymal retinoic acid binding protein; 1RBP, retinol binding protein; 1BEB,  $\beta$ -lactoglobulin; 2A2G,  $\alpha_2\mu$ -globulin.

No definite biological function has been ascribed to BLG although several proposals have been suggested. For example, from the structural and sequence similarities of BLG to RBP, it was thought that the protein might be involved in retinol transport from mother to neonate. Intestinal receptors were found that appeared specific for BLG.

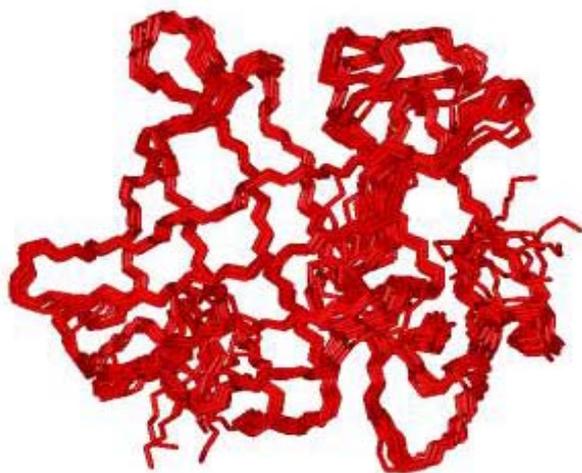
The protein is resistant to the proteolytic activity in the stomach (22). However, little endogenous retinol is found bound to BLG when it is first purified and the ligand most closely associated with the protein is palmitate. In milk,  $\alpha$ -lactalbumin, another small, abundant whey protein, is a modulator of  $\beta$ -galactosyl transferase, turning it into a lactose synthase. A modulation role has also been proposed for BLG by Perez et al. (23) who found that the protein appears to facilitate the activity of pre-gastric lipase by sequestering free fatty acids as they are produced.

BLG contains 162 amino acid residues, two disulphide bonds in addition to a free cysteine residue. The protein forms a dimer at neutral pH that dissociates into monomers below pH 3.0. The structure, obtained at neutral pH by X-Ray (2) (Figure 3.3), remains native in the monomeric state at acidic pH (Figure 3.4), as shown by NMR characterisation of the protein, performed within our and others laboratories (3-5).

In the native state BLG is a predominantly beta sheet protein consisting of a beta barrel of eight continuous antiparallel beta strands (A-H), an additional beta strand (I), involved in the dimer formation, and one major helix located at C-terminal end.



**Figure 3.3:** Representation of the X-ray structure of BLG obtained at pH 7 (2) (PDB code 1beb)



**Figure 3.4:** Superposition of the ten best NMR structures of BLG obtained at acidic pH (pH 2) (PDB code 1cj5) (3, 4, 6)

BLG shows unique folding characteristic. Although it is a predominantly beta-sheet protein, i) addition of trifluoroethanol to BLG was reported to induce a drastic conformational change to a predominantly alpha-helical structure (6); ii) the refolding kinetics measured by the far-UV circular dichroism (CD) indicated the accumulation of intermediates with non-native alpha-helices (6); iii) NMR equilibrium H/D exchange experiments suggested a transient nature for the intermediate identifying a slowest exchanging core localised in the F, G and H strands (8); iv) the study of peptide fragments of BLG (28) revealed the presence of native-like interactions in a peptide reproducing G and H strands, suggesting that this region could behave as folding nucleation site, recent refolding NMR studies provided a cartoon of the kinetic intermediate which showed persistent resistance to exchange in beta strands F, G and H, and weak protection for several residues in strand A. It was suggested that beta strand A could assume a non-native alpha-helix in the intermediate which thus seems to contain both native and non-native structures.

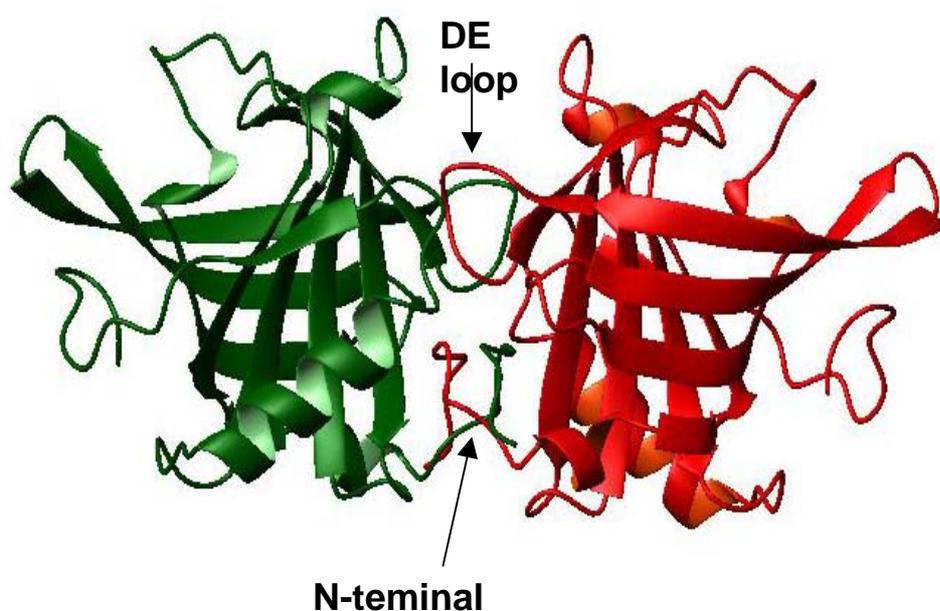
However, neither there is a clear picture of this non-native alpha helical intermediate nor it is clearly understood the role of the non-native alpha-helical structure in protein folding.

### *Folding studies*

The guiding principle for a deeper understanding of the folding mechanism of BLG has been the idea that only the close comparison of similar proteins, sharing the same folding topology and various degree of sequence identity (from 20 to 65%) could help in learning fundamental aspects of folding. In this line a comparative study was started in our laboratory on bovine and porcine beta lactoglobulins.

PLG contains 160 amino acid residues, two disulphide bonds and, differently from BLG, it does not have a free cysteine residue.

BLG and PLG in spite of the high sequence identity (62%) and similarity (83%), exhibit a different pH dependence of dimerisation and a different dimer interface (30). In BLG the ninth I strand and the AB loop participate to the dimer interface (Figure 3) while in the porcine protein the N terminal regions of two monomeric unit form a dimer through a domain swap (Figure 3.5).



**Figure 3.5:** X-ray structure of porcine beta-lactoglobulin obtained at pH 3 (PDB code 1exs)

A summary of the thermodynamic figures previously obtained by CD, NMR and fluorescence for BLG and PLG is reported in Table 1 and 2, to provide a view of the relative stability of the two proteins.

**Table 1** Thermodynamic parameters of PLG unfolding at 25 °C.

<b>Method</b>	<b>c<sub>mNI</sub></b>	<b>m<sub>NI</sub></b>	<b>ΔG<sup>0</sup><sub>NI</sub></b>	<b>c<sub>mIU</sub></b>	<b>m<sub>IU</sub></b>	<b>ΔG<sup>0</sup><sub>NU</sub></b>
<b>pH</b>	(M)	(kcal M <sup>-1</sup> mol <sup>-1</sup> )	(kcal mol <sup>-1</sup> )	(M)	(kcal M <sup>-1</sup> mol <sup>-1</sup> )	(kcal mol <sup>-1</sup> )
<b>pH 2</b>						
λ <sub>max</sub>	1.0 ± 0.1	2.2 ± 0.3	2.1 ± 0.3			
Global CD-fluo	1.1 ± 0.1	2.5 ± 0.5	2.8 ± 0.7	3.2 ± 0.2	1.4 ± 0.2	7.2 ± 0.6
NMR	0.7 ± 0.1	3.5 ± 0.1	2.3 ± 0.1			
<b>pH 6</b>						
λ <sub>max</sub>	1.3 ± 0.1	3.8 ± 0.5	5.0 ± 0.5			
Global CD-fluo	1.4 ± 0.1	3.0 ± 0.5	4.2 ± 0.5	3.0 ± 0.2	0.9 ± 0.2	6.9 ± 0.6
NMR	1.1 ± 0.1	4.2 ± 0.3	4.6 ± 0.5			

**Table II** Thermodynamic parameters of BLG unfolding at 25 °C.

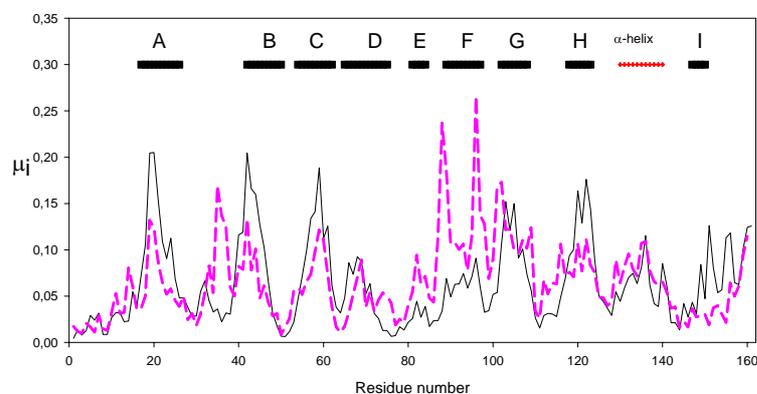
<b>Method</b>	<b>c<sub>mNI</sub></b>	<b>m<sub>NI</sub></b>	<b>ΔG<sup>0</sup><sub>NI</sub></b>	<b>c<sub>mIU</sub></b>	<b>m<sub>IU</sub></b>	<b>ΔG<sup>0</sup><sub>NU</sub></b>
	(M)	(kcal M <sup>-1</sup> mol <sup>-1</sup> )	(kcal mol <sup>-1</sup> )	(M)	(kcal M <sup>-1</sup> mol <sup>-1</sup> )	(kcal mol <sup>-1</sup> )
<b>pH 2</b>						
Global CD	3.2 ± 0.1	2.8 ± 0.5	9.0 ± 0.7	3.8 ± 0.2	1.0 ± 0.2	12.8 ± 0.6
				<b>c<sub>mNU</sub></b>	<b>m<sub>NU</sub></b>	<b>ΔG<sup>0</sup><sub>NU</sub></b>
				(M)	(kcal M <sup>-1</sup> mol <sup>-1</sup> )	(kcal mol <sup>-1</sup> )
<b>pH 6</b>						
Global CD				2.8 ± 0.1	4.1 ± 0.5	11.5 ± 0.5

In order to understand which are the key residues contributing to determine the characteristics of the folding mechanism for these two proteins, we turned to a computational approach, newly developed by Giorgio Colombo (CNR in Milano) with whom a collaboration was established for the study of the folding determinants. This computational approach is based on the energetic analysis of long

time scale molecular dynamics (MD) simulations around the native state. The basic idea of the method is that by means of MD simulations around the native conformation, describing the protein and the surrounding solvent at the all-atom level, an accurate and compact interaction energetic map of the native state of the protein can be generated (24). This map is simplified by means of an eigenvalue decomposition. The components of the eigenvector associated with the lowest eigenvalue identify those residues mostly responsible for the stability. The method was previously tested on model proteins (the alpha spectrin SH3 domain, the IgG binding domain of G protein on CI2): in all cases it seems to perform well and the comparison of these predictions with mutagenesis experiments provided excellent results (24).

The definition of a residue as determinant for the stabilization of protein native state / (hot site) on the basis of the components of the eigenvector was derived from the original paper (24). The most probable value of the eigenvector components distribution was calculated, and all the sites of the protein displaying an eigenvector component larger than the most probable value were defined as hot sites.

We performed a 6 ns simulation of BLG and PLG in explicit solvent, and the analysis of the normalised eigenvector associated with the lowest eigenvalue, as a function of residue number, basically identified how the stabilisation energy is distributed among different amino acids of the proteins (Figure 3.6). It was possible to identify the so-called “hot sites”, i.e. those residues building cluster of strong interactions during the dynamics. From the comparison of simulations performed on both BLG and PLG a different distribution of “hot sites” have been observed in the two proteins: in BLG the hot sites, represented in red in Figure 3.7a, are gathered in a central core region, while PLG shows a different hot sites distribution with a less extended contiguous folding core (Figure 3.7b).



**Figure 3.6** The eigenvectors ( $\mu_i$ ) associated with the largest eigenvalue in the all-atom MD simulations of BLG (black) and PLG (pink). Calculated on the last 2 ns of each simulation. Regions of secondary structure ( $\beta$ -strands and  $\alpha$ -helices) are indicated at the top of the diagram.



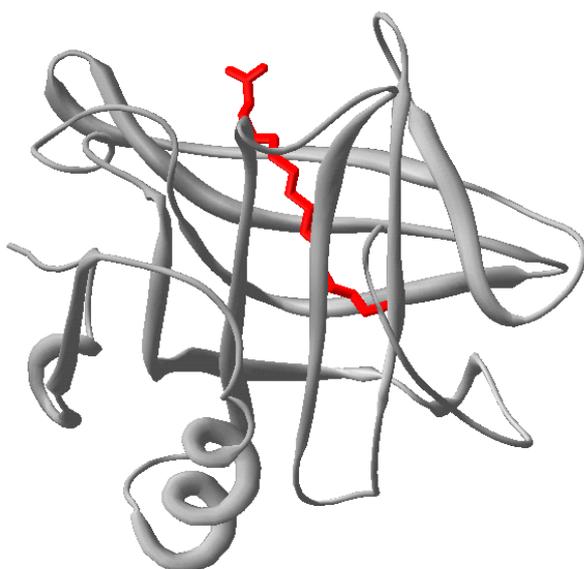
**Figure 3.7** Graphical representation of the distribution of eigenvector components (hot sites, in red in the structures) resulting from computation: BLG (a) and PLG (b).

The observed energetic differences, projected on the folding mechanism, suggested a different folding mechanism for these two highly similar proteins. In addition residues which appeared to be determinant for the stability (the hot sites) are located in highly mobile regions of PLG, such as the EF loop, thus predicting a minor stability for the porcine protein, in agreement with experimental CD data. Quantitative characterisation of the dynamic properties of BLG and PLG relied on principal component analysis of the covariance matrix of the positional fluctuations of the C-alpha atoms. The covariance matrix was built from the equilibrated portion of the trajectories, and its diagonalisation afforded the principal directions of the large amplitude concerted motions. Essential dynamics analysis showed that the principal overall motions of the two proteins are different, with PLG being more flexible, again in agreement with spectroscopic measurements.

In conclusion this computational approach predicted some protein properties, which were verified by spectroscopic techniques and afforded the rationale for mutant design in order to single out residues important for protein stability.

### *Interaction studies*

As far as the mode of binding of BLG is concerned, many papers have appeared in the literature with various proposals related to both the number (one or two) and the location (inside the barrel or in a surface cleft) of the binding sites. Two X-ray structures were reported for BLG complexed with palmitic (12) (Figure 3.8) and 12-bromo-dodecanoic acid (13). Both structures agree with the fatty acid located in a hydrophobic pocket within the calix of the protein (Figure 3.8).



**Figure3.8** X-ray structure of bovine BLG complexed with palmitic acid at neutral pH (PDB code: 1B00).

BLG is capable of binding ligands of a diverse nature. Table 3 shows a variety of the ligands that have been associated with the protein together with the method of measurement of the dissociation constant.

What is clear, however, is that BLG has an observable affinity for hydrophobic ligands like bromodo-decanoic acid (26), palmitic acid (27) and retinol.

One of the most used methods for determination of the interaction between ligand and protein has been the change in intrinsic fluorescence, thought to be largely that of Trp19 (25).

Selected ligand binding data for bovine  $\beta$ -lactoglobulin

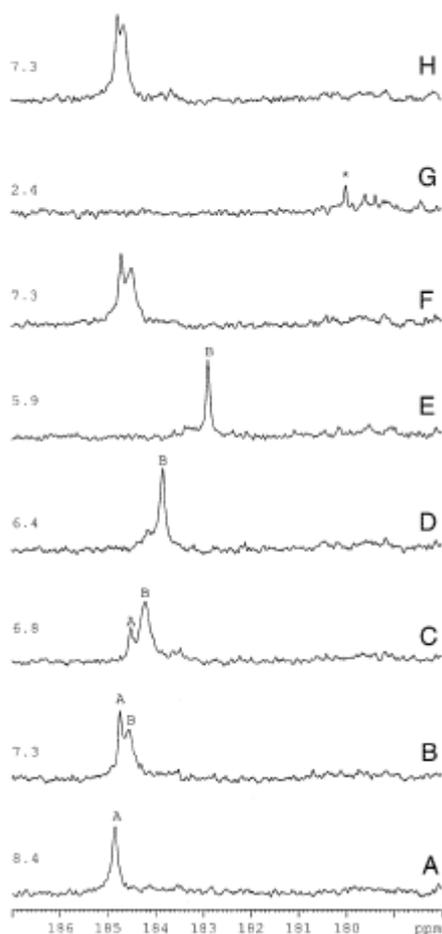
Ligand	$K_d$ (M)	$r^2$	pH	Method
Lauric	$7.0 \times 10^{-7}$	0.86	7.0	Fluorescence
Palmitate	$1.0 \times 10^{-7}$	0.9	7.0	Fluorescence
	$1.5 \times 10^{-6}$	1	7.4	Equilibrium dialysis
Stearate	$1.2 \times 10^{-7}$	1	7.0	Fluorescence
Nonanone	$4.1 \times 10^{-4}$	1	6.7	Equilibrium dialysis
Iodobutane	$3.6 \times 10^{-4}$	1	5.3	Equilibrium dialysis
SDS	$3.2 \times 10^{-6}$	1	7.5	Equilibrium dialysis
5-DSA <sup>b</sup>	$8.0 \times 10^{-7}$	1	7.0	EPR
	$4.0 \times 10^{-7}$	1	7.0	Fluorescence
Retinol	$1.5 \times 10^{-7}$	0.85	7.5	Fluorescence
	$2.0 \times 10^{-8}$	1	7.5	Fluorescence
	$3.6 \times 10^{-8}$	1	7.5	Affinity chromatography
	$6.7 \times 10^{-5}$	1	7.2	Equilibrium dialysis
Retinoic acid	$3.8 \times 10^{-8}$	1	8.0	Equilibrium dialysis
	$2.0 \times 10^{-7}$	1	7.0	Fluorescence
<i>cis</i> -Parinaric acid	$2.8 \times 10^{-8}$	0.8	7.0	Fluorescence
$\beta$ -Ionone	$6.0 \times 10^{-7}$	1	3.0	Fluorescence
Toluene	$2.2 \times 10^{-3}$	2	5.8	Equilibrium dialysis
PNP <sup>c</sup>	$3.1 \times 10^{-5}$	1	6.0	Fluorescence
ANS	$4.0 \times 10^{-5}$	1	7.5	Fluorescence
Ellipticene	$9.0 \times 10^{-8}$	0.9	8.5	Fluorescence
	$1.4 \times 10^{-6}$	0.5	8.0	Fluorescence
Bromophenol blue	$5.0 \times 10^{-4}$	1	9.4	Spectroscopy
Protoporphyrin	$4.0 \times 10^{-7}$	0.8	7.0	Fluorescence
Vitamin D <sub>2</sub>	$4.91 \times 10^{-9}$	1	7.0	Fluorescence
Cholesterol	$3.49 \times 10^{-8}$	1	7.0	Fluorescence
Tween 20	$1.1 \times 10^{-7}$	1	7.0	Fluorescence

**Table 3:** Data reported for BLG binding (25)

A solution study aimed at elucidating the pH dependence of the fatty acid binding and its mechanism was undertaken in the NMR laboratory.

BLG binds fatty acid within an internal cavity, with the carboxyl group pointing towards the solvent and work performed in our laboratory showed that the driving force for the binding is likely to be due to favourable electrostatic interactions between the carboxylate and charged lysines located at the opening of the protein cavity. <sup>13</sup>C NMR titration experiments performed on unlabelled BLG with <sup>13</sup>C<sub>1</sub> enriched palmitic acid (PA) have clearly pointed out that the amount of bound PA is drastically reduced upon decreasing pH and the ligand is completely released at pH 2. In the pH range 7.3-6.4 a conformational equilibrium was observed for the bound ligand (A and B forms in Figure 3.9) reflecting the dynamics of EF loop (85-90 region), triggered by the titration of E89 at anomalously high pKa (ca. 6.5) (paper 5 reported in Appendix) (Figure 3.9).

Our results clearly show for the first time that the protonation state of GLU 89 influences ligand binding in beta-lactoglobulins, forcing the EF loop to act as a mobile lid, hindering the access to the protein cavity when it is in close conformation.



**Figure 3.9.** Carboxyl region of the 125.7 MHz 1D proton decoupled  $^{13}\text{C}$  spectra acquired on the complex of singly  $^{13}\text{C}$  labeled  $[1-^{13}\text{C}]\text{PA}$  and BLG at 37 °C at (A) pH 8.4, (B) pH 7.3, (C) pH 6.8, (D) pH 6.4, (E) pH 5.9, (F) reversing the pH from (E) to pH 7.3, (G) lowering the pH from (F) to pH 2.4. The carboxyl resonance of PA is indicated with an asterisk, (H) reversing the pH from (G) to pH 7.3. A and B lettering refers to the two conformers observed for the bound PA.

It was shown that the highly similar porcine (PLG) (62% identity, 83% similarity) and equine (ELG) (58% identity, 74% similarity) beta-lactoglobulins had neither fatty acids physiologically bound nor the ability to bind them *in vitro* at neutral pH (ref). In addition previous study performed in our laboratory indicated that PLG is unable to bind palmitic acid in the pH range 2-8 (paper 5 in Appendix).

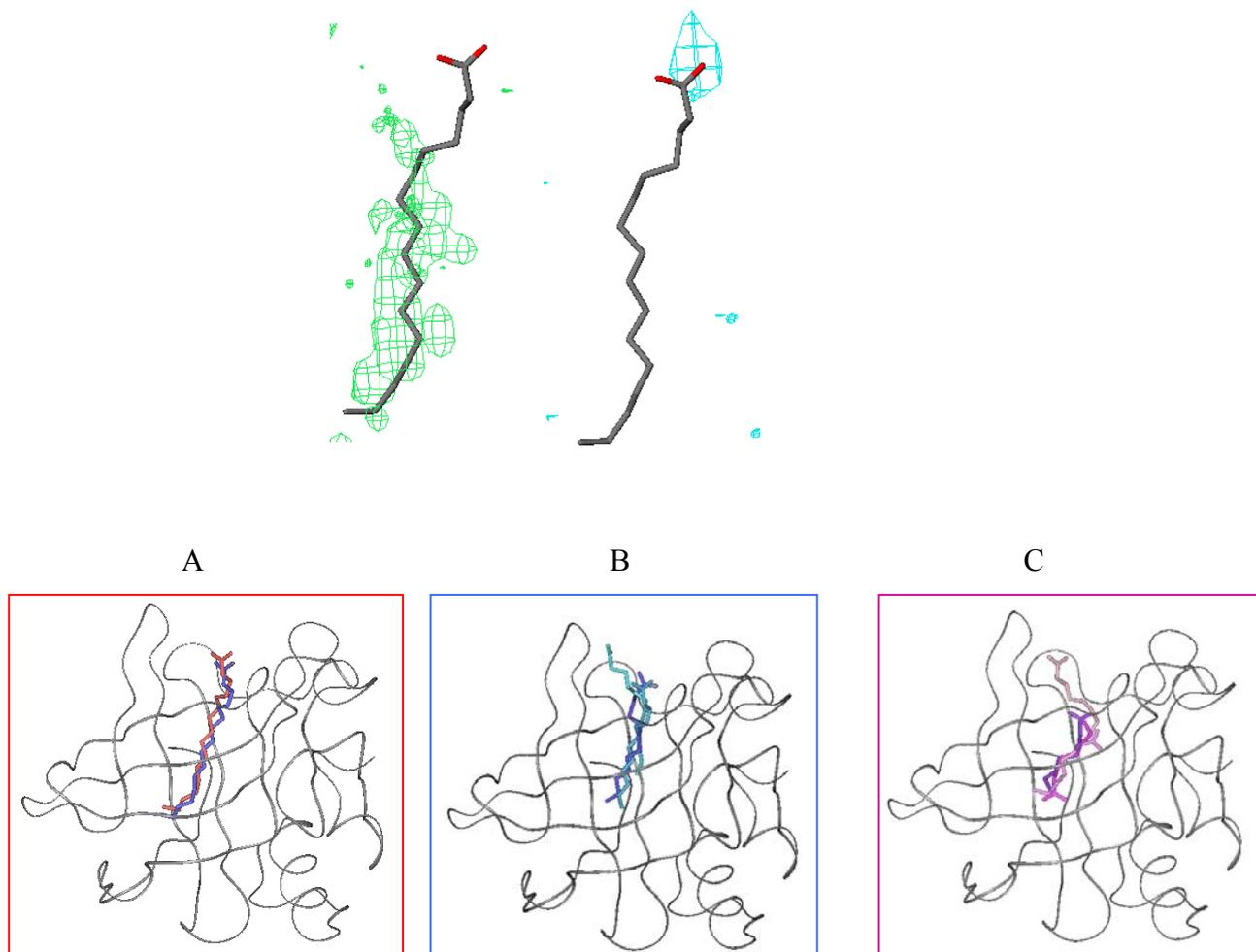
In order to better understand the mechanism of ligand binding I have studied i) the binding affinity of BLG for different fatty acids and ii) I have performed a comparative analysis of BLG and PLG binding capabilities.

Binding affinity of BLG for different chain length fatty acids was approached combining fluorescence studies, performed by the group of Giancarlo Baldini (University of Milano Bicocca)

with docking simulations experiments that I have implemented in the NMR Laboratory, after following the course “ESF training course on Molecular interactions: new frontiers for computational methods” held in Verona (20-25 July 2002)

The result of fluorescence studies, performed by exploiting the competition between fatty acids and ANS, a widely used fluorescent hydrophobic probe, revealed that in the pH range 6-8, which correlates with the calyx accessibility modulated by the conformation of the EF loop, fatty acids with longer aliphatic chains (palmitate and laurate) displayed larger affinities for BLG. These fluorescence data were complemented with docking simulations experiments, performed with the program GRID (29). The GRID method has been developed for determining energetically favourable binding sites for small chemical groups (probes) on a target molecule (protein) (29). The probe groups are small chemical entities, such as carboxy oxygens, water molecules, and hydrophobic probes, which are moved through a regular grid of points around the target molecule in order to calculate, at each point of the grid, an interaction energy, thus generating a molecular interaction field (MIF). The ligand molecule is represented as a collection of GRID probes, and MIFs calculated for each probe are used to define the ligand position with respect to the target and to estimate the binding energy of the ligand at its binding site. The docked solutions are ranked on the basis of the total energy of interaction with the target, evaluated by GRID forcefield, to estimate the binding affinity of each solution.

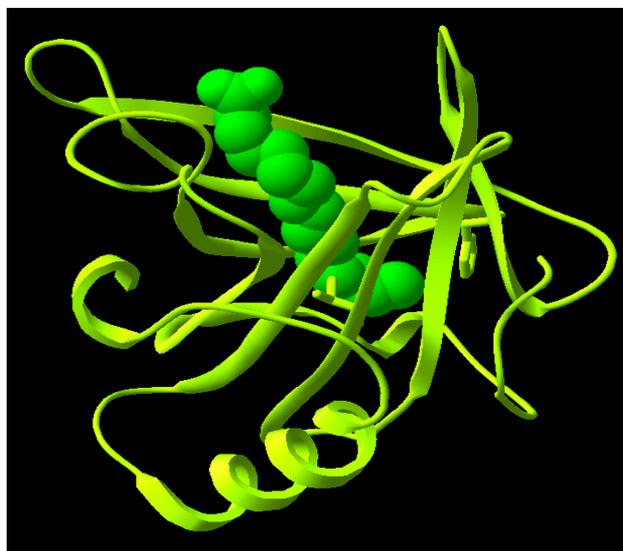
Initially I docked palmitic acid (PA), the most abundant BLG endogenous ligand, to all available BLG structures. GRID generated a docking solution for PA within the protein calyx only when the target BLG presented the EF loop in the open conformation (Figure 3.10). A very good agreement was observed for the positioning of both the aliphatic chain and the carboxyl tail of the docked and experimentally determined (PDB id: 1b0o) PA. The comparison of palmitic acid-binding energy (-14.8 Kcal/mole), with the most reliable docked solution obtained for lauric acid (-11.9 Kcal/mole), pointed to a lower affinity for this shorter fatty acid, in agreement with the fluorescence data showing a decrease of binding affinity with the shortening of the fatty-acid chain. When the shorter caprylic acid was docked to BLG, four solutions were obtained in which the aliphatic chains fit, for different extensions, the favorable hydrophobic interaction regions. However, the location of the carboxyl tail never fits the favorable electrostatic interaction regions, suggesting a very low-binding specificity of caprylic acid to BLG. Interestingly the obtained docking energies exhibited a very good agreement with the binding free energies detected by fluorescence.



**Figure 3.10** Graphical representation of molecular interaction fields in the region of palmitic acid binding site. Palmitic acid docking solution obtained from GRID program for interaction with BLG is shown in grey. The green and blue contours indicate regions of favourable interaction energies for hydrophobic (energy level= 0.5 Kcal/mole) and carboxyl oxygen (energy level = -6.2 Kcal/mole) probes, respectively (upper panel). Docking solutions (lower panels) obtained for interaction with BLG (PDB code 1bsy) with palmitic (A), lauric (B) and caprilic acid (C).

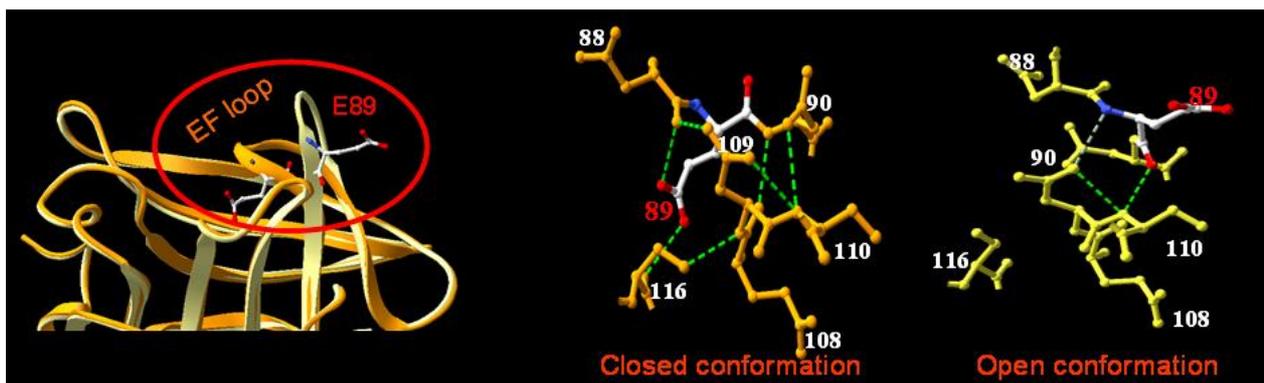
I have then performed docking experiments with the available PLG X-ray structure (PDB id: 1exs), obtained at pH 3.2, with the EF loop in the closed conformation. These simulations were never successful in locating PA within the protein cavity. I have then modelled the PLG structure with the EF loop in the open conformation and with this starting structure GRID could locate fatty acids within the hydrophobic pocket (Figure 3.11), clearly underlying the important role played by this loop in binding.

This result was important as it clearly pointed i) the role played by the EF loop in modulating the binding properties and ii) that only the comparative analysis of BLG and PLG binding mode would help in understanding the binding mechanism.



**Figure 3.11** Docking solution obtained from GRID program for interaction of PLG (PDB code 1exs) with palmitic

The inspection of BLG structure revealed that residues involved in the H-bond formation, stabilising the EF loop closed conformation are: D88, E89, N90, N109, S110 and S116 (Figure 3.12). The titration of E89 side-chain occurs at unusually high pH (ca.6.5) and indeed requires the loss of the H-bond pattern presented in figure 3.12, causing the fold back of EF loop, with a consequent solvent exposure of glutammic acid side-chain. If the conformational change of EF loop determines the binding ability, it is clear that all factors affecting E89 pK<sub>a</sub>, such as nature and charge of close residues, will modulate the pH of binding.



**Figure 3.12** The H-bonds stabilising the closed and open EF loop conformation are presented with green dotted lines.

pK<sub>a</sub> theoretical calculations were therefore performed starting from BLG and PLG experimentally determined structures and on the modelled structures of a few lactoglobulins (*lacc\_felca*, *laca\_canfa*, *lacb\_horse*, *laca\_horse*, *lacb\_macgi*) selected on the basis of their clustering in the phylogenetic tree.

The inspection of all beta-lactoglobulin structures allowed the identification of contributions due to mutations at the level of side-chains either participating to H-bonds or close in space to E89 side-chain. The analysis of the pK<sub>a</sub> data revealed that all those beta-lactoglobulins where the residues flanking E89 were mutated to a negatively charged residue exhibited an increased, calculated pK<sub>a</sub>. Thus E89 pK<sub>a</sub> computed for *lacb\_pig*, *lacc\_felca* and *lacb\_horse* were higher than those computed for *lacb\_bovin*, reflecting the mutation of the flanking N88 residue to a negatively charged aspartic acid. It is important to stress that *laca\_canfa*, showing a negative charged residue at position 88, similarly to *lacb\_pig*, *lacc\_felca* and *lacb\_horse*, exhibited a calculated pK<sub>a</sub> (5.64) similar to that obtained for *lacb\_bovin* (5.61). The analysis of the model showed that E88 side-chain in *laca\_canfa* pointed in the opposite direction with respect to E89, with the distance between the two carboxylic groups being the highest (8.5 Å) among all the analysed structures (in the range 3.5-6.0 Å), thus reducing electrostatic effects.

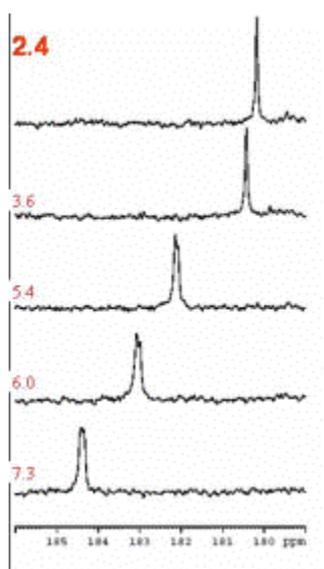
The strongest validation of these observations came from NMR titration experiments demonstrating that PLG is indeed able to bind PA but at pH higher than 8. The dependence of the amount of bound PA on pH can be fitted with a sigmoid curve with a transition midpoint shifted at pH 9.7, affording a clear indirect estimate of Glu89 pK<sub>a</sub>, in agreement with pK<sub>a</sub> computation.

Altogether these data have clearly indicated, for the first time, the molecular mechanism of ligand binding clarifying that the pH dependent conformational change of EF loop is a common feature of all beta-lactoglobulins and possibly of all lipocalins.

This analysis, while allowing the identification of structurally and functionally important electrostatic interactions in beta lactoglobulins, opened the way to the design of engineered proteins, characterised by binding capability in a selected pH range.

Indeed I started to analyse the binding properties of the mutant E89A, which was expressed from *Pichia pastoris* by the group of Marina Lotti (University of Milano Bicocca), collaborating with our laboratory for the setup of expression protocols from yeasts of beta-lactoglobulins and mutants production.

As shown in Figure 3.13 the titration of unlabelled BLGE89A with [1-<sup>13</sup>C] palmitic acid, followed by 1D <sup>13</sup>C NMR, revealed that this mutant was able to bind fatty acids in the pH range 2-8, differently from the native protein. Indeed when E89 is mutated to alanine, the EF loop, remains always in the open conformation.



**Figure 3.13:** Carboxyl region of the 125.7 MHz 1D proton decoupled <sup>13</sup>C spectra acquired on the complex of singly <sup>13</sup>C labeled [1-<sup>13</sup>C] PA and PLG at 37°C. The pH is indicated in red for each spectrum.

Interestingly the binding mechanism identified for beta-lactoglobulins is reminiscent of the mechanism of lipases interfacial activation associated with a conformational change, in which a lid, consisting of one  $\alpha$ -helix, opens up by rotating around its hinge regions, thus allowing ligand binding. As it was reported that: i) beta-lactoglobulins may play a role in increasing lipases activity, by removing free fatty acids and ii) liver pig triacylglycerol hydrolases exhibited activity at an

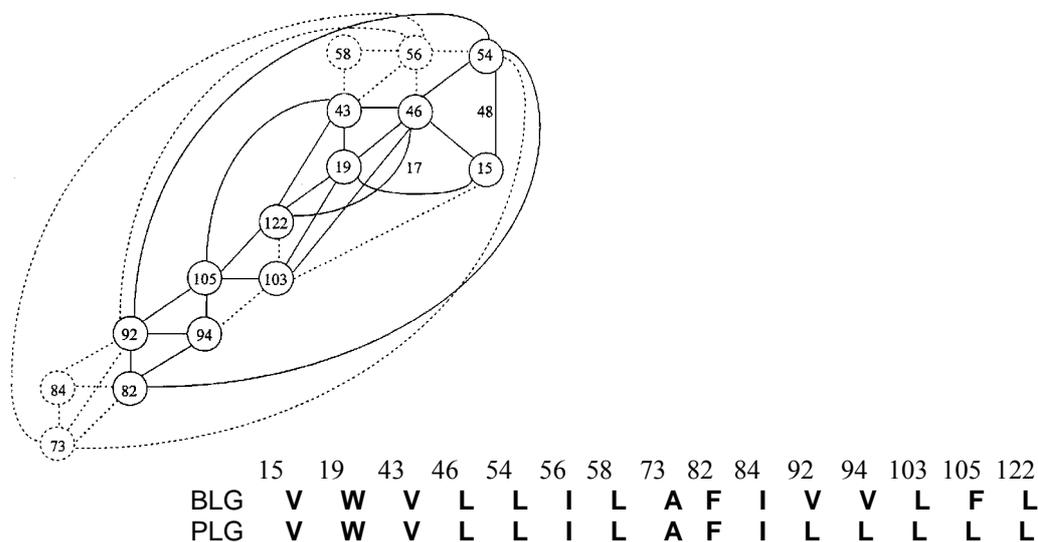
optimum alkaline pH of 8.5, it is likely that the unusually high pH of binding, detected for PLG, is functional to lipases activity. This result confirms the function of beta-lactoglobulins as related to activation of lipases.

### *Work in progress and future perspectives*

Several mutants of BLG and PLG have been designed and are under production or will be produced. The rationale of their design is described in the following:

The designed mutants are:

1. BLG K69E and PLG E69K: they will be used in binding studies to verify the proposed key role of K69 in the interaction with fatty acids. This analysis will also clarify the importance of electrostatic interactions as driving force for the binding.
2. PLG L105F, L92V, L94V: the different stability of porcine and bovine beta lactoglobulins may be ascribed to mutations in the hydrophobic cluster (Figure 3.14) (3) identified in BLG as an important element for the stability of the protein. Among the seventeen residues belonging to this cluster, only 105, 92 and 94 are conservatively mutated in porcine with respect to BLG. The three mutants will be studied by unfolding equilibrium experiments with the aim of obtaining new information about stability and folding properties of beta-lactoglobulins.



**Figure 3.14** Schematic summary of the connectivities defining the buried hydrophobic cluster. NOE observed among the side chains of the cluster residues are indicated by continuous lines, while predicted NOEs (on an X-ray basis) are indicated with dotted lines. Sequence alignment of BLG and PLG for the residues belonging to the hydrophobic cluster is reported.

3. BLG Q35K: we have observed that at pH 7 the dimer interface of BLG displays rather neutral electrostatic potential, while the same region in PLG is highly positive, mainly due to the positively charged residue K35. The mutation of the neutral Q35 into K in BLG should therefore produce the same high positive electrostatic potential and disfavour the dimerization in this position at any pH, as it occurs for the porcine protein. In this way it would be possible to obtain monomeric BLG at neutral pH.

**This project led to the publication of the following papers:**

- M. Collini, L. D'Alfonso, G. Baldini, M. Catalano, L. Ragona and H. Molinari, Competitive binding of fatty Acids and the fluorescent probe 1-8-anilinonaphtalene sulfonate to bovine beta-lactoglobulin, **Protein Sci.**, 12, 1596-603, 2003. (Paper N° 4 in the Appendix)
- L. Ragona, M. Catalano, R. Ugolini, F. Fogolari, L. Zetta and H. Molinari, EF loop conformational change triggers ligand binding in beta-lactoglobulins, **J. Biol. Chem.**, 278, 38840, 2003. (Paper N° 5 in the Appendix)
- S. Brocca, E. Pedrazzoli, L. Ragona, M. Catalano H. Molinari and M. Lotti, Heterologous expression of bovine and porcine b-lactoglobulins in *Pichia pastoris*: towards a comparative functional characterization, **J. Biotechnol.**, 109, 169-178, 2004. (Paper N° 6 in the Appendix)
- L. Ragona, G. Colombo, M. Catalano and H. Molinari, Determinants of protein stability and folding: comparative analysis of beta-lactoglobulins and liver basic fatty acid binding protein, **PROTEINS**, in press, 2005. (Paper N° 7 in the Appendix)

My contribution to paper N° 4 consisted in the docking simulations using the program GRID. I have actively participated to all the stages of the work leading to paper N° 5, while my contribution to paper N° 6 was limited to the NMR control of produced recombinant proteins. For paper N° 7 I contributed to the general analysis of MD runs and comparisons with experimental data.

## References

1. Perez, M.D, and Calvo, M., (1995), *J. Dairy Sci.*, **78**, 978-988
2. Brownlow, S., Morais, J.H., Cabral, Cooper, R., Flower, D.R., Yewdall, S.J., Polikarpov, I., North, T., and Sawyer, L.,(1997) *Structure* **5**,481-495.
3. Ragona, L., Pusterla, F., Zetta, L., Monaco, H., and Molinari, H., (1997) *Folding & Design* **2**, 281-290.
4. Fogolari, F., Ragona, L., Zetta, L., Romagnoli, S., de Kruif, K.and Molinari, H., (1998) *FEBS Lett.* **436**, 149-154.
5. Kuwata, K., Hoshino, M., Forge, V., Era, S., Batt, C., and Goto, Y., (1999) *Protein Sci.* **8**, 2541-2545.
6. Kuwata, K., Hoshino, M., Era, S., Batt, C., and Goto, Y., (1998) *J.Mol. Biol.* **283**, 731-739.
7. Dobson, C. M.,(1999)*Trends Biochem Sci.* **24**, 329-3.
8. Ragona, L., Fogolari, F., Romagnoli, S., Zetta, L., Maubois, J.L.and Molinari, H.,(1999) *J. Mol. Biol.* **293**,95 3-969.
9. Hosszu, L.P., Baxter, N.J., Jackson, G.S., Power, A., Clarke, A.C., Waltho, J.P., Craven, C.J and Collinge, J., (1999) *Nature Struct. Biol.* **6**, 740-743.
10. Di Pietro, S.M., Veerkamp, J.H. and Santomé, J.A.. (1999) *Eur. J. Biochem.* **259**, 127-134.
11. Schievano, E., Mammi, S.and Peggion, E., (1999) *Biopolymers*, **50**, 1-11.
12. Wu ,S.Y., Perez, M.D, Pujol, P. and Sawyer, L., (1999) *J. Biol. Chem.* **274**,170-174,.
13. Qin, B.Y., Creamer, L.K., Baker, E.N and Jameson, G.B., (1998) *FEBS Lett.***438**,272-278.
14. Zidek, L., Novotny, M.V. and Stone, M.J., (1999) *Nature Struct. Biol.* **6**, 1118-1121.
15. Fogolari, F., Ragona, L., Licciardi, S., Romagnoli, S., Michelutti, R., Ugolini, R. and Molinari, H., (2000), *Proteins: Structure, Function and Genetics.* **39**, 317-330.
16. Fogolari, F. and Briggs, J., (1997) *Chemical Physics Letters* **281**, 135-139.
17. Fogolari, F., Zuccato, P., Esposito, G. and Viglino, P., (1999) *Biophys. J.* **76**, 1-16.
18. Fogolari, F., Elcock, A.H., Esposito, G., Viglino, P., Briggs, J. and McCammon, J.A., (1997) *J. Mol. Biol.*, **267**, 368-381.
19. Baginski, M., Fogolari, F. and Briggs, J. M., (1997) *J. Mol. Biol.*, **274**, 253-267.
20. Fogolari, F., Esposito, G., Viglino, P., Briggs, J. and McCammon, J.A., (1997) *J. Mol. Biol.* **267**, 368-81.
21. Flower, D.R., North, A.C.T., Sansom, C. E., *Biochimica et Biophysica Acta* (2000), **1482** , 9-24
22. Sugai, S., Ikeguchi, M., Shimizu, A., (1999) *Int. J. Food Sci. Technol.* **34** 437-444.

23. Perez, M. D., Sanchez, L., Aranda, P., Ena, J. M., Oria, R. & Calvo, M., (1992) *Biochim. Biophys. Acta*, **1123**, 151–155
24. Tiana, G, Simona, F, De Mori, GM, Broglia RA, Colombo, G., (2004) *Protein Sci* **13**, 113-124.
25. Sawyer, L., Kontopidis, G., (2000) *Biochimica et Biophysica Acta* **1482** 136-148
26. Qin, B.Y., Creamer , L.K., Baker, E.N., Jameson, G.B., (1998) *FEBS Lett.*, **438**, 272-278.
27. Wu, S.Y., Perez, M.D, Puyol, P., Sawyer, L., (1999) *J. Biol. Chem.*, **274**, 170-174.
28. Ragona, L., Catalano, M., Zetta, L., Longhi, R., Fogolari, F., Molinari, H., (2002) *Biochemistry*, **41**, 2786-2796
29. Kastenholtz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E., Fox, T., (2000) *J. Med. Chem.*, **43**, 3033-3044.
30. Ugolini, R., Ragona, L., Silletti, E., Focolari, F., Visschers, R. W., Alting, A. C., Molinari, H., (2001) *Eur. J. Biochem.* **268**, 4477-88.

## CHAPTER 4

### 4.1 Bile Acid Binding Protein (cl-BABP)

Chicken liver basic fatty acid-binding protein (Lb-FABP), more accurately described as chicken liver bile acid binding protein (cL-BABP) for its role as bile acid transporter (paper number 8 in the appendix) belongs to the conserved multigene family of the intracellular lipid binding proteins (iLBP) having molecular masses around 15 kDa. Members of this family have been found throughout the animal kingdom, both in invertebrates and vertebrates. It is believed that individual genes of this ubiquitous gene family arose from an ancestral iLBP gene through gene duplication and diversification [1-3]. As iLBPs are not present in plants or fungi, it has been suggested that the ancestral gene has likely evolved after the animal kingdom had separated from plants and fungi, more than 1000 million years ago. The first gene duplication may have appeared well before the vertebrate/invertebrate divergence (1-3), and hence members of both branches are present in vertebrates and invertebrates. Subsequent duplications gave rise to the four subfamilies recognized today, which contain the different iLBPs found in vertebrates species (4) summarised in Table 4.1. Subfamily I, which comprises proteins specific for vitamin A derivatives, can be subdivided into the cellular retinoic acid-binding proteins (CRABP-I and II) and the cellular retinol binding proteins (CRBP-I, II, III, and IV) (5).

Subfamily II contains proteins with larger binding sites that allow binding of bulkier ligands, such as bile acids, eicosanoids, and heme. In addition to the intestinal bile acid-binding protein (I-BABP) and the liver-type (L-) FABP, this subfamily contains the basic liver-type (Lb-FABP), the only iLBP that is not expressed in mammals and which is the object of my research work. It is found in the liver of birds, fish, reptiles, and amphibians (6). It has been recently proposed to change the nomenclature basic liver-type, into the more general BABP (7).

Intestinal-type (I-) FABP is the only member in subfamily III, while the remaining proteins (heart-type (H-), adipocyte-type (A-), epidermal-type (E-), myelin-type (M-), testis-type (T-), and brain-type (B-) FABP) belong to subfamily IV.

The various FABPs are still named after the tissues in which they have been discovered, or are prominently expressed. It has been recognized, however, that such a classification is somewhat misleading, since no FABP is specific for a given tissue, and most tissues express various FABP-types. L-FABP, for example, is strongly expressed in the liver, and was thought to be the only FABP expressed in this tissue, at least in adult animals (recent studies have reported the presence of E-FABP, I-FABP, and A-FABP in liver as well (8,9)). But L-FABP is also found, albeit in far smaller concentrations, in the intestine, kidney, lung, and pancreas. In contrast, I-FABP and I-

BABP are confined to the digestive tract and not expressed prominently in other tissues (10). H-FABP and E-FABP are the most ubiquitously expressed iLBPs; the former is most prominent in cardiac and skeletal muscle, but also present in kidney, lung, mammary tissue, placenta, testis, stomach, ovary, and the brain (11). The latter protein is widely expressed in skin, lung, heart and skeletal muscle, kidney, testis, and adipose tissue, as well as in the brain and the retina. In addition to H- and E-FABP, the nervous system contains two other FABPs, B-FABP that is present in brain and retina, and the M-FABP that seems to be specific for the Schwann cells forming the myelin sheath (12,13). A-FABP was originally thought to be confined to adipocytes (14) but has been detected in macrophages as well (15). In most cases, the expression pattern is similar in all vertebrates, with the notable exception that non-mammalian vertebrates express in their livers a basic liver-type FABP (Lb-FABP), as already mentioned, that is different from the mammalian L-FABP.

Table 4.1 contains the classification reported in the literature for the different iLBPs.

Table 4.1:

FABPs as members of the intracellular lipid binding protein family

iLBP-type	Gene	Expression	Ligands
<i>Subfamily I</i>			
CRBP I, II, III, and IV		Ubiquitous in mammalian cells	Retinol
CRABP I and II		Ubiquitous in mammalian cells	Retinoic acids
<i>Subfamily II</i>			
L-FABP	<i>Fabp1</i>	Liver, intestine, kidney, lung, and pancreas	Long-chain fatty acids, acyl-CoAs, and heme
I-BABP	<i>Fabp6</i>	Ileum	bile acids
Lb-FABP	<i>Fabp10</i>	Fish and bird liver	Long-chain fatty acids
<i>Subfamily III</i>			
I-FABP	<i>Fabp2</i>	Intestine, liver	Long-chain fatty acids
<i>Subfamily III</i>			
H-FABP	<i>Fabp3</i>	Heart, skeletal muscle, brain, kidney, lung, mammary, placenta, testis, ovary, and stomach	Long-chain fatty acids
A-FABP	<i>Fabp4</i>	Adipose tissue and macrophages liver	Long-chain fatty acids
E-FABP	<i>Fabp5</i>	Skin, adipose tissue, lung, brain, heart, skeletal muscle, testis, retina, and kidney	Long-chain fatty acids
B-FABP	<i>Fabp7</i>	Brain, glia cells, and retina	Long-chain fatty acids and docosahexacenoic acid
M-FABP	<i>Fabp8</i>	Brain and Schwann cells	Long-chain fatty acids
T-FABP	<i>Fabp9</i>	Testis	Long-chain fatty acids

FABPs are highly soluble in water and bind ligands that have low solubility in water at neutral pH, mainly fatty acids but also bind lipids such as bile acids.

FABPs have generally been easy to purify and crystallize for X-ray analysis. In addition, the low molecular weight and low helix content make them very amenable to structural determinations by NMR. A recent literature survey found 11 NMR solution structures and 28 X-ray crystal structures of FABPs (not including the closely related intracellular retinoid-binding proteins) with very high resolution (16). Even though the homology of their amino acid sequences is variable and as low as 25%, all family members have a similar structural motif. The important structural features common to FABPs are illustrated by the solution structure of chicken liver bile acid binding protein that has been determined as described in paper 8 in appendix (Figure 4.1). FABPs have a common  $\beta$ -barrel structure, which is formed by two orthogonal five-stranded  $\beta$ -sheets. The binding pocket is located inside the barrel, the opening of which is framed on one side by the N-terminal helix–turn–helix domain. With few exceptions FABPs do not contain disulphide bridges.



**Figure 4.1** Ribbon drawing showing the DYANA structure of cl-BABP with the least violations of NOE distance constraints after simulated annealing and subsequent energy minimisation (PDB code 1MVG) (paper 8 in the appendix)

NMR studies can reveal aspects of the dynamics and/or conformational variability of the proteins. Specifically relaxation measurements of the protein backbone can reveal the dynamics in all parts of the protein and shed light on how ligands might enter and exit (17-19).

In this line my research approach to the study of recombinant cl-BABP was focussed on the determination of the dynamics property of this protein in its apo and holo form.

The procedure for the expression protocol to obtain doubly labelled ( $^{13}\text{C}$  and  $^{15}\text{N}$ ) cl-BABP was set up in the NMR Laboratory, starting from a clone provided by J. Foote (Fred Hutchinson Cancer Research Centre, Seattle WA 98109, USA).

Interestingly the recombinant protein exhibited different secondary shifts and flexibility, in its C-terminal end, with respect to the purified protein, previously studied by simple  $^1\text{H}$  NMR spectroscopy in our NMR Laboratory. A group of Japanese researchers identified and deposited in the Swiss-Prot data bank the sequence of a genetic variant of the protein displaying the mutation T91C. This mutation led to the formation of one disulphide bridge between C80 and C91 (H. Monaco, personal communication). As a result the purified and recombinant proteins are different, as the clone was that of the genetic variant containing a T at position 91.

A multiple sequence alignment among the other known non mammalian liver basic FABP indicated that position 91 only in few cases is a cysteine (6).

As a consequence of structural and stability differences between the recombinant and purified protein, I had to reassign again the NMR spectra. Furthermore, given the high flexibility of the recombinant protein, the structural assignment requested a high number of 3D heteronuclear experiments performed on the doubly labelled sample both in  $\text{H}_2\text{O}$  and in  $\text{D}_2\text{O}$  at 700 MHz. (For all the details of the NMR experiments, see Chapter 5).

All the work relative to this wide part of my thesis, concerning the assignment, the structure determination of apo cl-BABP and the dynamic measurements performed on the apo and holo proteins (complexed with bile acids) are described in the following paragraph (4.1). I have organised this part already in the form of a manuscript to be submitted, however inserting a lengthy discussion of the results and all the Tables and Figures necessary to clearly describe and report all the work done, especially on protein dynamics. Indeed it is important to stress that proteins are dynamic molecules that often undergo conformational changes while performing their specific functions, such as ligand binding. NMR spectroscopy can be used to monitor the dynamic behaviour of a protein at a multitude of specific sites. Protein movements on a broad range of timescales can be monitored using various types of NMR experiments, as described in Chapter 5.

My contribution to this research project started with a stage in Roma at the University of Tor Vergata to learn how to set up all the 3D experiments needed for structure determination of  $^{13}\text{C}$ - $^{15}\text{N}$  doubly labelled proteins, as described in Chapter 5. The knowledge acquired in the group of Prof. Daniel Cicero was brought back to our laboratory and all the 3D sequences were installed on the 500 MHz spectrometer of the University of Verona. The 3D spectra were reacquired on apo and holo cl-BABP.

I learned on the acquisition and analysis of relaxation data at the EMBO Course: “Multidimensional NMR in structural biology” that I have attended at il Ciocco (Lucca, August 11-16, 2002).

**Up to now this project has led to the publication of the paper:**

Vasile, F., Ragona, L., Catalano, M., Zetta, L., Perduca, M., Monaco, H., Molinari H., Letter to the editor: solution structure of chicken liver fatty acid binding protein, (2003) **J. of Biomolecular NMR**, 25, 157-160 (Paper N° 8 in the Appendix)

## References

1. Matarese, V., Stone, R.L., Waggoner, D. W., Bernlohr, D. A. (1989) *Prog Lipid Res*, **28**, 245-72.
2. Schleicher, C. H., Cordoba, O. L., Santome, J. A., Dell 'Angelica, E.C., (1995) *Mol Biol Int* **36**,1117-25.
3. Schaap, F. G., van der Vusse, G. J., Glatz, J. F. C., (2002) *Mol Cell Biochem*, **239**, 69-77.
4. Hanho, T., Lucke, C., Spener, F., (2002) *Mol Cell Biochem*, **239**, 45-54.
5. Ross, A. C., (1993) *FASEB J*, **7**, 317-27.
6. Di Pietro, S. M., Veerkamp, J. H., Santome, J. A.,. (1999) *Eur J Biochem.*, **259**, 127-34.
7. Nichesola, D., Perduca, M., Capaldi, S., Carrozzo, M. E., Rigetti, P. G., Monaco, H. L., (2004) *Biochemistry*, **43**, 14072-9.
8. Yu, L. R, Zeng, R., Shao, X. X., Wang, N., Xu, Y. H., Xia, Q.C., (2000) *Electrophoresis*, **21**, 3058-68.
9. Owada, Y., Suzuki, I., Noda, T., Kondo, H., (2002) *Mol Cell Biochem*, **239**, 83-6.
10. Agellon, L. B, Toth, M. J., Thomson, A. B (2002) *Mol Cell Biochem*, **239**, 79-82.
11. Zanotti, G., (1999) *Biochim Biophys Acta*, **1441**, 94-105.
12. Veerkamp, J.H., Zimmerman, A. W., (2001) *J Mol Neurosci*, **16**, 133-42.
13. Owada, Y., Kondo, H., Wiley-VCH; 2003, 253 .66.
14. Bernlohr, D. A, Simpson, M. A., Vogel Hertzal, A., Banaszak, L. J., (1997) *Annu Rev Nutr*, **17**, 277-303.
15. Fu, Y. C., Luo, N. L., Lopes-Virella, M. F., (2000) *J Lipid Res*, **41**, 2017-23.
16. Cohn, S. M., Simon, T. C., Roth, K. A., Birkenmeier, E. H., Gordon J. (1992) *J Cell Biol*, **119**, 27-44.
17. Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I., Spiegelman, B. M., (1994) *Genes Dev.*, **8**, 1224-34.
18. Issemann, I., Prince, R., Tugwood, J., Green, S., (1992) *Biochem Soc*, **20**, 824-7.
19. Schachtrup, C., Emmler, T., Bleck, B., Sandqvist, A., Spener, F., (2004) *Biochem J* [in press].
20. Lueke, C., Zhang, F., Hamilton, J. A., Sacchettini, J. C., Ruterjans, H., (2000) *Eur. J. Biochem.* **267**, 2929-2938.
21. Beringhelli, T., Goldoni, L., Capaldi, S., Bossi, A., Perduca, M., Monaco, H., (2001) *Biochemistry*, **40**, 12604-12611.
22. Luecke, C., Zhang, F., Ruterjans H., Hamilton, J. A., Sacchettini J. C., (1996) *Structure*, **4**, 785-800.

## **4.2 NMR structure of chicken liver bile acid binding protein and dynamics characterisation of its apo and holo forms**

M. Catalano<sup>1,2</sup>, L. Ragona<sup>2</sup>, M. Luppi<sup>1</sup>, D. Cicero<sup>3</sup>, J. Foote<sup>4</sup>, F. Fogolari<sup>1</sup>, L. Zetta<sup>2</sup> and H. Molinari<sup>1</sup>

<sup>1</sup>Dipartimento Scientifico e Tecnologico, Strada Le Grazie 15, 37134 Verona, Italy; <sup>2</sup>Laboratorio NMR, ISMAC, CNR, via Bassini 15, 20133 Milano, Italy; <sup>3</sup> Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 00133 Roma, Italy; <sup>4</sup>Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

Key words: NMR structure solution, Relaxation measurements, model-free analysis, reduced spectral density mapping, bile acid binding proteins.

Abbreviations: cl-BABP (chicken liver bile acid binding protein), het-NOE (heteronuclear Nuclear Overhauser Effect)

## Introduction

Bile acids are generated *in vivo* from cholesterol in the liver, and they undergo an enterohepatic circulation involving the small intestine, liver, and kidney. To understand the molecular mechanism of this transportation, it is essential to gain insight into the three-dimensional (3D) structures and dynamics behaviour of proteins involved in the bile acid recycling in free and complexed form and to compare them with homologous members of this protein family (Tochtrop et al., 2003).

As a first step in this direction we report here the solution structure and the dynamics of recombinant chicken liver bile acid binding protein (cl-BABP) at two pHs (7 and 5.6) at 25°C and compare the results with those previously reported by us at pH 5.6 and 37°C for the same protein purified from chicken liver (Vasile et al., 2003). It should be mentioned that the purified protein was initially referred to as liver basic fatty acid binding protein, due to its high isoelectric point. We have subsequently shown, on the basis of sequence alignment comparisons, that bile acids are expected to be the putative ligands of this protein, which was hence addressed as chicken liver bile acid binding protein (cl-BABP) (Nichesola et al., 2004; Vasile et al., 2003).

The purified and recombinant proteins display different  $^1\text{H}$  chemical shifts in a localised region, and different solution stability with time. We show here that recombinant apo chicken liver BABP does not exhibit pH dependent conformational changes in the pH range 5.6-7 and all the differences observed between the recombinant and purified protein have led to the awareness that the purified protein was a different genetic variant, displaying the T91C mutation which resulted in the presence of an additional stabilising disulphide bridge (position 80-91) (Monaco, personal communication). The presence of two genetic variants of this protein is consistent with a new sequence submitted to the EMBL/GenBank/DDBJ Databases (Murai, A., Kusumoto, K., Okumura, J. Regulation of gene expression of two liver type fatty acid binding proteins in chicken. Submitted may 2001 to the EMBL/GenBank/DDBJ Databases).

The results of our structural calculations are discussed here in relation to the family of homologous bile acid binding proteins. The dynamics behaviour of the apo and holo cl-BABP (complexed with

chenodeoxycholic acid) is addressed here, as it plays a very important role in protein function, in general, and, specifically, in iLBP family. Indeed these proteins, in spite of the highly conserved topology, display different stability and binding affinities which have been suggested to be determined by the dynamics properties (Bakowies and van Gunsteren, 2002).

## **Materials and methods**

*Protein expression and purification.* Recombinant cl-BABP was expressed as soluble protein in *E. coli* BL21 (DE3) bearing the recombinant plasmid pET24d. The plasmid was kindly provided by Dr. Jefferson Foote (Fred Hutchinson Cancer Research, Seattle, USA).

Transformed cells were grown on plates containing 50 µg/ml kanamicin. A 150 ml overnight culture was grown from a freshly streaked out plate and 15 ml were used to inoculate 1 L TB. These were incubated at 37 °C until cell density reached an OD<sub>600</sub> of 0.8. Protein expression was induced by addition of 0.5 mM isopropylthiogalactopyranoside (IPTG) and incubation continued overnight at 20 °C.

The cells were harvested by centrifugation for 30 minutes at 8000 x g. The pellets were collected and stored at -20°C to facilitate cell lysis, and re-suspended in lysis-buffer (50 mM Tris, 10% sucrose, 1 mM EDTA, 10 mM β-mercaptoethanol, pH 8.0). The solution was subjected to sonication and centrifuged at 7000 x g for 45 minutes, causing cellular debris to be pelleted.

The supernatant, containing cl-BABP, was loaded on a DEAE cellulose (Whatman) anion exchange column equilibrated with 50 mM Tris-acetate at pH 7.8. The same buffer was used for protein elution at 1 ml/min at 4 °C. Fractions containing Lb-FABP were concentrated using an Amicon concentrator and resolved on a Sephacryl S-100 HR (Amersham Biosciences) column equilibrated with 50 mM Tris-HCl, 0.2 M NaCl at pH 7.2. Elution was performed at 1ml/min at 4 °C.

cl-BABP was delipidated at 37 °C using a Lipidex 1000 column (Perkin Elmer), equilibrated with 10 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> at pH 7.4 (Glatz and Veerkamp, 1983). A maximum of 15 mg of cl-BABP at 1 mg/ml was loaded on a XK14-20 column (Amersham Biosciences). Elution was

performed at 0.2 ml/hour. The protein purity was demonstrated by the presence of a single band on SDS-PAGE. The protein yields were 90 mg/L of bacterial culture.

$^{15}\text{N}$  isotope labeling of cl-BABP was achieved using M9 minimal media containing 1gr/L  $^{15}\text{NH}_4\text{Cl}$ . and The protocol for the  $^{15}\text{N}$  expression was developed following protocols reported in the literature (Marley et al., 2001). *E. coli* were grown at 37 °C until cell density reached an  $\text{OD}_{600}$  of 0.8. Cells were pelleted by centrifugation at 5000 x g for 30 minutes at 4°C, then washed with a M9 solution without  $^{15}\text{NH}_4\text{Cl}$  and finally pelleted again by centrifugation. The cell pellet was re-suspended in M9 containing  $^{15}\text{NH}_4\text{Cl}$  and inoculated in 250 ml of  $^{15}\text{N}$  enriched minimal media. After one hour shaking at 37 °C to allow for the recovery of growth and the clearance of unlabeled metabolites, protein expression was induced by addition of 0.5 mM IPTG. After induction, incubation continued for 4 hours at 37 °C. The cell lysis and the purification procedure were the same as those for unlabelled cl-BABP. The extent of labeling was verified by MALDI mass analysis and the isotope incorporation was found to be more than 92%.  $^{15}\text{N}$ -cl-BABP was obtained in a yield of 50 mg/L of minimal media.  $^{13}\text{C}$   $^{15}\text{N}$  double labelling was obtained with the same procedure growing cells using M9 minimal media containing 1gr/L  $^{15}\text{NH}_4\text{Cl}$  and 8gr/L  $^{13}\text{C}$  enriched sucrose. The extent of labelling, verified by MALDI mass analysis was > 87% and yields of 25 mg/L of minimal media were obtained.

Commercial chenodeoxycholic acid (Sigma-Aldrich) was employed for the preparation of holo cl-BABP with a ligand to protein ratio 5:1, as previously described (Tochtrop et al., 2004).

*NMR experiments.* The NMR data were recorded on Bruker Avance 500, 600 and 700 MHz spectrometers equipped with pulse field gradient triple-resonance probes. 1 mM protein samples in phosphate buffer at pH 7 and 298K were employed for structure determination and relaxation measurements. Acetate buffer solutions were employed for the preparation of samples at pH 5.7.

The following triple resonance experiments (Pelton et al., 1991) were recorded on [ $^{15}\text{N}$ ,  $^{13}\text{C}$ ] apo and holo cl-BABP in  $\text{H}_2\text{O}$  at 700 MHz at pH 7.0: HNCA, HN(CO)CA [120 ( $^{15}\text{N}$ ) × 84 ( $^{13}\text{C}$ ) × 1024 ( $^1\text{H}$ ) complex data points, 4 scans], HNCO [120( $^{15}\text{N}$ ) × 52 ( $^{13}\text{C}$ ) × 1024 ( $^1\text{H}$ ) complex data points, 4

scans], CBCANH and CBCA(CO)NH [ $96 (^{15}\text{N}) \times 108 (^{13}\text{C}) \times 1024 (^1\text{H})$  complex data points, 8 scans]. On the doubly-labelled sample dissolved in  $\text{D}_2\text{O}$ , the following experiments (Powers et al., 1991) were recorded: HACACO [ $256 (^{13}\text{C}) \times 44 (^{13}\text{C}) \times 1024 (^1\text{H})$  complex data points, 4 scans], (H)CCH-COSY and (H)CCH-TOCSY [ $248(^{13}\text{C}) \times 172 (^{13}\text{C}) \times 1024 (^1\text{H})$  and  $200(^{13}\text{C}) \times 200 (^{13}\text{C}) \times 1024 (^1\text{H})$  complex data points, 4 scans, respectively], H(C)CH-COSY and H(C)CH-TOCSY [ $248(^{13}\text{C}) \times 148 (^1\text{H}) \times 1024 (^1\text{H})$  and  $148 (^1\text{H}) \times 200(^{13}\text{C}) \times 1024 (^1\text{H})$  complex data points, 4 scans, respectively]. Two NOESY-type 3D experiments were acquired (mixing 100 ms), one optimised for aliphatic and one for aromatic residues [ $80 (^{13}\text{C}) \times 116 (^1\text{H}) \times 1024 (^1\text{H})$  and  $88 (^{13}\text{C}) \times 120 (^1\text{H}) \times 1024 (^1\text{H})$  complex data points, 16 and 8 scans, respectively].

$^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY (Marion et al., 1989) [mixing 85 ms,  $160(^1\text{H}) \times 48 (^{15}\text{N}) \times 1024 (^1\text{H})$  complex data points] and  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY (Marion et al., 1989) [mixing 150 ms,  $224 (^1\text{H}) \times 48 (^{15}\text{N}) \times 1024 (^1\text{H})$  complex data points, 16 scans] and HNHA (Kuboniwa et al., 1994) [ $128 (^1\text{H}) \times 48 (^{15}\text{N}) \times 1024 (^1\text{H})$  complex data points, 16 scans] experiments were acquired on the  $^{15}\text{N}$  enriched cl-BABP at 500 MHz.

2D homonuclear TOCSY (mixing 70 ms) and NOESY (mixing 150 ms) were performed at 700 MHz on cl-BABP: data matrices contained  $2048 \times 1280$  and  $2048 \times 1024$  points for NOESY (32 scans) and TOCSY (48 scans), respectively. Water suppression was achieved using the excitation sculpting sequence (Prost et al., 2002) for TOCSY and WATERGATE (Piotto et al., 1992) for NOESY.

2D homonuclear TOCSY and 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were also performed on  $^{15}\text{N}$  enriched cl-BABP at pH 5.7 at 298 and 310K, i.e. in the same conditions as those previously reported for the purified protein (1), in order to carefully compare the two structures and to investigate the presence of pH dependent conformational equilibria. Spectra acquired at low pH were helpful to detect amide resonances in fast exchange with solvent at pH 7.

$^{15}\text{N}$  relaxation experiments (Farrow et al., 1994), run as water flip-back version, were acquired at 600 and 700 MHz both at pH 7 and 5.7 with  $128 (^{15}\text{N}) \times 2048 (^1\text{H})$  complex points. Eleven delays (2.5, 20, 60, 100, 150, 200, 300, 400, 600, 800, 1000 ms) were used for  $R_1$  measurements and nine delays (16.96, 33.92, 50.80, 67.84, 101.76, 135.68, 169.6, 220.48, 237.44 ms) for  $R_2$  measurements. The delay in the CPMG pulse train was set to 0.45 ms. Relaxation times were calculated *via* least squares fitting of peak intensities.

$^1\text{H}$ - $^{15}\text{N}$  NOE experiments were acquired with  $96 (^{15}\text{N}) \times 2048 (^1\text{H})$  complex points with an overall delay of 6 s (Farrow et al., 1994).

Relaxation measurements were identically performed at pH 7 for holo-cL-BABP complexed with chenodeoxycholic acid.

Data were processed with XWINNMR and NMRPipe (Delaglio et al., 1995) and analysed with NMRView 5.0.3 software package (Johnson, 2004).

*Structure calculation of apo cl-BABP:* Volume integration was performed on the 3D  $^{15}\text{N}$ - $^{13}\text{C}$  NOESY and  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY spectra using NMRView (Johnson, 2004). Peak volumes calibration was performed using the median method, a routine of NMRView program, and the obtained list of distances was used as input for DYANA (Guntert et al., 1997) calculations.  $\phi$  angle restraints were derived from  $J_{\text{HN},\text{H}\alpha}$  coupling constants estimated from 3D HNHA experiments (Kuboniwa et al., 1994).  $\phi$  angle restraints of  $139^\circ \pm 30^\circ$  for  $J_{\text{HN},\text{H}\alpha}$  coupling constants greater than 8.0 Hz and  $60^\circ \pm 30^\circ$  for  $J_{\text{HN},\text{H}\alpha}$  coupling constants smaller than 5.0 Hz were used as restraints.

Amide proton exchange rates were estimated from a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra performed at different times after dissolving the protein in  $\text{D}_2\text{O}$  (data not shown); hydrogen bond formation or solvent exclusion from amide protons was assumed to account for slow and medium exchange rate. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints. Each hydrogen bond was introduced as O-N distance of 2.00 Å and HN-O distance of 3.00 Å.

A subset of unambiguously assigned NOEs was used to calculate an initial ensemble of NMR structures, which was subsequently employed, in an iterative manner, to interpret the remaining ambiguous NOEs. Some further restraints were added at the list of unambiguous restraints and selected if they lowered the global rmsd. The restraints were re-examined to check for consistent violations. One hundred calculations were run employing DYANA (Guntert et al., 1997) and the 20 conformers with the lowest residual target function were analysed. The 20 final DYANA structures were further refined using the AMBER force field, as implemented in the program DYSCOVER (Molecular Simulations, San Diego, CA). Each structure was minimised performing 100 steps of steepest descent and 300 steps of conjugate gradient. The 10 structures with the lowest potential energy were selected for further analysis.

*<sup>15</sup>N relaxation parameters.* R<sub>1</sub> and R<sub>2</sub> relaxation rates were derived by fitting cross peak intensities, measured from the appropriate experiments, to a single exponential function using the Rate Analysis routine of NMRView program (Johnson, 2004). The heteronuclear NOE effect was calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation.

*Modelfree analysis.* Calculations of Lipari-Szabo motional parameters were performed with the program MOEDELFREE (version 4.0). Relaxation of amide <sup>15</sup>N nuclear spins in a diamagnetic protein can be generally described by dipolar coupling with directly attached protons and <sup>15</sup>N chemical shift anisotropy (Lipari and Szabo, 1981), as in [1-3]:

$$R_1 = (d^2/4) [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad [1]$$

$$R_2 = (d^2/8) [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (c^2/6) [4J(0) + 3J(\omega_N)] + R_{ex} \quad [2]$$

$$NOE = 1 + (d^2/4R_1) (\gamma_H/\gamma_N) [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad [3]$$

where  $d = (\mu_0 h (\gamma_H \gamma_N / 8\pi) \langle r_{NH}^{-3} \rangle$ ,  $c = \omega_N \Delta\sigma / 3^{1/2}$ ,  $\mu_0$  is the permeability of free space,  $h$  is Planck's constant,  $\gamma_H$  and  $\gamma_N$  are the gyromagnetic ratios of <sup>1</sup>H and <sup>15</sup>N, respectively,  $r_{NH} = 1.02 \text{ \AA}$ ,  $\omega_N$  and

$\omega_H$  are the Larmor frequencies of  $^{15}\text{N}$  and  $^1\text{H}$ , respectively and  $\Delta\sigma = -160$  ppm is the chemical shift anisotropy measured for  $^{15}\text{N}$  nuclei in helical polypeptide chains (Wei et al., 2001).

Slow conformational exchange ( $R_{\text{ex}}$ ) contributes to the transversal relaxation rate  $R_2$ . Spectral density functions  $J(\omega_i)$  at the angular frequency  $\omega_i$  depend on the internal motions of the  $^1\text{H}$ - $^{15}\text{N}$  bond vector and on the overall tumbling of the protein.

The extended Lipari-Szabo formalism proposes different spectral density functions (Table I), which depend upon  $S^2$  (the generalised motional order parameter),  $\tau_m$  (the overall correlation time of rotational diffusion),  $\tau_e$  (the effective correlation time) and  $R_{\text{ex}}$  (the rate of conformational exchange).

The MODELFREE analysis was performed fitting the relaxation parameters to the five models of Table I and selecting, for each residue, the model which best fits the data by minimising the target function of the sum-squared error (SSE) (equation 4).

$$\text{SSE} = (R_1 - R_1^*)^2 / \sigma_1^2 + (R_2 - R_2^*)^2 / \sigma_1^2 + (\text{hNOE} - \text{hNOE}^*)^2 / \sigma_{\text{hNOE}}^2 \quad [4]$$

where  $R_1$ ,  $R_2$  and hNOE are the experimental relaxation parameters,  $\sigma_1$ ,  $\sigma_1$  and  $\sigma_{\text{hNOE}}$  are the uncertainties on the experimental data and  $R_1^*$ ,  $R_2^*$  and hNOE\* are the back calculated relaxation parameters, according to the selected model. F-statistics, which measures the random statistical reduction in SSE by incorporation of additional parameters, was additionally used to check on the model selection (Mandel et al., 1995). A further check is the consistency between the selected model and the motional parameters (i.e.  $\tau_e > 20$  ps,  $R_{\text{ex}} > 0.5 \text{ s}^{-1}$  and  $S^2 > 0.9$ ).

The iterative optimisation of the overall rotational correlation time  $\tau_m$  was performed as described in (Feng et al., 1998; Li and Montelione, 1995).

Chemical exchange contributions ( $R_{\text{ex}}$ ) to  $R_2$ , as determined by CPMG spin-echo (20), can be derived from equation [6]:

$$R_2 = R_2^0 + R_{\text{ex}} \quad [6]$$

where  $R_{\text{ex}}$  can be written as in [7] (Peng and Wagner, 1995):

$$R_{\text{ex}} = \Phi \omega_N^2 \quad [7]$$

The factor  $\Phi$  depends on the type and the intrinsic rate constant of exchange process, on the population and chemical shift differences of participating sites and on the applied spin-echo period in the CPMG pulse train. For a given residue, the contribution of chemical exchange may be highlighted by a higher than average value monitored in a plot of  $R_2/R_1$  as a function of residue number.  $R_{ex}$  can be determined either with the Lipari Szabo approach, if the appropriate model can be selected, or from the spectral density  $J(0)$  at two fields (see later). Both approaches were used here in order to compare and to validate the results.

*Spectral density mapping.* Relaxation data have been analysed according to the reduced spectral density mapping, using two methods, as reported (Farrow et al., 1995). In the first method, values of the reduced spectral density function at each field strength (600 and 700 MHz) were calculated from equations [1-3] in the hypothesis that  $J(\omega_H \pm \omega_N) \cong J(\omega_H) \cong J(0.87\omega_H)$ , as in [8-10] :

$$J(0.87\omega_H) = [4/(5d^2)] (\gamma_H/\gamma_N) (NOE - 1) R_1 \quad [8]$$

$$J(\omega_N) = 4/(3d^2 + 4c^2) [ R_1 - 7/4 d^2 J(0.87\omega_H) ] \quad [9]$$

$$J(0) = (6R_2 - 3R_1 - 3.6 \sigma) / (3d^2 + 4c^2) \quad [10]$$

where symbols have the usual meaning, as described before and  $\sigma = (\gamma_H/\gamma_N) (NOE - 1) R_1$ .

In the hypothesis that the dynamics is limited to a single motion, the spectral density is given by a single Lorentzian and takes the form of equation [11] (Krizova et al., 2004):

$$J(\omega) = \frac{J(0)}{1 + 6.25(\omega J(0))^2} \quad [11]$$

A spectral density correlation graph can be built as reported in (Krizova et al., 2004), giving a qualitative description of the molecular motions. To determine the conformational exchange contributions, a quantitative analysis had to be performed.  $J_{eff}(0)$  may contain conformational exchange contributions, according to  $J_{eff}(0) = J(0) + \lambda R_{ex}$ . In the absence of  $R_{ex}$  contributions,  $J_{eff}(0) = J(0)$  and the obtained  $J_{eff}(0)$  values are field independent. In the presence of  $R_{ex}$ ,  $J_{eff}(0)$  values are expected to increase with field strength, as  $R_{ex} \propto \omega_N^2$ , and a plot of  $J(0)_{eff1}/J(0)_{eff2}$ , determined at

two field strengths, as a function of residue number, can monitor the exchange contribution on single residues.  $R_{ex}$  was estimated from equation [7] where  $\Phi$  is defined by equation [12] (Peng and Wagner, 1995):

$$\Phi = \frac{J_{\text{eff1}}(0) - J_{\text{eff2}}(0)}{\lambda_1 \omega_{1N}^2 - \lambda_2 \omega_{2N}^2} \quad [12]$$

where  $J_{\text{eff1}}(0)$  and  $J_{\text{eff2}}(0)$ ,  $\lambda_1$  and  $\lambda_2$  are determined at 700 and 600MHz, respectively.  $\lambda_1 = 0.27$  (ns/rad)<sup>2</sup>,  $\lambda_2 = 0.29$  (ns/rad)<sup>2</sup>, as calculated from  $\lambda = (3/2)[1/(3d^2 + c^2)]$ .

In the second approach we have calculated  $J(0)_{\text{eff}}$  using simultaneously the relaxation data at two field strengths, as previously reported (Farrow et al., 1995), and then derived the contribution of  $R_{ex}$  to  $R_2$  rate measured at 600MHz (equation 13):

$$J(0)_{\text{eff}} = \frac{1}{\beta} \left\{ \left[ \frac{1}{T_2^{700}} - \kappa \frac{1}{T_2^{600}} \right] - \left( \frac{3d^2}{8} \right) \left[ J(\omega_N^{700}) - \kappa J(\omega_N^{600}) \right] - \left( \frac{c_{700}^2}{2} \right) \left[ J(\omega_N^{700}) - J(\omega_N^{600}) \right] - \left( \frac{13d^2}{8} \right) \left[ J(0.955\omega_H^{700}) - \kappa J(\omega_H^{600}) \right] \right\}$$

$$R_{ex} = \frac{1}{T_2^{600}} - \left( \frac{d^2}{2} + \frac{2c_{600}^2}{3} \right) J(0) - \left( \frac{3d^2}{8} + \frac{c_{600}^2}{2} \right) J(\omega_N^{600}) - \left( \frac{13d^2}{8} \right) J(0.955\omega_H^{600})$$

$$\kappa = \left( \omega_H^{700} / \omega_H^{600} \right)^2, \beta = \left( \frac{d^2}{2} \right) (1 - \kappa)$$

$$J(\omega_N) = 4/(3d^2 + 4c^2) [ R_1 - 7/4 d^2 J(0.921\omega_H) ] \quad [13]$$

where  $J(0.921\omega_H)$  and  $J(0.955\omega_H)$  have been determined using a first order Taylor expansion of  $J(\omega)$  about  $J(0.87\omega_H)$  at the two field strengths (equation 14):

$$J(\varepsilon\omega_H) \approx J(0.87\omega_H) + (\varepsilon - 0.87)\omega_H J'(0.87\omega_H) \quad [14]$$

where  $\varepsilon = 0.921$  or  $0.955$ ;  $J'(0.87\omega_H) = [J(0.87\omega_H^{700}) - J(0.87\omega_H^{600})] / [0.87(\omega_H^{700} - \omega_H^{600})]$ .

All the data obtained with the described methods were compared in order to give a unique description of the protein dynamics.

## Results

*Assignment strategy.* To illustrate the spectra quality, strips derived from  $^{13}\text{C}$  edited NOESY spectra at pH 7 are reported in Figure 1. Backbone assignment, performed by a combination of classical 3D NMR experiments on the doubly labelled protein, as described in Material and Methods, was not straightforward especially for the C-terminal region of the protein corresponding to strands F, G, H and I. In this region, breaks in the process of assigning were caused by missing correlations due either to fast exchange of amide protons with solvent and/or to conformational exchange (see later). It has been therefore necessary to combine the standard 3D backbone assignment strategy with the sequential assignment strategy based on homonuclear 2D TOCSY/NOESY and 3D  $^1\text{H}$ - $^{15}\text{N}$  TOCSY/NOESY experiments. In addition 3D  $^1\text{H}$ - $^{15}\text{N}$  TOCSY/NOESY were performed at lower pH (5.6) to check whether amides in fast exchange with solvent at pH 7 were visible at lower pH. In this way it was possible to fill some gaps in the assignment, for example identifying  $\text{H}_\text{N}$  and  $\text{H}_\alpha$  proton frequencies and then detecting the corresponding spin systems in the (H)CCH-COSY/TOCSY and in the H(C)CH-COSY/TOCSY. In some cases the starting point was the identification, in non-crowded regions of  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra, of some residues with typical chemical shifts, such as, for example,  $\text{C}_\alpha$ - $\text{H}_\alpha$  of glycines,  $\text{C}_\beta$ - $\text{H}_\beta$  of threonines and  $\text{C}_\alpha$ - $\text{H}_\alpha$ / $\text{C}_\beta$ - $\text{H}_\beta$  of serines. These spin systems were then found in H(C)CH-COSY and TOCSY and in 2D  $^1\text{H}$  TOCSY and NOESY. As an example (Figure 2), T<sub>57</sub> and T<sub>59</sub> had missing resonances in  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra at pH 7, however their  $\text{C}_\beta$ - $\text{H}_\beta$  peaks were identified from  $^1\text{H}$ - $^{13}\text{C}$ -HSQC spectra and their complete spin system determined as described. In addition their resonances appeared at pH 5.6. In other cases it was possible to obtain the assignment on the basis of  $\text{C}_\alpha(i-1)$  frequency derived from HNCA spectra. The complete spin system of the considered  $i-1$  residue was then identified in the (H)CCH-COSY/TOCSY and in the H(C)CH-COSY/TOCSY and its sequential assignment confirmed in the homonuclear 2D spectra.

The  $^{15}\text{N}$  assignment of eight residues ( $\text{M}_{73}$ ,  $\text{V}_{82}$ ,  $\text{V}_{90}$ ,  $\text{S}_{93}$ ,  $\text{E}_{94}$ ,  $\text{K}_{85}$ ,  $\text{E}_{99}$ ,  $\text{Q}_{100}$ ) could not be performed.

Alltogether, following the different strategies described, most of the gaps in the assignment were filled and only five residues ( $\text{M}_{73}$ ,  $\text{V}_{90}$ ,  $\text{S}_{93}$ ,  $\text{E}_{99}$ ,  $\text{Q}_{100}$ ) remained completely unassigned. Indeed they were located in that region of the protein mostly affected by conformational exchange, as revealed by relaxation analysis (see later).

Following the same strategy described for apo cl-BABP, the  $^1\text{H}$  and  $^{15}\text{N}$  assignment of the complex with chenodeoxycholic acid was determined.

The complete assignment of apo and holo cl-BABP are reported as supplementary material (Table 2). The comparison of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra obtained for the apo protein at 298K at pH 5.6 and 7 did not show amides chemical shift changes, defined as  $\delta = [(\Delta\delta_{\text{HN}})^2 + \Delta\delta_{\text{N}}^2/25]/2]^{1/2}$ ,  $> 0,06$  ppm (Cicero et al., 2003) (Figure 3a), thus indicating the absence of backbone conformational rearrangement in this pH range. However  $^{15}\text{N}$  chemical shift differences  $> 0.1$  ppm were observed for some residues ( $\text{Q}_{11}$ ,  $\text{D}_{46}$ ,  $\text{F}_{62}$ ,  $\text{G}_{65}$ ,  $\text{K}_{66}$ ,  $\text{A}_{68}$ ,  $\text{I}_{70}$ ,  $\text{T}_{72}$ ,  $\text{C}_{80}$ ,  $\text{L}_{89}$ ,  $\text{T}_{91}$ ,  $\text{K}_{92}$ ,  $\text{F}_{96}$ ,  $\text{S}_{97}$ ,  $\text{H}_{98}$ ,  $\text{K}_{103}$ ,  $\text{G}_{104}$ ,  $\text{E}_{106}$ ,  $\text{T}_{110}$ ,  $\text{I}_{111}$ ,  $\text{G}_{115}$ ,  $\text{R}_{121}$ ,  $\text{K}_{123}$ ,  $\text{V}_{125}$ ) (Figure 3b) and this variation could held some relevance in view of the dynamics behaviour which will be discussed later.

The resonances of few amides, missing at pH 7, were visible at pH 5.7 as shown, for example, in Figure 4, reporting on the identification of residues  $\text{G}_{44}$  and  $\text{G}_{104}$ . The new detected amides were  $\text{T}_{57}$ ,  $\text{T}_{59}$ ,  $\text{E}_{67}$ ,  $\text{C}_{80}$ ,  $\text{S}_{97}$ ,  $\text{E}_{101}$  and  $\text{N}_{105}$ . It is worth mentioning that 3D  $^1\text{H}$ - $^{15}\text{N}$  TOCSY/NOESY spectra obtained at pH 5.7 revealed the presence of additional peaks which could be attributed to slowly exchanging conformations. Residues showing additional peaks were  $\text{S}_3$  (A strand),  $\text{G}_{44}$  (BC loop),  $\text{F}_{47}$  (C strand),  $\text{D}_{74}$  (EF loop),  $\text{A}_{85}$  (FG loop),  $\text{L}_{89}$  (G strand) and  $\text{G}_{104}$  (HI loop).

*Secondary structure determination of apo and holo cl-BABP.* Secondary  $^{13}\text{C}_\alpha$  and  $\text{H}_\alpha$  chemical shifts (Wishart and Sykes, 1994) are reported in Figure 5a,b for apo cl-BABP and secondary  $\text{H}_\alpha$  chemical shifts of the holo protein are shown in Figure 5c.

A schematic representation of the antiparallel beta-sheet structure of apo protein is reported in Figure 6. The average  $H_{\alpha}$  secondary shifts within each strands are reported for apo and holo protein in Figure 7a and b, respectively.

*Tertiary Structure calculation of apo cl-BABP.* A summary of experimental restraints used for structure calculation is reported in Table 3. The decision was taken to introduce in the calculation only totally unambiguous restraints, i.e. those correlations which were not affected by overlap in any spectra. In this way a set of 1000 non redundant NOEs was supplemented by (i) 26 distance restraints for 13 backbone hydrogen bonds defined on the basis of deuterium hydrogen exchange studies (data not shown) and (ii) 48  $\phi$  angle constraints derived from  $J_{HN-H\alpha}$  coupling constants. It should be stressed that this protein is highly flexible, as revealed both by H/D exchange and relaxation measurements, and several residues did not exhibit NOE correlations. The superposition of the best ten NMR structures, as obtained after DYANA molecular dynamics simulations followed by energy minimisation, reported in Figure 8, afford an RMSD<sub>bb</sub> value (calculated with respect to the mean) of  $1.86 \pm 0.33$  Å.

*<sup>15</sup>N relaxation data for apo-clBABP:* The  $R_2/R_1$  ratios at 700 and 600 MHz, as a function of residue number, are reported in Figure 9a-d at pH 7 and 5.7. Fast internal mobility will decrease the  $R_2/R_1$  ratio, while conformational exchange increases it. At pH 7, significantly high  $R_2/R_1$  ratios ( $> 9$ ) were found for residues E<sub>67</sub>, A<sub>68</sub>, T<sub>72</sub>, T<sub>81</sub>, L<sub>89</sub>, T<sub>91</sub>, F<sub>96</sub>, S<sub>97</sub>, V<sub>102</sub>, T<sub>110</sub>, I<sub>111</sub> while at pH 5.7 high  $R_2/R_1$  ratios ( $> 11$ ) were detected for L<sub>64</sub>, A<sub>68</sub>, I<sub>70</sub>, T<sub>72</sub>, C<sub>80</sub>, T<sub>81</sub>, L<sub>89</sub>, T<sub>91</sub>, S<sub>93</sub>, F<sub>96</sub>, S<sub>97</sub>, V<sub>102</sub>, T<sub>110</sub>, I<sub>111</sub>. Residues C<sub>80</sub> and S<sub>93</sub> have high  $R_2/R_1$  ratio at pH 5.7 and their resonances disappears at pH 7. Residue I<sub>70</sub> exhibits a high  $R_2/R_1$  ratio only at pH 5.7, while at both pHs F<sub>96</sub> and S<sub>97</sub> have the highest  $R_2/R_1$ . Heteronuclear NOE values lower than 0.7 were detected at pH 7 at the end of helix II and in the loop connecting helix II to strand B (A<sub>31</sub>, D<sub>33</sub>, I<sub>34</sub>, K<sub>35</sub>), in the CD and FG loops (Q<sub>56</sub>, A<sub>85</sub>, K<sub>88</sub>) in the E and H strands (E<sub>67</sub>, F<sub>96</sub>). At pH 5.7 NOE values lower than 0.7 were identified substantially at the same locations (L<sub>27</sub>, D<sub>33</sub>, I<sub>34</sub>, K<sub>35</sub>, R<sub>55</sub>, Q<sub>56</sub>, M<sub>73</sub>, V<sub>82</sub>, A<sub>85</sub>, K<sub>88</sub>, E<sub>94</sub>, K<sub>95</sub>) (Figure 9a).

<sup>15</sup>N relaxation data for *holo-clBABP*. The R<sub>2</sub>/R<sub>1</sub> ratios and heteronuclear NOEs at 700 and 600 MHz, as a function of residue number, are reported in Figure 10a,b for the holo protein at pH 7. Only residues R<sub>120</sub> and V<sub>125</sub>, at the C-terminal end of the protein, exhibited a high R<sub>2</sub>/R<sub>1</sub> ratio, while no residues exhibited NOE values lower than 0.7.

*Model Free analysis for apo cl-BABP*. Averaged R<sub>2</sub>/R<sub>1</sub> ratios were initially used to estimate an overall correlation time τ<sub>m</sub> (Farrow et al., 1994) which resulted to be 6.9±0.4 ns and 7.6±0.6 for data at 700 MHz at pH 7 and 5.7, respectively, and 7.2±0.3 ns and 7.0±0.4 ns for data at 600 MHz, at pH 7 and 5.7 respectively. The final optimised τ<sub>m</sub> were 7.1 ns at pH 7 and 7.4 ns at pH 5.7.

At 700 MHz and pH 7 all residues display order parameter values greater than 0.8 with a calculated S<sup>2</sup> average value of 0.92±0.04. The same holds at pH 5.7 (except for residues E<sub>67</sub> and A<sub>85</sub>) with a calculated S<sup>2</sup> average value of 0.94±0.05.

As a result of the global model analysis, at pH 5.7 thirtyseven residues exhibited τ<sub>e</sub>: Thirtytwo exhibited values in the range 26-190 ps (E<sub>12</sub>, A<sub>22</sub>, L<sub>23</sub>, E<sub>25</sub>, D<sub>26</sub>, I<sub>28</sub>, M<sub>30</sub>, D<sub>33</sub>, I<sub>34</sub>, K<sub>35</sub>, Q<sub>42</sub>, T<sub>53</sub>, Q<sub>56</sub>, T<sub>57</sub>, E<sub>67</sub>, A<sub>68</sub>, D<sub>69</sub>, I<sub>70</sub>, G<sub>75</sub>, K<sub>79</sub>, T<sub>81</sub>, A<sub>85</sub>, G<sub>87</sub>, K<sub>88</sub>, T<sub>91</sub>, F<sub>96</sub>, K<sub>103</sub>, G<sub>104</sub>, N<sub>105</sub>, Q<sub>106</sub>, G<sub>114</sub> and V<sub>125</sub>), three in the range 290-400 ps (L<sub>27</sub>, R<sub>32</sub>, R<sub>55</sub>) and two in the range 600-720ps (D<sub>74</sub>, K<sub>76</sub>). At pH 7 only twenty residues exhibit τ<sub>e</sub>: seventeen in the range 20-190ps (E<sub>12</sub>, E<sub>15</sub>, L<sub>27</sub>, K<sub>29</sub>, M<sub>30</sub>, A<sub>31</sub>, R<sub>32</sub>, I<sub>34</sub>, Q<sub>56</sub>, K<sub>76</sub>, L<sub>78</sub>, A<sub>85</sub>, K<sub>88</sub>, F<sub>96</sub>, K<sub>103</sub>, E<sub>106</sub>, V<sub>125</sub>), two residues in the range 700-850 ps (D<sub>33</sub> and K<sub>35</sub>) and one (D<sub>26</sub>) exhibiting a τ<sub>e</sub> value of 1657 ps. The data are better visualised in graphs and on the structure reported in Figure 11a,b,c.

At pH 5.7 residues L<sub>64</sub>, C<sub>80</sub>, L<sub>89</sub>, T<sub>91</sub>, F<sub>96</sub>, S<sub>97</sub>, T<sub>110</sub> and I<sub>111</sub> exhibited conformational exchange contributions (R<sub>ex</sub>) and at pH 7 residues affected by R<sub>ex</sub> were E<sub>67</sub>, A<sub>68</sub>, T<sub>72</sub>, T<sub>81</sub>, L<sub>89</sub>, T<sub>91</sub>, F<sub>96</sub>, S<sub>97</sub>, V<sub>102</sub>, T<sub>110</sub>, I<sub>111</sub> (Figure 12). Few residues (E<sub>67</sub>, T<sub>81</sub>, T<sub>91</sub> and F<sub>96</sub>) changed their dynamical behavior on going from neutral to acidic pH where they were characterised by motions in the ps to ns timescale (τ<sub>e</sub>). Indeed T<sub>91</sub>, and F<sub>96</sub> at pH 5.7 are better fitted by a model containing both τ<sub>e</sub> and R<sub>ex</sub>

contributions. F<sub>96</sub>, S<sub>97</sub> and I<sub>111</sub> exhibited higher R<sub>ex</sub> contributions at pH 5.7, in agreement with the higher R<sub>2</sub>/R<sub>1</sub> ratios observed.

*Model Free analysis for holo cl-BABP.* Correlation times of 7.3±0.4 ns and 7.1±0.5 were derived from R<sub>2</sub>/R<sub>1</sub> ratios at 700 and 600MHz, respectively at pH 7. The final optimised value was 7.2 ns. At 700 MHz and pH 7 a S<sup>2</sup> average value of 0.92±0.05 was calculated. As a result of the global model analysis, fourteen residues exhibited τ<sub>e</sub>: thirteen in the range 17-105 ps (A<sub>20</sub>, M<sub>30</sub>, A<sub>31</sub>, I<sub>34</sub>, Q<sub>42</sub>, V<sub>48</sub>, R<sub>32</sub>, R<sub>55</sub>, D<sub>69</sub>, K<sub>79</sub>, C<sub>80</sub>, F<sub>96</sub>, K<sub>103</sub>) and one (K<sub>95</sub>) exhibited a τ<sub>e</sub> of 1300 ps.

*Reduced Spectral Density mapping.* The values of J(0) and J(0.87ω<sub>H</sub>), derived from equations [8] and [10] (Material and Methods) for apo and holo cl-BABP at pH 7 and at two different fields showed that J(0.87ω<sup>600</sup><sub>H</sub>) is always greater than J(0.87ω<sup>700</sup><sub>H</sub>) thus validating the working hypothesis of a spectral density as a sum of Lorentzian functions.

Spectral density correlation graphs were built on the basis of equation [11] (Krizova et al., 2004). J(ω<sub>N</sub>), J(0.87ω<sub>H</sub>) vs J(0), as measured at 700MHz for the apo and holo proteins are shown in Figure 13 and 14, respectively. The R<sub>ex</sub> values were calculated according to equations [12] and [13] of Material and Methods (see later). The combined results indicated an R<sub>ex</sub> contribution, for apo protein, for residues G<sub>65</sub>, E<sub>67</sub>, A<sub>68</sub>, I<sub>70</sub>, T<sub>72</sub>, T<sub>81</sub>, T<sub>91</sub>, F<sub>96</sub>, H<sub>98</sub>, V<sub>102</sub>, V<sub>108</sub>, T<sub>110</sub>, I<sub>111</sub>, R<sub>120</sub> at pH 7 and for residues A<sub>68</sub>, I<sub>70</sub>, T<sub>72</sub>, C<sub>80</sub>, T<sub>81</sub>, L<sub>89</sub>, T<sub>91</sub>, S<sub>97</sub>, V<sub>102</sub>, V<sub>108</sub>, T<sub>110</sub>, I<sub>111</sub> and R<sub>120</sub> at pH 5.7. These values, together with those obtained from the model-free approach (Peng and Wagner, 1995) are reported in Figure 15, for comparison. In the holo-protein all R<sub>ex</sub> contributions vanished, except for the last C-terminal residue V<sub>125</sub>.

## Discussion

*Structural determination.* In this study we carried out a series of 3D heteronuclear NMR experiments in order to obtain the structural characterisation of the recombinant cl-BABP. As discussed, this protein and the purified one are two different genetic variants, differing for the residue at position 91, a threonine in the expressed protein and a cysteine in the protein purified from chicken liver.

All the set of NMR data led to an almost complete assignment of the backbone nuclei. The analysis of the average secondary  $H^\alpha$  chemical shifts, of the apo and holo proteins, indicated the presence of more secondary structure in the C-terminal end of the holo-protein (Figure 7).

The 3D structure, obtained for the apo protein (Figure 8), is typical of all the proteins of iILB family, showing a barrel made of ten antiparallel beta strands with an helix-turn-helix motif connecting strands A and B somehow closing the internal cavity.

The structural quality of the minimised structures was examined with the PROCHECK-NMR (Laskowski et al., 1993). Analysis of the backbone dihedral angles showed that 95 % of all non-glycine non-proline residues in apo cl-BABP fall within the additional allowed regions of conformational space. Considering that this analysis includes some poorly defined regions located in the C-terminal end, this result can be considered reasonable. The average pairwise backbone RMSD for the 10 best minimised NMR structures (Figure 8) is  $1.86 \pm 0.33$  Å. The distribution of distance restraints per residue (Figure 16a) accounts for the observed distribution of average global displacement, reported in Figure 16b. From this Figure it appears that the largest deviations ( $> 1.5$  Å) are clustered around residues 56-58 (N-terminal D strand), 72-81 (EF loop and F strand), 86-87 (FG loop), 92-100 (GH loop and H strand). The high backbone dispersion of certain segment of cl-BABP essentially corresponds to residues that show fewer distance restraints (Figure 16b), because of either conformational dispersion/mobility or lack of assignments.

At pH 5.7 few more amides could be detected, however the number of added restraints did not exceed the 5% of the total restraints obtained at pH 7, thus reinforcing the picture of a highly flexible molecule. (Figure 8).

It is worth stressing that the chemical shift analysis confirmed the absence of major conformational changes in the pH range 5.6-7. At pH 5.6 chemical shift differences higher than 0.2 ppm, between the purified and recombinant cl-BABP were observed for residues E<sub>67</sub> (DE loop), K<sub>79</sub>, C<sub>80</sub>, T<sub>81</sub> (F strand) A<sub>85</sub>, T<sub>91</sub> (G strand), N<sub>105</sub>, E<sub>106</sub>, M<sub>107</sub> and V<sub>108</sub> (I strand), consistent with the new evidence of the C80-C91 disulphide bridge in the purified protein.

In the apo form, the recombinant protein samples displayed a reduced stability, indicated by the appearance (after few days of measurement at room temperature) of additional amide signals belonging to a denatured form of the protein.

*Dynamics analysis.* To determine the presence of conformational mobility the internal backbone dynamics were studied for both apo and holo cl-BABP by measuring <sup>15</sup>N relaxation rates and steady state heteronuclear <sup>15</sup>N(<sup>1</sup>H) NOE at 700 and 600 MHz. The dynamic parameters ( $S^2$ ,  $\tau_e$ ,  $R_{ex}$ ,  $J(0)$ ,  $J(0.87\omega_H)$ ,  $J(\omega_N)$ ) were obtained for the apo protein at two pHs (5.7 and 7) and for the holo protein at pH 7, using both the model free and the reduced spectral density mapping approaches.

*Dynamics data of apo cl-BABP at pHs 5.7 and 7.* Viewed over the entire protein sequence, the average order parameter ( $S^2$ ) was substantially unchanged on going from pH 5.7 (0.94±0.05) to 7 (0.92±0.04). It is observed that a higher number of residues exhibited  $\tau_e$  contributions (picosecond to nanoseconds timescale) at pH 5.7 (Figure 11). Indeed at acidic pH not only residues in helixII and in F strand but also residues in CD loop and E, G, H strands showed fast dynamics (Figure 11). Several residues of the apo protein are affected by  $\mu$ s-ms dynamics, both at pH 5.7 and 7. A qualitative good agreement was observed between dynamics data obtained through the model free and reduced spectral mapping approach. All the residues affected by conformational exchange, as determined by the different methods at two pHs, are shown on the protein structure in Figure 17. It should be noted that these residues are located only on one face of the protein, at the level of strands

FGHI. Interestingly this face of the protein contains two pH titrable histidine side-chains, i.e. the partially exposed H<sub>83</sub>, located in F strand, and the partially buried H<sub>98</sub>, located in H strand and pointing towards the internal cavity. The observed dynamics might therefore be coupled to the exchange between the protonated and deprotonated states of these two histidine residues. Indeed nearly all the residues showing an R<sub>ex</sub> contribution at the two pHs examined are within 5Å from either H<sub>83</sub> or H<sub>98</sub> (Figure 17), with the only exception of E<sub>67</sub>, T<sub>72</sub> and V<sub>102</sub>, which are however in close contact with residues nearby H<sub>83</sub> and H<sub>98</sub>. This observation reinforces the suggestion that the equilibrium between protonated and deprotonated forms of histidines play a role in determining the observed dynamics. As shown in Figure 18, on going from pH 5.7 to pH 7, few peaks showed an enhanced increase in R<sub>1</sub>, while R<sub>2</sub> values seem to be pH independent in the range 5.7-7, with the noticeable exception of residue 97 and 108 showing a substantial increase of R<sub>2</sub> (indicating large changes in the μs to ms dynamics) at acidic pH. All these residues are close to H<sub>98</sub> which could play the most relevant role in the observed dynamics. The R<sub>2</sub> effect on H<sub>98</sub> itself could not be calculated as its relaxation rates could be measured only at pH 7. In line with this observation it was observed that H<sub>98</sub> exhibited a minor chemical shift change on going from pH 7 ( $\delta^H$  8.77,  $\delta^N$  125.66) to pH 5.7 ( $\delta^H$  8.8,  $\delta^N$  125.37) while a major change was observed on going from pH 5.7 to 4.8 ( $\delta^H$  8.85,  $\delta^N$  124.8). H<sub>83</sub> resonance, unambiguously assigned at pH 7 ( $\delta^H$  8.66,  $\delta^N$  124.02) could not be assigned at pH 5.7 as the resonance either was broadened beyond detection or moved away to very crowded regions, thus suggesting that pH 5.7 is somehow close to its pK<sub>a</sub>. These data are consistent with H<sub>98</sub> titrating at lower pH with respect to H<sub>83</sub>, in agreement with pK<sub>a</sub> theoretical calculations, obtained on the ten best DYANA structures, affording a mean pK<sub>a</sub> of 5.7 for H<sub>83</sub> and 4.65 for H<sub>98</sub>. It is important to underline that pK<sub>a</sub> calculation have to be considered not as an exact indication of the pK<sub>a</sub> but rather their relative values are indicative. Consistent with this hypothesis <sup>15</sup>N chemical shift changes ( $\delta^N$ ) > 0.1 ppm were observed, in the pH range 5.7-7, for residues close to the two histidines (Figure 3b).

At the same time on going from neutral to acidic pH an enhanced mobility on the picosecond to nanosecond timescale is observed for several residues located in the C-terminal face of the protein, and specifically some of the residues showing  $\tau_e$  contributions are within 5Å from H<sub>83</sub> and H<sub>98</sub> (T<sub>81</sub>, A<sub>85</sub>, G<sub>87</sub>, K<sub>88</sub>, T<sub>91</sub>, F<sub>96</sub>). A higher mobility is instead observed for residues located in helixII at neutral pH. The variation of protein dynamics in this pH range is also confirmed by the presence, at pH 5.7, of additional peaks assigned to residues S<sub>3</sub>, G<sub>44</sub>, F<sub>47</sub>, D<sub>74</sub>, A<sub>85</sub>, L<sub>89</sub>, G<sub>104</sub>, in agreement with protein segments exhibiting conformational exchange.

*Comparison of dynamics data of apo and holo cl-BABP at pH 7.* The average order parameter ( $S^2$ ) was substantially unchanged ( $0.92\pm 0.04$ ) upon binding with chenodeoxycholic acid. Residues exhibiting  $\tau_e$  contributions are substantially located in the same regions for the apo and holo forms, i.e. in helixII and in CD, FG and GH loops.

Most interestingly, upon binding, micro-to millisecond motions disappear, as graphically clearly represented in Figure 13-14, where the values of spectral density functions available at three frequencies can be visualized as points in a 3D space of dimensions  $J(0)$ ,  $J(\omega_N)$ , and  $J(0.87\omega_H)$ . In Figure 13-14 expansions of these correlation graphs are reported, in order to better visualise the residues lying outside the complete “cloud” of data points (Krizova et al., 2004). Most points are clustered within the Lorentzian limits, indicative of a rigid particle rotation (Barthe et al., 1999). Based on such graphs, amino acids can be classified into three basic categories. The first category can be described as flexible residues with little restriction of the backbone N–H vector motions. Such residues are manifested by low  $J(0)$  and high  $J(0.87\omega_H)$ . Noone of the analysed apo and holo cl-BABP residues can be classified in this way (Figures 13a and 14a). The second category represents largely rigid residues exhibiting no obvious sign of a slow conformational exchange. Most amino acids of both free and ligand-bound cl-BABP belonged to this category (Figures 13a and 14a). It is worth noting that a residue by residue analysis revealed that the points of the holo protein lie closer to the single-Lorentzian limit in most cases, indicating higher rigidity of holo cl-BABP compared to the apo protein. The third category represents residues experiencing

conformational exchange on  $\mu\text{s}$ – $\text{ms}$  time scale. The conformational exchange results in increasing values of  $J(0)$  without affecting  $J(\omega_N)$  or  $J(0.87\omega_H)$ . It is interesting to observe that, on going from apo to holo cl-BABP, the  $R_{\text{ex}}$  contributions vanish, as shown by the disappearance of points on the right of the lorentzian line (Figure 14). Most of the residues with vanished  $R_{\text{ex}}$  contribution in the holo protein map to the regions that exhibited secondary  $^{15}\text{N}$  shift changes  $> 0.5$  ppm upon binding at pH 7 (R<sub>32</sub>, I<sub>37</sub>, T<sub>72</sub>, K<sub>76</sub>, C<sub>80</sub>, T<sub>81</sub>, L<sub>89</sub>, T<sub>91</sub>, K<sub>92</sub>, F<sub>96</sub>, S<sub>97</sub>, H<sub>98</sub>, G<sub>104</sub>, V<sub>108</sub>, E<sub>109</sub>) (Figure 19). Our data are relative to the binding of chenodeoxycholic acid. Recently the X-ray structure has been reported for the purified cl-BABP (the genetic variant with the 80-91 disulphide bridge) in complex with cholates. All bile acids have been shown to bind BABP with a 2:1 stoichiometry (Tochtrop et al., 2003) with similar overall affinity and different cooperativity. Chenodeoxycholic and cholic acid differ for the presence of a hydroxyl group in position 12 of the steroid tetracycle. It has been recently shown, by NMR titration of human ileal BABP (I-BABP) with  $^{15}\text{N}$  and  $^{13}\text{C}$  enriched glycochenodeoxycholic acid (GCDA) and glycocholic acid (GCA), that steroid ring hydroxylation pattern governs binding selectivity (Tochtrop et al., 2004). Indeed when I-BABP is treated with either GCA or GCDA alone, the ligands bind to both sites, while when treated with an equimolar mixture of the two bile salts, GCDA bind exclusively to one site and GCA to the other, indicating an energetic rather than a steric site selectivity.

In light of these data, it is interesting to compare the residues exhibiting  $^{15}\text{N}$  shifts upon chenodeoxycholic acid binding to recombinant cl-BABP with those in closest contact with cholates, T<sub>72</sub>, K<sub>76</sub>, F<sub>96</sub>, C<sub>91</sub>, H<sub>98</sub>, Q<sub>100</sub> (for the first cholates molecule) and Y<sub>14</sub>, L<sub>21</sub>, I<sub>34</sub>, T<sub>53</sub>, R<sub>55</sub>, Q<sub>56</sub> and M<sub>73</sub> (for the second cholates molecule), as derived from the X-ray structure of purified holo cl-BABP (Nichesola et al., 2004). From this comparison it appears that only residues in close contact with the first cholates molecule are among those exhibiting i)  $\Delta\delta^{15}\text{N}$  shift  $> 0.5$  ppm upon binding and ii)  $R_{\text{ex}}$  contributions.

It is possible to hypothesise that residues exhibiting the exchange contribution (closed to the first site) determine the selectivity for the binding of the first ligand molecule, thus shifting the equilibrium

towards a conformation capable of binding the second ligand molecule, in agreement with the different selectivity observed for the two sites. This mechanism could explain the high cooperativity observed for ligand binding.

Dynamics data also indicated that at neutral pH the  $\tau_e$  distribution was substantially similar for the apo and holo protein, while an increased flexibility on ps-ns timescale was observed at acidic pH for apo protein. This increased flexibility may be functional to ligand binding and release: indeed the vicinity of negatively charged membranes could induce some pH lowering (Fiori et al., 1999) with a consequent increase in protein flexibility.

Alltogether a disorder to order transition may be envisaged for cL-BABP upon binding. The intrinsic flexibility observed for the apo cL-BABP could allow to recognize a large number of biological targets, while the specific dynamic transition could contribute to determine the specificity for a given ligand. Thus the relaxation data suggest a correlation between motional properties of atoms and their role in ligand binding.

NMR titration experiments with different bile acids are in progress to further investigate the cooperativity of ligand binding.

## Figure captions

Figure 1. Strips taken at the indicated  $^{13}\text{C}$  frequency from a  $^{13}\text{C}$  edited NOESY at 700 MHz showing: the NOEs between  $\text{Y}_{14}\text{H}_\alpha$  and the  $i+3$   $\text{F}_{17}\text{H}_\beta$ , in the alfa helix region, and long range NOEs with  $\text{L}_{118}$  (a) NOEs between  $\text{V}_{82}\text{H}_\gamma$  and its nearby protons (b), to illustrate the quality of data.

Figure 2.  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of doubly labelled 1mM cl-BABP, pH 7, 298 K and 700 MHz. Negative peaks, as selected on multiplicities basis are in blue (a); the region corresponding to threonines  $\text{H}_\beta$ - $\text{C}_\beta$ , highlighted with a square in (a), is expanded in (b).

Figure 3. Amide chemical shift differences ( $\delta$ , ppm) between cl-BABP at pH 7 and 5.6 at 298K.  $\delta = [(\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}}^2/25)/2]^{1/2}$  averages the effect on  $^1\text{H}$  and  $^{15}\text{N}$  shifts. Grey bars indicate residues assigned on the basis of 3D  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TOCSY/NOESY and 2D conventional homonuclear spectra (a);  $\delta_{\text{N}}(\text{pH}7) - \delta_{\text{N}}(\text{pH}5.7)$  as a function of residue number (b). Dotted lines at  $\pm 0.1$  ppm are drawn to easily identify residues exhibiting higher shifts.

Figure 4. Selected regions of the  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of 1mM cl-BABP at 500.13 MHz and 298 K, at pH 7 (a), and pH 5.6 (b).

Figure 5. Secondary  $\text{H}_\alpha$  chemical shifts of apo cl-BABP (pH= 7, 298 K): positive bars correspond to extended structure, negative bars to helical structure (a); secondary  $\text{C}_\alpha$  chemical shifts: positive bars correspond to helical structure and negative bars to extended structure (b); secondary  $\text{H}_\alpha$  chemical shifts of holo cl-BABP (pH= 7, 298 K) (c).

Figure 6. Schematic representation of the antiparallel beta sheet structure in cl-BABP. Only backbone atoms and sequence numbers of amino acids are shown. The distance restraints and H-

bonds used for structure calculation are depicted in red and blue, respectively. Capital letters indicate the strands.

Figure 7. Average secondary  $H_{\alpha}$  chemical shifts for each secondary structure element of cl-BABP apo (black) and holo (red), as obtained from data recorded at 700 MHz, pH7 298 K.

Figure 8. Superposition of the final ten best minimised DYANA structures of cl-BABP.

Figure 9.  $^{15}\text{N}$   $R_2/R_1$  rates at pH 5,7 of apo cl-BABP (a), and heteronuclear NOE effects (b) obtained at two different fields: 600 MHz (black) and 700 MHz (red). The same rates (c) and heteronuclear NOE effects (d) are reported at pH 7.

Figure 10.  $^{15}\text{N}$   $R_2/R_1$  rates at pH 7 of holo cl-BABP(a), and heteronuclear NOE effects (b) obtained at two different fields: 600 MHz (black) and 700 MHz (red).

Figure 11.  $\tau_e$  contributions, as a function of residue number, as derived from model free approach at pH 7 (red) and 5.7 (black) (a). Residues affected by  $\tau_e$  at pH 7 and 5.7 are highlighted on the protein structures (b) and (c), respectively. Three color codes are used at both pHs, cyano (21-190 ps), blue (700-850 ps), purple (1657 ps) at pH 7 (b) and yellow (26-190 ps), orange (290-400 ps), red (600-720 ps) at pH 5.7 (c).

Figure 12.  $R_{ex}$  contributions derived for apo cl-BABP, as a function of residue number, from model free approach at pH 7 (red) and at pH 5.7 (black).

Figure 13. Graphical analysis of spectral density values (Krizova et al., 2004) derived for apo cl-BABP at 700 MHz:  $J(0.87\omega_H)$  vs.  $J(0)$  (upper panel) and expansion of the same graph to visualize

residues outside the “cloud” of data points (lower panel), (a);  $J(\omega_N)$  vs  $J(0)$  (upper panel) and expansion of the same graph to visualize residues outside the “cloud” of data points (lower panel), (b).

Figure 14. Graphical analysis of spectral density values (Krizova et al., 2004) derived for holo cl-BABP at 700 MHz:  $J(0.87\omega_H)$  vs.  $J(0)$  (upper panel) and expansion of the same graph to visualize residues outside the “cloud” of data points (lower panel), (a);  $J(\omega_N)$  vs  $J(0)$  (upper panel) and expansion of the same graph to visualize residues outside the “cloud” of data points (lower panel), (b).

Figure 15.  $R_{ex}$  values obtained from equation [12] (blue), [13] (green) and from model-free approach (red) at pH 7(a).

Figure 16. Number of restraints per residue, as obtained from the program DYANA (Guntert et al., 1997) (a); distribution of average global displacement: grey bars indicate missing assignment (due to  $^{15}\text{N}$  line broadening) and red bars indicate residues without long range NOEs.

Figure 17. All the residues exhibiting an  $R_{ex}$  contribution as derived from Lipari-Szabo (Peng and Wagner, 1995) and reduced spectral density mapping (Farrow et al., 1995) at pH 7 (left) and pH 5.7 (right) are reported on the structure. In left and right panel, respectively, cyan and in yellow indicate residues detected by Lipari-Szabo while blue and orange indicate additional residues detected by reduced spectral density mapping.

Figure 18.  $R_1$  (a) and  $R_2$  (b) obtained for apo cl-BABP at 700 MHz, 298K, pH 5.7 (red) and pH 7 (blue).

Figure 19.  $\Delta\delta^{15}\text{N}$  [ $\text{abs}(\delta^{15}\text{N apo} - \delta^{15}\text{N holo})$ ] (pH 7) as a function of residue number. Dotted line at + 0.5 ppm is drawn to identify residues exhibiting higher shifts.

#### References

Bakowies, D., and van Gunsteren, W. F. (2002). Simulations of apo and holo-fatty acid binding protein: structure and dynamics of protein, ligand and internal water. *J Mol Biol* 315, 713-736.

Barthe, P., Chiche, L., Declerck, N., Delsuc, M. A., Lefevre, J. F., Malliavin, T., Mispelter, J., Stern, M. H., Lhoste, J. M., and Roumestand, C. (1999). Refined solution structure and backbone dynamics of  $^{15}\text{N}$ -labeled C12A-p8MTCP1 studied by NMR relaxation. *J Biomol NMR* 15, 271-288.

Cicero, D. O., Melino, S., Orsale, M., Brancato, G., Amadei, A., Forlani, F., Pagani, S., and Paci, M. (2003). Structural rearrangements of the two domains of *Azotobacter vinelandii* rhodanese upon sulfane sulfur release: essential molecular dynamics,  $^{15}\text{N}$  NMR relaxation and deuterium exchange on the uniformly labeled protein. *Int J Biol Macromol* 33, 193-201.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry* 33, 5984-6003.

Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995). Spectral density function mapping using  $^{15}\text{N}$  relaxation data exclusively. *J Biomol NMR* 6, 153-162.

Feng, W., Tejero, R., Zimmerman, D. E., Inouye, M., and Montelione, G. T. (1998). Solution NMR structure and backbone dynamics of the major cold-shock protein (CspA) from *Escherichia coli*: evidence for conformational dynamics in the single-stranded RNA-binding site. *Biochemistry* 37, 10881-10896.

Fiori, S., Renner, C., Cramer, J., Pegoraro, S., and Moroder, L. (1999). Preferred conformation of endomorphin-1 in aqueous and membrane-mimetic environments. *J Mol Biol* 291, 163-175.

Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273, 283-298.

Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278, 313-352.

Krizova, H., Zidek, L., Stone, M. J., Novotny, M. V., and 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. *J Biomol NMR* 28, 369-384.

Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994). Measurement of HN-H  $\alpha$  J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J Biomol NMR* 4, 871-878.

Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231, 1049-1067.

Li, Y. C., and Montelione, G. T. (1995). Human type-alpha transforming growth factor undergoes slow conformational exchange between multiple backbone conformations as characterized by nitrogen-15 relaxation measurements. *Biochemistry* 34, 2408-2423.

Lipari, G., and Szabo, A. (1981). Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry* 20, 6250-6256.

Mandel, A. M., Akke, M., and Palmer, A. G., 3rd (1995). Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246, 144-163.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989). Overcoming the overlap problem in the assignment of <sup>1</sup>H NMR spectra of larger proteins by use of three-dimensional heteronuclear <sup>1</sup>H-<sup>15</sup>N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* 28, 6150-6156.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20, 71-75.

Nichesola, D., Perduca, M., Capaldi, S., Carrizo, M. E., Righetti, P. G., and Monaco, H. L. (2004). Crystal structure of chicken liver basic Fatty Acid-binding protein complexed with cholic acid. *Biochemistry* 43, 14072-14079.

Pelton, J. G., Torchia, D. A., Meadow, N. D., Wong, C. Y., and Roseman, S. (1991).  $^1\text{H}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  NMR signal assignments of IIIgIc, a signal-transducing protein of *Escherichia coli*, using three-dimensional triple-resonance techniques. *Biochemistry* 30, 10043-10057.

Peng, J. W., and Wagner, G. (1995). Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry* 34, 16733-16752.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2, 661-665.

Powers, R., Clore, G. M., Bax, A., Garrett, D. S., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. M. (1991). Secondary structure of the ribonuclease H domain of the human immunodeficiency virus reverse transcriptase in solution using three-dimensional double and triple resonance heteronuclear magnetic resonance spectroscopy. *J Mol Biol* 221, 1081-1090.

Prost, E., Sizun, P., Piotto, M., and Nuzillard, J. M. (2002). A simple scheme for the design of solvent-suppression pulses. *J Magn Reson* 159, 76-81.

Tochtrop, G. P., Bruns, J. L., Tang, C., Covey, D. F., and Cistola, D. P. (2003). Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein. *Biochemistry* 42, 11561-11567.

Tochtrop, G. P., DeKoster, G. T., Covey, D. F., and Cistola, D. P. (2004). A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J Am Chem Soc* 126, 11024-11029.

- Vasile, F., Ragona, L., Catalano, M., Zetta, L., Perduca, M., Monaco, H., and Molinari, H. (2003). Solution structure of chicken liver basic fatty acid binding protein. *J Biomol NMR* 25, 157-160.
- Wei, Y., Lee, D. K., and Ramamoorthy, A. (2001). Solid-state  $(^{13}\text{C})$  NMR chemical shift anisotropy tensors of polypeptides. *J Am Chem Soc* 123, 6118-6126.
- Wishart, D. S., and Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol* 239, 363-392.

Table 1. Spectral density functions of the models used for the  $^{15}\text{N}$  relaxation data analysis

Model	Spectral density function	Optimized parameters
1	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2)\}$	$S^2$
2 <sup>a</sup>	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + (1-S^2)\tau'_e/(1 + \omega^2\tau_e'^2)\}$	$S^2, \tau_e$
3	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2)\}$ $1/T_{2(\text{obs})} = 1/T_2 + R_{\text{ex}}$	$S^2, R_{\text{ex}}$
4	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + (1-S^2)\tau'_e/(1 + \omega^2\tau_e'^2)\}$ $1/T_{2(\text{obs})} = 1/T_2 + R_{\text{ex}}$	$S^2, \tau_e, R_{\text{ex}}$
5 <sup>b</sup>	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + S_f^2(1-S_s^2)\tau'_s/(1 + \omega^2\tau_s'^2)\}$	$S_f^2, S_s^2, \tau_e$

<sup>a</sup> $\tau_e' = \tau_m \tau_e / (\tau_m + \tau_e)$ .

<sup>b</sup> $\tau_s' = \tau_m \tau_s / (\tau_m + \tau_s)$ ;  $S^2 = S_f^2 S_s^2$ .

**Table 2** Analysis of the 10 best structures obtained for apo cl-BABP.

Conformational restraints and structural parameters determined using the DYANA program and AMBER forcefield implemented in the DISCOVER program.

---

(a) Restraints	
Number of upper limit distance restraints	1000
Number of hydrogen bond restraints	13
Number of torsion angle ( $\phi$ ) restraints	48

---

(b) DYANA	
Target function ( $\text{\AA}^2$ )	$2.07 \pm 0.46$
Average number of upper restraint violation $> 0.25 \text{\AA}$ per structure	0
Maximum violation ( $\text{\AA}$ )	0
Average number of angle restraint violations $> 5^\circ$ per structure	0
Maximum violation (degrees)	0
RMSD (backbone atoms) (3-125)	$1.47 \pm 0.22$
RMSD (heavy atoms) (3-125)	$2.16 \pm 0.18$

---

(c) DISCOVER (AMBER forcefield)	
Total energy (kcal/mol)	$-402 \pm 18$
Bond energy (kcal/mol)	$21 \pm 1$
Angle energy (kcal/mol)	$141 \pm 2$
Torsion angle (kcal/mol)	$160 \pm 6$
Out of plane energy (kcal/mol)	$3.9 \pm 0.4$
Hydrogen bond energy (kcal/mol)	$-44 \pm 2$
Lennard-Jones energy (kcal/mol)	$-362 \pm 14$
Coulomb energy (kcal/mol)	$-321 \pm 12$
Restraining potential energy (kcal/mol)	$67 \pm 12$
Average number of upper restraint violations $> 0.25 \text{\AA}$	0
Maximum violation ( $\text{\AA}$ )	0
Average number of angle restraint violations $> 5^\circ$ per structure	0
Maximum violation (degree)	0
RMSD (backbone atoms) (3-125)	$1.86 \pm 0.04$
RMSD (heavy atoms)	$2.94 \pm 0.21$

---

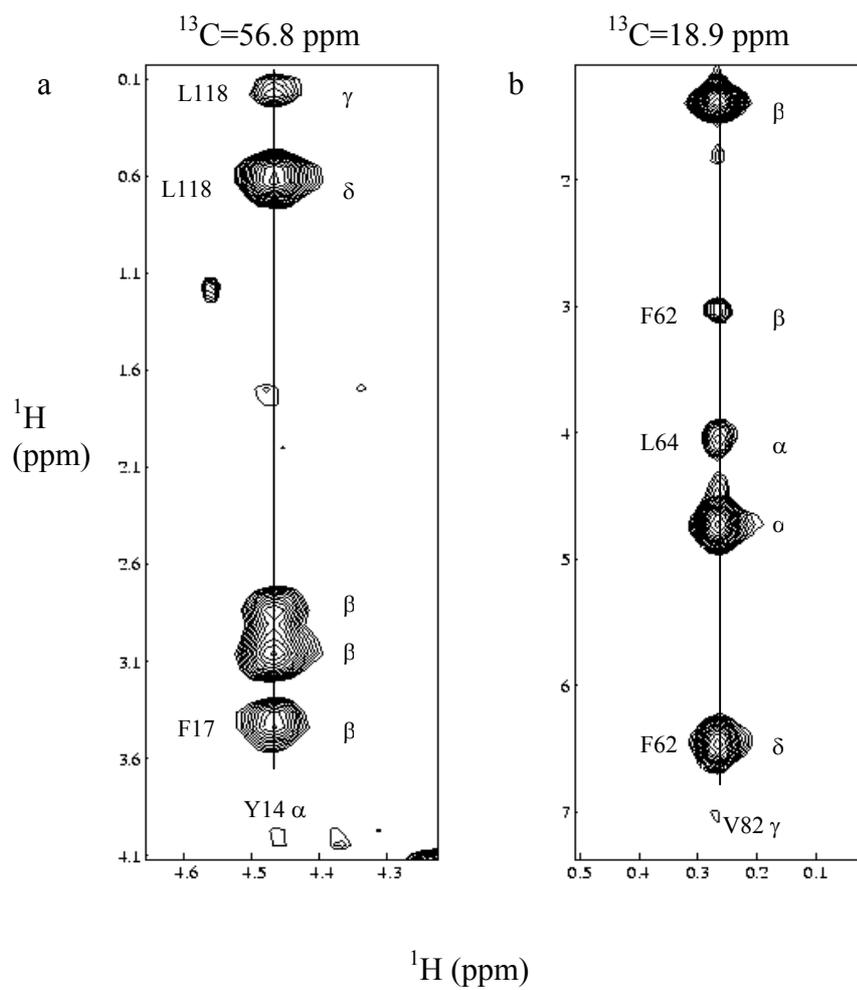


Figure 1

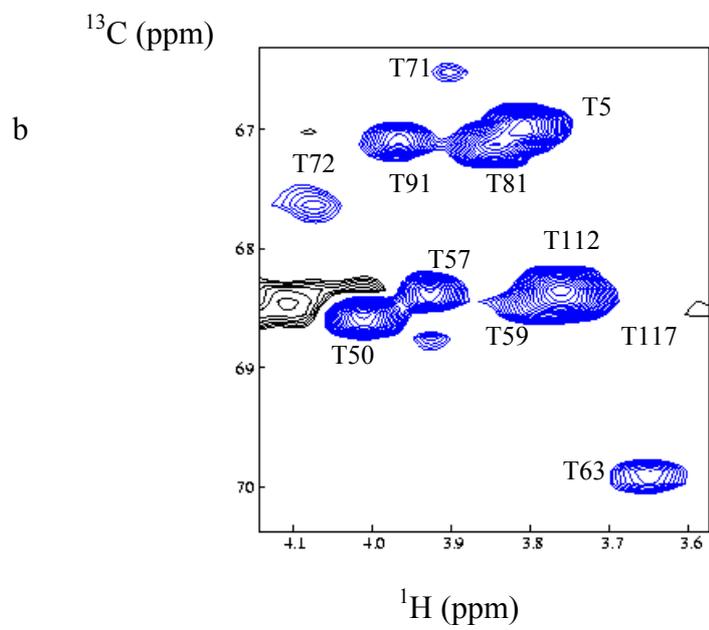
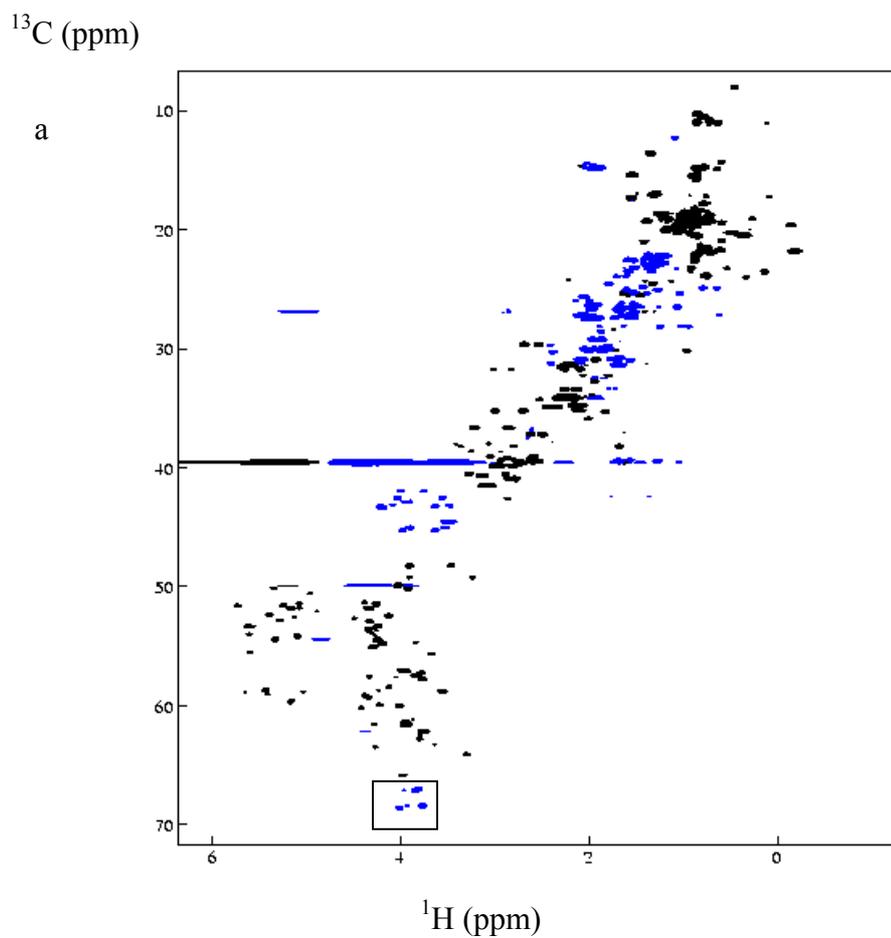
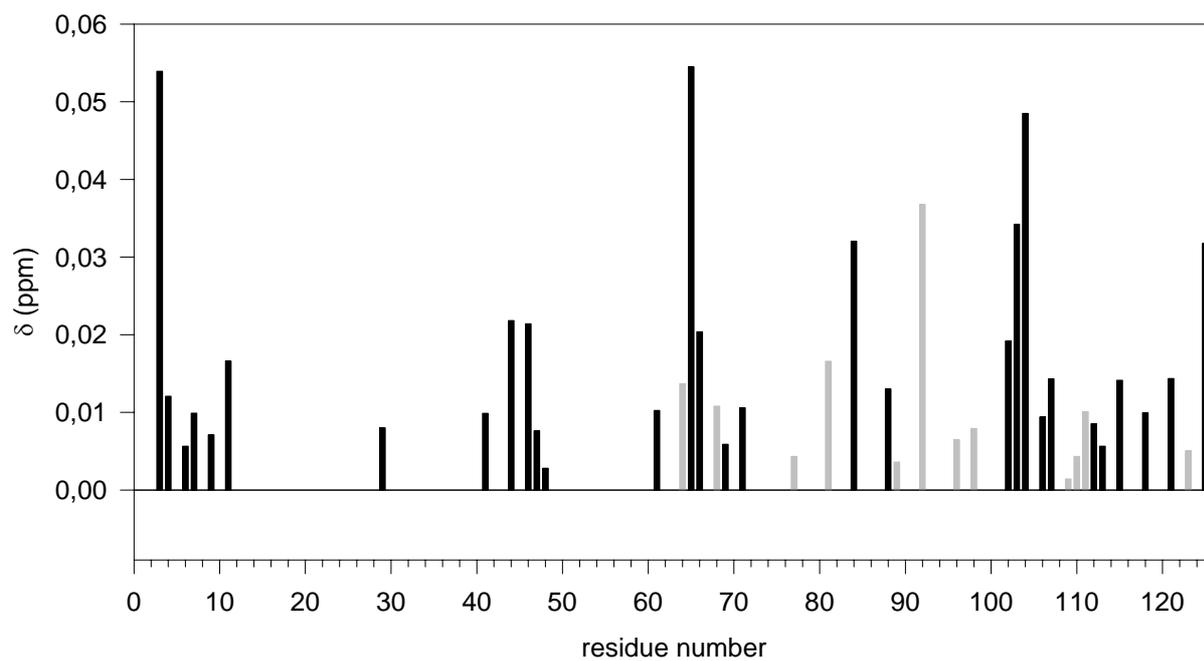


Figure 2

a



b

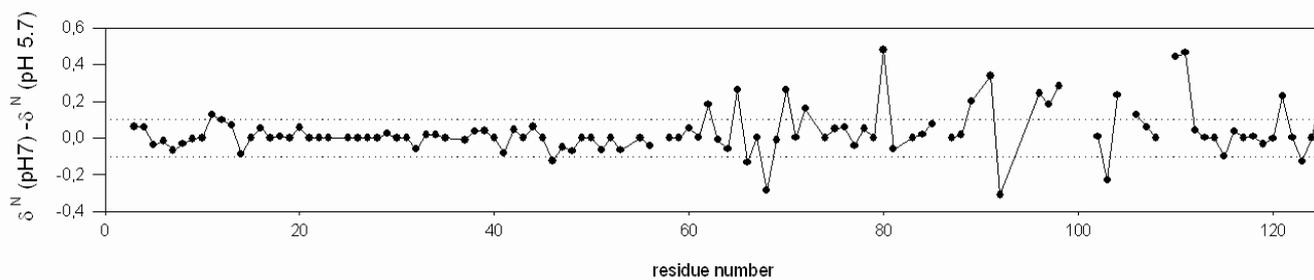


Figure 3

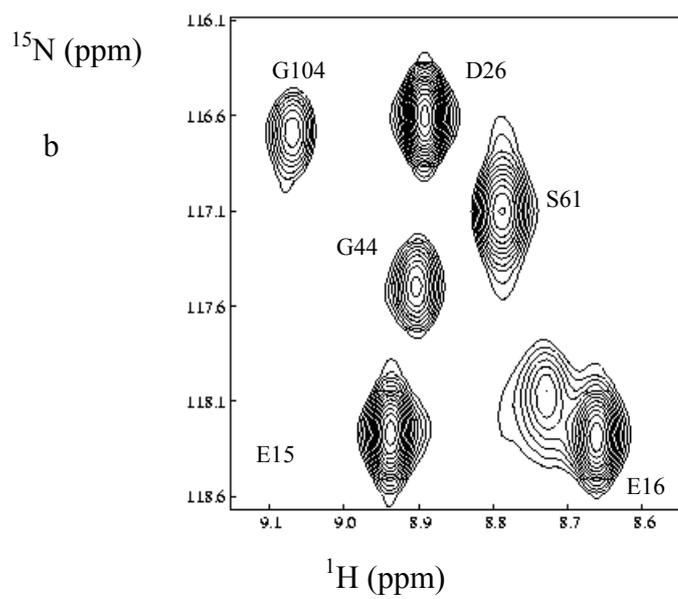
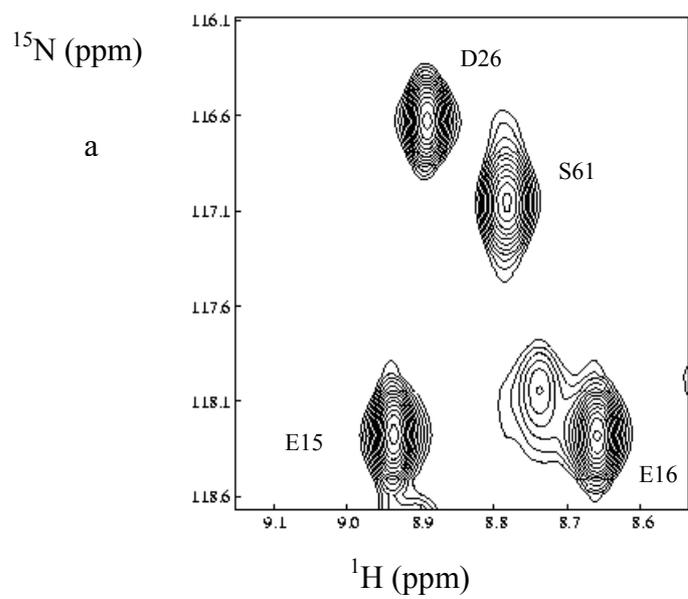


Figure 4

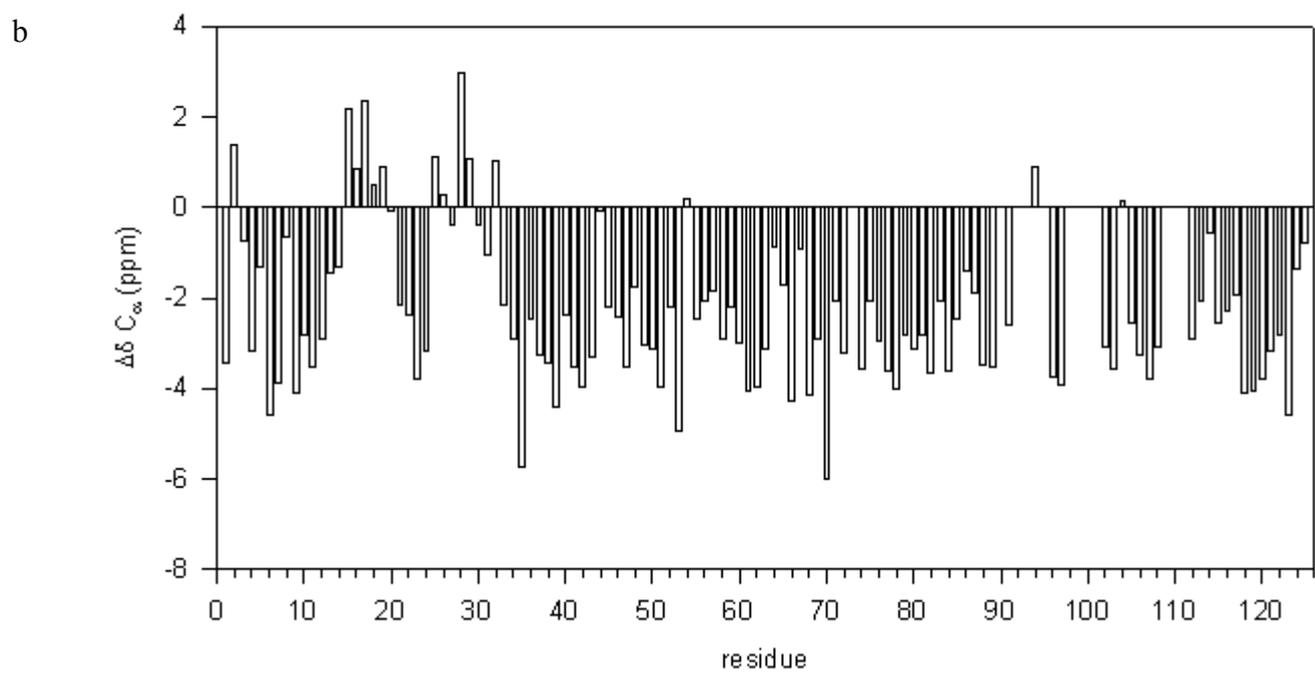
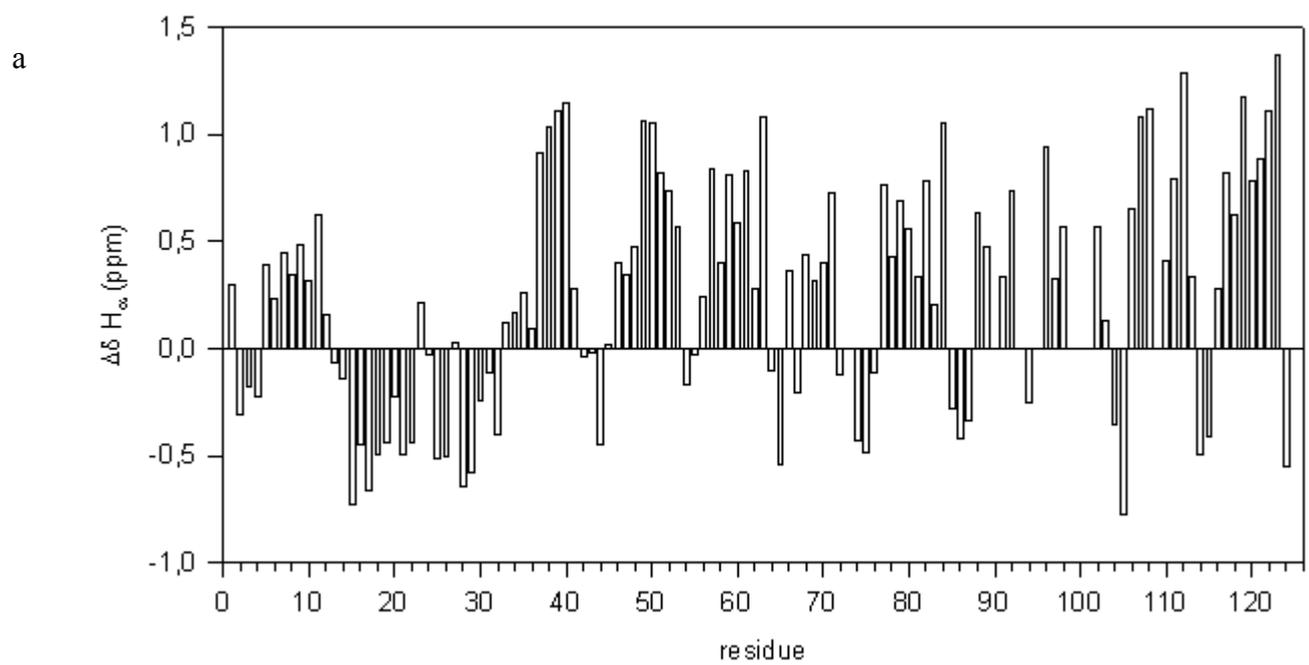


Figure 5

c

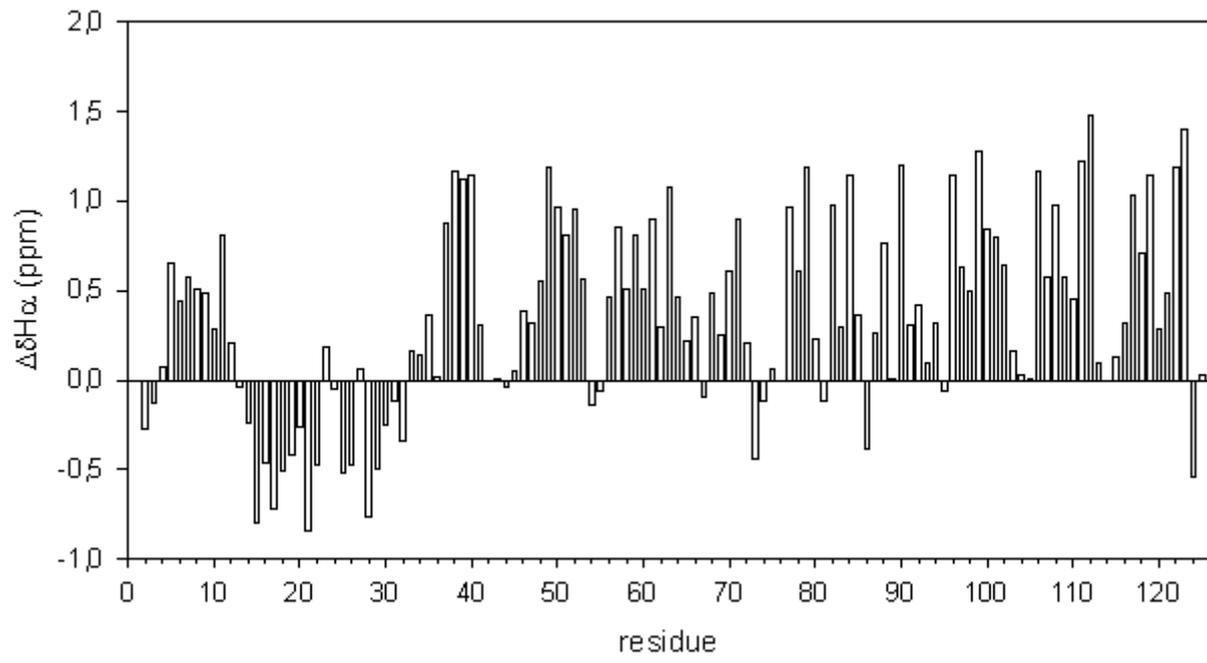


Figure 5

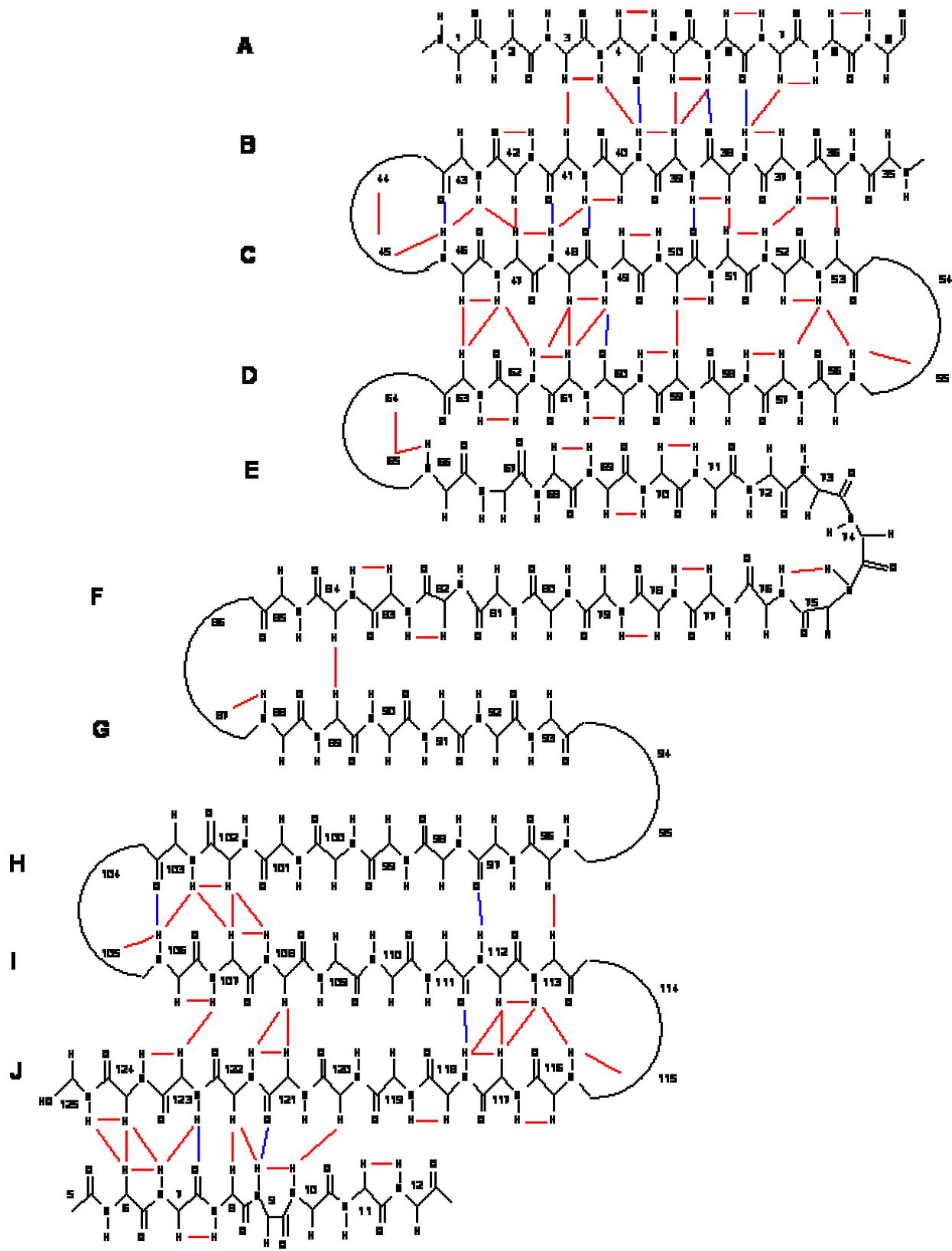


Figure 6

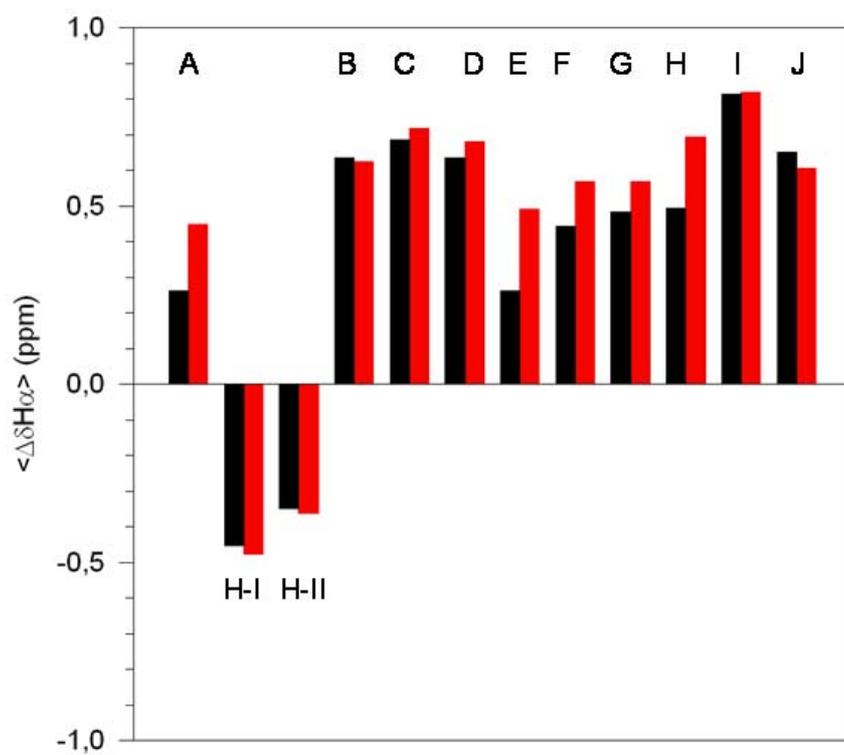


Figure 7

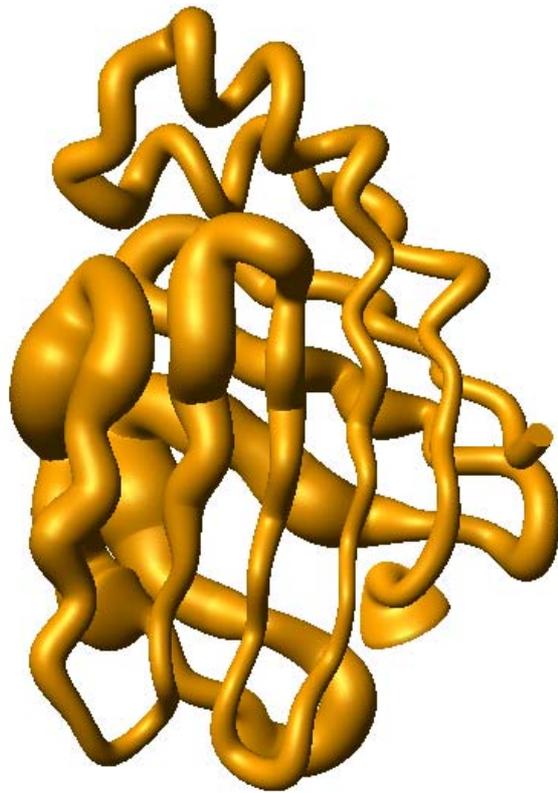


Figure 8

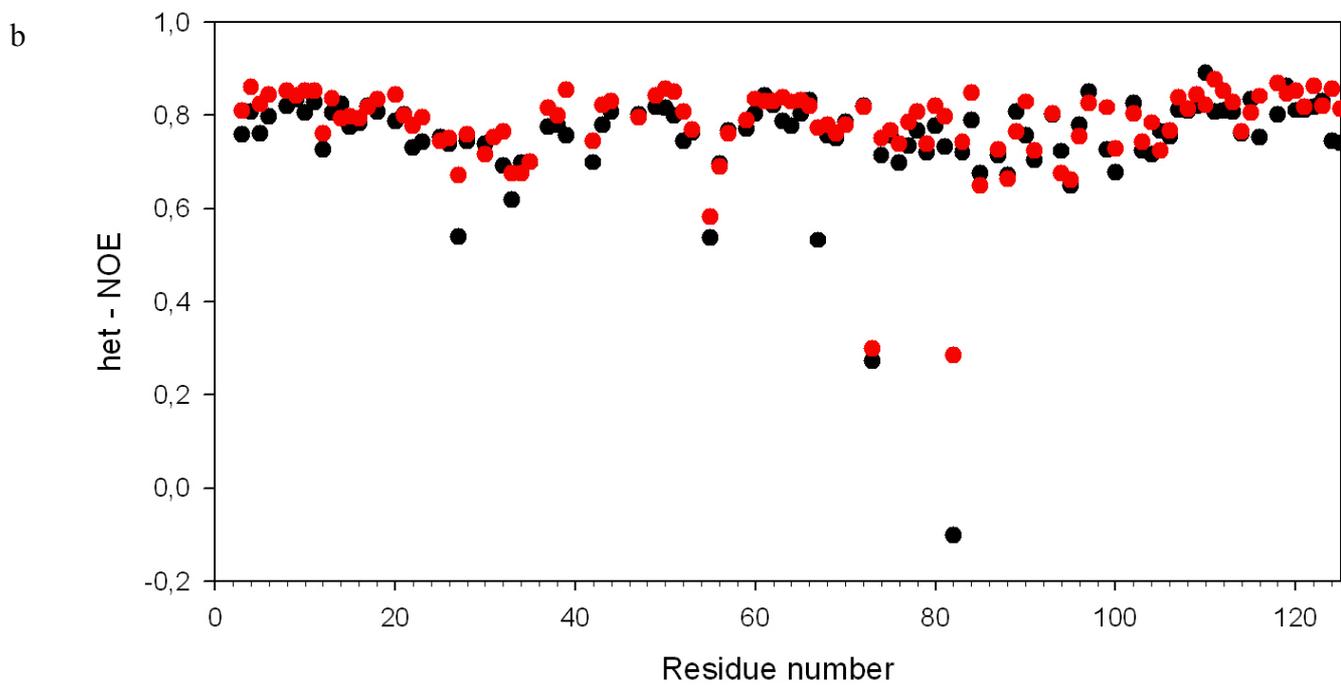
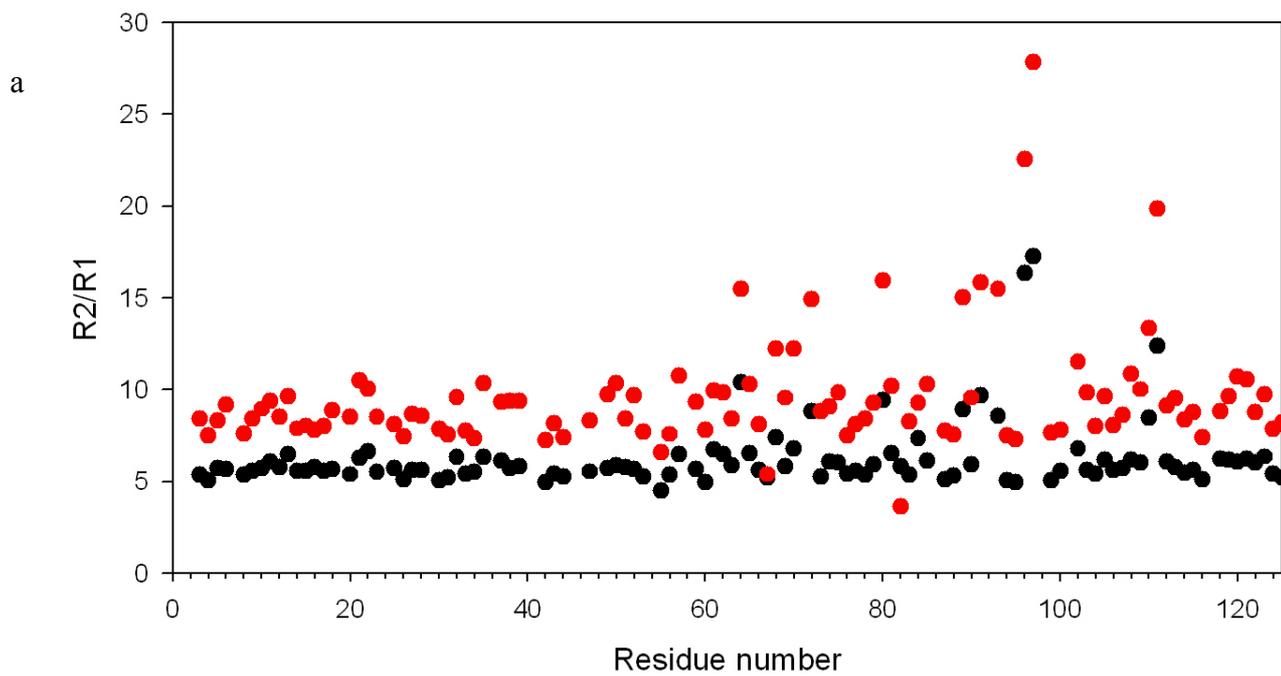
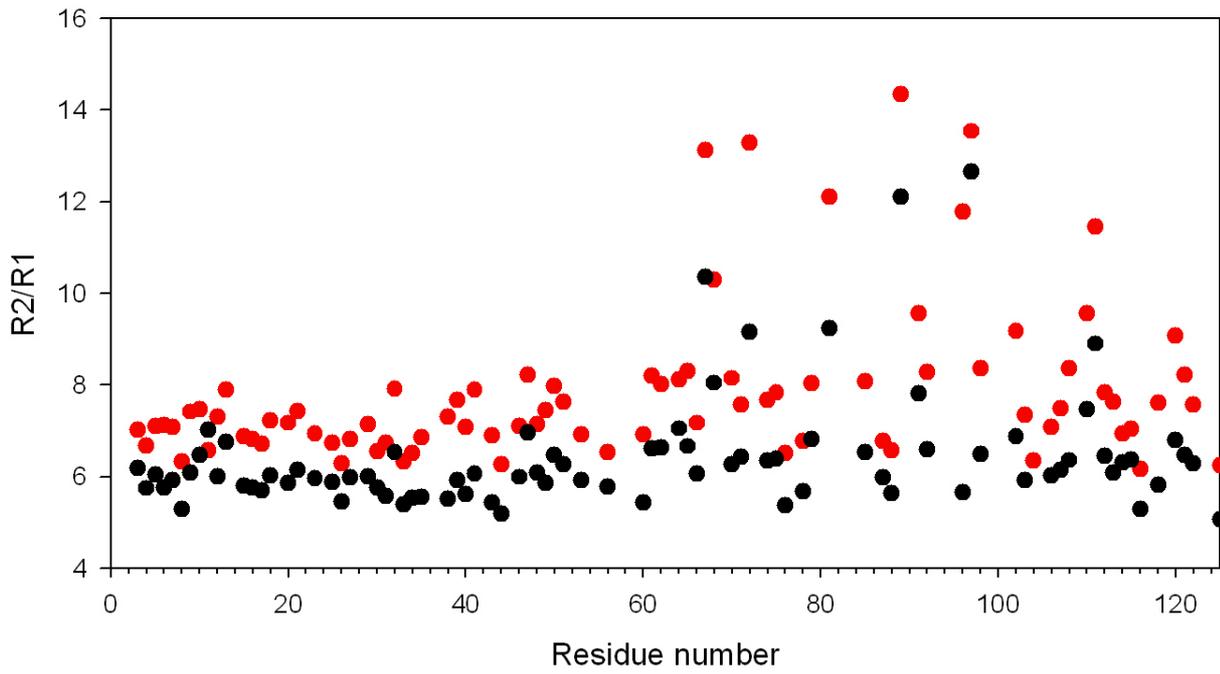


Figure 9

c



d

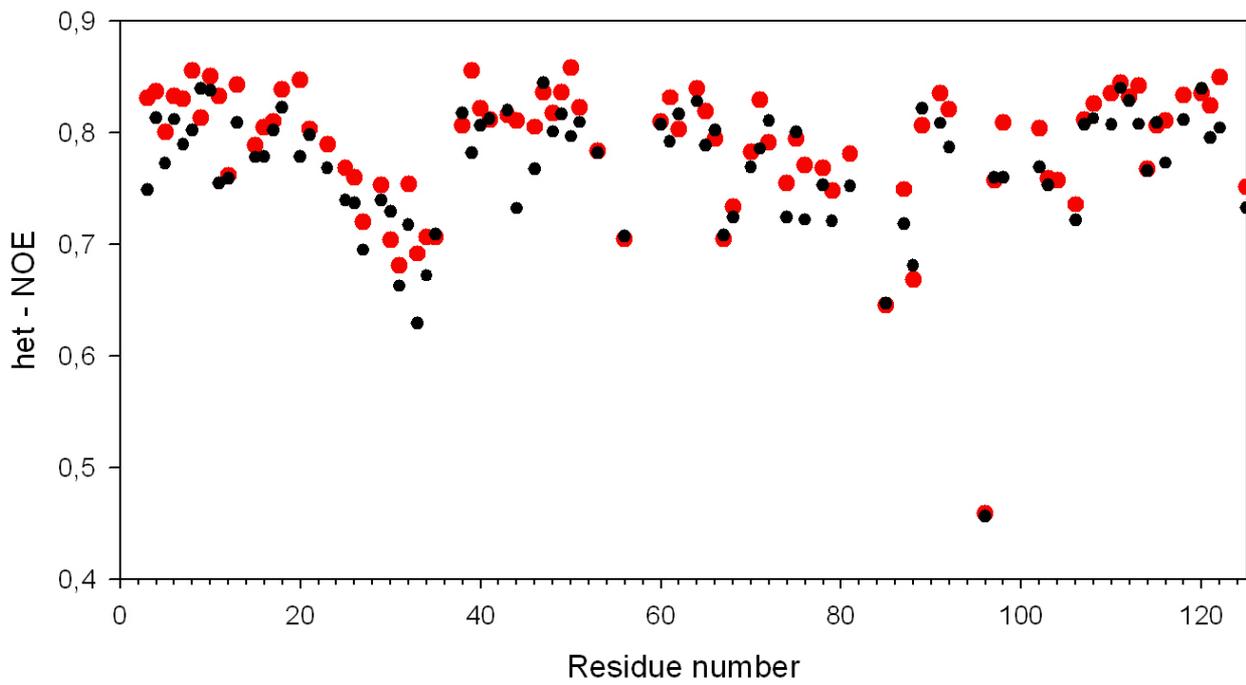


Figure 9

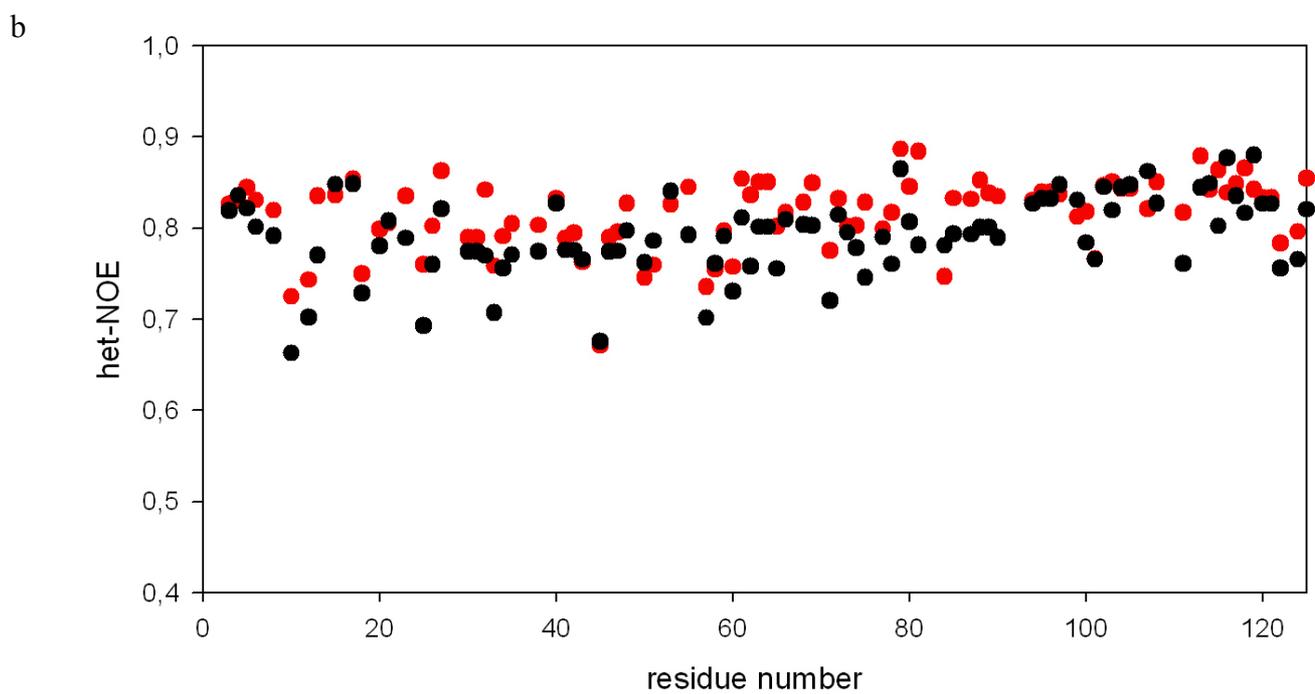
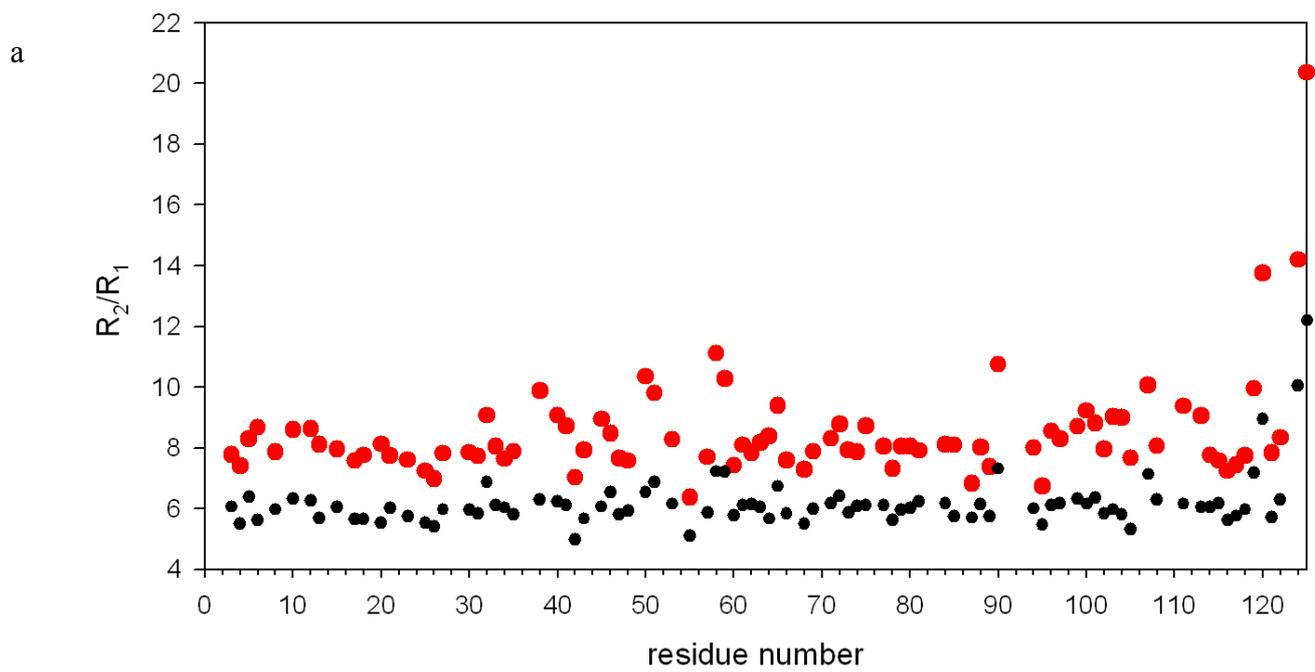
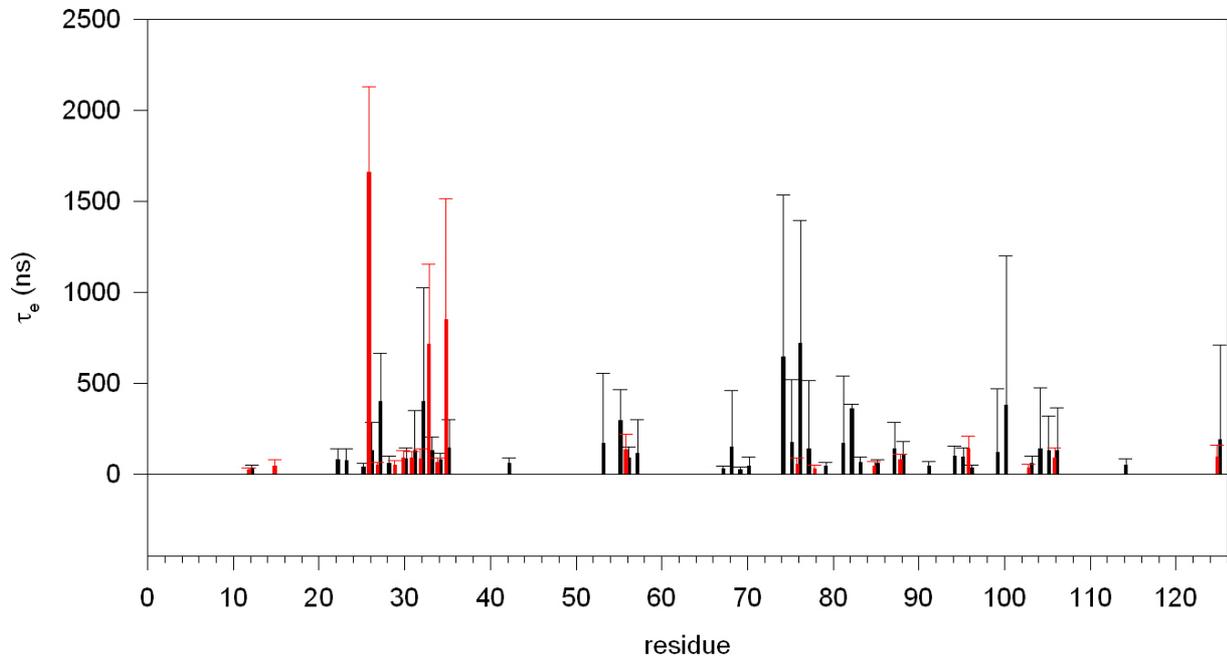
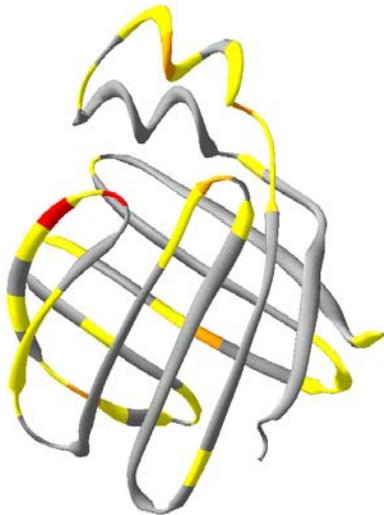


Figure 10

a



b



c

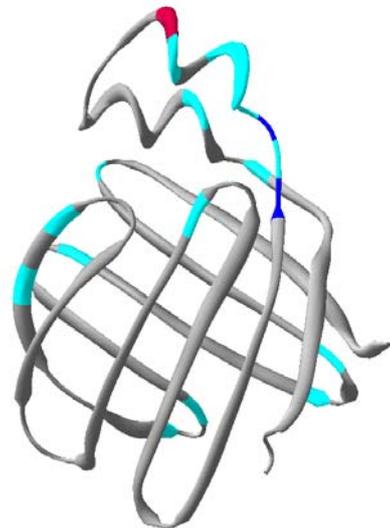


Figure 11

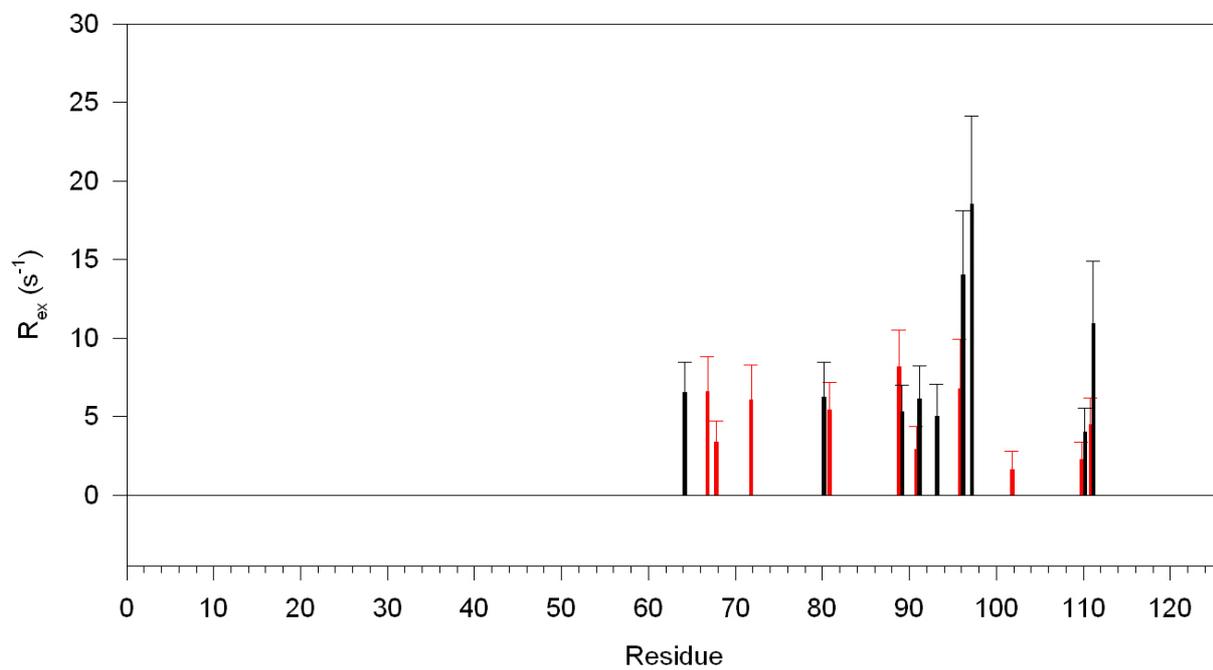


Figure 12

a

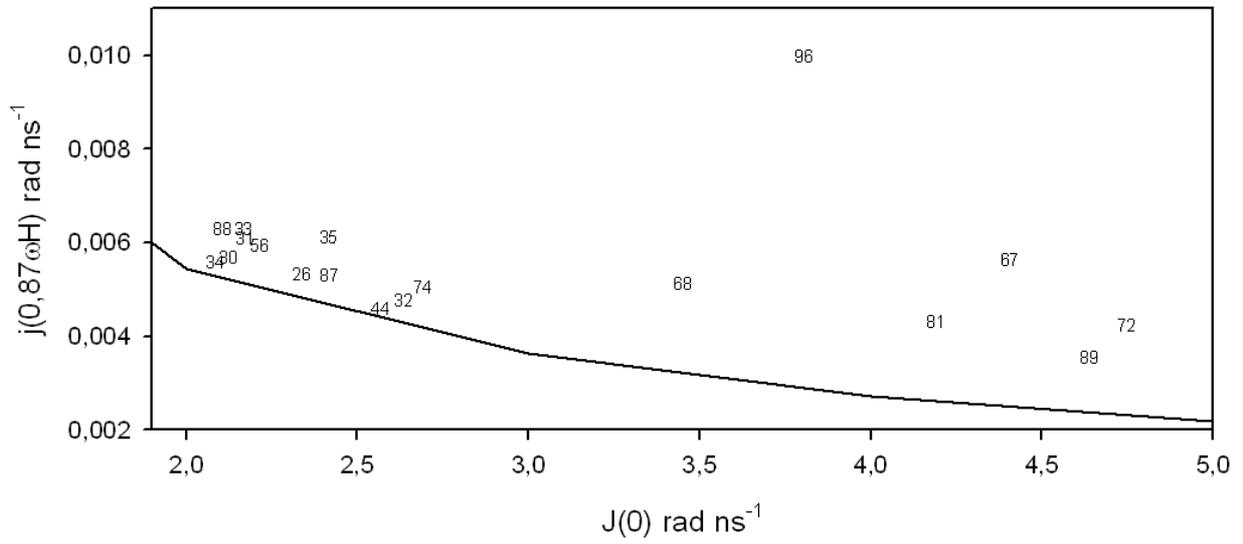
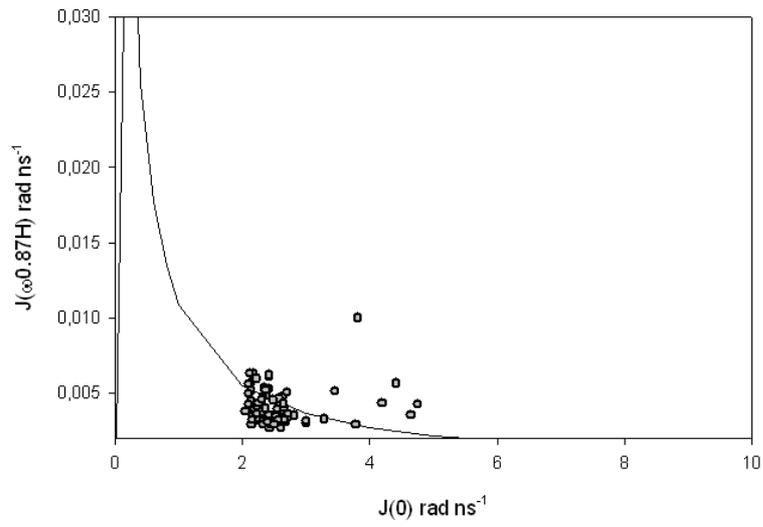


Figure 13

b

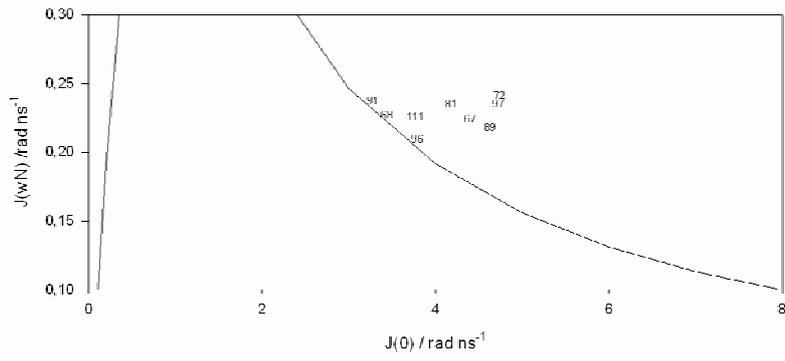
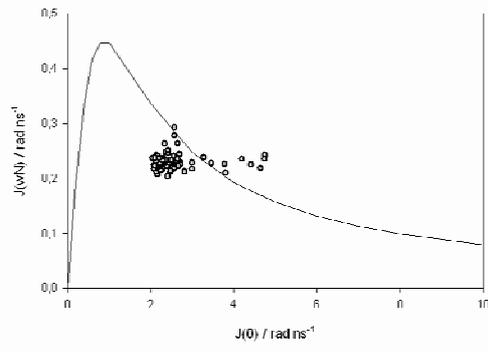


Figure 13

a

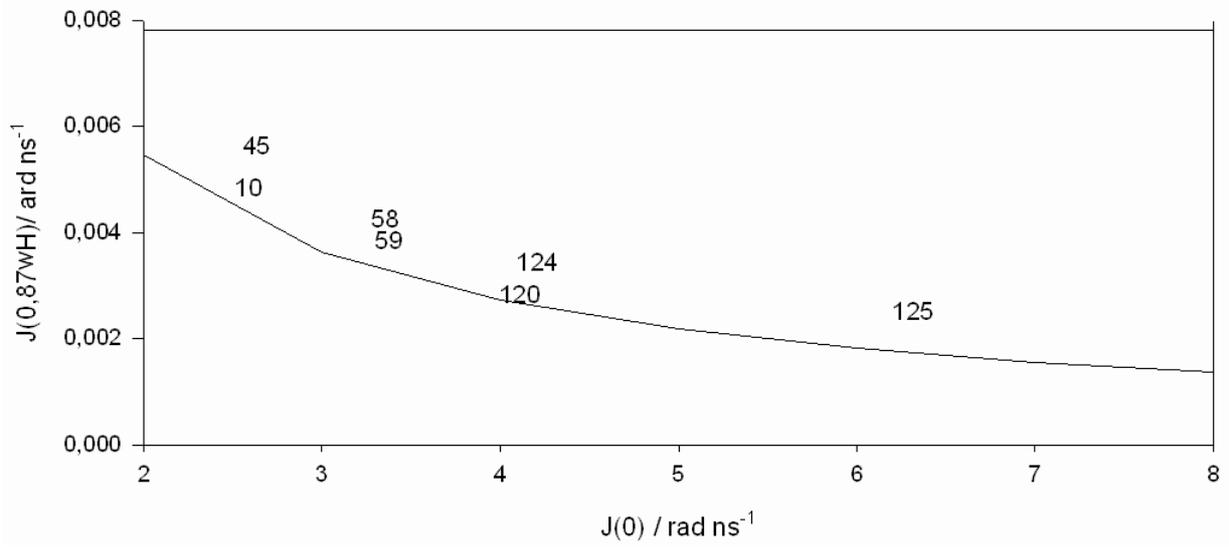
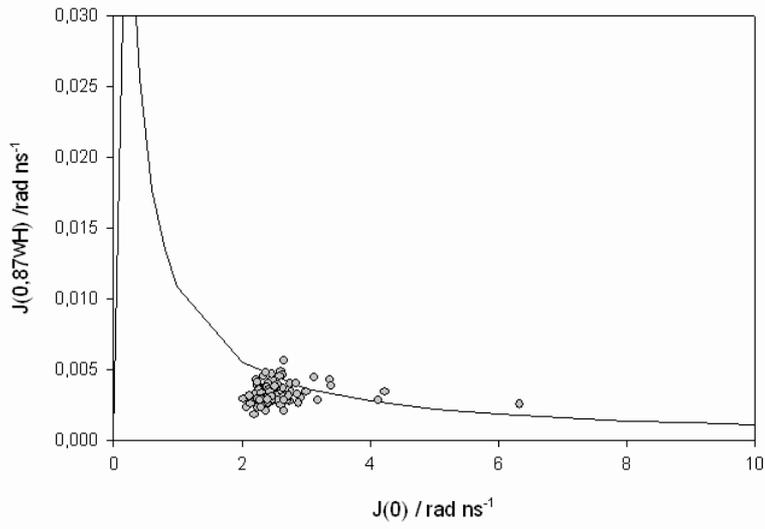


Figure 14

b

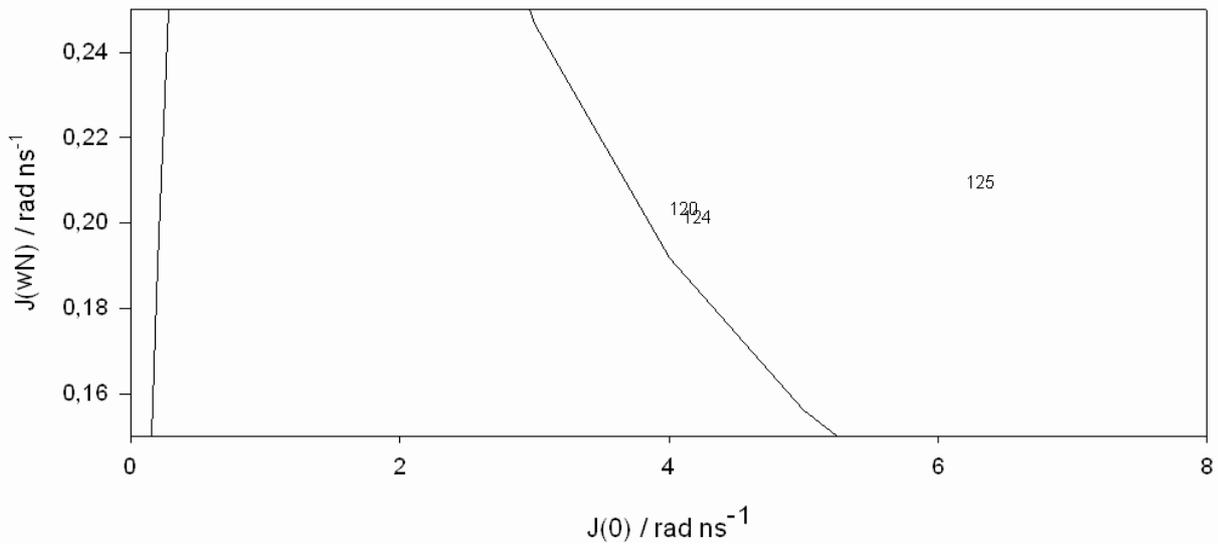
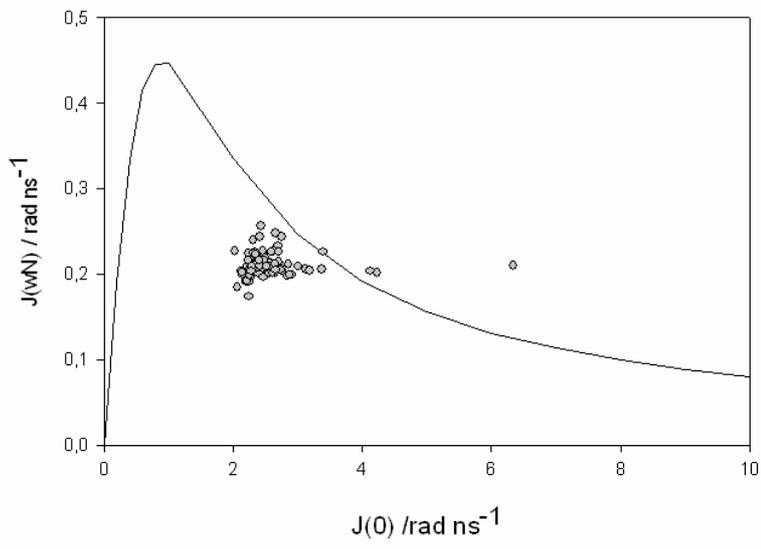


Figure 14

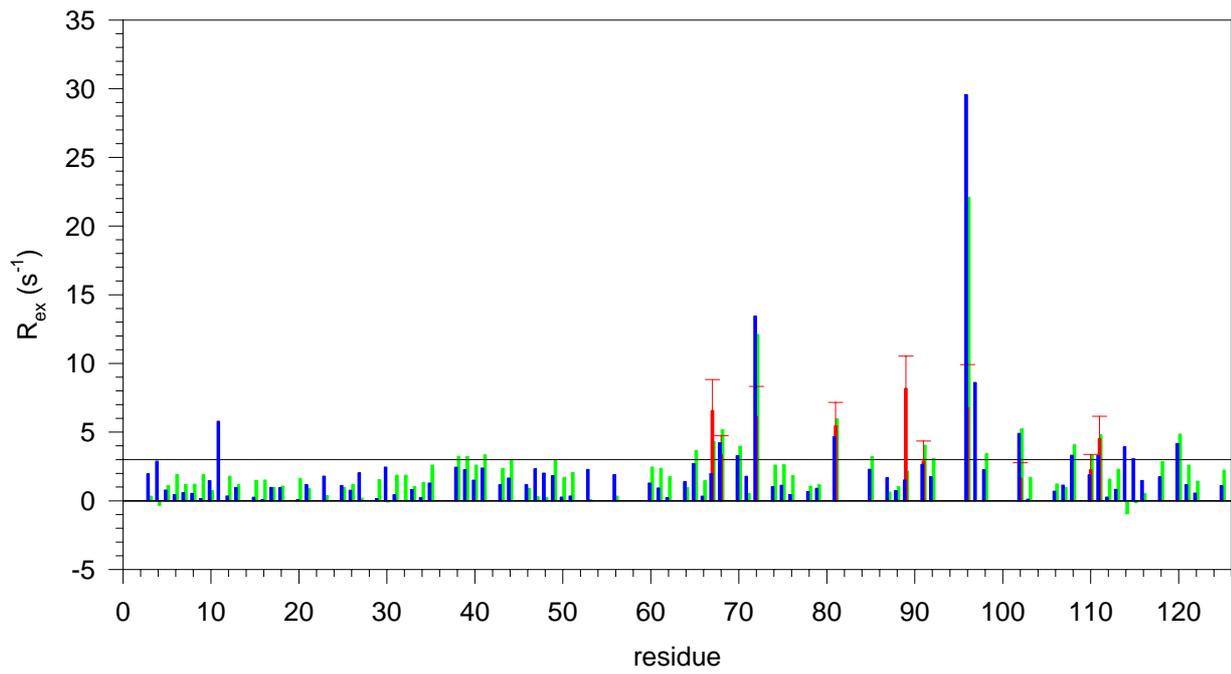


Figure 15

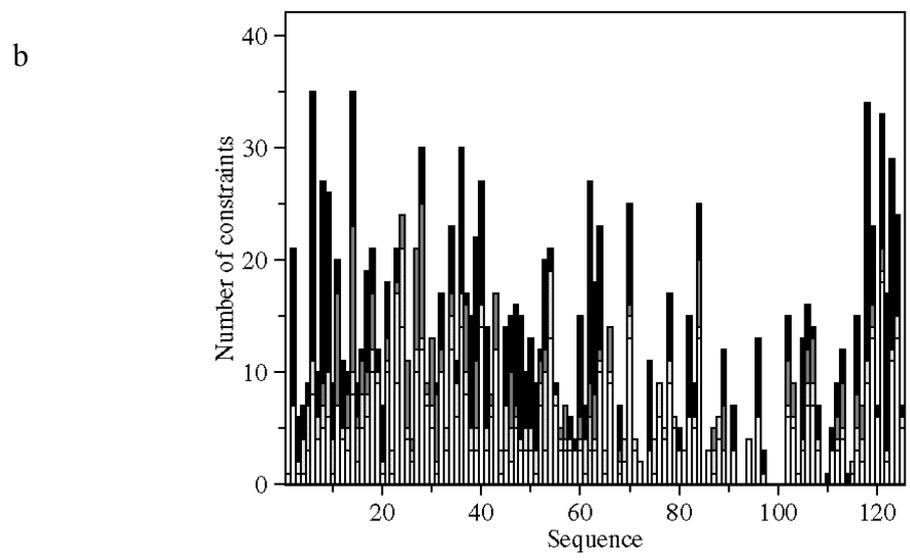
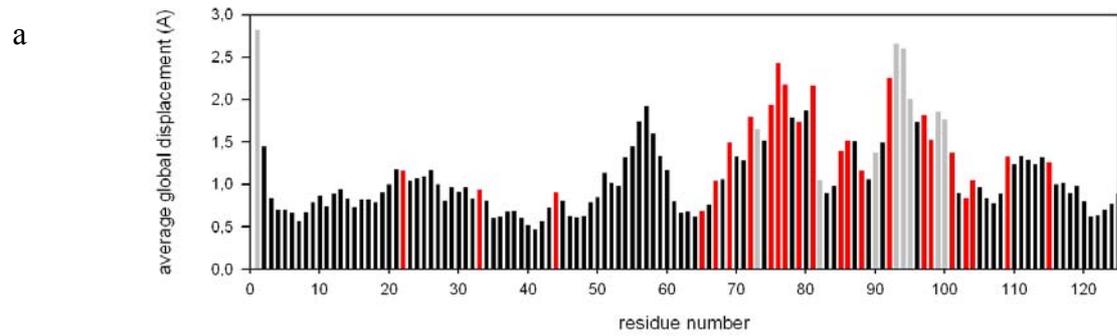


Figure 16

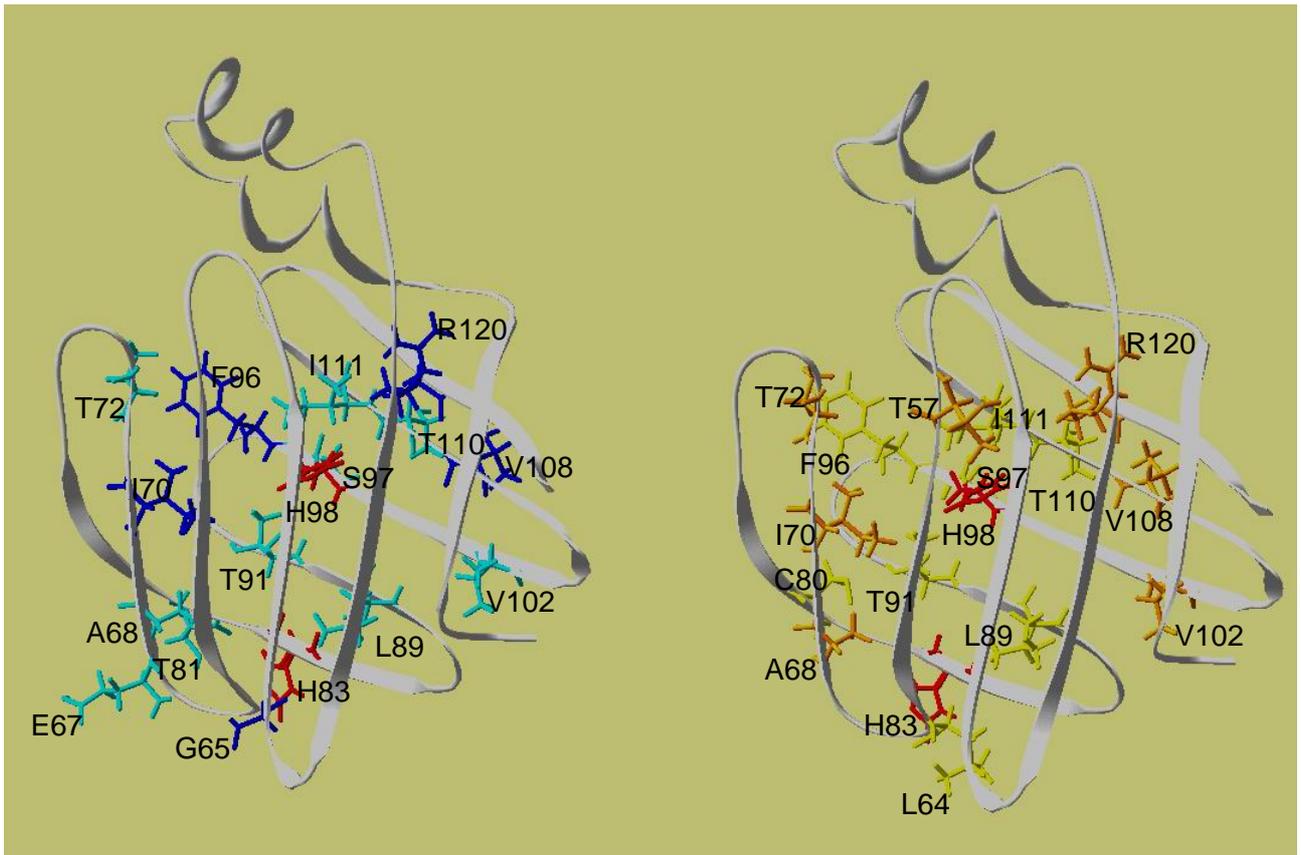


Figure 17

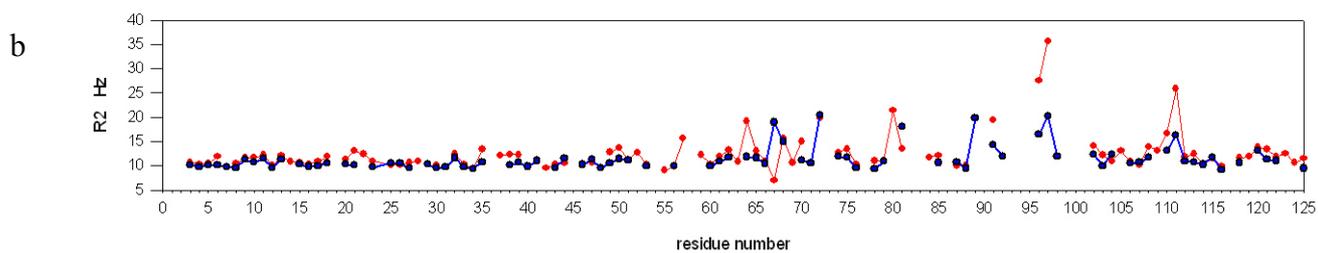
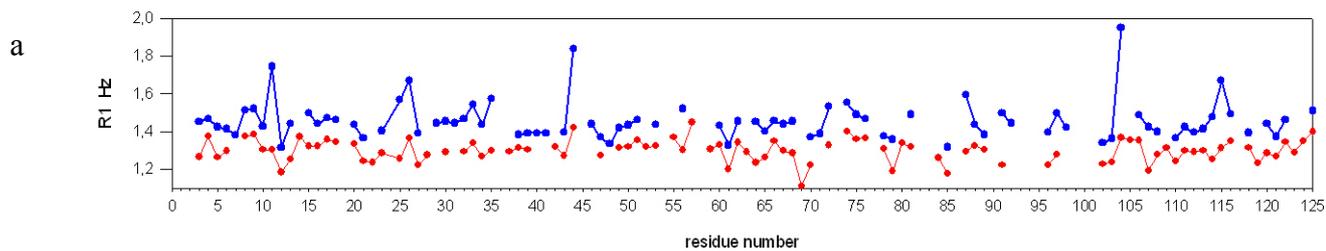


Figure 18

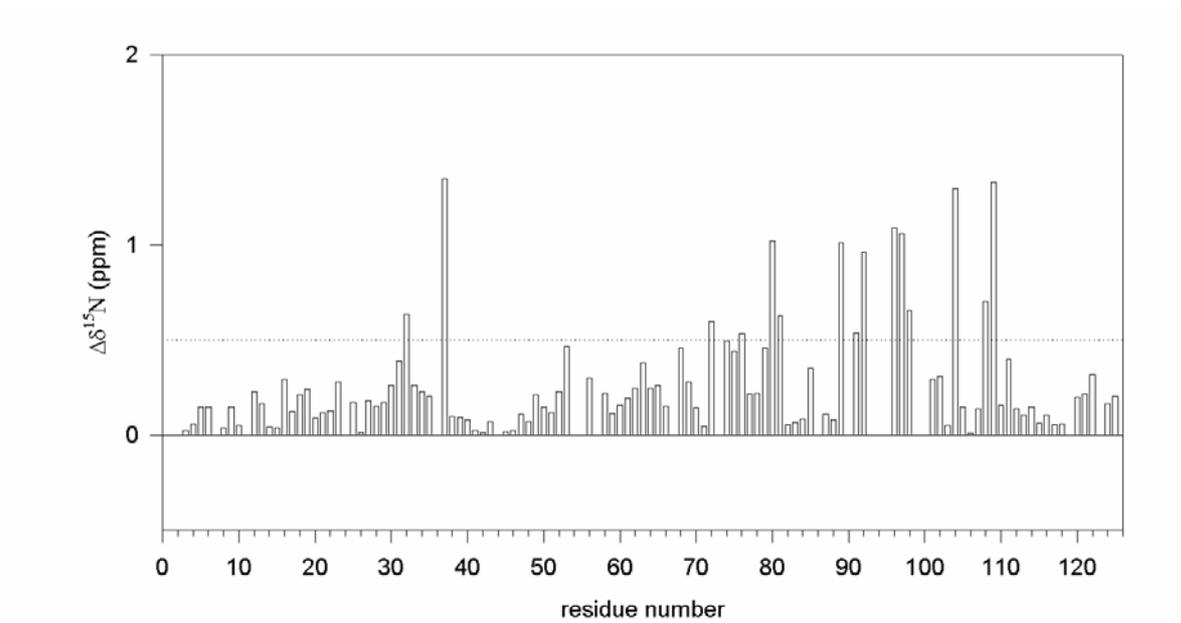


Figure 19

**Table 1: Chemical shift of apo cl-BABP at pH 7**

Residue number	$\delta C\alpha$	$\delta CO$	$\delta H\alpha$	$\delta HN$	$\delta N$
1	49,061		4,645		
2	59,104	173,533	4,37		
3	57,758	172,054	4,333	8,024	111,379
4	65,514	178,679	3,758	9,079	111,766
5	60,509	180,027	4,743	8,366	117,922
6	52,917	172,726	4,918	9,484	127,91
7	51,816	173,819	4,815	9,371	123,72
8	61,546	173,931	4,276	8,286	131,504
9	54,139	168,783	5,089	9,504	124,186
10	49,674	172,831	4,667	7,341	126,041
11	52,161	171,804	4,989	8,9	120,435
12	53,723	173,709	4,445	9,269	124,906
13	51,675	173,801	4,681	9,215	118,884
14	56,835	174,712	4,481	7,882	118,255
15	58,781	176,221	3,55	8,957	118,286
16	57,478	176,923	3,843	8,684	118,39
17	60,049	173,637	4,002	8,183	123,588
18	55,6	177,308	3,657	8,501	118,021
19	57,108	178,43	3,926	7,913	117,705
20	52,399	176,893	4,124	7,957	125,759
21	52,936	173,652	3,657	7,253	116,977
22	50,127	174,638	3,915	7,984	118,582
23	51,234	172,007	4,385	7,068	119,817
24	60,126	175,519	4,459		
25	57,744	175,743	3,776	8,811	122,577
26	54,484	176,072	4,255	8,908	116,65
27	54,716	177,166	4,155	7,261	120,788
28	64,069	175,205	3,284	7,887	121,589
29	57,257	176,191	3,776	7,858	117,527
30	55,028	175,265	4,268	7,236	115,577
31	51,436	177,968	4,258	8,059	121,089
32	57,019	173,622	3,989	7,991	116,728
33	52,041	173,511	4,882	8,01	117,394
34	58,472	173,123	4,114	7,579	120,904
35	50,477	171,802	4,679	8,769	129,463
36	60,854	171,931	4,584		
37	57,843	173,085	4,853	8,038	118,703
38	58,773	172,592	5,039	9,574	128,961
39	52,209	173,744	5,374	9,624	128,503
40	58,713	168,843	5,119	9,582	125,803
41	52,456	171,068	4,654	9,238	128,102
42	51,729	172,054	4,334	8,241	125,858
43	52,922	173,607	4,317	8,799	130,378
44	45,019	171,447	3,519	8,926	117,572
45	51,995	171,918	4,777	8,812	127,669

46	51,771	172,571	5,212	8,022	120,443
47	54,179	174,04	4,986	9,2	122,089
48	60,451	173,085	4,427	8,473	121,452
49	59,115	173,518	5,014	9,316	128,201
50	59,087	171,147	5,432	9,727	128,055
51	54,634		5,325	9,335	122,484
52	53,947	173,352	5,1	9,64	126,713
53	56,831	170,576	4,926	8,453	116,286
54	63,478	175,047	4,27		
55	53,778	173,548	4,352	8,009	111,721
56	53,637	171,7	4,61	7,589	118,164
57	59,979	170,998	5,22	8,835	123,732
58	59,354	172,279	4,337	8,832	127,672
59	59,596	170,727	5,153	8,953	126,018
60	50,112	170,5	5,349	9,223	125,541
61	54,356	170,372	4,837	8,798	116,964
62	54,037		4,92	8,166	118,554
63	58,745		5,431	9,29	119,189
64	54,24		4,019	9,054	125,198
65	43,366	171,947	4,186	8,845	108,444
66	51,934		4,776	7,807	120,498
67	59,908		4,087	8,674	124,06
68	48,37		4,793	9,166	127,715
69	51,3		5,077	8,467	121,504
70	55,251	172,951	4,354	8,913	126,502
71	59,803		5,078	9,448	123,923
72	58,6		4,226	8,947	119,871
74	50,619	173,966	4,308	8,299	114,872
75	43,17	171,947	4,076	8,147	108,544
76	53,271	171,462	4,244	7,44	121,621
77	52,575	174,002	5,124	7,94	118,149
78	51,109	173,484	4,599	9,137	125,576
79	53,391		5,048	8,499	124,285
80	55,078		5,253	9,213	124,456
81	59		4,685	8,596	113,251
82	58,56		4,59		
83	52,918	170,1	4,86	8,689	124,067
84	51,51	173,88	5,227	8,722	123,12
85	49,033	174,897	4,068	9,095	129,345
86	51,7		4,327	8,226	119,864
87	43,212	170,708	3,636	8,592	103,64
88	52,736	172,488	4,999	7,889	120,696
89	51,594		4,55	8,471	126,526
90					
91	59,2		4,78	8,677	125,718
92	53,033			8,849	125,383
93	57,4		4,254		

94	57,5		4,039		
95					
96	53,98		5,603	7,477	120,632
97	54,4		4,83	8,148	119,584
98				8,774	125,656
99					
100					
101				8,66	122,592
102	59,14	172,935	4,517	8,773	126,884
103	52,666	174,306	4,496	8,948	129,431
104	45,263	171,53	3,613	9,104	116,974
105	50,572	171,043	4,959	8,914	125,509
106	53,277	171,343	5,56	8,102	120,347
107	51,632	173,778	5,13	8,788	123,819
108	59,134		4,699	8,974	126,784
109				9,243	127,815
110	59,317			8,799	120,02
111	58,564	171,416		9,501	132,311
112	58,881	171,138	5,638	9,296	123,203
113	55,668	173,413	5,039	9,078	126,117
114	44,506	172,92	3,475	8,49	115,845
115	42,528	171,316	3,558	8,482	105,727
116	59,952	172,009	4,215	8,018	123,893
117	59,884	171,068	5,172	8,702	123,519
118	51,015	172,262	4,772	9,588	132,538
119	57,047		5,164	7,856	125,628
120	52,233		5,2	9,228	126,573
121	52,74	172,228	5,28	8,899	126,509
122	55,483	170,559	5,595	9,305	117,876
123	51,6	173,169	5,728	8,35	118,553
124	54,694	174,084	3,839	8,284	125,842
125	61,403	178,767	3,965	8,32	129,753

**Table 2: Chemical shift of holo cl-BABP at pH 7**

Residue number	$\delta H\alpha$	$\delta HN$	$\delta N$
1			
2	4,385	8,651	
3	4,377	7,957	111,121
4	4,042	9,05	111,329
5	5,004	8,166	117,622
6	5,138	9,638	127,288
7	4,942	9,308	121,713
8	4,454	8,239	131,198
9	5,086	9,51	123,1
10	4,637	7,286	125,612
11	5,181	8,749	
12	4,501	9,004	123,746
13	4,714	9,097	117,955
14	4,363	7,825	118,06
15	3,494	8,958	118,469
16	3,831	9,08	118,4
17	3,945	8,324	123,373
18	3,658	8,275	116,897
19	3,947	7,709	116,284
20	4,083	7,92	126,353
21	3,322	7,386	116,658
22	3,878	7,805	118,054
23	4,357	6,849	118,164
24	4,394		
25	3,772	8,821	123,508
26	4,284	8,898	116,486
27	4,233	7,189	122,018
28	3,191	7,884	120,493
29	3,866	7,777	116,383
30	4,272	7,445	116,869
31	4,232	8,556	120,186
32	4,038	7,767	120,747
33	4,921	8,246	115,94
34	4,091	7,291	120,352
35	4,719	8,961	130,356
36	4,458		
37	4,824	9,067	126,455
38	5,162	9,447	128,677
39	5,414	9,543	128,023
40	5,098	9,642	125,84
41	4,678	9,279	127,946
42	4,367	8,219	125,867
43	4,369	8,819	129,777
44	3,928	8,906	

45	4,809	8,835	127,547
46	5,146	8,028	120,272
47	4,979	9,289	122,469
48	4,503	8,515	121,687
49	5,142	9,353	126,782
50	5,318	9,788	128,873
51	5,307	9,171	122,057
52	5,317	9,343	126,279
53	4,915	9,062	116,853
54	4,304		
55	4,32	8,263	112,964
56	4,833	7,763	116,255
57	5,209	8,486	117,952
58	4,461	8,917	126,046
59	5,164	8,962	126,687
60	5,261	9,108	124,527
61	5,397	8,819	115,667
62	4,959	8,225	116,938
63	5,434	9,764	120,056
64	4,638	9,158	126,844
65	4,186	9,087	107,534
66	4,715	7,901	121,14
67	4,196	8,821	
68	4,839	9,297	130,875
69	5,015	8,301	119,871
70	4,56	9,09	126,874
71	5,254	9,369	123,727
72	4,554	8,717	115,842
74	4,079	10,353	119,507
75	4,639	8,692	112,418
76	4,033	7,58	107,675
77	4,359	8,092	120,165
78	5,329	8,039	119,594
79	4,784	9,168	124,011
80	5,547	8,344	121,248
81	4,883	8,652	117,847
82	4,236	8,621	117,688
83	4,931	8,59	123,6
84	4,93	8,59	123,7
85	5,317	8,77	122,607
86	4,714	9,42	130,434
87	4,365	8,556	
88	4,237	8,53	102,9
89	5,122	7,868	119,978
90	4,178	9,4	131,944
91	5,15	9,441	123,604
92	4,658	9,379	127,153

93		9,09	118,4
94	4,591	9,23	122,1
95	4,61	8,347	118,239
96	4,297	8,296	114,3
97	5,81	6,991	113,347
98	5,134	8,72	112,603
99	5,132	9,56	128,139
100	5,57	7,991	125,683
101	5,21	9,226	126,01
102	5,09	8,883	125,276
103	4,587	8,81	124,855
104	4,517	8,924	129,103
105	4	9,434	125,971
106	4,761	8,802	130,299
107	5,456	8,092	120,165
108	5,092	8,932	123,975
109	4,931	8,577	122,155
110	4,866	8,71	133,5
111	4,806	8,754	120,913
112	5,168	9,418	129,545
113	5,833	9,5	123,1
114	4,753	9,11	126,7
115	3,597	8,483	116,729
116	4,1	8,55	105,5
117	4,273	8,135	123,918
118	5,388	8,76	123,492
119	4,884	9,592	132,093
120	5,095	7,759	125,138
121	4,662	8,99	127,304
122	4,87	9,126	125,423
123	5,685	9,368	118,218
124	5,757	8,724	120,475
125	3,844	8,147	124,811
	3,977	8,179	128,54

Table 3 NMR NOE restraints

1 ALA QB	1 ALA HA	2.932
2 PHE HA	2 PHE HB3	4.127
2 PHE HA	2 PHE HB2	3.818
2 PHE HB2	2 PHE HB3	3.079
2 PHE HA	2 PHE HD1	3.901
2 PHE HB3	2 PHE HD1	4.281
2 PHE HB2	2 PHE HD1	4.294
2 PHE HB3	2 PHE HE1	4.723
3 SER HA	3 SER HB2	3.276
4 GLY HA2	4 GLY HA1	2.558
5 THR HB	5 THR HA	3.535
5 THR QG2	5 THR HA	3.391
5 THR QG2	5 THR HB	2.898
6 TRP HA	6 TRP HB3	3.708
6 TRP HA	6 TRP HB2	3.708
6 TRP HB3	6 TRP HD1	4.733
6 TRP HB2	6 TRP HD1	5.016
6 TRP HA	6 TRP HE1	5.500
6 TRP HB3	6 TRP HE1	5.874
6 TRP HB2	6 TRP HE1	5.874
6 TRP HA	6 TRP HE3	4.369
7 GLN HA	7 GLN HB3	3.373
7 GLN HA	7 GLN HB2	3.410
7 GLN HG3	7 GLN HA	4.006
7 GLN HA	7 GLN HG2	3.347
8 VAL HA	8 VAL HB	4.427
8 VAL QG1	8 VAL HA	3.824
8 VAL QG1	8 VAL HB	3.512
8 VAL QG2	8 VAL HA	4.214
8 VAL QG2	8 VAL HB	3.662
9 TYR HA	9 TYR HB3	2.930
9 TYR HA	9 TYR HB2	2.855
9 TYR HB3	9 TYR HB2	2.929
9 TYR HA	9 TYR HD1	4.300
9 TYR HB3	9 TYR HD1	3.727
9 TYR HB2	9 TYR HD1	3.613
10 ALA QB	9 TYR HD1	4.048
10 ALA QB	9 TYR HE1	4.219
10 ALA QB	10 ALA HA	2.865
8 VAL QG2	11 GLN HG3	3.642
11 GLN HG2	8 VAL QG2	4.427
11 GLN HA	11 GLN HB3	3.153
11 GLN HA	11 GLN HB2	2.899
11 GLN HB3	11 GLN HB2	3.002
11 GLN HA	11 GLN HG3	3.082
11 GLN HA	11 GLN HG2	3.146
11 GLN HB3	11 GLN HG2	3.995
11 GLN HG3	11 GLN HG2	2.915
12 GLU- HA	12 GLU- HB3	3.259
12 GLU- HA	12 GLU- HB2	3.056
12 GLU- HA	12 GLU- HG3	3.351

12 GLU- HA	12 GLU- HG2	3.423
13 ASN HA	13 ASN HB3	3.252
13 ASN HA	13 ASN HB2	3.267
13 ASN HB2	13 ASN HB3	3.035
11 GLN HA	14 TYR HB3	4.665
11 GLN HB3	14 TYR HB3	4.624
11 GLN HB2	14 TYR HB3	4.339
14 TYR HB2	11 GLN HA	5.461
11 GLN HB3	14 TYR HB2	4.701
11 GLN HB2	14 TYR HB2	4.696
11 GLN HB3	14 TYR HD1	5.446
11 GLN HB2	14 TYR HD1	5.467
14 TYR HA	14 TYR HB3	3.802
14 TYR HA	14 TYR HB2	3.487
14 TYR HA	14 TYR HD1	3.265
14 TYR HB3	14 TYR HD1	4.152
14 TYR HB2	14 TYR HD1	4.008
14 TYR HA	14 TYR HE1	4.858
14 TYR HB3	14 TYR HE1	5.858
14 TYR HB2	14 TYR HE1	5.858
15 GLU- HA	15 GLU- HB2	2.938
15 GLU- HG3	15 GLU- HA	3.236
13 ASN HA	16 GLU- HB2	3.705
16 GLU- HB2	13 ASN HB3	3.133
16 GLU- HB2	13 ASN HB2	4.271
16 GLU- HA	16 GLU- HB2	2.794
16 GLU- HA	16 GLU- HG3	3.574
16 GLU- HG3	16 GLU- HB2	3.008
16 GLU- HG2	16 GLU- HA	3.804
16 GLU- HG2	16 GLU- HB2	3.008
17 PHE HB3	14 TYR HA	4.756
14 TYR HA	17 PHE HB2	3.931
17 PHE HA	17 PHE HB3	3.305
17 PHE HA	17 PHE HB2	3.664
17 PHE HB2	17 PHE HB3	3.209
17 PHE HA	17 PHE HD1	3.661
17 PHE HB3	17 PHE HD1	3.876
17 PHE HB2	17 PHE HD1	3.834
18 LEU QD1	14 TYR HE1	4.201
18 LEU HG	14 TYR HD1	4.292
18 LEU HG	14 TYR HE1	4.164
15 GLU- HA	18 LEU HG	3.598
18 LEU HA	18 LEU HB3	3.707
18 LEU HA	18 LEU HB2	3.471
18 LEU HB2	18 LEU HB3	3.202
18 LEU QD1	18 LEU HA	3.275
18 LEU QD1	18 LEU HB3	3.927
18 LEU QD1	18 LEU HB2	3.599
18 LEU HA	18 LEU HG	4.396
18 LEU HG	18 LEU HB3	3.181
18 LEU HG	18 LEU HB2	3.492
18 LEU HG	18 LEU QD1	3.242
19 LYS+ HA	19 LYS+ HB2	2.764
19 LYS+ HD2	19 LYS+ HA	3.432
19 LYS+ HB2	19 LYS+ HE3	3.743
19 LYS+ HA	19 LYS+ HG3	3.295

19 LYS+ HG3	19 LYS+ HB2	2.884
19 LYS+ HG3	19 LYS+ HE3	3.623
19 LYS+ HA	19 LYS+ HG2	3.443
19 LYS+ HB2	19 LYS+ HG2	2.635
19 LYS+ HG3	19 LYS+ HG2	2.260
20 ALA QB	20 ALA HA	2.880
18 LEU HA	21 LEU HG	3.796
21 LEU HA	21 LEU HB3	3.235
21 LEU HA	21 LEU HB2	3.499
21 LEU HB2	21 LEU HB3	2.806
21 LEU QD1	21 LEU HA	3.625
21 LEU QD1	21 LEU HB2	4.338
21 LEU QD1	21 LEU QD2	4.272
21 LEU HG	21 LEU HA	3.438
21 LEU HG	21 LEU HB3	2.867
21 LEU HG	21 LEU HB2	3.265
21 LEU QD1	21 LEU HG	3.599
22 ALA QB	22 ALA HA	2.673
21 LEU HG	23 LEU QD1	4.161
23 LEU HA	23 LEU HB2	3.109
23 LEU HA	23 LEU QD1	3.579
23 LEU QD1	23 LEU HB2	3.281
23 LEU QD2	23 LEU HA	3.335
23 LEU HB2	23 LEU QD2	3.443
23 LEU HG	23 LEU HA	3.323
23 LEU HB2	23 LEU HG	2.962
23 LEU QD1	23 LEU HG	3.228
23 LEU QD2	23 LEU HG	3.215
23 LEU HA	24 PRO HD3	2.830
24 PRO HD3	23 LEU HB2	3.614
24 PRO HD3	23 LEU QD2	4.115
23 LEU HA	24 PRO HD2	3.056
24 PRO HD2	23 LEU HB2	3.756
24 PRO HD2	23 LEU QD2	4.318
24 PRO HA	24 PRO HB3	2.899
24 PRO HA	24 PRO HB2	2.801
24 PRO HB2	24 PRO HB3	2.327
24 PRO HD3	24 PRO HA	4.800
24 PRO HB3	24 PRO HD3	3.899
24 PRO HB2	24 PRO HD3	4.342
24 PRO HA	24 PRO HD2	4.801
24 PRO HB3	24 PRO HD2	4.139
24 PRO HB2	24 PRO HD2	4.372
24 PRO HD3	24 PRO HD2	2.791
24 PRO HA	24 PRO HG2	4.800
24 PRO HG2	24 PRO HB2	2.979
24 PRO HD3	24 PRO HG2	2.876
24 PRO HD2	24 PRO HG2	3.170
25 GLU- HB3	25 GLU- HA	3.317
25 GLU- HA	25 GLU- HB2	2.884
25 GLU- HA	25 GLU- HG3	3.815
25 GLU- HA	25 GLU- HG2	3.775
26 ASP- HA	26 ASP- HB3	2.723
26 ASP- HA	26 ASP- HB2	2.727
27 LEU QD1	24 PRO HG2	3.686
24 PRO HD3	27 LEU HG	4.152

24 PRO HD2	27 LEU HG	3.921
27 LEU HA	27 LEU HB3	3.018
27 LEU HA	27 LEU HB2	3.124
27 LEU HA	27 LEU QD1	4.042
27 LEU HB2	27 LEU QD1	4.064
27 LEU QD2	27 LEU HA	3.416
27 LEU HB3	27 LEU QD2	3.853
27 LEU HB2	27 LEU QD2	4.062
27 LEU HG	27 LEU HA	3.467
27 LEU QD1	27 LEU HG	3.086
27 LEU QD2	27 LEU HG	3.224
15 GLU- HG3	28 ILE QG2	3.684
18 LEU HG	28 ILE HA	3.739
18 LEU HG	28 ILE QG2	3.748
19 LYS+ HA	28 ILE QD1	3.690
19 LYS+ HE3	28 ILE QD1	5.629
25 GLU- HA	28 ILE HB	2.708
25 GLU- HA	28 ILE QD1	3.587
28 ILE QD1	25 GLU- HG3	5.436
25 GLU- HG2	28 ILE QD1	6.000
28 ILE HG13	25 GLU- HA	5.219
25 GLU- HA	28 ILE HG12	3.557
28 ILE HA	28 ILE HB	3.896
28 ILE HA	28 ILE QD1	3.646
28 ILE QD1	28 ILE HB	3.239
28 ILE HA	28 ILE HG13	3.158
28 ILE HB	28 ILE HG13	3.082
28 ILE HA	28 ILE HG12	3.657
28 ILE HB	28 ILE HG12	3.385
28 ILE QD1	28 ILE HG12	2.902
28 ILE HG13	28 ILE HG12	3.155
28 ILE QG2	28 ILE HA	3.523
28 ILE QG2	28 ILE HB	2.920
28 ILE QG2	28 ILE HG12	3.636
29 LYS+ HB2	26 ASP- HA	3.205
29 LYS+ HA	29 LYS+ HB2	2.620
29 LYS+ HD2	29 LYS+ HA	3.463
29 LYS+ HA	29 LYS+ HE2	3.918
29 LYS+ HG3	29 LYS+ HA	3.578
29 LYS+ HB2	29 LYS+ HG3	2.982
29 LYS+ HA	29 LYS+ HG2	3.471
29 LYS+ HG2	29 LYS+ HB2	3.205
27 LEU HA	30 MET HB2	3.126
30 MET HB2	27 LEU QD2	3.893
30 MET HG3	27 LEU QD1	5.968
27 LEU HA	30 MET HG2	4.171
30 MET HG2	27 LEU QD1	5.613
27 LEU QD2	30 MET HG2	5.263
30 MET HA	30 MET HB2	3.000
30 MET HG3	30 MET HA	3.516
30 MET HG3	30 MET HB2	3.121
30 MET HA	30 MET HG2	3.415
30 MET HG2	30 MET HB2	3.047
18 LEU HG	31 ALA QB	3.349
31 ALA QB	28 ILE HA	3.840
28 ILE QD1	31 ALA QB	4.021

31 ALA QB	28 ILE HG13	4.086
31 ALA QB	28 ILE QG2	4.679
31 ALA HA	31 ALA QB	2.990
32 ARG+ HA	14 TYR HD1	3.795
32 ARG+ HA	14 TYR HE1	3.529
15 GLU- HG3	32 ARG+ HD3	4.293
32 ARG+ HD2	15 GLU- HG3	4.901
32 ARG+ HA	18 LEU HG	3.823
28 ILE QG2	32 ARG+ HD3	4.778
32 ARG+ HD2	28 ILE QG2	5.110
32 ARG+ HA	32 ARG+ HB3	3.104
32 ARG+ HA	32 ARG+ HB2	2.956
32 ARG+ HA	32 ARG+ HD3	4.334
32 ARG+ HB3	32 ARG+ HD3	3.841
32 ARG+ HD2	32 ARG+ HA	4.844
32 ARG+ HG3	32 ARG+ HA	4.342
32 ARG+ HD3	32 ARG+ HG3	2.961
32 ARG+ HG2	32 ARG+ HA	4.053
32 ARG+ HG2	32 ARG+ HD3	3.679
32 ARG+ HD2	32 ARG+ HG2	2.930
33 ASP- HA	33 ASP- HB3	3.123
33 ASP- HA	33 ASP- HB2	2.930
33 ASP- HB3	33 ASP- HB2	2.272
34 ILE HG13	31 ALA HA	4.608
31 ALA HA	34 ILE HG12	4.004
34 ILE HA	34 ILE HB	3.616
34 ILE HA	34 ILE QD1	3.503
34 ILE HB	34 ILE QD1	3.291
34 ILE HA	34 ILE HG13	3.717
34 ILE HB	34 ILE HG13	3.649
34 ILE QD1	34 ILE HG13	3.469
34 ILE HG12	34 ILE HA	3.670
34 ILE HG12	34 ILE HB	3.272
34 ILE QD1	34 ILE HG12	2.992
34 ILE HG12	34 ILE HG13	2.531
34 ILE QG2	34 ILE HA	3.566
34 ILE QG2	34 ILE HB	3.017
35 LYS+ HA	8 VAL QG2	3.928
35 LYS+ HE2	8 VAL QG2	5.318
35 LYS+ HA	35 LYS+ HB3	3.207
35 LYS+ HA	35 LYS+ HB2	3.188
35 LYS+ HE2	35 LYS+ HA	3.717
35 LYS+ HA	35 LYS+ HG3	3.257
35 LYS+ HA	35 LYS+ HG2	3.571
35 LYS+ HE2	35 LYS+ HG2	3.414
36 PRO HA	8 VAL QG1	4.442
8 VAL HB	36 PRO HD3	4.400
8 VAL QG1	36 PRO HD3	4.021
8 VAL QG2	36 PRO HD3	4.542
8 VAL HB	36 PRO HD2	4.910
36 PRO HD2	8 VAL QG1	4.610
8 VAL QG2	36 PRO HD2	4.568
8 VAL QG1	36 PRO HG2	3.603
35 LYS+ HA	36 PRO HD3	3.137
35 LYS+ HA	36 PRO HD2	2.959
36 PRO HA	36 PRO HB3	3.099

36 PRO HA	36 PRO HB2	2.977
36 PRO HB2	36 PRO HB3	2.623
36 PRO HD3	36 PRO HA	4.800
36 PRO HB3	36 PRO HD3	4.564
36 PRO HD3	36 PRO HB2	3.928
36 PRO HD2	36 PRO HA	4.800
36 PRO HB3	36 PRO HD2	4.475
36 PRO HD2	36 PRO HB2	4.294
36 PRO HD3	36 PRO HD2	3.014
36 PRO HA	36 PRO HG2	4.800
36 PRO HG2	36 PRO HB2	3.215
36 PRO HD3	36 PRO HG2	3.351
36 PRO HD2	36 PRO HG2	3.327
37 ILE HA	37 ILE HB	3.346
37 ILE QD1	37 ILE HA	3.981
37 ILE QD1	37 ILE HB	3.023
37 ILE HG12	37 ILE HA	3.464
37 ILE HG12	37 ILE HB	2.505
37 ILE HG12	37 ILE QD1	3.064
37 ILE QG2	37 ILE HA	3.458
37 ILE QG2	37 ILE HB	2.918
38 VAL HB	6 TRP HB3	3.548
38 VAL HB	6 TRP HB2	3.453
38 VAL HB	38 VAL HA	3.455
38 VAL HA	38 VAL QG2	3.153
38 VAL HB	38 VAL QG2	3.050
39 GLU- HA	5 THR HA	2.820
5 THR HB	39 GLU- HA	4.823
39 GLU- HA	6 TRP HD1	4.704
37 ILE QD1	39 GLU- HB3	4.432
37 ILE QD1	39 GLU- HB2	4.421
39 GLU- HG3	37 ILE QD1	4.647
39 GLU- HG3	37 ILE HG12	4.126
39 GLU- HG2	37 ILE QD1	4.812
39 GLU- HG2	37 ILE HG12	4.108
39 GLU- HA	39 GLU- HB3	3.076
39 GLU- HA	39 GLU- HB2	3.022
39 GLU- HG2	39 GLU- HA	3.696
40 ILE QD1	2 PHE HA	4.495
40 ILE QD1	2 PHE HB3	4.740
40 ILE QD1	2 PHE HB2	4.149
40 ILE QG2	2 PHE HB3	4.281
40 ILE QG2	2 PHE HB2	3.984
40 ILE QG2	2 PHE HD1	5.348
3 SER HA	40 ILE HB	3.846
40 ILE QG2	3 SER HA	3.714
40 ILE HA	40 ILE HB	4.394
40 ILE QD1	40 ILE HA	4.352
40 ILE QD1	40 ILE HB	3.987
40 ILE HA	40 ILE HG13	3.766
40 ILE HG13	40 ILE HB	4.873
40 ILE QD1	40 ILE HG13	3.434
40 ILE HA	40 ILE HG12	3.709
40 ILE HG12	40 ILE HB	3.868
40 ILE QD1	40 ILE HG12	3.384
40 ILE HG12	40 ILE HG13	3.330

40 ILE HA	40 ILE QG2	3.421
40 ILE QG2	40 ILE HB	3.528
40 ILE QG2	40 ILE HG13	4.212
40 ILE QG2	40 ILE HG12	3.959
41 GLN HA	3 SER HA	4.315
41 GLN HA	3 SER HB2	3.875
41 GLN HA	41 GLN HB3	2.954
41 GLN HA	41 GLN HB2	3.216
41 GLN HG2	41 GLN HA	3.409
42 GLN HA	42 GLN HB3	3.252
42 GLN HA	42 GLN HB2	3.325
42 GLN HB2	42 GLN HB3	3.045
42 GLN HG3	42 GLN HA	3.635
42 GLN HG2	42 GLN HA	4.115
42 GLN HG3	42 GLN HG2	2.893
43 LYS+ HA	43 LYS+ HB3	2.945
43 LYS+ HA	43 LYS+ HB2	2.961
43 LYS+ HD3	43 LYS+ HA	3.634
43 LYS+ HD2	43 LYS+ HA	4.512
43 LYS+ HB3	43 LYS+ HE2	4.006
43 LYS+ HB2	43 LYS+ HE2	4.060
43 LYS+ HG3	43 LYS+ HA	3.492
43 LYS+ HB3	43 LYS+ HG3	3.101
43 LYS+ HB2	43 LYS+ HG3	2.892
43 LYS+ HG3	43 LYS+ HE2	3.290
43 LYS+ HG2	43 LYS+ HA	3.821
43 LYS+ HG2	43 LYS+ HE2	4.078
44 GLY HA1	44 GLY HA2	2.632
45 ASP- HA	45 ASP- HB3	3.212
45 ASP- HA	45 ASP- HB2	3.106
45 ASP- HB2	45 ASP- HB3	2.667
43 LYS+ HB3	46 ASP- HA	4.532
46 ASP- HA	43 LYS+ HB2	3.931
46 ASP- HA	46 ASP- HB3	3.126
46 ASP- HA	46 ASP- HB2	3.282
47 PHE HA	42 GLN HA	3.172
47 PHE HA	47 PHE HB3	3.803
47 PHE HA	47 PHE HB2	3.282
47 PHE HA	47 PHE HD1	4.037
47 PHE HB3	47 PHE HD1	4.429
47 PHE HB2	47 PHE HD1	3.972
41 GLN HB3	48 VAL QG2	3.965
41 GLN HB2	48 VAL QG2	4.160
48 VAL HA	48 VAL HB	3.958
48 VAL QG2	48 VAL HA	2.972
48 VAL HB	48 VAL QG2	2.950
49 VAL HA	49 VAL HB	3.826
49 VAL QG2	49 VAL HA	3.475
49 VAL QG2	49 VAL HB	3.281
39 GLU- HB3	50 THR HB	3.403
39 GLU- HB2	50 THR HB	3.550
39 GLU- HG3	50 THR HB	4.338
50 THR QG2	39 GLU- HB3	3.692
50 THR QG2	39 GLU- HB2	3.939
50 THR HB	50 THR HA	3.905
50 THR QG2	50 THR HA	3.289

50 THR QG2	50 THR HB	2.889
51 SER HA	38 VAL HA	3.132
51 SER HA	38 VAL QG2	3.363
51 SER HB2	38 VAL HA	4.140
38 VAL QG2	51 SER HB2	3.836
51 SER HA	51 SER HB2	2.654
52 LYS+ HA	52 LYS+ HB3	3.242
52 LYS+ HA	52 LYS+ HB2	3.095
52 LYS+ HD2	52 LYS+ HA	3.085
52 LYS+ HD2	52 LYS+ HE2	2.677
52 LYS+ HA	52 LYS+ HG2	3.489
52 LYS+ HB3	52 LYS+ HG2	3.396
52 LYS+ HB2	52 LYS+ HG2	3.243
34 ILE QD1	53 THR HA	4.317
34 ILE QG2	53 THR HB	3.593
53 THR QG2	34 ILE QG2	3.936
36 PRO HA	53 THR HA	4.521
36 PRO HA	53 THR QG2	3.719
36 PRO HB3	53 THR QG2	3.573
53 THR QG2	36 PRO HB2	3.520
53 THR QG2	36 PRO HG2	3.546
53 THR HA	53 THR HB	2.996
53 THR HA	53 THR QG2	3.645
53 THR QG2	53 THR HB	3.136
54 PRO HD3	34 ILE QD1	4.604
54 PRO HD2	34 ILE QD1	4.388
53 THR HA	54 PRO HD3	3.238
54 PRO HD3	53 THR HB	3.353
53 THR QG2	54 PRO HD3	4.283
53 THR HA	54 PRO HD2	3.018
54 PRO HD2	53 THR HB	3.522
53 THR QG2	54 PRO HD2	4.385
54 PRO HB3	54 PRO HA	3.080
54 PRO HA	54 PRO HB2	2.934
54 PRO HB2	54 PRO HB3	2.810
54 PRO HB2	54 PRO HD3	4.772
54 PRO HA	54 PRO HD2	4.625
54 PRO HD2	54 PRO HB2	4.137
54 PRO HA	54 PRO HG3	3.709
54 PRO HG3	54 PRO HB2	3.008
54 PRO HD3	54 PRO HG3	3.885
54 PRO HA	54 PRO HG2	4.058
54 PRO HG2	54 PRO HB2	3.625
54 PRO HD3	54 PRO HG2	3.869
54 PRO HG2	54 PRO HD2	3.788
55 ARG+ HD2	34 ILE QD1	5.008
55 ARG+ HA	55 ARG+ HB3	2.767
55 ARG+ HA	55 ARG+ HB2	2.828
55 ARG+ HD2	55 ARG+ HA	3.607
55 ARG+ HD2	55 ARG+ HB3	3.264
55 ARG+ HD2	55 ARG+ HB2	2.769
55 ARG+ HA	55 ARG+ HG2	2.877
55 ARG+ HD2	55 ARG+ HG2	2.588
56 GLN HA	56 GLN HB2	2.944
56 GLN HA	56 GLN HG3	3.610
56 GLN HA	56 GLN HG2	3.655

57 THR HA	52 LYS+ HB2	4.467
52 LYS+ HA	57 THR QG2	4.162
57 THR HB	57 THR HA	3.433
57 THR QG2	57 THR HA	2.891
57 THR HB	57 THR QG2	3.091
51 SER HB2	58 VAL HB	3.457
51 SER HB2	58 VAL QG2	3.608
58 VAL HA	58 VAL HB	3.526
58 VAL QG2	58 VAL HA	3.050
58 VAL HB	58 VAL QG2	3.043
59 THR HA	50 THR HA	2.867
59 THR HB	59 THR HA	3.346
59 THR QG2	59 THR HA	2.942
59 THR QG2	59 THR HB	2.886
60 ASN HB3	49 VAL HB	4.899
49 VAL HB	60 ASN HB2	4.362
60 ASN HA	60 ASN HB3	3.213
60 ASN HA	60 ASN HB2	3.562
60 ASN HB3	60 ASN HB2	3.251
61 SER HA	48 VAL HA	2.799
48 VAL HA	61 SER HB2	3.920
48 VAL QG2	61 SER HB2	2.964
61 SER HA	61 SER HB2	2.638
47 PHE HB3	62 PHE HD1	5.117
47 PHE HB2	62 PHE HD1	4.820
49 VAL QG2	62 PHE HD1	3.715
60 ASN HB3	62 PHE HD1	4.865
60 ASN HB2	62 PHE HD1	5.097
61 SER HA	62 PHE HD1	4.726
62 PHE HA	62 PHE HB2	2.549
62 PHE HA	62 PHE HD1	4.100
62 PHE HB2	62 PHE HD1	3.726
63 THR QG2	45 ASP- HB3	3.812
63 THR QG2	45 ASP- HB2	3.654
46 ASP- HA	63 THR HA	2.805
63 THR HA	46 ASP- HB3	4.463
46 ASP- HB2	63 THR HA	5.046
46 ASP- HA	63 THR HB	5.067
46 ASP- HA	63 THR QG2	3.761
63 THR HA	47 PHE HD1	5.185
63 THR HA	47 PHE HE1	5.010
63 THR HA	63 THR HB	3.813
63 THR QG2	63 THR HA	3.496
63 THR QG2	63 THR HB	3.052
45 ASP- HB3	64 LEU HB2	4.275
45 ASP- HB2	64 LEU HB2	4.041
45 ASP- HA	64 LEU HG	3.300
64 LEU HG	45 ASP- HB3	4.976
64 LEU HG	45 ASP- HB2	4.284
64 LEU QD1	62 PHE HD1	5.000
64 LEU HA	64 LEU HB3	4.228
64 LEU HB2	64 LEU HA	4.276
64 LEU HB2	64 LEU HB3	3.121
64 LEU QD1	64 LEU HA	3.746
64 LEU QD1	64 LEU HB3	3.777
64 LEU QD1	64 LEU HB2	3.937

64 LEU HG	64 LEU HA	4.708
64 LEU HG	64 LEU HB3	3.487
64 LEU HG	64 LEU HB2	3.223
64 LEU QD1	64 LEU HG	3.337
65 GLY HA2	65 GLY HA1	2.779
63 THR QG2	66 LYS+ HA	5.332
66 LYS+ HB3	63 THR HB	3.626
63 THR HB	66 LYS+ HB2	3.675
66 LYS+ HD2	63 THR HB	3.884
66 LYS+ HA	66 LYS+ HB3	2.941
66 LYS+ HA	66 LYS+ HB2	3.073
66 LYS+ HD2	66 LYS+ HA	3.477
66 LYS+ HE2	66 LYS+ HB3	4.676
66 LYS+ HE2	66 LYS+ HD2	2.849
66 LYS+ HA	66 LYS+ HG3	3.060
66 LYS+ HB3	66 LYS+ HG3	3.379
66 LYS+ HE2	66 LYS+ HG3	2.912
66 LYS+ HG2	66 LYS+ HA	3.372
62 PHE HA	68 ALA QB	3.963
68 ALA QB	62 PHE HB2	3.568
68 ALA QB	62 PHE HD1	5.069
68 ALA HA	68 ALA QB	3.141
69 ASP- HA	69 ASP- HB3	3.349
69 ASP- HA	69 ASP- HB2	3.237
70 ILE QD1	60 ASN HA	5.560
70 ILE QD1	60 ASN HB3	4.071
70 ILE QD1	60 ASN HB2	4.559
70 ILE QG2	60 ASN HB3	5.181
70 ILE QG2	60 ASN HB2	5.650
70 ILE QD1	62 PHE HA	5.704
70 ILE QD1	62 PHE HB2	4.343
70 ILE QD1	62 PHE HD1	4.236
70 ILE QG2	62 PHE HD1	6.000
70 ILE QD1	68 ALA QB	4.543
70 ILE HA	70 ILE HB	3.829
70 ILE QD1	70 ILE HA	3.491
70 ILE QD1	70 ILE HB	4.443
70 ILE HG13	70 ILE HA	4.677
70 ILE HG13	70 ILE HB	3.724
70 ILE QD1	70 ILE HG13	3.670
70 ILE HA	70 ILE HG12	4.545
70 ILE HG12	70 ILE HB	3.720
70 ILE QD1	70 ILE HG12	3.744
70 ILE HA	70 ILE QG2	3.704
70 ILE QG2	70 ILE HB	3.729
70 ILE QG2	70 ILE HG13	4.040
70 ILE QG2	70 ILE HG12	4.258
71 THR HB	71 THR HA	3.950
71 THR QG2	71 THR HA	3.457
71 THR QG2	71 THR HB	3.059
72 THR HA	72 THR QG2	4.061
72 THR HB	72 THR QG2	4.936
74 ASP- HB3	21 LEU HB3	5.020
21 LEU HB2	74 ASP- HB3	5.023
21 LEU HG	74 ASP- HB3	4.752
74 ASP- HB2	21 LEU HB3	4.822

21 LEU HG	74 ASP- HB2	4.673
23 LEU QD2	74 ASP- HB3	4.134
23 LEU QD1	74 ASP- HB2	5.300
23 LEU QD2	74 ASP- HB2	4.782
74 ASP- HA	74 ASP- HB3	3.430
74 ASP- HA	74 ASP- HB2	3.466
74 ASP- HB3	74 ASP- HB2	2.935
75 GLY HA2	75 GLY HA1	2.476
76 LYS+ HA	76 LYS+ HB2	2.849
76 LYS+ HD2	76 LYS+ HA	3.874
76 LYS+ HA	76 LYS+ HE2	5.208
76 LYS+ HD2	76 LYS+ HE2	2.926
76 LYS+ HG2	76 LYS+ HA	3.619
76 LYS+ HD2	76 LYS+ HG2	3.118
77 LYS+ HA	77 LYS+ HB2	2.853
77 LYS+ HA	77 LYS+ HG2	3.406
77 LYS+ HB2	77 LYS+ HG2	2.949
77 LYS+ HG2	77 LYS+ HD2	4.028
78 LEU HB3	78 LEU HA	4.161
78 LEU HA	78 LEU HB2	2.862
78 LEU HA	78 LEU QD1	4.659
78 LEU HB3	78 LEU QD1	4.881
78 LEU HB2	78 LEU QD1	4.639
78 LEU HA	78 LEU QD2	3.577
78 LEU HG	78 LEU HA	4.213
78 LEU QD1	78 LEU HG	3.616
78 LEU QD2	78 LEU HG	3.297
79 LYS+ HA	79 LYS+ HB3	3.284
79 LYS+ HA	79 LYS+ HB2	3.970
79 LYS+ HG2	79 LYS+ HA	3.231
79 LYS+ HB3	79 LYS+ HG2	4.197
79 LYS+ HB2	79 LYS+ HG2	4.040
80 CYS HA	80 CYS HB3	3.902
80 CYS HA	80 CYS HB2	3.629
80 CYS HB2	80 CYS HB3	3.069
81 THR HA	81 THR HB	4.245
81 THR HA	81 THR QG2	3.987
81 THR HB	81 THR QG2	3.254
82 VAL HB	62 PHE HD1	4.666
82 VAL QG1	62 PHE HB2	5.544
82 VAL QG1	62 PHE HD1	5.373
82 VAL QG2	62 PHE HB2	5.422
82 VAL QG2	62 PHE HD1	3.957
82 VAL QG2	64 LEU HA	5.032
68 ALA QB	82 VAL QG1	5.707
82 VAL HB	82 VAL HA	4.793
82 VAL QG1	82 VAL HA	4.679
82 VAL QG1	82 VAL HB	4.328
82 VAL QG2	82 VAL HA	3.698
82 VAL QG2	82 VAL HB	3.519
82 VAL QG1	82 VAL QG2	4.826
83 HIS HA	64 LEU HB3	4.247
83 HIS HA	64 LEU HB2	4.383
64 LEU QD1	83 HIS HA	3.798
83 HIS HA	83 HIS HB3	3.020
83 HIS HA	83 HIS HB2	3.255

83 HIS HB3	83 HIS HB2	2.674
83 HIS HB3	83 HIS HD2	5.247
83 HIS HB2	83 HIS HD2	5.205
84 LEU QD1	47 PHE HE1	5.021
84 LEU QD1	47 PHE HZ	4.624
84 LEU QD2	47 PHE HZ	4.426
84 LEU HB3	64 LEU QD1	5.843
84 LEU HB2	64 LEU QD1	4.967
84 LEU HA	84 LEU HB3	3.456
84 LEU HA	84 LEU HB2	3.204
84 LEU HB2	84 LEU HB3	2.638
84 LEU HA	84 LEU QD1	3.885
84 LEU QD1	84 LEU HB3	3.492
84 LEU HB2	84 LEU QD1	4.279
84 LEU HA	84 LEU QD2	3.528
84 LEU HB3	84 LEU QD2	4.195
84 LEU HB2	84 LEU QD2	4.077
84 LEU HG	84 LEU HA	4.195
84 LEU HG	84 LEU HB3	3.360
84 LEU QD1	84 LEU HG	3.463
84 LEU QD2	84 LEU HG	3.589
86 ASN HB3	86 ASN HA	3.130
86 ASN HA	86 ASN HB2	2.870
86 ASN HB3	86 ASN HB2	2.200
87 GLY HA1	84 LEU QD1	4.650
87 GLY HA2	84 LEU QD1	4.828
87 GLY HA2	87 GLY HA1	2.355
88 LYS+ HA	88 LYS+ HB3	3.068
88 LYS+ HA	88 LYS+ HB2	3.155
88 LYS+ HG3	88 LYS+ HA	3.544
88 LYS+ HG2	88 LYS+ HA	3.978
89 LEU QD1	2 PHE HB2	6.000
89 LEU QD1	2 PHE HD1	5.472
89 LEU QD1	2 PHE HE1	4.397
89 LEU QD1	62 PHE HD1	4.383
84 LEU HA	89 LEU HA	3.310
89 LEU QD1	84 LEU HA	5.172
84 LEU QD1	89 LEU QD1	6.000
84 LEU QD2	89 LEU QD1	5.668
89 LEU HB2	89 LEU HA	5.196
89 LEU QD1	89 LEU HA	4.302
89 LEU HG	89 LEU HA	4.440
91 THR QG2	80 CYS HB3	4.086
91 THR QG2	80 CYS HB2	3.760
91 THR HB	82 VAL QG1	5.940
91 THR QG2	82 VAL QG1	4.869
91 THR HB	91 THR HA	3.756
91 THR QG2	91 THR HA	3.130
91 THR HB	91 THR QG2	3.727
94 GLU- HA	94 GLU- HB3	3.100
94 GLU- HA	94 GLU- HB2	2.848
94 GLU- HA	94 GLU- HG3	3.600
94 GLU- HG2	94 GLU- HA	3.544
96 PHE HB3	78 LEU QD1	4.782
78 LEU QD2	96 PHE HB3	5.379
78 LEU QD1	96 PHE HB2	5.511

78 LEU QD2	96 PHE HB2	6.000
78 LEU QD1	96 PHE HD1	4.978
78 LEU QD2	96 PHE HD1	4.829
96 PHE HA	96 PHE HB3	3.624
96 PHE HA	96 PHE HB2	3.202
96 PHE HB2	96 PHE HB3	2.935
96 PHE HA	96 PHE HD1	3.559
96 PHE HB3	96 PHE HD1	3.925
96 PHE HB2	96 PHE HD1	3.649
97 SER HB2	97 SER HA	3.466
102 VAL QG2	2 PHE HA	4.492
102 VAL QG2	2 PHE HD1	4.532
102 VAL QG2	2 PHE HE1	5.037
102 VAL HB	102 VAL HA	3.926
102 VAL HA	102 VAL QG1	4.406
102 VAL HB	102 VAL QG1	3.922
102 VAL HA	102 VAL QG2	4.215
102 VAL HB	102 VAL QG2	3.747
102 VAL QG2	102 VAL QG1	4.594
103 LYS+ HB3	103 LYS+ HA	3.401
103 LYS+ HA	103 LYS+ HB2	2.712
103 LYS+ HA	103 LYS+ HG2	3.283
103 LYS+ HB3	103 LYS+ HG2	3.350
103 LYS+ HB2	103 LYS+ HG2	3.079
104 GLY HA2	104 GLY HA1	2.462
105 ASN HA	6 TRP HH2	4.245
105 ASN HA	6 TRP HZ3	4.592
105 ASN HA	105 ASN HB3	3.040
105 ASN HA	105 ASN HB2	3.178
105 ASN HB2	105 ASN HB3	2.502
106 GLU- HA	6 TRP HE3	4.224
106 GLU- HA	6 TRP HZ3	3.878
106 GLU- HG3	6 TRP HZ3	5.119
106 GLU- HG2	6 TRP HZ3	5.320
106 GLU- HA	106 GLU- HB3	3.346
106 GLU- HB2	106 GLU- HA	3.704
106 GLU- HB2	106 GLU- HB3	2.440
106 GLU- HG3	106 GLU- HA	3.836
106 GLU- HG2	106 GLU- HA	4.028
106 GLU- HG2	106 GLU- HB2	3.331
106 GLU- HG3	106 GLU- HG2	2.364
107 MET HA	102 VAL HA	3.167
102 VAL HB	107 MET HA	4.892
107 MET HA	102 VAL QG2	4.320
102 VAL QG2	107 MET HG3	4.803
107 MET HA	107 MET HB3	4.005
107 MET HA	107 MET HB2	3.650
107 MET HB3	107 MET HB2	4.735
107 MET HA	107 MET HG3	3.418
107 MET HG3	107 MET HB3	4.269
107 MET HA	107 MET HG2	3.594
107 MET HG2	107 MET HB3	4.508
108 VAL HB	108 VAL HA	3.423
108 VAL QG2	108 VAL HA	3.326
108 VAL HB	108 VAL QG2	2.999
111 ILE HG12	111 ILE HA	4.319

111 ILE HG12	111 ILE QD1	3.999
111 ILE HG12	111 ILE QG2	3.802
112 THR HB	112 THR HA	3.785
112 THR QG2	112 THR HA	3.505
112 THR QG2	112 THR HB	3.089
20 ALA QB	113 PHE HD1	4.588
20 ALA QB	113 PHE HE1	4.778
96 PHE HA	113 PHE HA	3.762
113 PHE HA	113 PHE HB3	3.954
113 PHE HB2	113 PHE HA	4.113
113 PHE HB3	113 PHE HB2	2.993
113 PHE HA	113 PHE HD1	4.611
20 ALA QB	114 GLY HA2	4.577
115 GLY HA2	115 GLY HA1	2.470
116 VAL QG2	16 GLU- HB2	3.739
17 PHE HA	116 VAL HB	3.949
116 VAL QG2	17 PHE HA	3.457
116 VAL QG2	17 PHE HB3	4.577
116 VAL QG2	17 PHE HB2	4.120
20 ALA QB	116 VAL HB	4.157
20 ALA QB	116 VAL QG2	4.100
116 VAL HB	113 PHE HD1	4.548
116 VAL QG2	113 PHE HD1	4.301
116 VAL HA	116 VAL HB	3.567
116 VAL HA	116 VAL QG2	3.205
116 VAL HB	116 VAL QG2	3.142
112 THR HA	117 THR HA	3.088
117 THR QG2	112 THR HA	4.663
117 THR QG2	117 THR HA	3.479
117 THR QG2	117 THR HB	3.006
11 GLN HA	118 LEU QD1	4.615
118 LEU QD1	11 GLN HB3	3.563
118 LEU QD1	11 GLN HB2	4.094
118 LEU HA	13 ASN HB3	4.077
118 LEU HA	13 ASN HB2	3.662
118 LEU QD1	14 TYR HA	3.593
118 LEU QD1	14 TYR HB3	3.844
118 LEU QD1	14 TYR HB2	3.849
118 LEU QD1	14 TYR HD1	3.729
118 LEU QD1	14 TYR HE1	5.173
14 TYR HA	118 LEU HG	4.438
118 LEU HG	14 TYR HB3	4.004
118 LEU HG	14 TYR HB2	4.576
118 LEU HG	14 TYR HD1	3.660
118 LEU HG	14 TYR HE1	4.525
118 LEU QD1	17 PHE HA	5.405
17 PHE HB3	118 LEU QD1	5.144
118 LEU QD1	17 PHE HB2	4.794
17 PHE HB3	118 LEU HG	5.229
118 LEU HG	17 PHE HB2	4.992
118 LEU HA	118 LEU HB3	3.425
118 LEU HA	118 LEU HB2	3.278
118 LEU HB3	118 LEU HB2	3.583
118 LEU HA	118 LEU QD1	3.570
118 LEU QD1	118 LEU HB3	4.073
118 LEU HG	118 LEU HA	4.589

118 LEU HB3	118 LEU HG	4.001
118 LEU HB2	118 LEU HG	4.123
118 LEU QD1	118 LEU HG	3.244
119 ILE HB	12 GLU- HB3	3.264
119 ILE HB	12 GLU- HB2	3.395
12 GLU- HG3	119 ILE HB	4.880
119 ILE HB	12 GLU- HG2	4.761
119 ILE QG2	12 GLU- HG3	5.210
12 GLU- HG2	119 ILE QG2	5.208
119 ILE HA	110 THR HA	3.702
119 ILE HB	119 ILE HA	4.145
119 ILE QD1	119 ILE HA	3.856
119 ILE QD1	119 ILE HB	3.533
119 ILE HA	119 ILE HG13	3.658
119 ILE HG13	119 ILE HB	3.691
119 ILE QD1	119 ILE HG13	3.097
119 ILE HG12	119 ILE HA	3.679
119 ILE QD1	119 ILE HG12	3.337
119 ILE HG12	119 ILE HG13	2.528
119 ILE QG2	119 ILE HA	3.513
119 ILE QG2	119 ILE HB	3.209
119 ILE HG13	119 ILE QG2	3.600
119 ILE QG2	119 ILE HG12	3.540
8 VAL QG2	120 ARG+ HD2	4.270
120 ARG+ HA	120 ARG+ HB3	3.040
120 ARG+ HA	120 ARG+ HB2	3.367
120 ARG+ HB2	120 ARG+ HD2	3.789
120 ARG+ HG2	120 ARG+ HA	3.819
120 ARG+ HG2	120 ARG+ HB2	3.072
120 ARG+ HG2	120 ARG+ HD2	3.384
121 ARG+ HD3	9 TYR HE1	5.140
121 ARG+ HD2	9 TYR HE1	5.033
10 ALA HA	121 ARG+ HD3	4.185
10 ALA QB	121 ARG+ HD3	3.813
121 ARG+ HD2	10 ALA HA	5.653
10 ALA QB	121 ARG+ HD2	4.269
105 ASN HA	124 ARG+ HG3	4.254
121 ARG+ HA	108 VAL HA	3.273
121 ARG+ HG3	108 VAL HA	5.162
121 ARG+ HG2	108 VAL HA	4.852
121 ARG+ HD3	119 ILE QG2	5.123
119 ILE QG2	121 ARG+ HD2	5.453
121 ARG+ HA	121 ARG+ HB3	3.349
121 ARG+ HA	121 ARG+ HB2	3.347
121 ARG+ HB3	121 ARG+ HB2	3.050
121 ARG+ HA	121 ARG+ HD3	4.967
121 ARG+ HD3	121 ARG+ HB3	3.801
121 ARG+ HD3	121 ARG+ HB2	3.498
121 ARG+ HA	121 ARG+ HD2	4.703
121 ARG+ HD2	121 ARG+ HB3	3.632
121 ARG+ HD2	121 ARG+ HB2	3.561
121 ARG+ HD2	121 ARG+ HD3	2.524
121 ARG+ HA	121 ARG+ HG3	3.878
121 ARG+ HG3	121 ARG+ HB2	4.075
121 ARG+ HD3	121 ARG+ HG3	3.351
121 ARG+ HD2	121 ARG+ HG3	3.460

121 ARG+ HA	121 ARG+ HG2	3.534
121 ARG+ HG2	121 ARG+ HB2	4.137
121 ARG+ HD3	121 ARG+ HG2	3.676
121 ARG+ HD2	121 ARG+ HG2	3.551
122 SER HB3	6 TRP HB3	4.236
6 TRP HB2	122 SER HB3	5.543
122 SER HB3	6 TRP HE3	4.808
122 SER HB2	6 TRP HB3	4.421
6 TRP HB2	122 SER HB2	5.017
122 SER HB2	6 TRP HE3	5.118
122 SER HA	8 VAL HA	3.206
8 VAL HA	122 SER HB3	3.982
8 VAL QG1	122 SER HB3	4.647
122 SER HB3	8 VAL QG2	5.432
8 VAL HA	122 SER HB2	3.635
8 VAL QG1	122 SER HB2	4.103
122 SER HB2	8 VAL QG2	4.752
122 SER HA	122 SER HB3	3.737
122 SER HA	122 SER HB2	3.581
123 LYS+ HA	6 TRP HE3	4.061
123 LYS+ HA	6 TRP HZ3	4.013
123 LYS+ HB3	9 TYR HD1	4.820
123 LYS+ HB3	9 TYR HE1	4.819
123 LYS+ HB2	9 TYR HD1	4.948
123 LYS+ HB2	9 TYR HE1	4.603
123 LYS+ HD2	9 TYR HD1	4.597
123 LYS+ HE2	9 TYR HD1	4.714
123 LYS+ HE2	9 TYR HE1	4.389
123 LYS+ HG3	9 TYR HD1	4.555
123 LYS+ HG3	9 TYR HE1	4.783
123 LYS+ HG2	9 TYR HD1	4.933
123 LYS+ HG2	9 TYR HE1	4.922
123 LYS+ HB3	123 LYS+ HA	3.450
123 LYS+ HA	123 LYS+ HB2	3.086
123 LYS+ HD2	123 LYS+ HA	4.528
123 LYS+ HD2	123 LYS+ HE2	3.679
123 LYS+ HA	123 LYS+ HG3	3.779
123 LYS+ HB3	123 LYS+ HG3	3.143
123 LYS+ HB2	123 LYS+ HG3	3.206
123 LYS+ HE2	123 LYS+ HG3	2.796
123 LYS+ HG2	123 LYS+ HA	4.580
123 LYS+ HG2	123 LYS+ HB2	3.497
123 LYS+ HE2	123 LYS+ HG2	3.325
6 TRP HA	124 ARG+ HA	3.413
105 ASN HA	124 ARG+ HB3	4.229
105 ASN HB3	124 ARG+ HB3	5.055
105 ASN HB2	124 ARG+ HB3	5.086
124 ARG+ HB2	105 ASN HA	4.371
105 ASN HB3	124 ARG+ HB2	4.314
105 ASN HB2	124 ARG+ HB2	4.167
124 ARG+ HA	124 ARG+ HB3	3.839
124 ARG+ HB2	124 ARG+ HA	3.828
124 ARG+ HB2	124 ARG+ HB3	2.899
124 ARG+ HD3	124 ARG+ HB3	3.991
124 ARG+ HD3	124 ARG+ HB2	3.769
124 ARG+ HD2	124 ARG+ HB3	3.865

124 ARG+ HB2	124 ARG+ HD2	3.539
124 ARG+ HA	124 ARG+ HG3	3.841
124 ARG+ HG3	124 ARG+ HB2	4.149
124 ARG+ HD3	124 ARG+ HG3	3.886
124 ARG+ HD2	124 ARG+ HG3	3.624
124 ARG+ HG2	124 ARG+ HA	3.947
124 ARG+ HG2	124 ARG+ HD2	3.582
125 VAL HA	125 VAL HB	2.923
125 VAL HA	125 VAL QG1	3.261
125 VAL HB	125 VAL QG1	3.154
125 VAL QG2	125 VAL HA	3.588
125 VAL QG2	125 VAL HB	2.853
4 GLY HN	3 SER HA	2.761
5 THR HN	4 GLY HA1	3.350
5 THR HN	4 GLY HA2	3.299
5 THR HN	6 TRP HD1	3.952
6 TRP HN	5 THR HA	3.651
7 GLN HN	6 TRP HA	3.044
8 VAL HN	7 GLN HA	3.345
9 TYR HN	8 VAL HA	3.356
9 TYR HN	10 ALA HN	3.178
13 ASN HN	12 GLU- HA	3.400
13 ASN HN	14 TYR HN	3.976
15 GLU- HN	14 TYR HN	3.547
16 GLU- HN	15 GLU- HA	4.244
15 GLU- HN	16 GLU- HN	3.160
16 GLU- HN	17 PHE HN	3.628
18 LEU HN	17 PHE HN	3.708
18 LEU HN	19 LYS+ HN	3.674
21 LEU HN	20 ALA HN	3.181
23 LEU HN	22 ALA HA	3.238
22 ALA HN	23 LEU HN	3.327
25 GLU- HN	24 PRO HA	2.647
26 ASP- HN	27 LEU HN	3.315
27 LEU HN	28 ILE HN	3.136
30 MET HN	29 LYS+ HN	3.419
31 ALA HN	30 MET HN	3.207
34 ILE HN	33 ASP- HA	3.667
34 ILE HN	33 ASP- HN	3.234
35 LYS+ HN	34 ILE HA	2.961
37 ILE HN	36 PRO HA	2.830
38 VAL HN	37 ILE HA	2.842
6 TRP HN	39 GLU- HA	3.733
39 GLU- HN	38 VAL HA	2.792
4 GLY HN	40 ILE HN	3.720
40 ILE HN	39 GLU- HA	2.775
41 GLN HN	40 ILE HA	2.827
42 GLN HN	41 GLN HA	2.779
45 ASP- HN	44 GLY HA1	3.000
45 ASP- HN	44 GLY HA2	4.000
43 LYS+ HN	46 ASP- HN	4.500
46 ASP- HN	45 ASP- HA	3.796
45 ASP- HN	46 ASP- HN	3.008
47 PHE HN	46 ASP- HA	2.547
41 GLN HN	48 VAL HN	3.757
49 VAL HN	48 VAL HA	2.815

50 THR HN	49 VAL HA	3.156
51 SER HN	50 THR HA	2.832
37 ILE HN	52 LYS+ HN	4.185
52 LYS+ HN	51 SER HA	2.899
53 THR HN	52 LYS+ HA	2.967
53 THR HN	56 GLN HN	4.408
56 GLN HN	55 ARG+ HA	3.780
53 THR HN	57 THR HA	4.221
58 VAL HN	57 THR HA	2.941
61 SER HN	60 ASN HA	2.649
62 PHE HN	47 PHE HN	3.877
62 PHE HN	48 VAL HA	3.923
62 PHE HN	61 SER HA	2.721
47 PHE HN	63 THR HA	3.028
63 THR HN	62 PHE HA	3.278
65 GLY HN	64 LEU HA	3.722
65 GLY HN	66 LYS+ HN	4.327
69 ASP- HN	68 ALA HA	3.298
70 ILE HN	69 ASP- HA	2.697
71 THR HN	70 ILE HA	3.366
76 LYS+ HN	75 GLY HA1	4.357
76 LYS+ HN	75 GLY HA2	4.105
75 GLY HN	76 LYS+ HN	3.610
78 LEU HN	77 LYS+ HA	2.976
79 LYS+ HN	78 LEU HA	3.134
84 LEU HN	83 HIS HA	2.852
88 LYS+ HN	87 GLY HA2	4.140
87 GLY HN	88 LYS+ HN	4.003
103 LYS+ HN	102 VAL HA	2.630
103 LYS+ HN	106 GLU- HN	3.770
106 GLU- HN	105 ASN HA	3.618
107 MET HN	106 GLU- HA	3.089
108 VAL HN	107 MET HA	2.601
113 PHE HN	112 THR HA	2.952
113 PHE HN	116 VAL HN	3.727
115 GLY HN	116 VAL HN	4.020
113 PHE HN	117 THR HA	3.594
117 THR HN	116 VAL HA	2.544
118 LEU HN	112 THR HA	4.101
118 LEU HN	117 THR HA	3.046
119 ILE HN	118 LEU HA	3.391
10 ALA HN	121 ARG+ HN	4.206
9 TYR HN	122 SER HA	3.627
122 SER HN	121 ARG+ HA	3.007
7 GLN HN	123 LYS+ HN	4.207
107 MET HN	123 LYS+ HA	4.324
7 GLN HN	124 ARG+ HA	3.951
124 ARG+ HN	123 LYS+ HA	2.861
125 VAL HN	6 TRP HA	4.018
125 VAL HN	124 ARG+ HA	2.852
2 PHE HD1	89 LEU QD1	4.771
2 PHE HZ	102 VAL QG2	4.278
6 TRP HN	38 VAL O	2.00
6 TRP N	38 VAL O	3.00
7 GLN HN	123 LYS+ O	2.00
7 GLN N	123 LYS+ O	3.00

9 TYR HN	121 ARG+ O	2.00
9 TYR N	121 ARG+ O	3.00
38 VAL HN	6 TRP O	2.00
38 VAL N	6 TRP O	3.00
39 GLU- HN	50 THR O	2.00
39 GLU- N	50 THR O	3.00
40 ILE HN	4 GLY O	2.00
40 ILE N	4 GLY O	3.00
41 GLN HN	48 VAL O	2.00
41 GLN N	48 VAL O	3.00
46 ASP- HN	43 LYS+ O	2.00
46 ASP- N	43 LYS+ O	3.00
48 VAL HN	41 GLN O	2.00
48 VAL N	41 GLN O	3.00
49 VAL HN	60 ASN O	2.00
49 VAL N	60 ASN O	3.00
106 GLU- HN	103 LYS+ O	2.00
106 GLU- N	103 LYS+ O	3.00
112 THR HN	97 SER O	2.00
112 THR N	97 SER O	3.00
118 LEU HN	111 ILE O	2.00
118 LEU N	111 ILE O	3.00

## CHAPTER 5

### *NMR Methods*

#### 5.1 Introduction

Theoretical analysis of an NMR experiment requires calculation of the signal observed following a sequence of radio frequency (r. f.) pulses and delays. The initial state of the spin system is described by the equilibrium density operator. Evolution of the density operator through the sequence of pulses and delays is calculated using Liouville-von Neumann equation:

$$\frac{d\sigma(t)}{dt} = -i[H, \sigma(t)]$$

The hamiltonian  $H$  consists of the Zeeman, scalar coupling, and r. f. pulse terms that govern evolution of the density operators. The expectation value of the observed signal at the desired time is calculated as the trace of the product of the density operator and the observation operator corresponding to the observable magnetisation.

It is possible to describe the NMR experiment using the product operator approach. It is convenient to concentrate on operators themselves, rather than to solutions of the Shroedinger equation. For weakly coupled spin systems the most straightforward choice is the use of product operators constructed from spin angular momentum operators  $I_x, I_y, I_z$ . These components are also closely related to macroscopic magnetisation. A system consisting of two spin  $\frac{1}{2}$  nuclei,  $I$  and  $S$ , can be described by 16 product operators:

$\frac{1}{2} 1$	the unity operator
$I_z, S_z,$	longitudinal magnetisation
$I_x, I_y, S_x, S_y,$	In-phase transverse magnetisation
$2I_zS_x, 2I_zS_y, 2I_xS_z, 2I_yS_z,$	Anti-phase transverse magnetisation
$2I_xS_x, 2I_yS_x, 2I_yS_y,$	two spin coherence
$2I_zS_z,$	longitudinal two-spin order

the various product operators are transformed by pulses, spin-spin couplings and chemical shifts.

The rules for these transformations are:

chemical shift evolution:

$$I_z \xrightarrow{\Omega t I_z} I_z$$

$$I_{x,y} \xrightarrow{\Omega t I_z} I_{x,y} \cos \pi \Omega t \pm I_{y,x} \sin \pi \Omega t$$

Where  $\Omega = \omega_0 - \omega_{rf}$  is the chemical shift frequency relative to the carrier frequency ( $\omega_{rf}$ ) and  $\omega_0$  is the larmor frequency.

effect of rf pulses:

$$I_z \xrightarrow{\beta_{y,x}^0} I_z \cos \beta \pm I_{x,y} \sin \beta$$

$$I_{x,y} \xrightarrow{\beta_{x,y}^0} I_{x,y}$$

$$I_{x,y} \xrightarrow{\beta_{y,x}^0} I_{x,y} \cos \beta \pm I_z \sin \beta$$

where  $\beta$  is the flip angle and the subscript indicates the phase of the pulse.

scalar evolution:

$$I_z \xrightarrow{\pi J_{IS} t 2 I_z S_z} I_z$$

$$2 I_x S_y \xrightarrow{\pi J_{IS} t 2 I_z S_z} 2 I_x S_y$$

$$I_{x,y} \xrightarrow{\pi J_{IS} t 2 I_z S_z} I_{x,y} \cos \pi J_{IS} t \pm 2 I_{y,x} S_z \sin \pi J_{IS} t$$

$$2 I_{x,y} S_z \xrightarrow{\pi J_{IS} t 2 I_z S_z} 2 I_{x,y} S_z \cos \pi J_{IS} t \pm 2 I_{y,x} \cos \pi J_{IS} t$$

where  $J_{IS}$  is the coupling constant between spin I and S.

Product operators can be used to analyse NMR experiments using a simple set of rules.

## 5.2 Heteronuclear NMR experiments

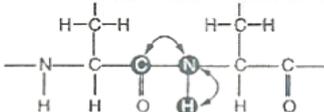
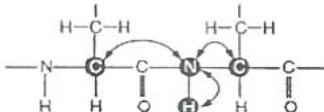
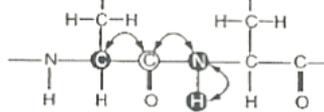
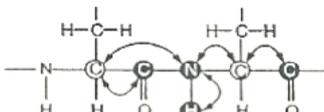
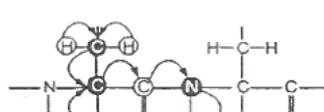
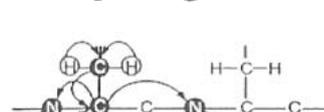
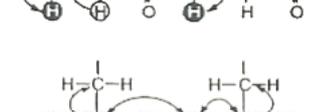
The traditional assignment strategy for proteins that are not isotopically enriched (1) makes use of a combination of COSY/TOCSY and NOESY 2D spectra. However resonance overlap becomes so severe in 2D spectra of proteins with more than about 80 residues, that it is necessary to turn to  $^{13}\text{C}$  and  $^{15}\text{N}$  labelling, in order to carry out NMR structure determination.

The heteronuclear triple-resonance experiments (2) correlate backbone  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $^1\text{H}^{\alpha}$ ,  $^{13}\text{C}^{\alpha}$  and  $^{13}\text{CO}$  (and side chain  $^1\text{H}$  and  $^{13}\text{C}$ ) spins using one bond ( $^1\text{J}$ ) or two bond ( $^2\text{J}$ ) scalar couplings which are largely independent of conformation. Also, since the  $^1\text{J}$  ( $^2\text{J}$ ) couplings are generally larger than line width of nuclei under consideration, the coherence transfer via these couplings remain efficient for relatively large molecules with short transverse relaxation times and can compete with loss of magnetisation as a result of relaxation during the pulse sequence.

Table 5.1 summarizes the correlations that are observed, and the scalar couplings utilized for coherence transfer, in several useful triple resonance experiments. These experiments offer alternative ways to establish sequential backbone connectivities. At least two and often more independent pathways can be found to support a given sequential assignment, without any knowledge of the spin system type. The nomenclature for these triple-resonance experiments reflect the magnetisation transfer pathway of the experiment: nuclei that are involved in the magnetisation transfers define the name of the experiment. Spins, whose chemical shifts are not evolved, are put in parentheses. For an out-and-back type experiment, where magnetisation of a spin is transferred to a remote spin and then back the same way, only the first half of the magnetisation transfer is used for the name.

The NMR pulse sequences for the assignment of  $^{13}\text{C}$ ,  $^{15}\text{N}$ , labelled proteins are composed of a large number of pulses that have to be applied at three or four different frequencies in order to excite  $^1\text{H}$ ,  $^{15}\text{N}$ , aliphatic  $^{13}\text{C}$  and carbonyl  $^{13}\text{C}$  resonances.

Table 5.1 Common triple resonance experiments used for sequential assignment and mentioned in this thesis.

Experiment	Correlations observed	Magnetization transfer	Relative S/N [%]
HNCO	$H^N(i)-N(i)-CO(i-1)$		100
HNCA	$H^N(i)-N(i)-C^\alpha(i)$ $H^N(i)-N(i)-C^\alpha(i-1)$		50 15
HN(CO)CA	$H^N(i)-N(i)-C^\alpha(i-1)$		58
HN(CA)CO	$H^N(i)-N(i)-CO(i)$ $H^N(i)-N(i)-CO(i-1)$		13 4
CBCA(CO)N H	$C^\beta(i)-C^\alpha(i)-N(i-1)-H^N(i-1)$		13/9 $\alpha/\beta$
CBCANH	$C^\beta(i)-C^\alpha(i)-N(i)-H^N(i)$ $C^\beta(i)-C^\alpha(i)-N(i+1)-H^N(i+1)$		4/1.7 $\alpha/\beta$ 1.3/0.5 $\alpha$ $\beta$
HNCACB	$H^N(i)-N(i)-C^\alpha(i)-C^\beta(i)$ $H^N(i)-N(i)-C^\alpha(i-1)-C^\beta(i-1)$		4/1.7 $\alpha/\beta$ 1.3/0.5 $\alpha$ $\beta$
HNCAHA	$H^N(i)-N(i)-H^\alpha(i)$ $H^N(i)-N(i)-H^\alpha(i-1)$		30

Perturbing the populations of stationary states within a dipolar coupled spin system causes a dipolar time dependent changes in the intensities of dipolar-coupled resonance signals via the heteronuclear Overhauser effect (NOE). Dipolar cross-relaxation is an extremely useful mixing process in multidimensional NMR spectroscopy, because the efficiency of mixing depends on the distance between interacting spins. Thus, through-space, and through-bond, magnetisation transfer generates cross peaks in the NOE mixing process of the NOESY experiments (3). As NOESY experiments contain information on inter-nuclear distances they are the most powerful and important technique available for structural investigation of biomolecules by NMR spectroscopy. Homonuclear 2D NOESY spectra are commonly used to measure distances between protons in small proteins but spectral overlap become serious with increasing number of residues.

Three-dimensional heteronuclear-edited NMR experiments resolve cross-peaks between  $^1\text{H}$  spins according to the chemical shift of the heteronuclei bonded directly to the  $^1\text{H}$  spins.

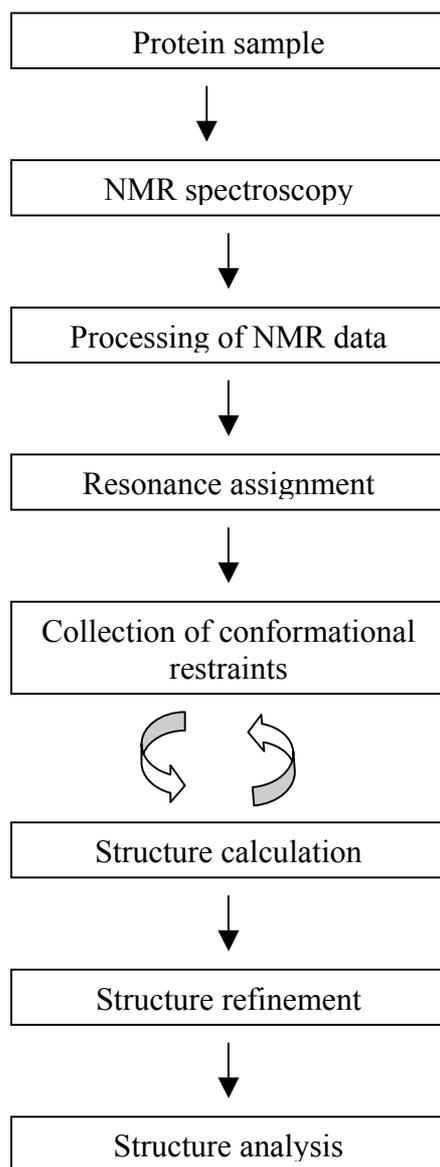
In my thesis work I performed heteronuclear experiments for structure calculation of cl-BABP protein.

The following 3D experiments were acquired at 700 MHz on  $^{13}\text{C}$ - $^{15}\text{N}$  doubly enriched cl-BABP: HNCA, HNC0, HNC0CA, CBCACONH, CBCANH. The following spectra were acquired on the sample in  $\text{D}_2\text{O}$ : HACACO, hCCH-TOCSY, hCCH-COSY, HcCH-TOCSY, HcCH-COSY, HcCH-NOESY.

2D- HSQC, 3D- HSQC-TOCSY, 3D-HSQC-NOESY were acquired, at 500 and 700 MHz, on  $^{15}\text{N}$  enriched cl-BABP.

### **5.3 Structure calculation**

The idea of computer-aided structure calculation is to convert distance- and torsion-angle-data (constraints) into a visible structure, following the procedure outlined in Scheme 5.1.

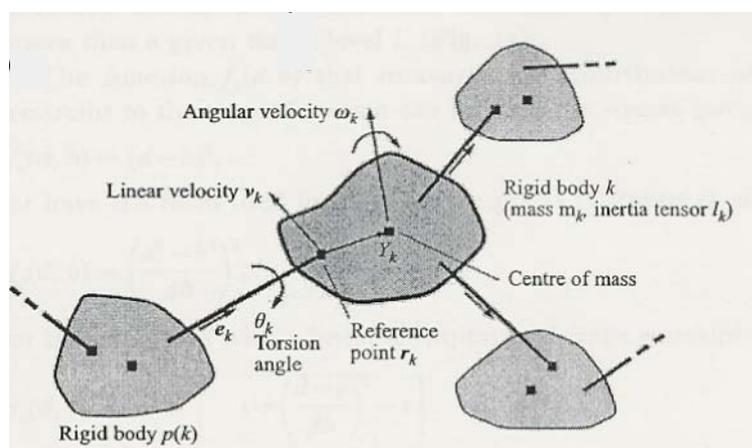


**Scheme 5.1**

The program used for structure calculation is DYANA (DYnamics Algorithm for Nmr Applications) (4), it performs simulated annealing by molecular dynamics in torsion angle space and uses a fast recursive algorithm to integrate the equations of motions.

A torsion angle dynamics is a molecular dynamics simulations using torsion angles instead of Cartesian coordinates as degree of freedom. For torsion angle dynamics calculations with DYANA the molecule is represented as a tree structure consisting of a base rigid body that is fixed in space and  $n$  rigid bodies, which are connected by  $n$  rotatable bonds. The degrees of freedom are exclusively torsion angles, i.e. rotations about single bonds, so the number of degrees of freedom is reduced respect to the Cartesian coordinates. Each rigid body is made up of one ore several mass

points (atoms) with variable relative positions. The tree structure starts from a base, typically from the N-terminus of the polypeptide chain, and terminates with “leaves” at the ends of the side chains and the C-terminus. The rigid bodies are numbered from 0 to  $n$ . Each rigid body with a number  $k > 1$  has a single nearest neighbour in the direction towards the base, which has a number  $p(k) < k$  (Scheme 5.2). The torsion angle between the rigid bodies  $p(k)$  and  $k$  is denoted by  $\theta_k$ . The conformation of the molecule is uniquely specified by the values of all torsion angles,  $\theta = (\theta_1, \dots, \theta_n)$ . For each rotatable bond,  $\mathbf{e}_k$  denotes the unit vector in the direction of the bond, and  $\mathbf{r}_k$  is the position vector of its end point. All these vectors are referred to an inertial frame of reference that is fixed in space.



**Scheme 5.2** Tree structure formed by torsion angles of a molecule.

In DYANA program the potential energy of the molecule is called target function  $V$ . The target function  $V \geq 0$  is defined such that  $V=0$  if and only if all experimental distance restraints are fulfilled and all non-bonded atom pairs satisfy a check for the absence of steric overlap. It measures restraints violation such that  $V(\theta') < V(\theta)$  whenever a conformation  $\theta$  satisfies the restraint more closely than another conformation  $\theta'$ .

The integration of equation of motion in torsion angle dynamics is reported in reference 4.

The potential energy landscape of a protein is complex and studded with many local minima. Because the temperature, i.e. the kinetic energy, determines the maximal height of energy barriers that can be overcome in a molecular dynamics simulation, the temperature schedule is important for the success and efficiency of a simulated annealing calculation. The simulated annealing protocol used by the program DYANA is reported in ref 4.

## 5.4. The relaxation

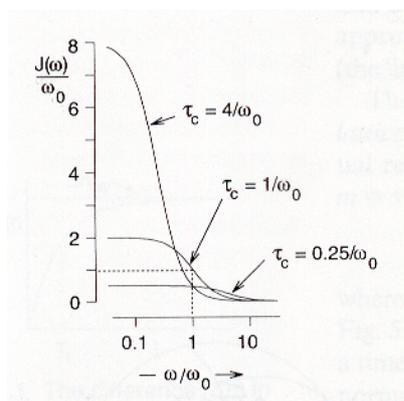
In NMR experiments, radio-frequency pulses disturb the equilibrium of the spin systems. Relaxation is the process by which equilibrium is regained through interaction of the spins with the thermal molecular environment.

The two relaxation processes, the spin-lattice relaxation and the spin-spin relaxation, are reviewed in this chapter together with a description of the applications of relaxation experiments in the protein study.

### 5.4.1. Dipole-dipole interaction and rotational motion in liquid

The mechanism of nuclear spin relaxation lies in magnetic interactions, the most important being dipolar coupling. Every nucleus with non-zero spin quantum number has a magnetic dipole and therefore behaves like a small bar magnet, producing a local magnetic field,  $B_{\mu}$ . The interaction of a nucleus with neighbours nuclei through the local magnetic fields produced by the magnetic moments is known as dipolar coupling. The dipolar coupling between two nuclei depends on the separation  $r$  and on  $\theta$ , the angle between the inter-nuclear vector and the static field. As the molecules translate, rotate and vibrate,  $r$  and  $\theta$  vary in a complicated way causing the interaction to fluctuate rapidly. Thus the dipolar coupling, modulated by molecular motions, causes nuclear spins to experience time-dependent local magnetic fields which, if they contain a component at the Larmor frequency, can induce the transition which return spins to equilibrium.

The frequency with which the local magnetic fields  $B_{\mu}$  fluctuate depends on the correlation time  $\tau_c$ . The correlation time indicates how long  $B_{\mu}$  maintains the same position and orientation before changing them due to the random collisions among the molecules in the solution. Rapid fluctuations have a small value of  $\tau_c$ , while slow fluctuations have a large value of  $\tau_c$ . The spectrum of the molecular motions, that depends on the value of  $\tau_c$ , is described by the spectral density  $J(\omega)$  (Figure 5.1) ( $\omega$  is the angular frequency in  $\text{radian s}^{-1}$ ).



**Figure 5.1.** The spectral density  $J(\omega)$  drawn for three values of the correlation time  $\tau_c$ .

$J(\omega)$  can be interpreted as the probability of finding a component of the random motion at a particular frequency. The integral of  $J(\omega)$  over all frequencies is a constant, independent of  $\tau_c$ . The relation between  $J(\omega)$  and  $\tau_c$  is:

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2\tau_c^2} \quad (5.1)$$

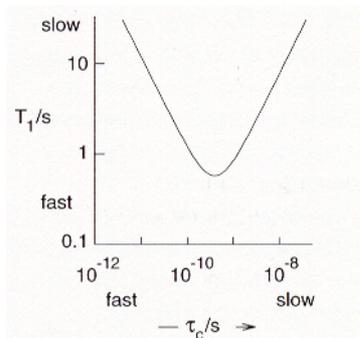
#### 5.4.2. Spin-lattice relaxation

Spin-lattice relaxation is caused by fluctuating local fields which induce nuclei to flip amongst their available spin states. The rate of this process,  $T_1^{-1}$ , depends on the probability that the local field have a component oscillating at the appropriate frequency, namely the Larmor frequency  $\omega_0$ .  $T_1^{-1}$  is proportional to the spectral density  $J(\omega_0)$  as follows:

$$\frac{1}{T_1} = \gamma^2 \langle B_\mu^2 \rangle J(\omega_0) \quad (5.2)$$

where  $\langle B_\mu^2 \rangle$  is the mean square value of the local field.

Figure 2.1 shows that  $J(\omega_0)$  is small for  $\tau_c^{-1}$  much smaller than  $\omega_0$ , or much larger than  $\omega_0$  and reaches a maximum when  $\tau_c^{-1}$  matches the Larmor frequency ( $\omega_0\tau_c = 1$ ). This behaviour is summarised in Figure 5.2. For rapidly tumbling molecules with  $\omega_0\tau_c \ll 1$  (left-hand side of Figure 5.2),  $J(\omega_0) \approx 2\tau_c$  and the relaxation gets slower as the mean tumbling rate is increased. Conversely, slowly tumbling molecules have  $\omega_0\tau_c \gg 1$  (right-hand side of Figure 5.2) and  $J(\omega_0) \approx 2/\omega_0^2\tau_c$ , so that the relaxation accelerates as the tumbling speeds up. The maximum relaxation rate occurs for  $\omega_0\tau_c = 1$  (minimum  $T_1$ ), at which point  $J(\omega_0) = 1/\omega_0$ .



**Figure 5.2.** The dependence of the spin-lattice relaxation time  $T_1$  on the correlation time  $\tau_c$ . The regions of the graph corresponding to fast and slow tumbling and fast and slow relaxation are indicated.

For macromolecules that rotate slowly relative to  $\omega_0$ , so that  $\omega_0\tau_c \gg 1$ ,  $T_1$  increases proportionally to  $\tau_c$ .

Physically, spin-lattice relaxation couples the spins (very weakly) to the motion of the molecule that carry them and so provides a pathway for the exchange of energy between the spin system and its surroundings. The energy absorbed or released in the course of the spin relaxation is transferred from or to the motions of the molecules, causing a slight cooling or warming of the lattice. Since the spin energies are small compared to the rotational, vibrational and translational energy of the molecules in solution, nuclear spins are relaxed with a non measurable small change in the temperature of the sample.

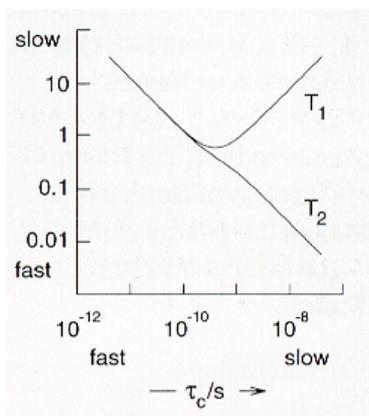
### 5.4.3. Spin-spin relaxation

The spin-spin relaxation is the process through which the  $xy$  magnetisation created by a radio frequency pulse decay to zero by the randomisation of the individual spins. The local fluctuating magnetic fields  $B_\mu$  are responsible of the  $T_2$  relaxation. These fields indeed produce small time-dependent variations in the precession frequencies of individual spins which lead to loss of phase-coherence in the sample.

The relation between  $T_2$  and the spectral density  $J(\omega)$  is:

$$\frac{1}{T_2} = \frac{1}{2} \gamma^2 \langle B_\mu^2 \rangle J(\omega_0) + \frac{1}{2} \gamma^2 \langle B_\mu^2 \rangle J(0) \quad (5.3)$$

The motional dependence of  $T_2$  is shown in Figure 5.3, together with the  $T_1$  behaviour.  $T_2$  increases as the tumbling gets faster; while in the slow motion limit ( $\omega_0\tau_c \gg 1$ ), typical of macromolecules, is simply inversely proportional to the correlation time. The two relaxation times,  $T_1$  and  $T_2$ , are identical in the extreme narrowing conditions  $\omega_0\tau_c \ll 1$ .



**Figure 5.3.** Dependence of  $T_1$  and  $T_2$  on correlation time  $\tau_c$ . The regions of the graph corresponding to fast and slow tumbling and fast and slow relaxation are indicated.

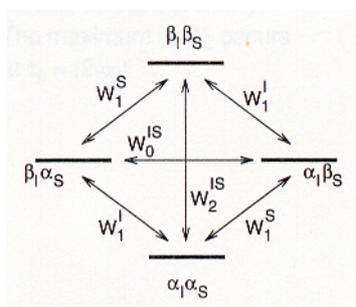
In general,  $T_1$  is greater than  $T_2$  and this is because there are additional causes of loss of transverse magnetization, the main of whose is that the static magnetic field  $B_0$  is not uniform throughout the sample. Thus, if we divide the sample in small regions such that the field is uniform in regions

known as isochromats, then the total magnetization is the sum of all these regions, each of which contributes a precessing vector which differs slightly in frequency. Consequently the transverse magnetization of the sample as a whole is reduced and the transverse relaxation is faster ( $T_2$  is shorter).

#### 5.4.4. Nuclear Overhauser effect

The dipole-dipole interaction between nuclei close together in the space causes a particular relaxation path known as cross relaxation that is the origin of the Nuclear Overhauser effect (NOE). The NOE is the fractional change in intensity of one NMR line when another resonance is selectively perturbed. A description of the phenomenon is given below.

Consider a pair of dipolar coupled  $\frac{1}{2}$  spins, I and S. Denoting the  $+1/2$  and  $-1/2$  spin state of each spin with  $\alpha$  and  $\beta$  respectively, the four possible combination of the two spins I and S are usually represented as  $\alpha_I\alpha_S$ ,  $\alpha_I\beta_S$ ,  $\beta_I\alpha_S$  and  $\beta_I\beta_S$  (Figure 5.4). At thermal equilibrium, the relative population of the four states can be described, to a good approximation as  $1+2\Delta$  ( $\alpha_I\alpha_S$ ),  $1-2\Delta$  ( $\beta_I\beta_S$ ),  $1$  ( $\alpha_I\beta_S$ ) and  $1$  ( $\beta_I\alpha_S$ ), where  $\Delta = \frac{1}{2}\hbar\omega_0/kT$



**Figure 5.4.** Energy levels for a pair of  $\frac{1}{2}$  spins nuclei I and S, showing the six possible relaxation pathways.

In Figure 5.4, six relaxation pathways are indicated:

- $W_1^I$  and  $W_1^S$  correspond to the single spin flipping  $\alpha_I \leftrightarrow \beta_I$  and  $\alpha_S \leftrightarrow \beta_S$  and are the spin-lattice relaxation processes.
- $W_0^{IS}$  and  $W_2^{IS}$  correspond to the simultaneous flipping of both I and S  $\alpha_I\alpha_S \leftrightarrow \beta_I\beta_S$  (both spins flipping in the same direction) and  $\alpha_I\beta_S \leftrightarrow \beta_I\alpha_S$  (I and S flipping in opposite directions). These relaxation processes are known as cross relaxation.

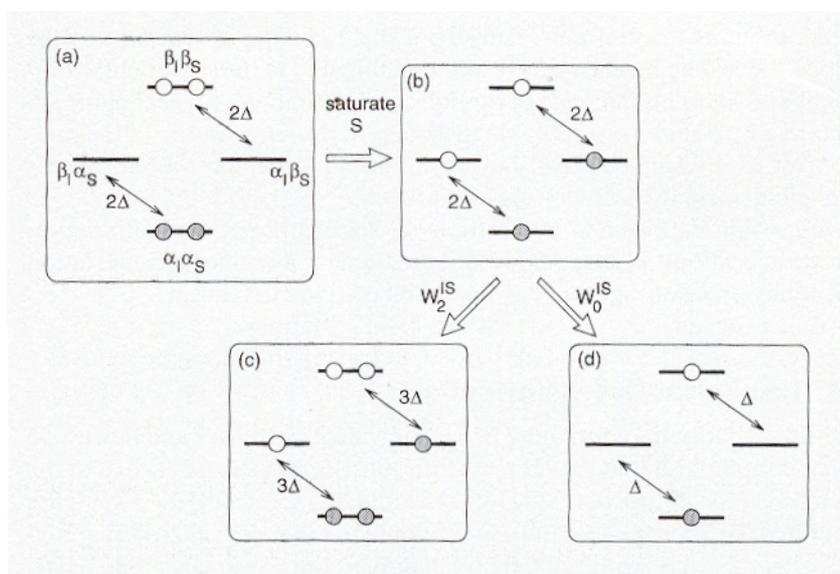
Cross relaxation comes about because the chaotic molecular motion, combined with the mutual dipolar interaction, causes the fluctuating local fields experienced by I and S to be correlated, with the results that the nuclei can undergo simultaneous spin-flips. The  $W_0^{IS}$  and  $W_2^{IS}$  processes are

extra pathways that allow the spin state populations to return to equilibrium following some disturbance.

The NOE effect can be explained through the following experiment. Imagine that the S transitions are saturated, i.e. the population  $\alpha_I\alpha_S$  and  $\alpha_I\beta_S$ , and the population  $\beta_I\alpha_S$  and  $\beta_I\beta_S$  are equalised, by the application of an appropriate radio frequency field. This has no effect on the population difference across the I transitions ( $\alpha_I\alpha_S \leftrightarrow \beta_I\alpha_S$  and  $\alpha_I\beta_S \leftrightarrow \beta_I\beta_S$ ).

Two different processes are now analysed (Figure 5.5):

1. All the relaxation pathways are insignificant except  $W_2^{IS}$ . This relaxation route transfers population between  $\alpha_I\alpha_S$  and  $\beta_I\beta_S$  and restores the equilibrium population of these two states,  $1+2\Delta$  and  $1-2\Delta$ . The population difference across the I transitions is now  $3\Delta$ , and therefore the intensity of the I signal has increased by 50%. Cross relaxation has transferred magnetization from the saturated spin S to its dipolar-coupled spin I.
2. All the relaxation pathways are insignificant except  $W_0^{IS}$ . This relaxation route transfers population between  $\beta_I\alpha_S$  and  $\alpha_I\beta_S$  and restores the equilibrium population of these two states (both unity). The population difference across the I transitions is now  $\Delta$ , and therefore the intensity of the I signal has decreased by 50%.



**Figure 5.5.** Spin state populations for a pair of neighbouring  $\frac{1}{2}$  spin nuclei I and S. Shaded circles indicate a population excess of  $\Delta$ ; open circles, a population deficit of  $\Delta$ . (a) Thermal equilibrium. (b) Effect of saturating both transitions of spin S. (c) Effect of  $W_2^{IS}$  cross relaxation. (d) Effect of  $W_0^{IS}$  cross relaxation.

The NOE can be quantified by the parameter  $\eta$ :

$$\eta = \frac{i - i_0}{i_0} \quad (5.4)$$

where  $i$  is the perturbed NMR intensity of the spin I and  $i_0$  its normal intensity.

The maximum homonuclear NOE can be  $\frac{1}{2}$  and the minimum  $-1$ . In reality, neither  $W_2^{IS}$  nor  $W_0^{IS}$  dominates the other relaxation pathways and  $\eta$  is somewhere between the extremes. The parameter  $\eta$  has the same sign as  $W_2^{IS} - W_0^{IS}$ .

Both the relaxation rates  $W_2^{IS}$  and  $W_0^{IS}$  are related to the spectral density  $J(\omega)$  as follows:

$$W_2^{IS} \approx J(2\omega_0) \quad \text{and} \quad W_0^{IS} \approx J(0) \quad (5.5)$$

Without giving all the mathematical details, it occurs that the proton-proton NOE,  $\eta$ , should be positive for fast motions ( $\omega_0\tau_c \ll 1$ ) and negative for slow motions ( $\omega_0\tau_c \gg 1$ ). The change of sign occurs when  $W_2^{IS} = W_0^{IS}$ , at which point the effect of the two cross relaxation pathways cancel; this happens when  $\omega_0\tau_c \approx 1$ .

NOE are also observable for heteronuclear pairs of spins.

#### 5.4.5 Relaxation experiments:

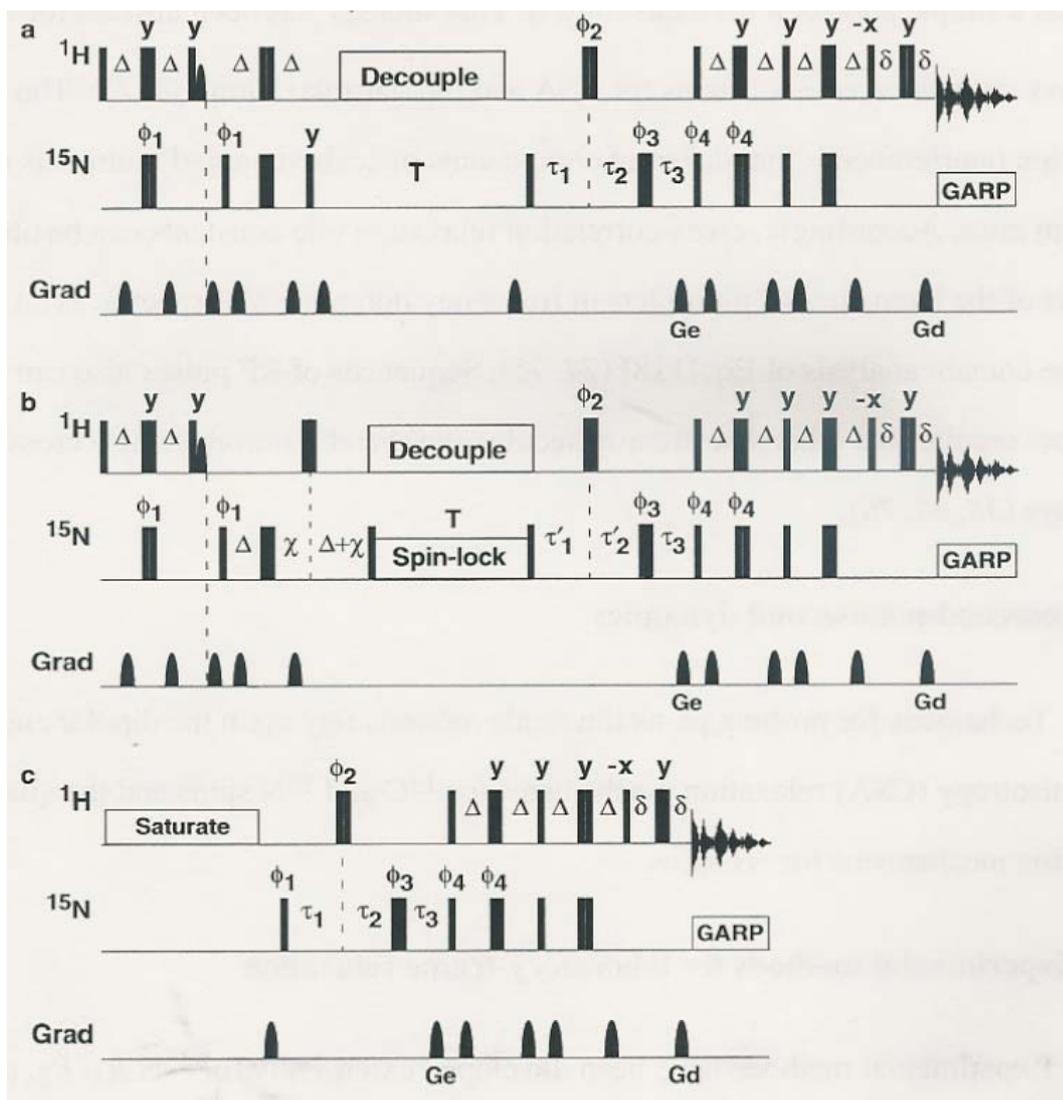
NMR has the unique capacity to investigate dynamics properties of molecules over a range of different time scales with atomic resolution (5).

Investigations of dynamic processes by high-resolution solution-state NMR spectroscopy can be categorized on the basis of the correlation times for experimentally-accessible motional process: laboratory frame nuclear spin relaxation measurements sensitive to pico second to nanosecond time scales, line shape analysis and rotating frame nuclear spin relaxation measurements sensitive to microsecond to millisecond time scale, and magnetization exchange spectroscopy sensitive to millisecond to second time scales. (6, 7-11).

Pulse sequences for measuring spin relaxation usually consists of five building blocks: preparation, relaxation, frequency labelling, mixing, and acquisition. The preparation period consists of a DEPT (12), INEPT (13, 14) or NOE (15, 16) transfer step from protons to the nucleus of interest. The density operator present after the preparation period provides the initial condition for the relaxation period,  $T$ . Chemical shifts are recorded during the  $t_1$  frequency-labelling period to generate the indirect dimension of the two-dimensional NMR spectrum. The desired heteronuclear coherence is

transferred to proton magnetization using reverse DEPT, INEPT, sensitivity enhanced (17, 18) polarization transfer sequences during mixing period. The relaxation-encoded, frequency-labelled transverse proton magnetization is recorded during the  $t_2$  acquisition period. The relaxation rate constant measured in a given experiment depends on the initial density operator and on any manipulation on the density operator during T. In most experiments, the relaxation period T, is increased parametrically in a time series of two-dimensional (2D) NMR spectra (19). In some cases such as steady state NOE measurements, T is fixed and the manner in which radio frequency pulses are applied during T is changed between different 2D spectra. Two-dimensional Fourier transformation yields a 2D frequency domain spectrum in which the relaxation information is encoded in the intensities or line shapes of the resonance signals.

Pulse sequences for measuring backbone amide  $^{15}\text{N}$  spin relaxation in proteins are shown in Figure 5.6. The inversion recovery technique (20) is used to measure the spin lattice relaxation rate constant for longitudinal magnetization,  $R_1$ . The Carr-Purcell-Meiboom-Gill (CPMG) (21, 22) technique is used to measure the spin-spin relaxation rate constant for transverse magnetization,  $R_2$ ; and the steady state NOE technique (23) is used to measure the  $^1\text{H}$ - $^{15}\text{N}$  NOE. The  $R_1$  and  $R_2$  experiments use refocused INEPT sequences for the preparation period and all three experiments use sensitivity-enhanced polarization transfer sequences for the mixing period. Decoupling of the  $^1\text{H}$  spins during the relaxation period T is used to suppress  $^1\text{H}$ - $^{15}\text{N}$  dipolar cross-relaxation and  $^{15}\text{N}$  CSA(chemical shift anisotropy)/ $^1\text{H}$ - $^{15}\text{N}$  dipolar relaxation interference in the  $R_1$  experiment and  $^{15}\text{N}$  CSA/ $^1\text{H}$ - $^{15}\text{N}$  dipolar relaxation interference in the  $R_2$  experiment.



**Figure 5.6:** Experimental technique for  $^{15}\text{N}$  (a)  $R_1$ , (b)  $R_2$  and (c) NOE spin relaxation measurements using two-dimensional, proton-detected pulse sequences.  $R_1$  and  $R_2$  intensity decay curves are recorded by varying the relaxation period  $T$  in a series of two dimensional experiments. The NOE is measured by recording one spectrum with saturation of  $^1\text{H}$  magnetisation and one spectrum without saturation. Narrow and wide bars depict  $90^\circ$  and  $180^\circ$  pulse respectively; water flip back pulses are shown as rounded bars (24). All pulses are x-phase unless otherwise indicated. Decoupling during acquisition is achieved with the GARP sequence (25); decoupling during  $T$  is performed using a train of cosine modulated  $180^\circ$  pulses designed to leave the water magnetization unperturbed (26). Spin lock during  $R_2$  experiment is performed using a continuous wave RF field centred in the spectral range of interest. Saturation during NOE experiments is performed using WALTZ-16 (27) or a train of  $^1\text{H}$  pulses.

Delays are  $\Delta=1/(4J_{\text{XH}})$ ,  $\chi=1/(2w_1)$ ,  $\tau_1=\Delta+\tau_1/2$ ,  $\tau_2=\Delta+\tau_2/2$ ,  $\tau_2=(1-2\Delta/t_{1\text{max}})t_1/2$ ,  $\tau_3=(1-t_1/t_{1\text{max}})\Delta$ ,  $t_{1\text{max}}$  is the maximum value of the  $t_1$  labelling period,  $\tau'_{1,2}=\tau_{1,2}+\chi$ . The delay  $\chi$  is used to orient

the initial magnetisation along the direction of the effective field in the rotating reference frame (28). The phase cycle is  $\phi_1 = x, -x$ ;  $\phi_2 = 4(x), 4(-x)$ ;  $\phi_3 = x, x, y, y, -x, -x, -y, -y$ ;  $\phi_4 = x$ ; receiver =  $x, -x, -x, x$ . The unlabelled gradients are used to suppress unwanted coherences and pulse imperfections. Gradient selection is achieved with gradients Gd and Ge. Echo/antiecho signals are recorded in separate experiments by inverting the amplitude of Ge and phase  $\phi_4$ .

#### 5.4.6 Relaxation data analysis:

$R_1$  and  $R_2$  relaxation rates are derived by fitting cross peak intensities, measured from the appropriate experiments (described in 5.4.5), to a single exponential function using the Rate Analysis routine of NMRView program (29). The heteronuclear NOE effect are calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation.

#### *Modelfree analysis*

Calculations of Lipari-Szabo motional parameters are performed with the program MOEDLFREE (version 4.0). Relaxation of amide  $^{15}\text{N}$  nuclear spins in a diamagnetic protein can be generally described by dipolar coupling with directly attached protons and  $^{15}\text{N}$  chemical shift anisotropy (30), as in [1-3]:

$$R_1 = (d^2/4) [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad [1]$$

$$R_2 = (d^2/8) [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (c^2/6) [4J(0) + 3J(\omega_N)] + R_{ex} \quad [2]$$

$$\text{NOE} = 1 + (d^2/4R_1) (\gamma_H/\gamma_N) [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad [3]$$

where  $d = (\mu_0 h (\gamma_H \gamma_N / 8\pi) \langle r_{\text{NH}}^{-3} \rangle$ ,  $c = \omega_N \Delta\sigma / 3^{1/2}$ ,  $\mu_0$  is the permeability of free space,  $h$  is Planck's constant,  $\gamma_H$  and  $\gamma_N$  are the gyromagnetic ratios of  $^1\text{H}$  and  $^{15}\text{N}$ , respectively,  $r_{\text{NH}} = 1.02 \text{ \AA}$ ,  $\omega_N$  and  $\omega_H$  are the Larmor frequencies of  $^{15}\text{N}$  and  $^1\text{H}$ , respectively and  $\Delta\sigma = -160 \text{ ppm}$  is the chemical shift anisotropy measured for  $^{15}\text{N}$  nuclei in helical polypeptide chains (31).

The extended Lipari-Szabo formalism proposes different spectral density functions (Table I), which depend upon  $S^2$  ( the generalised motional order parameter),  $\tau_m$  (the overall correlation time of rotational diffusion),  $\tau_e$  (the effective correlation time) and  $R_{ex}$  (the rate of conformational exchange).

Table I. Spectral density functions of the models used for the  $^{15}\text{N}$  relaxation data analysis

Model	Spectral density function	Optimized parameters
1	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2)\}$	$S^2$
2 <sup>a</sup>	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + (1-S^2)\tau_e'/(1 + \omega^2\tau_e'^2)\}$	$S^2, \tau_e$
3	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2)\}$ $1/T_{2(\text{obs})} = 1/T_2 + R_{\text{ex}}$	$S^2, R_{\text{ex}}$
4	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + (1-S^2)\tau_e'/(1 + \omega^2\tau_e'^2)\}$ $1/T_{2(\text{obs})} = 1/T_2 + R_{\text{ex}}$	$S^2, \tau_e, R_{\text{ex}}$
5 <sup>b</sup>	$J(\omega) = 2/5\{S^2\tau_m/(1 + \omega^2\tau_m^2) + S_f^2(1-S_s^2)\tau_s'/(1 + \omega^2\tau_s'^2)\}$	$S_f^2, S_s^2, \tau_e$

$$^a\tau_e' = \tau_m \tau_e / (\tau_m + \tau_e).$$

$$^b\tau_s' = \tau_m \tau_s / (\tau_m + \tau_s); S^2 = S_f^2 S_s^2.$$

The MODELFREE analysis was performed fitting the relaxation parameters to the five models of Table I and selecting, for each residue, the model which best fits the data by minimising the target function of the sum-squared error (SSE) (equation 4).

$$\text{SSE} = (R_1 - R_1^*)^2 / \sigma_1^2 + (R_2 - R_2^*)^2 / \sigma_1^2 + (\text{hNOE} - \text{hNOE}^*)^2 / \sigma_{\text{hNOE}}^2 \quad [4]$$

where  $R_1$ ,  $R_2$  and hNOE are the experimental relaxation parameters,  $\sigma_1$ ,  $\sigma_1$  and  $\sigma_{\text{hNOE}}$  are the uncertainties on the experimental data and  $R_1^*$ ,  $R_2^*$  and hNOE\* are the back calculated relaxation parameters, according to the selected model. F-statistics, which measures the random statistical reduction in SSE by incorporation of additional parameters, was additionally used to check on the model selection (32). A further check is the consistency between the selected model and the motional parameters (i.e.  $\tau_e > 20$  ps,  $R_{\text{ex}} > 0.5 \text{ s}^{-1}$  and  $S^2 > 0.9$ ).

The iterative optimisation of the overall rotational correlation time  $\tau_m$  was performed as described in (33, 34).

Chemical exchange contributions ( $R_{\text{ex}}$ ) to  $R_2$ , as determined by CPMG spin-echo (21), can be derived from equation [6]:

$$R_2 = R_2^0 + R_{\text{ex}} \quad [6]$$

where  $R_{\text{ex}}$  can be written as in [7] (35):

$$R_{\text{ex}} = \Phi \omega_N^2 \quad [7]$$

The factor  $\Phi$  depends on the type and the intrinsic rate constant of exchange process, on the population and chemical shift differences of participating sites and on the applied spin-echo period in the CPMG pulse train. For a given residue, the contribution of chemical exchange may be highlighted by an higher than average value monitored in a plot of  $R_2/R_1$  as a function of residue number.  $R_{\text{ex}}$  can be determined either with the Lipari Szabo approach, if the appropriate model can

be selected, or from the spectral density  $J(0)$  at two fields (see later). Both approaches were used here in order to compare and to validate the results.

### *Spectral density mapping.*

Relaxation data have been analysed according to the reduced spectral density approach (36) using two methods, as reported (36). In the first method, values of the reduced spectral density function at each field strength (600 and 700 MHz) were calculated from equations [1-3] in the hypothesis that  $J(\omega_H \pm \omega_N) \cong J(\omega_H) \cong J(0.87\omega_H)$ , as in [8-10] :

$$J(0.87\omega_H) = [4/(5d^2)] (\gamma_H/\gamma_N) (\text{NOE} - 1) R_1 \quad [8]$$

$$J(\omega_N) = 4/(3d^2 + 4c^2) [ R_1 - 7/4 d^2 J(0.87\omega_H) ] \quad [9]$$

$$J(0) = (6R_2 - 3R_1 - 3.6 \sigma) / (3d^2 + 4c^2) \quad [10]$$

where symbols have the usual meaning, as described before and  $\sigma = (\gamma_H/\gamma_N) (\text{NOE} - 1) R_1$ .

In the hypothesis that the dynamics is limited to a single motion, the spectral density is given by a single Lorentzian and takes the form of equation [11] (37):

$$J(\omega) = \frac{J(0)}{1 + 6.25(\omega J(0))^2} \quad [11]$$

A spectral density correlation graph can be built as reported in (37), giving a qualitative description of the molecular motions. To determine the conformational exchange contributions, a quantitative analysis had to be performed.  $J_{\text{eff}}(0)$  may contain conformational exchange contributions, according to  $J_{\text{eff}}(0) = J(0) + \lambda R_{\text{ex}}$ . In the absence of  $R_{\text{ex}}$  contributions,  $J_{\text{eff}}(0) = J(0)$  and the obtained  $J_{\text{eff}}(0)$  values are field independent. In the presence of  $R_{\text{ex}}$ ,  $J_{\text{eff}}(0)$  values are expected to increase with field strength, as  $R_{\text{ex}} \propto \omega_N^2$ , and a plot of  $J(0)_{\text{eff1}}/J(0)_{\text{eff2}}$ , determined at two field strengths, as a function of residue number, can monitor the exchange contribution on single residues.  $R_{\text{ex}}$  was estimated from equation [7] where  $\Phi$  is defined by equation [12] (35):

$$\Phi = \frac{J_{\text{eff1}}(0) - J_{\text{eff2}}(0)}{\lambda_1 \omega_{1N}^2 - \lambda_2 \omega_{2N}^2} \quad [12]$$

where  $J_{\text{eff1}}(0)$  and  $J_{\text{eff2}}(0)$ ,  $\lambda_1$  and  $\lambda_2$  are determined at 700 and 600MHz, respectively.  $\lambda_1 = 0.27$  (ns/rad)<sup>2</sup>,  $\lambda_2 = 0.29$  (ns/rad)<sup>2</sup>, as calculated from  $\lambda = (3/2)[1/(3d^2 + c^2)]$ .

In the second approach lated  $J(0)_{\text{eff}}$  can be calculated using simultaneously the relaxation data at two field strengths, as previously reported (36), and the contribution of  $R_{\text{ex}}$  to  $R_2$  rate measured at 600MHz can be derived (equation 13):

$$\begin{aligned}
J(0)_{\text{eff}} &= \frac{1}{\beta} \left\{ \left[ \frac{1}{T_2^{700}} - \kappa \frac{1}{T_2^{600}} \right] - \left( \frac{3d^2}{8} \right) \left[ J(\omega_N^{700}) - \kappa J(\omega_N^{600}) \right] - \left( \frac{c_{700}^2}{2} \right) \left[ J(\omega_N^{700}) - J(\omega_N^{600}) \right] - \left( \frac{13d^2}{8} \right) \left[ J(0.955\omega_H^{700}) - \kappa J(\omega_H^{600}) \right] \right\} \\
R_{\text{ex}} &= \frac{1}{T_2^{600}} - \left( \frac{d^2}{2} + \frac{2c_{600}^2}{3} \right) J(0) - \left( \frac{3d^2}{8} + \frac{c_{600}^2}{2} \right) J(\omega_N^{600}) - \left( \frac{13d^2}{8} \right) J(0.955\omega_H^{600}) \\
\kappa &= \left( \omega_H^{700} / \omega_H^{600} \right)^2, \beta = \left( \frac{d^2}{2} \right) (1 - \kappa) \\
J(\omega_N) &= 4/(3d^2 + 4c^2) [ R_1 - 7/4 d^2 J(0.921\omega_H) ] \tag{13}
\end{aligned}$$

where  $J(0.921\omega_H)$  and  $J(0.955\omega_H)$  have been determined using a first order Taylor expansion of  $J(\omega)$  about  $J(0.87\omega_H)$  at the two field strengths (equation 14):

$$J(\varepsilon\omega_H) \approx J(0.87\omega_H) + (\varepsilon - 0.87)\omega_H J'(0.87\omega_H) \tag{14}$$

where  $\varepsilon = 0.921$  or  $0.955$ ;  $J'(0.87\omega_H) = [J(0.87\omega_H^{700}) - J(0.87\omega_H^{600})] / [0.87(\omega_H^{700} - \omega_H^{600})]$ .

All the data obtained with the described methods can be compared in order to give a unique description of the protein dynamics.

## References

1. Wüthrich, K. (1986) Three-dimensional protein structure by NMR, *in NMR of proteins and nucleic acids*, pp 176-199, John Wiley & Sons, Inc., New York.
2. Cavanagh, J., Fairbrother, j., w., Palmer, A. G., Skelton N. J., Protein NMR spectroscopy; principle and practice. Academic Press (1996).
3. Sklenar, V., Piotto, M., Leppik, R., Saudek, (1993) *J. Magn. Res. Series A*, **102**, 241-245.
4. Guentert, P., (1998) *Quarterly Reviews of biophysics*, **31**, 145-237
5. Palmer, A.G, Williams, J., McDermott, A., (1996) *J.Phys.Chem.* 100, 13293-13310.
6. Palmer A. G., Kroenke, C. D., Loria, J. P., (2001) *Meth. Enzymology.* **339**, 204-238.
7. Daragan, V. A., Mayo, K. H., (1997) *Prog. Nucl. Magn. Reson. Spectrosc.* **31**, 63-105.
8. Fisher, M. W., Majumdar, A., Zuiderweg, E. R., (1998) *Progress in nuclear magnetic resonance* **33**, 207-272.
9. Kay, L. E., (1998) *Nat. struct Biol* **55**, 513-517.
10. Palmer, A. G., (1997) *Curr. Opin. Struct. Biol.* **7**, 732-737.
11. Palmer, A. G., (2001) *Annu. Rev. Biophys. Biomolec. Struct.* **30**, 129-155.

12. Doddrell, D. M. Pegg, D. T., Bendall, M. R., (1982) *J. Magn. Reson.* **48**, 323-327.
13. Burum, D.p., Ernst, R. R., (1980) *J. Magn. Reson.* **39**, 163-168.
14. Morris, G. A., Freeman, R. (1979) *J. Am. Chem. Soc.* **101**, 760-762.
15. Kay, L. E., Torchia, D. A., Bax, A., (1989) *Biochemistry* **28**, 8972-8979.
16. Kay, L., Bull, T. E., Nicholson, L. K., Griesinger, C., Schwalbe, H., Bax, A., Torchia, D. A., (1992) *J. Magn. Reson.* **100**, 538-558.
17. Kay, L., Keifer, P., Saarinen, T., (1992) *J. Am. Chem. Soc.* **114**, 10663-10665.
18. Palmer, A. G., Cavanagh, J., Wright, P.E, Rance, M., (1991) *J. Magn. Reson.* **93**, 151-170.
19. Skelton, N. J., Akke, M., Koerdel, J., Palmer, A. G., Rance, M., Chazin, W. J., (1993) *J. Magn. Reson.* **102**, 253-264.
20. Vold, R. L., Waugh, J. S., Jlein, M. P., Phelps, D. E., (1968) *J. Chem. Phys.* **48**, 3831-3832.
21. Carr, H. Y., Purcell, E. M., (1954) *Phys. Rev.* **94**, 688-691.
22. Meiboom, S., Gill, D., (1958) *Rev. Sci. Instrum* **29**, 688-691.
23. Noggle, J. H., Shirmer, R. E., (1971) *The Nuclear Overhauser effect: Chemical applications*, Academic press, New York.
24. Grzesiek, S., Bax, A., (1993) *J. Am. Chem. Soc.* **115**, 12593-94.
25. Shaka, A. J., Barker, P. B., Freeman, R., (1985) *J. Magn. Reson.* **64**, 547-552.
26. Farrow, N. A., Pawson, T., Singer, A. U., Pascal, S. M. et al., (1994) *Biochemistry* **33**, 5984-6003.
27. Shaka, A. J., Keeler, J., Frenkiel, T., Freeman, R., (1983) *J. Magn. Reson.* **52**, 335-338.
28. Yamazaki, T., Muhandram, R., Kay, L.E., (1994) *J. Am. Chem. Soc.* **116**, 8266-78.
29. Johnson B. A., Blevins R. A. (1994) *J. Biomol. NMR* **4**, 603-614.
30. Abragam, 1961, *Principles of Nuclear magnetism*, Oxford Science Publications.
31. Hiyama Y., Niu C.-H., Silverton J. V., Bavoso A. and Torchia D. A., (1988) *J. Am. Chem. Soc.* **110**, 2378-2383.
32. Mandel A. M., Akke M., Palmer III A., G. (1995) *J. Mol. Biol.* **246**, 144-163
33. Feng W, Tejero R, Zimmerman DE, Inouye M, Montelione GT. (1998) *Biochemistry* **37**, 10881-96
34. Li YC, Montelione GT. (1995) *Biochemistry.* **34**, 2408-23.
35. Peng JW, Wagner G. (1995) *Biochemistry.* **34**, 16733-52.
36. Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. (1995) *J Biomol NMR.* **6**, 153-62.
37. Krizova H, Zidek L, Stone MJ, Novotny MV, Sklenar V (2004) *J Biomol NMR.* **28**, 369-84.

# Anticarcinogenic Bowman Birk Inhibitor Isolated from Snail Medic Seeds (*Medicago scutellata*): Solution Structure and Analysis of Self-Association Behavior<sup>†</sup>

M. Catalano,<sup>‡</sup> L. Ragona,<sup>\*,‡</sup> H. Molinari,<sup>‡,§</sup> A. Tava,<sup>||</sup> and L. Zetta<sup>‡</sup>

Istituto per lo Studio delle Macromolecole, Laboratorio NMR, CNR, Via Ampère 56, 20131, Milano, Italy, Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, Strada Le Grazie, 37134 Verona, Italy, and Istituto Sperimentale Colture Foraggere, Viale Piacenza 29, 26900 Lodi, Italy

Received September 6, 2002; Revised Manuscript Received December 16, 2002

**ABSTRACT:** The high-resolution three-dimensional structure of a Bowman Birk inhibitor, purified from snail medic seeds (*Medicago scutellata*) (MSTI), has been determined in solution by <sup>1</sup>H NMR spectroscopy at pH 5.6 and 27 °C. The structure of MSTI comprises two distinct symmetric domains each composed of a three-stranded  $\beta$ -sheet containing a VIb type loop, where the active sites are located. A characteristic geometry of three aromatic residues confers stability to this protein, and we observe that this feature is conserved in all the Bowman Birk inhibitors of known structure. The two active domains exhibit different conformational features: the second domain displays higher flexibility and hydrophobicity with respect to the first one, and these properties have been correlated to a lower trypsin inhibitory specificity, in agreement with titration studies that have shown a stoichiometric ratio MSTI:trypsin of 1:1.5. NMR analysis indicated that MSTI undergoes self-association at concentrations higher than 2 mM, and the residues involved in this mechanism are localized at opposite faces of the molecule, having the highest positive and negative potential, respectively, thus indicating that electrostatic intermolecular interactions are the driving forces for MSTI association. Most of the residues affected by self-association are highly conserved in BBIs from different seeds, suggesting a functional relevance for these charged superficial patches, possibly involved in the interaction with other enzymes or macromolecules, thus triggering anti-carcinogenic activity.

The Bowman-Birk inhibitors (BBI)<sup>1</sup> are small serine protease inhibitors found in seeds of legumes and in many other plants (*I*). Typically, their molecular masses range between 6 and 9 kDa, and they contain seven disulfide bonds with a prominent role in the stabilization of their active configurations. Most BBIs exist in various isoforms (2). All members of the BBI family have two tandem homology regions on the same polypeptide chain, each comprising a consensus motif of three  $\beta$ -strands with a kinetically independent reactive site on the outermost loop. It has been shown that splitting the homology regions by partial peptic digestion yields two active fragments (3) and hence BBIs have been called “double-headed” inhibitors for their capability of inhibiting simultaneously and independently two different serine proteases. Most BBIs inhibit trypsin at the first reactive site (N-terminal) and chymotrypsin at the second

reactive site (C-terminal). The amino acid residues around the reactive site are usually designed as P3, P2, P1, P1', P2', and P3'. P1 residue is involved in the so-called “primary” contact region and confers inhibitory specificity: arginine and lysine for trypsin, leucine, phenylalanine, and tyrosine for chymotrypsin, and alanine for elastase (*I*). However, in contrast to primary specificity, trypsinolysis of soybean BBI (BBI-I) occurs not only in the predicted trypsin inhibitory domain but also in the predicted chymotryptic inhibitory domain (4).

Despite extensive studies of BBIs, only a few three-dimensional structures have been solved by X-ray or by NMR. They include structures of tracy soybean (5), peanut (6), pea seeds (2), and soybean inhibitor (7, 8). The X-ray structures of trypsin complexes of inhibitors from azuki bean (9), mung bean (10), and soybean (11) have also been reported. Recently a novel 14 amino acid residue cyclic peptide, cyclo (RCKTSIPPICFPDG), isolated from sunflower seeds, was found to be a potent inhibitor of trypsin ( $K_d = 100$  pM), and the X-ray structure of its complex with bovine trypsin showed both sequence and conformational similarity with the trypsin-reactive loop of the Bowman-Birk family of serine protease inhibitors (12). This inhibitor, however, is unique in being monofunctional, cyclic, and far shorter than inhibitors belonging to this family. It has been suggested that the high potency of this peptide is likely to arise from the considerable structural rigidity achieved through its cyclic nature which is further stabilized by a single internal disulfide bond.

<sup>†</sup> This work was supported by the Italian National Research Council “Progetto Finalizzato Biotecnologie” and “Progetto Agenzia 2000”.

\* To whom correspondence should be sent: Istituto per lo Studio delle Macromolecole, Laboratorio NMR, CNR, Via Ampère 56, 20131, Milano, Italy. Tel.: 0039 02 70643554. Fax: 0039 02 70643557. E-mail: laura.ragona@ismac.cnr.it.

<sup>‡</sup> Istituto per lo Studio delle Macromolecole.

<sup>§</sup> Università degli Studi di Verona.

<sup>||</sup> Istituto Sperimentale Colture Foraggere.

<sup>1</sup> Abbreviations: BBIs, Bowman-Birk inhibitors; BBI-I, Bowman-Birk inhibitor from soybean; MSTI, inhibitor from *Medicago scutellata* seed; PI-II, Bowman-Birk inhibitor from tracy soybean; PsTI-IVb, Bowman-Birk inhibitor from pea seed; TC, temperature coefficients; TSP, 3-(trimethylsilyl)-propionic acid-d<sub>4</sub> sodium salt.

The 11-residue peptide NleCTASIPPQCY, designed to reproduce the short  $\beta$ -sheet segment that forms the reactive site loop of BBIs, cyclized via a disulfide bridge, retains the "canonical conformation" typical of BBIs, indicating that this sequence represents an independent structural  $\beta$ -hairpin motif, and maintains the biological activity of the native protein (13). The 11-residue synthetic fragments SCTKSIxyQCY, where x and y were systematically substituted with A or P, were found to be potent inhibitors only when a cis-proline was present at position x (14).

It has been suggested that BBIs have antitumoral and radio-protective activity, and it is still unknown whether these activities are correlated to the inhibitory activity of trypsin/chymotrypsin or other enzymes or to the capability of interacting with other macromolecules (11, 15–21). In this respect research has been focalized into three main lines, namely: (i) cellular proteases involved in carcinogenesis and possibly inhibited by BBIs; (ii) the effect of the inhibitors on the nuclear excision repair system, through the stabilization of TP53 protein (22, 23), or the expression regulation of some protooncogenes such as c-myc and c-fos, overexpressed in many cancer cell lines (24); and (iii) in vivo inhibition of extra-cellular trypsin, secreted by many cancer cells and involved, with other metalloproteases, in the proteolysis of the extracellular matrix, thus favoring metastasis (16). The effort of the research along these lines stresses the importance of structural studies which may provide a link between specific structural features and antitumoral function.

Here we report a NMR structural study of a Bowman Birk trypsin inhibitor purified from *Medicago Scutellata* seeds (MSTI, 62 a.a.) (25). MSTI can inhibit the catalytic activity of bovine  $\beta$ -trypsin but, different from other BBIs, does not exhibit antichymotryptic activity.

Preliminary investigations showed that MSTI has cytotoxic activity on human breast carcinoma (MCF7) and human cervical carcinoma (HeLa) cell lines and improves the cytotoxic effects of cisplatin (26). These MSTI properties could be relevant for the treatment of cisplatin resistant tumoral cell lines. Interestingly, MSTI anticarcinogenic property indicates that antichymotryptic activity is not a strict requirement for antitumoral activity as often reported in the literature (27, 28).

Structural features, charge distribution and hydrophobicity of MSTI will be discussed in the present paper and compared to data available for other BBIs in order to elucidate common structural features of this class of proteins and to identify regions of the molecule relevant for its anticarcinogenic activity.

## MATERIAL AND METHODS

**Sample Preparation.** MSTI (MW 6926) was purified from *Medicago scutellata* seeds following the procedure previously reported (25). The stoichiometric ratio of MSTI:trypsin complex was evaluated on the pure MSTI sample following the procedure reported in reference (25).

NMR structural studies of MSTI were performed in aqueous and deuterated 20 mM sodium phosphate buffer at pH 5.6, in the temperature range 17–47 °C, and employing sample concentrations ranging from 0.6 to 2.0 mM, at 500 MHz. At concentrations lower than 1.2 mM, no amide

chemical shift change was observed, while at 2.0 mM amide shifts higher than 0.2 ppm were observed, probably due to self-association. NMR experiments performed at 500, 700, and 800 MHz, on 1.2 mM samples, were employed for structural characterization.

**NMR Spectroscopy.** NMR spectra were acquired on Bruker AVANCE 500, 700, and 800 MHz spectrometers.

Standard homonuclear DQF-COSY (29), TOCSY (30), and NOESY (31) experiments were recorded. Data matrixes contained  $4096 \times 1024$  points in  $f_2$  and  $f_1$ , respectively, for data acquired at 700 and 800 MHz, and  $2048 \times 512$  points for data acquired at 500 MHz. Water suppression was achieved with gradients by 3–9–19 pulse sequence (32) or using excitation sculpting sequence (33).

At 500 MHz, TOCSY experiments were acquired with two isotropic mixing times, 35 and 50 ms, while mixing times of 50 and 125 ms were employed for NOESY experiments.

At 700 and 800 MHz, TOCSY experiments were acquired with 50 ms isotropic mixing, and 60 and 125 ms mixing times were employed for NOESY experiments.

For amide exchange rate measurements, 12 sequential homonuclear 2D TOCSY spectra with 50 ms were acquired. The duration of each TOCSY experiment was 4 h.

Spectra were processed with XWINNMR. Analysis of spectra and cross-peaks volumes were performed using XEASY software (34).

All chemical shifts were referenced to the methyl resonance of 3-(trimethylsilyl)-propionic acid- $d_4$  sodium salt (TSP).

Secondary chemical shift maps were obtained using as reference the random coil chemical shifts reported by Wishart (35).

**Hydrogen Exchange Rates.** The hydrogen exchange rates ( $k_{ex}$ ) were determined by fitting cross-peak volumes to a first-order exponential decay,  $y = e^{-k_{ex}t}$ , where y represents the measured cross-peak volume of the resonances  $H^\alpha-H^N$  and t is the time in minutes (36). The time point for an experiment was taken to be the time elapsed from the addition of  $D_2O$  buffer to initiation of that NMR experiment plus half the duration of the acquisition. The cross-peak volumes were normalized to the cross-peak volume of the non exchangeable aromatic protons  $H_{2,6}/H_4$  of F56.

All the data were fitted with the program Sigmaplot (Jandel Scientific).

**Temperature Coefficient.** Temperature coefficients were measured for amide proton chemical shifts by analyzing TOCSY spectra acquired at 10° intervals from 17 to 47 °C. Chemical shifts versus temperature data were fitted to a linear equation with the program Sigmaplot (Jandel Scientific).

**Structure Calculations.** Calculated structures were obtained from restrained molecular dynamics simulations using DYANA (37) followed by energy minimization using the AMBER force field as implemented by the program DISCOVER (Molecular Simulations, San Diego, CA). Volume integration was performed on the NOESY spectrum acquired at 800 MHz with mixing time of 125 ms. The calibration of the peak volumes was performed using caliba, a routine of DYANA package, and the obtained list of distance restraints was used as input for DYANA calculations.

$\phi$  angle restraints were derived from  $^3J_{NH}$  constants estimated by the separation of extrema in the dispersive and absorptive plots of DQF-COSY spectra (38).  $\phi$  angle

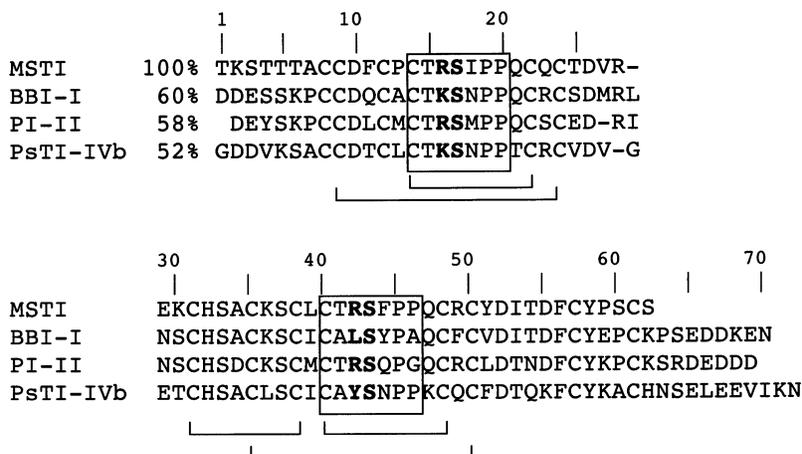


FIGURE 1: Amino acid sequence alignment of MSTI to three BBIs of known structure. Sequences of inhibitors from *Medicago Sculellata* (MSTI), soybean (BBI-I), tracy soybean (PI-II), and pea seed (PsTI-IVb) are shown. The percentage of sequence identity is reported. Residues in P1 and P1' position are shown in bold, and a box defines the active loop of the inhibitors. The pattern of disulphide bonds is drawn on each tandem region. BBIs present two additional disulfide bonds involving C8–C61 and C12–C57 (MSTI numbering), not shown in the figure.

restraints of  $-139^\circ \pm 30^\circ$  for  $^3J_{\text{HN-H}\alpha}$  coupling constants greater than 8.0 Hz and  $-60^\circ \pm 30^\circ$  for  $^3J_{\text{HN-H}\alpha}$  coupling constants smaller than 5.0 Hz were used for calculations.

Amide proton exchange rates were estimated from residual amide proton signals observed in a TOCSY spectrum recorded at 27 °C and pH 5.6 after the sample was exchanged through AMICON in deuterated phosphate buffer. Hydrogen bond formation or solvent exclusion from the amide protons was assumed to account for the slow and medium exchange rate amide protons. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints. Each hydrogen bond was introduced as O–N distance of 2.00 Å and HN–O distance of 3.00 Å.

One hundred calculations were run employing DYANA starting from random polypeptide conformations, and the 20 conformers with the lowest residual target function values were analyzed. The restraints were re-examined in view of consistent violations and relaxed where necessary. This procedure was repeated until no consistent violations were found in half or more of the structures. After this step 100 of new DYANA calculations were started, and the 20 structures with the lowest target functions were further refined using the AMBER force field. Each structure was minimized performing 100 steps of steepest descent and 300 steps of conjugate gradient. The 15 structures with the lowest potential energy were selected for further analysis. RMSD values were calculated over the range of residues showing medium and long-range distance restraints (i.e., between amino acids at positions  $i$  and  $j$ , with  $i - j \geq 2$ ), that is, between residue 6 and 61.

The INSIGHTII program (Molecular Simulations, San Diego, CA) was used to visually observe sets of structures, and the MOLMOL program (39) was employed for calculation and display of electrostatic potential surfaces.

## RESULTS

*Assignment of Spin Systems.* The assignment of MSTI spectra was not straightforward due to the presence of (i) two symmetrical motives, namely, C<sub>14</sub>TRSI**PP**QC<sub>22</sub> and

C<sub>40</sub>TRSF**PP**QC<sub>48</sub>, which represent the two tryptic inhibitory domains (Figure 1), and (ii) 14 cysteines, all involved in the following disulfide bridges C8–C61, C9–C24, C12–C57, C14–C22, C31–C38, C35–C50, and C40–C48 as previously determined (25). In addition the MSTI primary sequence presented 32 AMX spin systems that caused ambiguity in the first step of the residue assignment. Extensive overlap was observed at 500 and 700 MHz also for residues outside the two symmetric regions, and acquisition of NMR spectra at 800 MHz was necessary to complete the assignment of spin systems.

The starting points for the assignment were two isolated spin systems with amide chemical shift at low fields (11.52 and 11.54 ppm) corresponding to D10 and K36, respectively. The assignment of A7, showing  $\beta$  protons at high fields (0.50 ppm), and of some cysteine spin systems showing downfield shifted H $\alpha$  (C12, C14, C22, C24, C38, C40, C48, and C50) could be easily performed. All these cysteines showed H $\alpha$ –H $\alpha$  NOEs, useful starting points for the assignment of  $\beta$ -sheet regions of MSTI (see later). The assignment of the 500 MHz spectra at 27 °C was in a first step interrupted in the loop regions (residues 16–18 and 41–44) and at the level of residues Q21, Q23, K30, C31, H32, C35, Q47, F56, C61, and S62 due to extensive overlap in the amide region. To resolve these ambiguities, experiments were performed at different temperatures, in the range 17–47 °C, but the shifts induced were not sufficient to disentangle the overlapping signals.

At 700 MHz the overlapping signals were better resolved but only some hypothesis could be put forward relative to residues in the loop regions.

TOCSY and NOESY spectra were therefore performed at 800 MHz; many new NOEs could be unambiguously identified and the assignment could be completed. S17 and S43, located in the two homologous loops, exhibited completely overlapped spin systems which could be distinguished on the basis of the NOE correlations S17H $\beta$ –I18HN and S43H $\beta$ –F44HN. In fact, S43H $\beta$ –F44HN NOE, belonging to an intensely crowded region at 500 MHz, could be clearly identified at 800 MHz. Also at this field, S17HN–I18HN

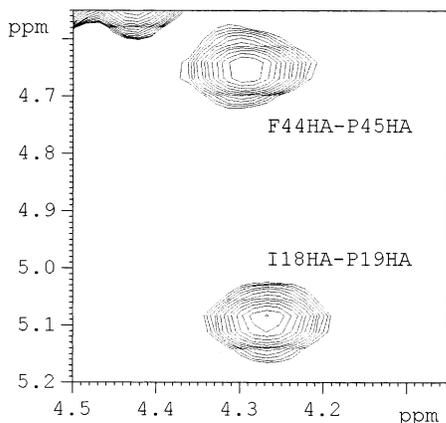


FIGURE 2: Selected region of NOESY spectrum of MSTI at 27 °C and pH 5.6 acquired at 800 MHz showing the NOE connectivities I18H $\alpha$ –P19H $\alpha$  and F44H $\alpha$ –P45H $\alpha$ .

and S43HN–F44HN cross-peaks could be assigned. R16 and R42 exhibited as well overlapping spin systems, except for the H $\alpha$  protons which could be identified only at 800 MHz. The assignment of C35, showing the HN–H $\alpha$  cross-peak superimposed to those of S43 and S17, was obtained at 800 MHz, on the basis of the identification of the following NOEs: C35H $\alpha$ –K36HN, C35HN–K36HN, and A34H $\beta$ 's–C35HN. At 800 MHz, H32 could be assigned since it was possible to identify H<sub>2</sub> and H<sub>4</sub> long-range NOEs with the side chains of Y58 and A34. The higher dispersion and sensitivity in the region 7.10–7.60 ppm and 8.10–8.80 ppm, at 800 MHz, permitted the assignment of K30 and C31, which were identified through the sequential NOEs K30H $\beta$ 's–HNC31 and K30H $\gamma$ 's–C31HN. At 800 MHz, the nearly degenerate H<sub>2,6</sub> and H<sub>3,5</sub> (7.37 and 7.35 ppm respectively) of F11 were unambiguously identified, together with the NH<sub>2</sub> resonances of all arginines, glutamines, and asparagines.

The observation of sequential  $d_{\alpha\delta}(i,i+1)$  and  $d_{\alpha\delta'}(i,i+1)$  (trans), or  $d_{\alpha\alpha}(i,i+1)$  (cis) connectivities in NOESY spectra

at short mixing times (60 ms) allowed the determination of the geometry of the X-Pro bonds. On these basis, I18–P19 and F44–P45 were shown to have cis geometry (Figure 2), while all the remaining X-Pro bonds were trans.

The complete assignment, as obtained at 800 MHz, is available as Supporting Information and has been deposited in BMRB (id code: BMRB–5617).

The assignments were tested using the automated assignment algorithm NOAH (40). The long-range ambiguous NOEs were peaked without assignment and used as input in a further NOAH run. The assignment suggested by NOAH was then analyzed, and a few reasonable NOESY peak assignments were added.

The assignment obtained at 800 MHz at 27 °C in turn allowed the assignment of the spectra acquired at 500 MHz at different temperatures (17, 37, 47 °C) and of the spectra acquired at a higher protein concentration (2 mM) at 27 °C (see later).

**Secondary Structure.** A summary of NOE connectivities and  ${}^3J_{\text{HN-H}\alpha}$  coupling constants is reported in Figure 3, while secondary chemical shifts are reported in Figure 4. Conformational shift maps are in good agreement with the secondary structure determined on the basis of NOEs. Two triple stranded  $\beta$ -sheets, located at the level of residues 11–15 (A strand), 21–25 (B strand), 53–55 (F strand), and 27–29 (C strand), with 37–41 (D strand) 47–51 (E strand) defining the first and second domain, respectively, could be identified on the basis of secondary shifts, canonical nonsequential  $d_{\alpha\alpha}(i,j)$ ,  $d_{\text{NN}}(i,i)$  and  $d_{\alpha\text{N}}(i,j)$  NOEs and  ${}^3J_{\text{HN-H}\alpha}$  couplings (Figure 5). Each  $\beta$ -sheet is characterized by a  $\beta$ -hairpin made of five-residue pairs of antiparallel  $\beta$ -sheets and a five-residue tight type VIb turn, together with an additional short strand forming a triple stranded  $\beta$ -sheet.

Each type VIb turn includes a P1–P1' reactive site conferring the tryptic inhibitory activity, corresponding to peptide bonds R16–S17 and R42–S43, and contains a cis peptide X-Pro bond (I18–P19 for the first  $\beta$ -sheet and



FIGURE 3: Summary of short and medium range upper distance limits and  ${}^3J_{\alpha\text{N}}$  values. The thickness of the bars indicates the relative normalized intensity of the NOE.  ${}^3J_{\alpha\text{N}}$ 's are three-bond coupling constants between H $\alpha$  and HN, where the symbols represent the following:  $\blacktriangledown$  for  ${}^3J_{\alpha\text{N}} > 8$  Hz,  $\blacktriangle$  for  ${}^3J_{\alpha\text{N}} < 5$  Hz. In the row labeled  $k_{\text{NH}}$ , filled circles identify residues which present slow amide proton exchange rates (see Table 1).

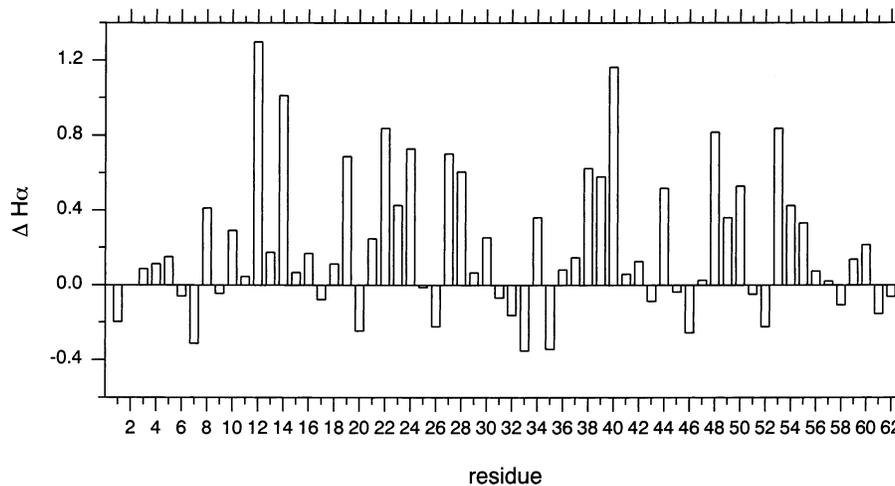


FIGURE 4: Secondary H $\alpha$  chemical shifts of MSTI at 27 °C and pH 5.6.

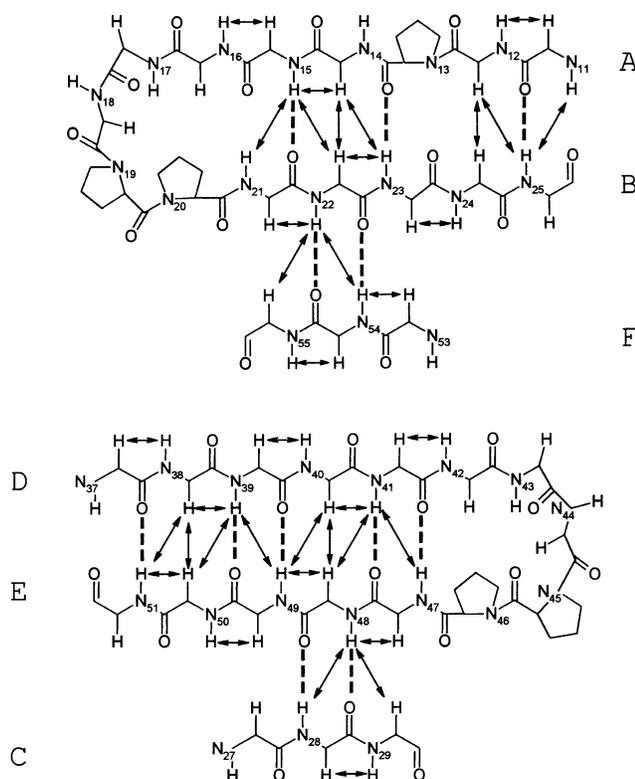


FIGURE 5: Secondary structure contacts and hydrogen bonds for MSTI at 27 °C and pH 5.6.  $\beta$ -structure elements A–F are indicated together with observed inter-strand connectivities (arrows). Dotted lines indicate H-bonds present in the final calculated structures.

F44–P45 for the second  $\beta$ -sheet) that is required for inhibitory activity (14). The following observations are consistent with a type VIb turn: (i) coupling constants  $^3J_{\text{HN-H}\alpha} \leq 5$  Hz were measured for residues T15 and T41; (ii) strong NOEs  $d_{\text{ON}}(i, i+3)$  were observed between P20H $\alpha$  and S17HN, in the first active loop, and between P46H $\alpha$  and S43HN in the second active loop (41). The presence of two turn conformations was also confirmed by negative secondary shifts for residues S17 and P20 in the first active loop and S43 and P46 in the second one (Figure 4).

A few half-turns could be identified at the level of residues 24–26 (N-terminal  $\beta$ -sheet) and 50–52 (C-terminal  $\beta$ -sheet) on the basis of C24H $\alpha$ –D26HN and C50H $\alpha$ –D52HN NOEs complemented by negative secondary shifts observed for D26

and D52 (Figures 3 and 4). The presence of H32H $\alpha$ –A34HN and S33HN–A34HN NOEs suggests that the polypeptide chain is bent, between the two domains, in a type I-like turn.

**Amide Exchange Rates and Temperature Coefficients.** Hydrogen/deuterium exchange experiments were performed at 27 °C and pH 5.6 in order to gain information on the hydrogen-bonded amides. The fitting of the measured H $\alpha$ –HN cross-peak volumes at different times after addition of D<sub>2</sub>O allowed the estimate of amide exchange rates ( $k_{\text{ex}}$ ) for 14 residues (Table 1). The results are consistent with the  $\beta$ -sheet structure identified on the basis of NOE connectivities: all amide protons of residues located in the strand regions show low  $k_{\text{ex}}$  with the exception of T15, T41, and Q47.

Amide temperature coefficient and amide exchange rate are, to some extent, complementary tools to predict hydrogen bond donors, and a combined use of the two approaches produces a far more reliable indicator of hydrogen bonding than either alone (42). Temperature coefficients are easily measured for all residues, do not display a pH dependence as amide exchange, and are not strongly influenced by surface exposure as exchange rates. A general criterion is that all of the amide protons displaying slow exchange rates and temperature coefficients (TC) more positive than  $-4.5$  ppb/K could be located in hydrogen-bonded regular secondary structures. Chemical shifts variations of MSTI as a function of temperature are reported in Table 1.

Exchange rates and TC are in good agreement. The criterion of TC more positive than  $-4.5$  ppb/K (within experimental error) identifies more amide protons involved in H-bond, some of them possibly corresponding to exposed H-bonded residues. Only two residues, namely, C48 and T54, show low  $k_{\text{ex}}$  and TC around 6 ppb/K, and they could be involved in less stable H-bonds stabilizing the short third strand.

The indication of the presence of H-bonds, as identified by the two techniques, was exploited for structure calculations. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints, as described in Material and Methods.

**Tertiary Structure.** A summary of experimental restraints used for structure calculations is reported in Table 2. A set of 726 NOEs was supplemented by (i) 21 distance restraints for the 7 disulfide bridges, (ii) 22 distance restraints for 11

Table 1: Temperature Coefficients and Amide Exchange Rates for MSTI at pH 5.6<sup>a</sup>

residue	NH $\Delta\delta/\Delta T$	$k_{\text{ex}}$ ( $10^{-5} \text{ min}^{-1}$ )
S3	-6.0 ± 2.0	
T4	-8.7 ± 0.5	
T5	-6.5 ± 0.9	
T6	<b>-2.9</b> ± 2.9	
A7	-8.6 ± 0.5	
C8	-5.5 ± 0.9	
C9	<b>-3.7</b> ± 0.5	154 ± 12
D10	<b>-4.2</b> ± 0.5	
F11	<b>-2.7</b> ± 0.6	46 ± 2
C12	-5.8 ± 0.4	
C14	-8.8 ± 0.6	
T15	<b>-4.2</b> ± 0.6	
R16/R42	-7.7 ± 0.5	
S17/S43	<b>-1.1</b> ± 0.8	
I18	-9.8 ± 0.6	
Q21	<b>-2.6</b> ± 0.6	196 ± 17
C22	<b>-5.0</b> ± 0.6	36 ± 12
Q23	<b>-0.9</b> ± 0.5	123 ± 5
C24	<b>-2.7</b> ± 0.6	
T25	<b>-3.7</b> ± 0.3	154 ± 8
D26	<b>-1.0</b> ± 0.7	
V27	-7.9 ± 0.7	
R28	<b>-5.0</b> ± 0.3	11.9 ± 0.3
E29	<b>-3.1</b> ± 0.6	
K30	<b>-3.7</b> ± 0.7	
C31	-10.0 ± 1.0	
H32	-6.1 ± 0.6	
S33	-6.9 ± 0.8	
A34	<b>-3.1</b> ± 0.8	
C35	<b>-2.6</b> ± 0.7	
K36	<b>-4.6</b> ± 0.5	
S37	<b>-2.3</b> ± 0.5	135 ± 5
C38	-9.2 ± 0.9	
L39	<b>-3.0</b> ± 0.6	76 ± 3
C40	-9.0 ± 0.7	
T41	<b>-4.8</b> ± 0.7	
R42/R16	-7.7 ± 0.4	
S43/S17	<b>-1.1</b> ± 0.8	
F44	-10.0 ± 0.7	
Q47	<b>-1.8</b> ± 0.1	
C48	-6.2 ± 0.3	118 ± 10
R49	<b>-1.5</b> ± 0.2	8 ± 1
C50	<b>-4.9</b> ± 0.6	
Y51	<b>-4.5</b> ± 0.5	126 ± 11
D52	<b>-1.1</b> ± 0.6	27 ± 10
I53	-7.5 ± 0.9	
T54	-5.8 ± 0.6	16 ± 1
D55	<b>-2.9</b> ± 0.3	
F56	<b>-2.7</b> ± 0.9	
C57	-7.7 ± 0.3	
Y58	-15.4 ± 0.3	
S60	-8.9 ± 0.6	
C61	<b>-1.1</b> ± 0.8	
S62	-6.4 ± 0.9	

<sup>a</sup> Temperature coefficients values lower than 5.0 ppb/K are shown in bold. Amide exchange rates were measured at 27 °C.

backbone hydrogen bonds defined on the basis of deuterium hydrogen exchange studies and temperature coefficient values, and (iii) 29  $\phi$  angle restraints derived from  $^3J_{\text{N}\alpha}$  coupling constants. The superposition of the 15 best structures is reported in Figure 6a, and the structures have been deposited in the PDB (PDB code 1MVZ). The structure of MSTI comprises two distinct domains composed by a three stranded  $\beta$ -sheet and the active site, located in a VIB type loop. The combination of the tight turn and the antiparallel  $\beta$ -sheet forms the characteristic  $\beta$ -hairpin inhibitory domain common to other Bowman Birk inhibitors (Figure 6b). The first domain comprises residues 11–25 and 53–55, while

Table 2: Structural Statistics of the 15 Best Structures Obtained for MSTI

(a) Restraints	
total number of distance restraints	726
number of intraresidual restraints	257
number of sequential restraints	264
number of medium-range restraints	22
number of long-range restraints	183
number of torsion angle restraints	29
number of hydrogen bonds	11
(b) DYANA	
20 Structures	
target function ( $\text{\AA}^2$ )	1.84 ± 0.19
$\langle \text{rmsd} \rangle_{\text{bb}}$ (6–61) ( $\text{\AA}$ )	0.75 ± 0.18
$\langle \text{rmsd} \rangle_{\text{heavy}}$ (6–61) ( $\text{\AA}$ )	1.51 ± 0.17
(c) DISCOVER	
15 Structures	
total potential energy (kcal/mol)	-115 ± 10
$\langle \text{rmsd} \rangle_{\text{bb}}$ (6–61) ( $\text{\AA}$ )	0.84 ± 0.09
$\langle \text{rmsd} \rangle_{\text{heavy}}$ (6–61) ( $\text{\AA}$ )	1.60 ± 0.12
$\langle \text{rmsd} \rangle_{\text{bb}}$ (11–15, 21–25, 53–55) ( $\text{\AA}$ )	0.40 ± 0.05
$\langle \text{rmsd} \rangle_{\text{bb}}$ (37–41, 47–51, 27–29) ( $\text{\AA}$ )	0.57 ± 0.09

the second involves residues 27–29 and 37–51. The  $\beta$ -hairpins are planar and extended, and each strand is held in place by the disulfide cross-link across the sheet, namely, C14–C22 for the first domain and C40–C48 for the second domain. The mentioned disulfide bridges are functionally relevant since they hold the two peptide chains together after P1–P1' peptide bond is cleaved by the protease (1). The other disulfide bridges connect residues belonging to one domain with residues at the border of the same domain (C9–C24 and C12–C57 for the first domain and C31–C38 and C35–C50 for the second one) or connect two residues at the border of the same domain (C8–C61). The regions of antiparallel  $\beta$ -sheets are quite well defined with  $\langle \text{rmsd} \rangle_{\text{bb}}$  (11–15, 21–25, 53–55) = 0.40 ± 0.05 Å for the first domain and  $\langle \text{rmsd} \rangle_{\text{bb}}$  (37–41, 47–51, 27–29) = 0.57 ± 0.09 Å for the second domain. The type VIB turns are less defined, due to the lack of NOEs especially for residues R16–S17 and R42–S43 which only show intra and sequential NOEs. The mentioned peptide bonds correspond to reactive sites and a hyper-exposed P1 residue (R in both domains) is a characteristic and important functional feature of the canonical conformation (43). The salt bridge between the basic P1 of the substrate with the acidic D189 in the interior of the S1 pocket of trypsin is responsible of the narrow preference of this enzyme for basic residues in position P1 (1).

Twelve backbone H-bonds typical of  $\beta$  structure were present in all the final structures (Figure 5). T15 and T41 amides are hydrogen bonded to backbone oxygen of residues Q21 and Q47, respectively. The two amides are located close to the first and second active sites, in a symmetrical position. They show NOEs typical of  $\beta$ -sheet structure, temperature coefficients lower than -5 ppb/K, but their hydrogen exchange rates were not measurable at pH 5.6. The fast exchange of the two amides underlines the exposure of the inhibitory region and the low stability of the H-bond pattern close to the active loops.

The amide proton of Q21 (P5' position), belonging to the first domain, forms H-bond with the side-chain oxygen ( $\text{O}\gamma$ ) of T15 (P2 position) in all the structures. Q47 (P5' position of the second domain) presents a low-temperature coefficient but a fast  $k_{\text{ex}}$ , and its amide forms H-bonds with  $\text{O}\gamma$  of T41 (P2) only in 6 out of 15 structures. In the remaining struc-

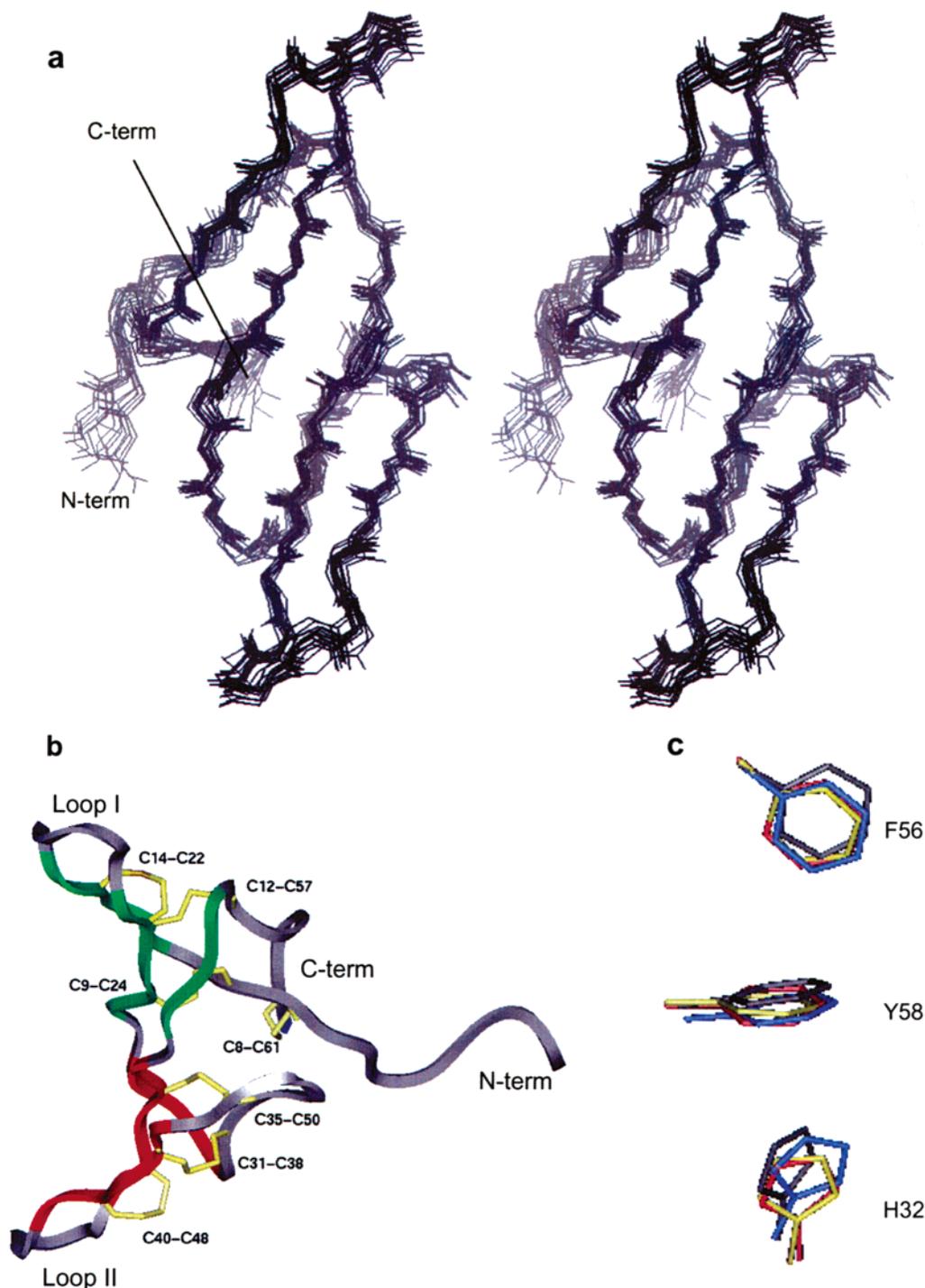


FIGURE 6: Stereoview of backbone superposition of the 15 final structures of MSTI. Superposition was performed in the range of residues 6–61 (a). Lowest-energy MSTI structure drawn in ribbon representation. Residues 11–25 and 53–55, corresponding to the first three-stranded  $\beta$ -sheet, are colored in green; residues 27–29 and 37–51, corresponding to the second three-stranded  $\beta$ -sheet, are colored in red. The active loops, the disulfide bridges and N and C-termini are indicated (b). Superposition of residues belonging to the aromatic cluster present in all BBIs of known structure (MSTI, gray; BBI-I, yellow; PI-II, red; PsTI-IVb, blue) (c).

tures, the side-chain oxygen of T41 was involved in H-bond with the amide proton or the side-chain oxidrile of S43 (P1'). Both Q47 and S43 amides have low-temperature coefficients but do not show a significant protection from exchange. A possible interchange of donors for O $\gamma$  T41 can be hypothesized, suggesting the presence of higher mobility in the second domain loop region with respect to the first one.

The lack of inter-domain disulfide bridges (all disulfide bridges are indeed localized either within a domain or between a residue of one domain and a residue at the border

of the same domain) (see Figure 6) allows for flexibility of the spatial orientation of the two domains relative to each other, as reflected by the lack of NOEs between the two domains. These observations underline the double headed feature of this inhibitor, requiring mobility of one domain relative to the other for its precise adjustment during the simultaneous inhibition of two trypsin molecules (2).

*Self-Association.* Spectra of MSTI were performed at different protein concentrations ranging from 0.6 to 2.0 mM in order to characterize the self-association behavior. The

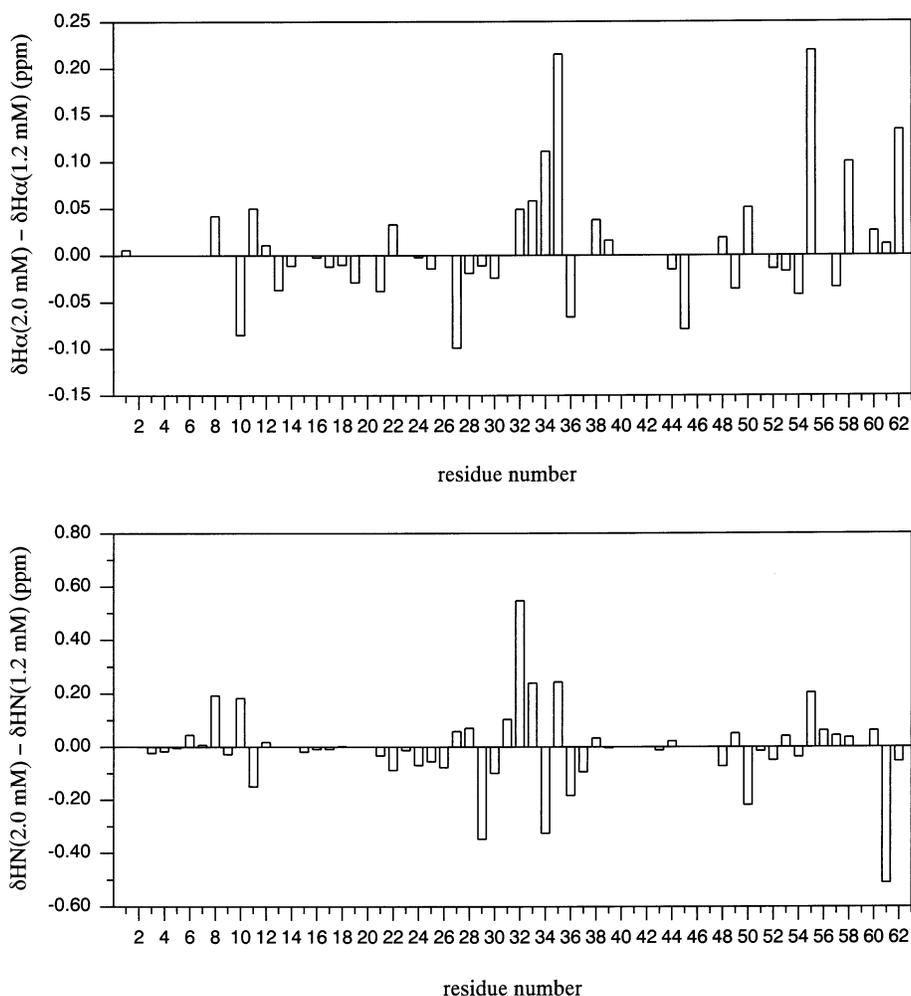


FIGURE 7:  $H\alpha$  (upper panel) and  $HN$  (lower panel) chemical shift differences measured for 1.2 and 2.0 mM MSTI samples.

assignment of the spectra at different protein concentrations was easily extended from that obtained for 1.2 mM sample. The comparison of  $H\alpha$  conformational shifts at 1.2 mM and 2.0 mM indicates that the secondary structure of the domains is preserved upon association.

At 2.0 mM protein concentration a few amide,  $\alpha$ , and side-chain protons shifts were observed (Figure 7), clearly due to self-association. The residues showing amide shifts greater than 0.1 ppm and/or  $H\alpha$  or side-chain hydrogen shifts greater than 0.05 ppm are listed: C8, D10, F11, V27, R28, E29, K30, C31, H32, S33, A34, C35, K36, S37, C48, C50, D55, F56, C57, C61. It is clear that all residues belonging to the loop connecting the two domains are affected by aggregation together with some residues located in the N and C-terminal regions and in the C and E strands of the second domain.

## DISCUSSION

Despite extensive studies on inhibitory and anticarcinogenic activity of BBIs, only few three-dimensional structures have been solved, namely, those of tracy soybean (5) (PI-II, PDB codes 1PI2), soybean (7, 8) (BBI-I, PDB codes 2BBI and 1K9B), and pea seeds (2) (PsTI-IVb, PDB code 1PBI) inhibitors.

The topology of MSTI, as determined by  $^1H$  NMR spectroscopy and molecular dynamics calculations, is common to all known Bowman Birk inhibitors and comprises

two distinct domains each composed by a triple stranded  $\beta$ -sheet and the active site, located in a VIb type turn. The structural comparison of MSTI with all BBIs of known structure performed through the VAST search service (<http://www.ncbi.nlm.nih.gov/Structure/VAST>) indicated that PsTI-IVb has the highest structural similarity, exhibiting an RMSD for alignable regions of 1.3 Å. The RMSD with other BBIs was slightly higher and comprised between 1.5 and 1.7 Å. The presence of an unusual buried cluster of charged side-chains, located in the inter-domain region, has been identified in the X-ray structure of BBI-I (8, 44). The cluster of charged side-chains is also present in MSTI, involving D26, R28, and H32 residues, while the same geometry cannot be established in PI-II and PsTI-IVb due to local sequence mutations at the level of R28.

Interestingly, three aromatic residues, highly conserved in all BBIs (Figure 1), namely, H32, F56, and Y58, form in MSTI an aromatic cluster showing a typical geometry in which pairs of aromatic residues are almost orthogonal to each other and with centroids at distances minor than 7 Å. This geometry is known to confer thermostability to many proteins (45), in agreement with the observation that MSTI shows high-temperature stability. The aromatic cluster is present in all the other BBIs of known structure (Figure 6c) conferring, together with the H-bond network and the seven disulfide bridges, a high stability to this class of inhibitors.

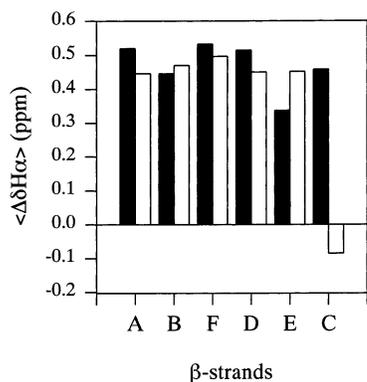


FIGURE 8: Comparison of average secondary chemical shifts of MSTI (black bars) and BBI-I (white bars). The results are reported for strands belonging to the first (A, B, and F) and the second domain (D, E, and C).

A comparison of NMR structural data can be performed between MSTI and BBI-I, sharing the highest sequence similarity and whose NMR solution structure has been previously determined (46).  $H_{\alpha}$  secondary shifts show a similar behavior for the two proteins with only few exceptions attributed to the mutation of the residue itself or of its neighbors. It is, however, interesting to note that the main difference occurs at the level of the third strand of the second domain (C strand). The average  $H_{\alpha}$  secondary shifts ( $\langle \Delta H_{\alpha} \rangle$ ) within each strand were calculated for MSTI and BBI-I and reported in Figure 8. BBI-I C strand shows a negative  $\langle \Delta H_{\alpha} \rangle$ , suggesting that this strand has a very low stability. Indeed,

the authors stated that the second domain (antichymotryptic) presents fewer stabilizing tertiary interactions with respect to the first one. The two  $\beta$ -sheets of MSTI display comparable stability on the basis of  $H_{\alpha}$  secondary shifts, a few differences were, however, evidenced in the active loop regions based on the analysis of the 15 final structures. Our data indicate that in the first domain a hydrogen bond between the hydroxyl group of threonine, in P2 position of the active loop, and the amide proton of residue in P5' position is formed in all the structures. In the second domain, this H-bond is present only in six structures out of 15 suggesting a higher mobility. Threonines in P2 position (T15 and T41) are highly conserved in BBIs, implying that these residues are functionally important for inhibitory properties.

It is reported that when BBIs interacts with the enzyme, the intramolecular hydrogen bond involving T15  $O_{\gamma}$  stabilizes the active loop also favoring the correct positioning of T15 in order to maximize the hydrophobic interactions with H57 and L99 residues of trypsin (11). Further studies on amino acid variations in P2 position of cyclic peptides, encompassing the reactive site loops, demonstrated that a threonine confers both the lowest dissociation constant and the slowest hydrolysis rate (47), as expected for a good inhibitor. All these data underline the utmost importance of P2 threonine, whose side chain can provide both hydrophobic interactions via its methyl group and H-bonds via the hydroxyl group.

Protease recognition mechanism is triggered by the distribution of hydrophilic and hydrophobic residues in the

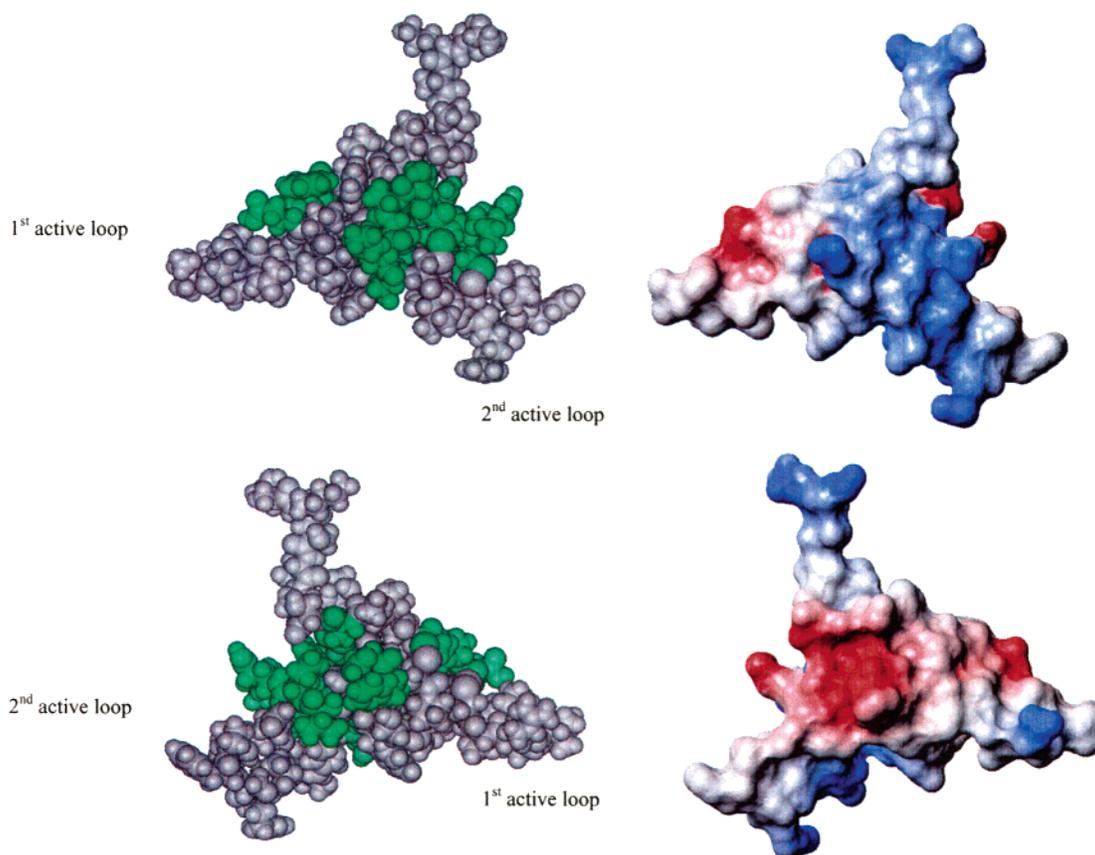


FIGURE 9: Self-association interfaces and electrostatic potential surfaces of MSTI. Two views of MSTI are offered in the lower and upper panel (one rotated by 180 degrees with respect to the other about z-axis). Residues affected by self-association are colored in green on CPK representation of the molecule (left panel); electrostatic potential surfaces were calculated by MOLMOL program (39) (right panel). The location of the first and second active loop in the two views are indicated.

domain regions. Hydrophilic bridges and conformational rigidity are characteristic of narrow specificity against trypsin, while high hydrophobicity and significant conformational mobility, typical of antichymotryptic domains, confer a broad inhibition specificity (11). Indeed, it is reported that BBI-I second domain is able to interact either with chymotrypsin or with trypsin (4). Polar and charged side chains are predominant in both domains of MSTI as expected for trypsin inhibition specificity, anyway the second domain presents a higher hydrophobicity. As an example the proline residue, in P4 position of the first domain, is replaced in the second domain by the longer hydrophobic side chain of L39. The same change takes place in BBI-I, where P4 position of the first domain (anti-tryptic) is a short hydrophobic side-chain A13, replaced in the anti-chymotryptic by an isoleucine (I40). This observation suggests a low specificity of MSTI second domain for trypsin, supported as well by the structural results pointing at its lower conformational rigidity. In agreement with our hypothesis, titration studies have shown a stoichiometric ratio MSTI:trypsin of 1:1.5. Moreover, the same  $K_d$  value of  $1 \times 10^{-10}$  M was measured for MSTI (25), presenting two antitryptic domains, and for BBI-I, presenting one antitryptic and one antichymotryptic domain (11).

Preliminary investigations have shown that MSTI has cytotoxic activity, reducing of about 20% the cellular survival of both MCF7 and HeLa, and improves the cisplatin effect on a few tumoral cell lines. (26). Since MSTI presents two antitryptic domains, these data indicate that antichymotryptic activity is not a strict requirement for antitumoral effect, as often reported in the literature. In addition it has been shown that inhibitory activity and antitumoral activity are not correlated. Indeed, after cleavage of BBI-I and linearization of the two resulting peptides, carrying the chymotrypsin and trypsin inhibitory sites, respectively, both fragments, lacking inhibitory activity, act radio-protectively (48). In agreement with these results, a recent analysis of a few nona-peptides, reproducing the second active loop of BBI-I, suggested that radio-protective and inhibitory properties are inherent to different residues (49). In this line we propose that the less rigid second domain of MSTI might bear some relevance relative to the recognition mechanism and to the antitumoral activity.

We further analyzed self-association behavior of MSTI collecting NMR data at different protein concentrations. The residues affected by self-association are localized in two well defined regions on opposite faces of the molecule (Figure 9, left panel). The calculation of the electrostatic potential surfaces shows that these two faces present an opposite potential (Figure 9, right panel), and we suggest that electrostatic interactions between them may play an important role in MSTI association. Residues involved in MSTI self-association, located at the border of the domains and in C and E strands (second domain), are highly conserved in BBIs from different seeds suggesting a functional relevance of the two superficial patches. The presence of self-association has also been reported for PsTI-IVb (2), PI-II (5) and BBI-I (7), but the structural data for the dimer have been reported only for PsTI-IVb. Its dimer interface comprises an extensive hydrogen-bond network together with hydrophobic contacts among residues belonging to both domains and to the longer C-terminal tail. It is important to stress that different residues

are involved in self-association of MSTI and PsTI-IVb. Residues involved in PsTI-IVb self-association are not conserved in BBI family, at variance with those of MSTI. From these data we infer that residues of MSTI affected by self-association may represent those regions of the protein capable of interacting with other enzymes or macromolecules, which could be of the highest importance in antitumoral activity.

The identification of association/interaction interfaces will be further investigated to correlate the association properties of MSTI with its antitumoral activity.

## ACKNOWLEDGMENT

We gratefully acknowledge the technical support offered by the Large Scale Facility of Firenze (CERM) for the acquisition of NMR spectra at 700 and 800 MHz. The authors thank A. Lanza and R. Iori for helpful discussions and F. Greco and G. Zannoni for technical assistance.

## SUPPORTING INFORMATION AVAILABLE

List of proton chemical shifts for MSTI at 27 °C and pH 5.6. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES

- Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem.* 49, 593–626.
- Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J., and Brunie, S. (1999) *J. Mol. Biol.* 285, 1195–207.
- Odani, S., and Ikenaka, T. (1978) *J. Biochem.* 83, 747–53.
- Jensen, B., Unger, K. K., Uebe, J., Gey, M., Kim, Y. M., and Flecker, P. (1996) *J. Biochem. Biophys. Methods* 33, 171–85.
- Chen, P., Rose, J., Love, R., Wei, C. H., and Wang, B. C. (1992) *J. Biol. Chem.* 267, 1990–1994.
- Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S., and Ikenazaka, T. (1993) *J. Mol. Biol.* 234, 722–734.
- Werner, M. H., and Wemmer, D. E. (1992) *Biochemistry* 31, 999–1010.
- Voss, R. H., Ermler, U., Essen, L. O., Wenzl, G., Kim, Y. M., and Flecker, P. (1996) *Eur. J. Biochem.* 242, 122–31.
- Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S., and Takahashi, K. (1986) *J. Biochem.* 100, 1637–1646.
- Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur. J. Biochem.* 212, 549–555.
- Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J. Mol. Biol.* 298, 477–491.
- Lockett, S., Garcia, R. S., Barker, J. J., Konarev, A. V., Shewry, P. R., Clarke, A. R., and Brady, R. L. (1999) *J. Mol. Biol.* 290, 525–33.
- Brauer, A. B., Kelly, G., McBride, J. D., Cooke, R. M., Matthews, S. J., and Leatherbarrow, R. J. (2001) *J. Mol. Biol.* 306, 799–807.
- Brauer, A. B. E., Domingo, G. J., Cooke, R. M., Matthews, S. J., and Leatherbarrow, R. J. (2002) *Biochemistry* 41, 10608–10615.
- Kennedy, A. R. (1998) *Am. J. Clin. Nutr.* 68, 1406S–1412S.
- Miyata, S., Koshikawa, N., Yasumitsu, H., and Miyazaki, K. (2000) *J. Biol. Chem.* 275, 4592–4598.
- Sorsa, T., Salo, T., Koivunen, E., Tyynelä, J., Kontinen, Y. T., Bergmann, U., Tuuttila, A., Niemi, E., Teronen, O., Heikkilä, P., Tschesche, H., Leinonen, J., Osman, S., and Stenman, U. H. (1997) *J. Biol. Chem.* 272, 21067–21074.
- Miyata, S., Miyagi, Y., Koshikawa, N., Nagashima, Y., Kato, Y., Yasumitsu, H., Hirahara, F., Misugi, K., and Miyazaki, K. (1998) *Clin. Exp. Metastasis* 16, 613–622.
- Ware, J. H., Wan, X. S., Rubin, H., Schechter, N. M., and Kennedy, A. R. (1997) *Arch. Biochem. Biophys.* 344, 133–138.
- Ware, J. H., Wan, X. S., and Kennedy, A. R. (1999) *Nutr. Cancer* 33, 174–177.
- Billings, P. C., St. Clair, W., Owen, A. J., and Kennedy, A. R. (1998) *Cancer Res.* 48, 1798–17802.

22. Dittmann, K. H., Gueven, N., Mayer, C., Ohneseit, P., Zell, R., Begg, A. C., and Rodemann, H. P. (1998) *Radiat. Res.* 150, 648–655.
23. Dittmann, K. H., Gueven, N., Mayer, C., and Rodemann, H. P. (1998) *Int. J. Radiat. Biol.* 74, 225–230.
24. St. Clair, W. H., Billings, P. C., and Kennedy, A. R. (1990) *Cancer Lett.* 52, 145–152.
25. Ceciliani, F., Tava, A., Iori, R., Mortarino, M., Odoardi, M., and Ronchi, S. (1997) *Phytochemistry* 44, 393–398.
26. Lanza, A., Tava, A., Zetta, L., Di Francesco, S., Robustelli della Cuna, F. S., and Robustelli della Cuna G. (2001) *Proc. XXVII Congresso Nazionale Oncolog.*, 151.
27. Kennedy, A. R. (1998) *Am. J. Clin. Nutr.* 68, 1406–1412.
28. Kennedy, A. R., and Wan, X. S. (2002) *Prostate* 50, 125–133.
29. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc.* 104, 6800–6801.
30. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* 66, 355–360.
31. Sklenar, V., Piotto, M., Leppik, R., and Sudek, V. (1993) *J. Magn. Reson., Ser A* 102, 241–245.
32. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* 2, 661–665.
33. Hwang, T. L., and Shaka, A. J. J. (1995) *J. Magn. Reson. ser A* 112, 275–279.
34. Bartels, C., Xia, T., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biom. NMR* 5, 1–10.
35. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR* 5, 67–81.
36. Wagner, G., and Wüthrich, K. (1982) *J. Mol. Biol.* 160, 343–361.
37. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* 273, 283–298.
38. Kim, Y., and Prestegard, J. H. (1989) *J. Magn. Reson.* 84, 9–13.
39. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* 14, 51–5, 29–32.
40. Mumenthaler, C., Güntert, P., Braun, W., and Wüthrich, K. (1997) *J. Biomol. NMR* 10, 351–362.
41. Richardson, J. S. (1981) *Adv. Protein Chem.* 34, 167–339.
42. Baxter, N. J., and Williamson, M. P. (1997) *J. Biomol. NMR* 9, 359–369.
43. Apostoluk, W., and Otlewski, J. (1998) *Proteins* 32, 459–474.
44. Bode, W., Greyling, H. J., Huber, R., Otlewski, J., and Wilusz, T. (1989) *FEBS Lett.* 242, 285–292.
45. Burley, S. K., and Petsko, G. A. (1985) *Science* 229, 23–28.
46. Werner, M. H., and Wemmer, D. E. (1991) *Biochemistry* 30, 3356–3364.
47. McBride, J. D., Brauer, A. B., Nievo, M., and Leatherbarrow, R. J. (1998) *J. Mol. Biol.* 282, 447–458.
48. Gueven, N., Dittmann, K., Mayer, C., and Rodemann, H. P. (1998) *Cancer Lett.* 125, 77–82.
49. Dittmann, K. H., Gueven, N., Mayer, C., and Rodemann, H. P. (2001) *Protein Eng.* 14, 157–160.

BI020576W

## Effects of the *Medicago scutellata* Trypsin Inhibitor (MsTI) on Cisplatin-induced Cytotoxicity in Human Breast and Cervical Cancer Cells

ANNALISA LANZA<sup>1</sup>, ALDO TAVA<sup>2</sup>, MADDALENA CATALANO<sup>3</sup>,  
LAURA RAGONA<sup>3</sup>, ILARIA SINGUAROLI<sup>1</sup>,

FRANCESCO SAVERIO ROBUSTELLI DELLA CUNA<sup>4</sup> and GIOACCHINO ROBUSTELLI DELLA CUNA<sup>1</sup>

<sup>1</sup>Divisione di Oncologia Medica, Fondazione Salvatore Maugeri, via Ferrata 8, 27100 Pavia;

<sup>2</sup>Istituto Sperimentale per le Colture Foraggere, viale Piacenza 29, 26900 Lodi;

<sup>3</sup>Istituto per lo Studio delle Macromolecole, Laboratorio NMR, CNR, via Ampère 56, 20131 Milano;

<sup>4</sup>Servizio di Farmacia, Fondazione Salvatore Maugeri, via per Montescano 4, 27040 Montescano, Italy

**Abstract.** Background: Snail medic (*Medicago scutellata* L.) seeds exhibit a significantly higher content of a trypsin inhibitor than other *Medicago* species. This inhibitor belongs to the Bowman-Birk family of serine protease inhibitors (BBI) and exhibits a good sequence homology with the BBI from soybean, while presenting some differences. It has been suggested that BBIs have antitumoral and radio-protective activity. Materials and Methods: In order to assess whether the inhibitor from *Medicago scutellata* (MsTI) seeds show similar properties to those of BBI from soybean with respect to potentiation of cisplatin-induced cytotoxicity, we evaluated the effects of MsTI on cisplatin-induced cell killing in MCF7 human breast carcinoma cells and HeLa human cervical carcinoma cells. Results: The 24-hour treatment of MsTI in the cell culture medium decreased the clonogenic survival of MCF7 and HeLa cells in a dose-dependent manner and enhanced cisplatin-induced cytotoxicity. The presence of MsTI during the entire incubation period reduced the  $D_{37}$  of cisplatin by 40% in both the cell lines. Conclusion: MsTI could be an useful agent for the potentiation of cisplatin-mediated cancer treatment.

Naturally occurring proteinase inhibitors are low molecular weight proteins that can repress the activity of proteinases by binding into the active site of the enzyme. In plants a high concentration of protease inhibitors are found in tissues that are particularly vulnerable to foreign proteases

Correspondence to: Dr. Annalisa Lanza, Divisione di Oncologia Medica, Fondazione Salvatore Maugeri, via Ferrata 8, 27100 Pavia, Italy. Tel: +39-0382-592060, Fax: +39-0382-592094, e-mail: alanza@fsm.it

Key Words: Serine protease inhibitor, *Medicago scutellata*, cytotoxicity.

such as seeds, tubers and bulbs, particularly in *Gramineae*, *Leguminosae* and *Solanaceae* (1). Seeds from *Leguminosae* contain a number of serine proteinase inhibitors which belong to at least two distinct inhibitor families: the Kunitz and the Bowman-Birk trypsin inhibitor (2). The possible role suggested for seed proteinase inhibitors includes the control of endogenous proteinases in the metabolism of storage proteins and during seed dormancy (3). Moreover, serine proteinase inhibitors appear to function as a part of the plant defensive system against pests via inhibition of their proteolytic enzymes (4). Finally, due to their high content of cysteine residues, they may also be seen as storage proteins, although present in small amount, which are degraded during germination and seedling growth (5).

The trypsin inhibitor from *M. scutellata* seeds (MsTI) was extracted, purified and its aminoacid sequence and inhibitory activity reported (6). The complete aminoacid sequence of MsTI, exactly defined by Catalano *et al.* (7), consists of 62 residues corresponding to a molecular mass of 6925 Da. MsTI is an inhibitor of trypsin, since the apparent Kd for the formation of the trypsin-MsTI complex is  $1.8 \times 10^{-9}$  M, but exhibits no measurable inhibitory activity toward chymotrypsin. MsTI belongs to the Bowman-Birk inhibitory family, containing seven disulphide bridges and two P1 reactive sites. More exactly MsTI belongs to the D-II Bowman-Birk inhibitory group, in which both residues involved in serine-proteinase interaction are Arg or Lys and which is double headed for trypsin, but not for chymotrypsin or elastase. Titration studies have shown a stoichiometric ratio MsTI:trypsin of 1:1.5. Homology comparison of the primary structure of MsTI with that of other inhibitors shows 81% sequence identity with a Bowman-Birk wound-induced trypsin inhibitor from *M. sativa* leaves and about 60% sequence identity with the well known soybean-derived

Bowman-Birk Inhibitor, BBI. MsTI is the first serine protease inhibitor to be isolated from seeds of *Medicago* species.

The structure of MsTI, determined in solution by magnetic resonance spectroscopy (NMR) at pH 5.6 and 27°C, comprises two distinct symmetric domains each composed of a three-stranded  $\beta$ -sheet in which the active sites are located. A characteristic geometry of three aromatic residues confers stability to this protein and this feature is conserved in all the Bowman-Birk inhibitors of known structure (7).

Many different protease inhibitors, especially BBI, were found to be highly effective inhibitors of carcinogenesis (8-10). BBI for example, the protease inhibitor most extensively studied as an anticarcinogenic agent against human cancer cells, inhibits chemical carcinogen- and radiation-induced malignant transformation *in vitro* (11), and suppresses cancer development in experimental animals (12, 13). Additionally, BBI could be a useful agent for treatment of prostate diseases (14), enhance the cytotoxicity induced by cisplatin in human lung carcinoma cells (15) and potentiate cisplatin-induced growth inhibition and cytotoxicity in different human tumor cancer cells, especially those that developed drug resistance to cisplatin (16, 17).

The clonogenic survival of normal human fibroblasts is not adversely affected by the inhibitor; in contrast BBI protects normal fibroblasts from a radiation-induced reduction in cell survival. The radio-protective effect of BBI seems to be the result of an improved nucleotide excision repair mechanism (18). Although the mechanisms for the cytotoxicity induced by protease inhibitors are unknown, several hypotheses can be postulated. Protease inhibitors may inhibit some serine proteases involved in certain vital functions of cancer cells, such as the processing of growth factor, cell surface receptors, transcription factors and invasion (19-21). Proteinase inhibitors may also reduce clonogenic survival by suppressing the expression of *c-fos* and *c-myc* proto-oncogenes (22-24). In any case it is still unknown whether the proteinase inhibitor properties (antitrypsin and/or antichymotrypsin) are mandatory for the exhibition of an antitumoral activity. Additionally, the pivotal role of proteinases in biological regulation makes them important targets for therapeutic interventions and for identification of new inhibitors.

Preliminary investigations have shown that MsTI has cytotoxic effects on MCF7 (human breast adenocarcinoma) and HeLa (human cervical carcinoma) and improves the cisplatin effects on both cell lines (25). The present study was undertaken to characterise *in vitro* the anticellular activity of MsTI on the basis of the antitumoral properties described for BBI from soybean, and to investigate whether the MsTI shows similar properties to those of BBI with respect to the potentiation of cisplatin-induced cytotoxicity in human cancer cells MCF7 and HeLa.

## Materials and Methods

**Cell lines and cell culture.** The human carcinoma cell lines used in this study are MCF7, a breast adenocarcinoma cell line and HeLa, a cervical carcinoma cell line, both obtained from the Bank of Biological Material Interlab Cell Line Collection (ICLC) of Genova, Italy. All cells were maintained as monolayer cultures at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM, EuroClone, Pero, Milano, Italy) or Minimum Essential Medium (MEM EBSS, Sigma Chemical, St. Louis, MO, USA), for MCF7 or HeLa cells, respectively, supplemented with 10% fetal bovine serum (FBS, EuroClone) and subcultured by trypsinization.

**Chemicals.** Bowman-Birk Inhibitor (BBI) and cis-Dichlorodiammineplatinum(II) (cisplatin) were purchased from Sigma Chemical. BBI was dissolved in medium at a concentration of 10 mg/ml, sterilised by filtration through a 0.2  $\mu$ m filter (Millipore S.p.A., Vimodrone, MI, Italy) and aliquoted at -20°C. Cisplatin was dissolved in dimethylformamide (DMF, Sigma Chemical) at a concentration of 10 mg/ml as a stock solution, diluted in medium and filtered through a 0.2  $\mu$ m filter (Millipore S.p.A.) before use. MsTI was purified from *Medicago scutellaria* seed by ion-exchange chromatography, gel filtration chromatography and reverse phase HPLC as described in Cecilian *et al.*, (6). The lyophilised high purity grade protein was dissolved in phosphate buffer at a concentration of 7.5 mg/ml, sterilised by filtration through a 0.2  $\mu$ m filter (Millipore S.p.A.) and aliquoted at -20°C.

**Colony-forming assay.** The dose-response relationships for MsTI to suppress the clonogenic survival of MCF7 and HeLa were determined by the colony-forming assay. In order to perform the experiments, the cells were seeded in duplicate in 100-mm tissue culture dishes at a density of 2000 or 600 cells per dish, for MCF7 cells or HeLa cells, respectively, and incubated at 37°C overnight. The medium was then changed, the control medium or the medium containing MsTI or BBI to reach final concentrations of 0 (control), 10, 20, 40 and 80  $\mu$ g/ml was added to dishes and the cells incubated for 24 hours. Then, the cells were washed and cultured in fresh complete medium without a medium change until visible colonies were formed in the dishes (14 days for MCF7 cells and 9 days for HeLa cells). The cell colonies were then fixed with methanol, stained with Coomassie Brilliant Blue G (Sigma Chemical) and colonies with >50 cells were counted.

A separate series of experiments were performed to assess the effect of MsTI on the clonogenic survival of MCF7 and HeLa cells exposed to cisplatin for a short period of time (2 hours). In order to perform these experiments, MCF7 and HeLa cells were seeded in 100-mm tissue culture dishes at 2000 (for MCF7) and 600 (for HeLa) cells per dish and incubated at 37°C overnight. The medium was then changed, MsTI was added to the medium in one half of the dishes at a final concentration of 50  $\mu$ g/ml and the cells were incubated for an additional period of 24 hours as pre-treatment. At the end of the pre-treatment period, cisplatin stock solution was added to the medium at the final concentrations of 0, 2, 4, 6, 8  $\mu$ g/ml for MCF7 cells and 0, 0.5, 1, 1.5, 2  $\mu$ g/ml for HeLa cells. The cells were exposed to cisplatin in culture for 2 hours, then washed with serum-free medium and incubated in medium with or without MsTI until visible colonies were formed in the dishes (14 days for

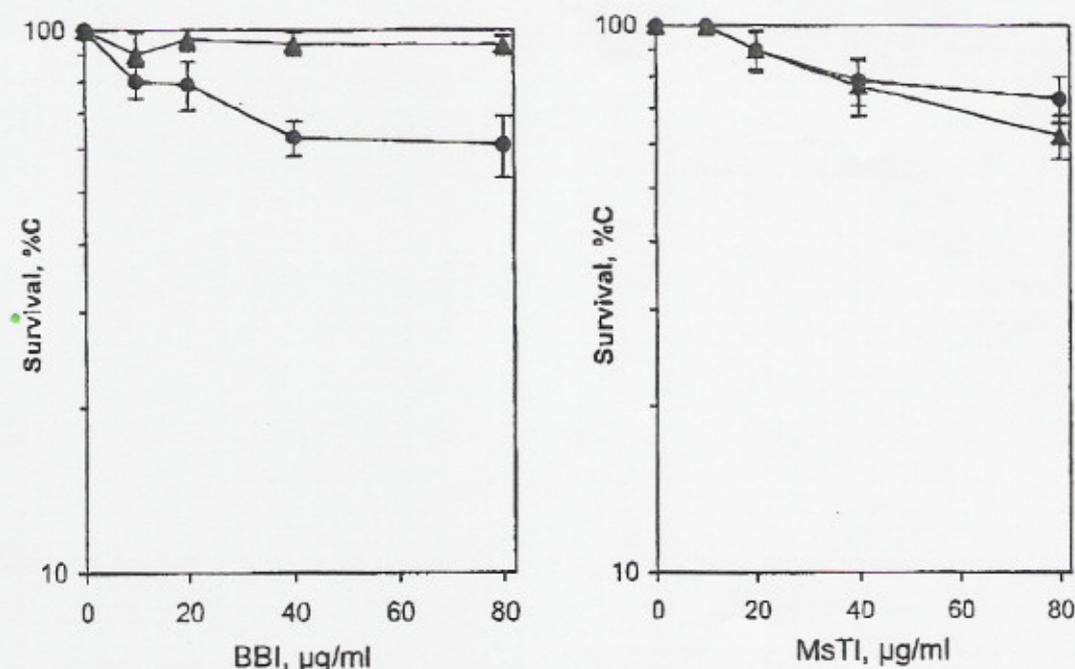


Figure 1. Effect of 24-hour treatment with BBI and MsTI on clonogenic survival of human breast adenocarcinoma MCF7 (●) and human cervical carcinoma HeLa (▲) cells. Cells were plated in tissue culture dishes and 18 hours after plating were exposed to various concentrations of BBI or MsTI for a period of 24 hours. At the end of this period, the cells were washed and cultured in complete medium for 6 and 11 days for HeLa and MCF7 cells, respectively. Cell colonies were fixed, stained and counted. Experiments were carried out in duplicate. Values are means of three experiments  $\pm$  SD.

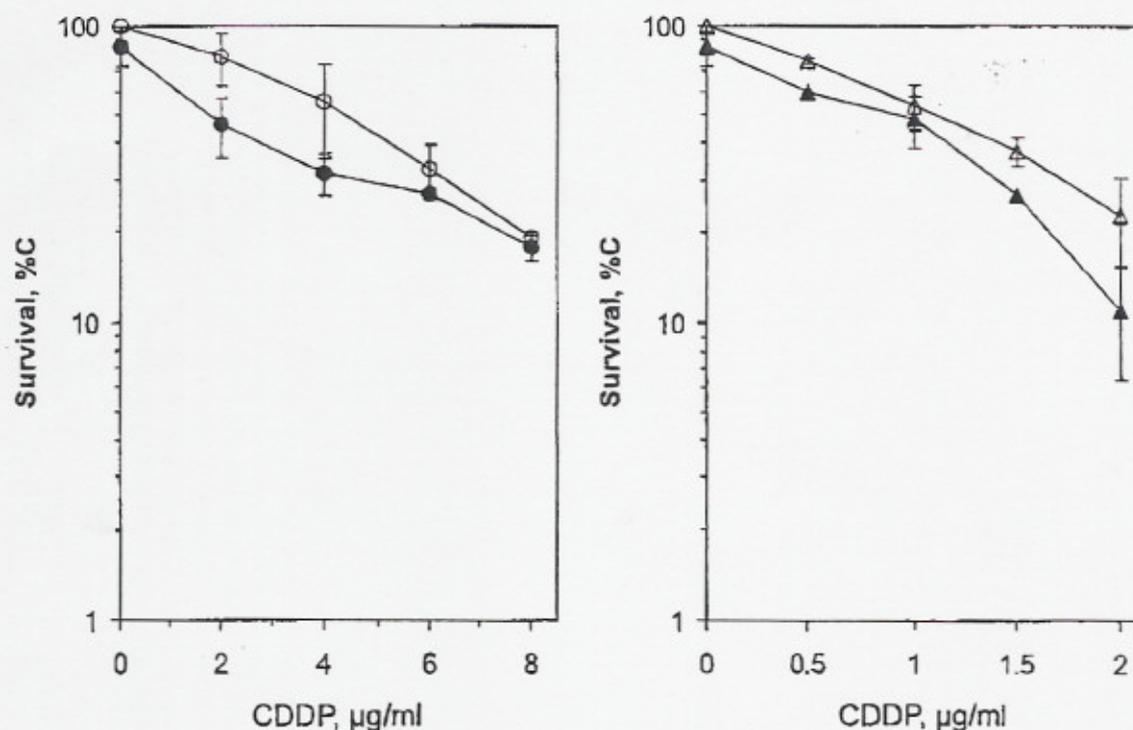


Figure 2. Clonogenic survival of cisplatin-treated MCF7 (left) and HeLa (right) cells cultured in control medium (open symbols) and medium containing MsTI at 24-hour pre-treatment (closed symbols). After 18 hours from plating, MsTI was added to the medium in one-half of the dishes at a final concentration of 50 µg/ml and the cells were incubated for an additional period of 24 hours. At the end of this incubation period, cisplatin was added to the medium at various concentrations for 2 hours, then the cells were washed and cultured in control medium for 6 and 11 days for HeLa and MCF7 cells, respectively. At the end of the incubation period, cell colonies were stained and counted. Experiments were carried out in duplicate and values are means of at least three experiments  $\pm$  SD.

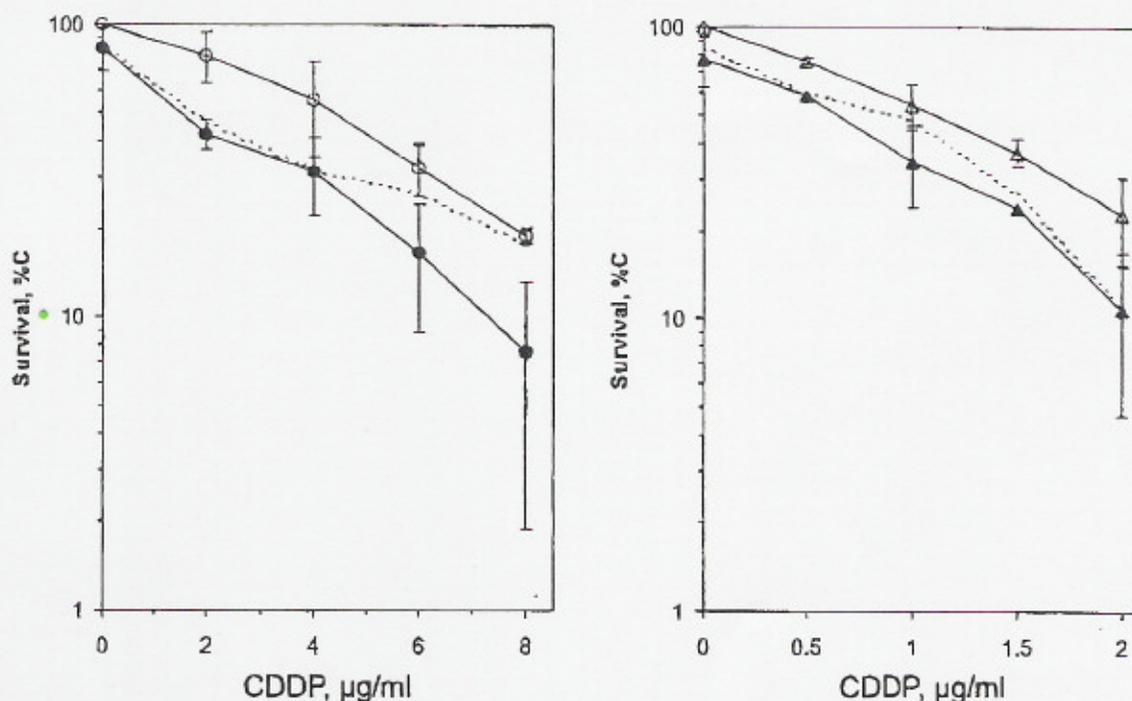


Figure 3. Clonogenic survival of cisplatin-treated MCF7 (left) and HeLa (right) cells cultured in control medium (open symbols) and medium containing MsTI for the entire incubation period (closed symbols). MCF7 cells and HeLa cells were plated. After 18 hours, MsTI was added to the medium in one-half of the dishes at a final concentration of 50 µg/ml and the cells were incubated for an additional period of 24 hours as pre-treatment. At the end of this incubation period, cisplatin was added to the medium at various concentrations for 2 hours, then the cells were washed and cultured in control medium or medium containing 50 µg/ml MsTI for 6 and 11 days for HeLa and MCF7 cells, respectively. At the end of the incubation period, cell colonies were stained and counted. Experiments were carried out in duplicate and values are means of at least three experiments  $\pm$  SD. The dotted lines show the survival curves of the pre-treatment series of experiments of Figure 2.

MCF7 cells and 9 days for HeLa cells). The cell colonies were then fixed with methanol, stained with Coomassie Brilliant Blue G (Sigma Chemical) and colonies with >50 cells were counted.

All survival points were done in duplicate and experiments were conducted a minimum of three times. Survival fraction was calculated by dividing the plating efficiency of drug-treated cells by that of untreated control cells. The  $D_{37}$  value was calculated as the drug concentration resulting in 37% survival of colony-forming ability.

**Cell cycle analysis.** A quantitative measure of cell cycle distribution was obtained by flow cytometric analysis of DNA histograms, as described in Pallavicini *et al.* (26). Briefly, MCF7 cells were seeded in 100-mm tissue culture dishes at 90,000 cells/ml in DMEM with 10% FBS. After overnight incubation to ensure attachment, the medium was changed and the control medium or the medium containing 50 µg/ml MsTI or 100 µg/ml BBI was added. After a 24-hour pre-treatment period, cells were exposed to 4 µg/ml cisplatin for 2 hours, then washed with serum-free medium, collected by trypsinization and centrifuged at 1300 rpm for 7 min. After centrifugation, the cell pellet was placed in ice and resuspended in 1 ml of phosphate buffer saline (PBS) 0.5 mM EDTA, fixed by addition of 3 ml of cold absolute ethanol and incubated at  $-20^{\circ}\text{C}$ . After ethanol fixation, cells were washed in PBS, resuspended with 1 ml of 2.5 µg/ml propidium iodide staining solution in PBS with 12.5 µg RNase A (1 mg/ml) and

incubated overnight at  $4^{\circ}\text{C}$ . Cell fluorescence was measured in a FACSCAN flow cytometer (Becton Dickinson, Mountain View, USA). Data were recorded in a Hewlett-Packard computer after electronic gating of cell debris and nuclear aggregates. The percentage of MCF7 nuclei in the different phases of the cell cycle ( $G_0/G_1$ , S and  $G_2/M$ ) was calculated from the histogram of the DNA fluorescence area. A minimum of  $10^4$  cells/sample were analysed.

## Results

In Figure 1 is reported the effect of 24-hour treatment with BBI or MsTI on the clonogenic survival of human breast adenocarcinoma MCF7 and human cervical carcinoma HeLa cells. The presence of BBI in the culture medium showed minimum clonogenic survival inhibition of HeLa cells up to 80 µg/ml and a dose-related inhibition of MCF7 cell survival only up to 40 µg/ml (Figure 1, left). Interestingly, in the presence of MsTI, the inhibition of the clonogenic survival of both cell lines was observed up to 80 µg/ml (Figure 1, right). Exposure of both cell lines to a higher concentration of BBI or MsTI (160 µg/ml) did not cause a further reduction of cell survival (data not shown).

Table I. Colony-forming ability of MCF7 and HeLa cells after exposure to cisplatin in the absence or in the presence of MsTI.

	Cisplatin D <sub>37</sub> , µg/ml	
	MCF7 cells	HeLa cells
Cisplatin	5.8	1.6
Cisplatin + MsTI 24h (50 µg/ml)	3.5	1.3
Cisplatin + MsTI (50 µg/ml)	3.5	1.0

D<sub>37</sub> is the cisplatin dose required to reduce survival to 37%.

The effect of treatment by MsTI in combination with cisplatin on the clonogenic survival of MCF7 and HeLa cells was determined in two colony formation assay experiments as shown in Figures 2 and 3. In the first series of experiments (Figure 2), the cells were exposed to a pre-treatment of 24 hours with 50 µg/ml MsTI before the 2-hour exposure to increasing concentrations of cisplatin, 2, 4, 6, 8 µg/ml for MCF7 cells or 0.5, 1, 1.5, 2 µg/ml for HeLa cells, respectively. The surviving fractions of MCF7 and HeLa cells in the presence of 50 µg/ml MsTI alone in the cell culture medium were reduced by 15% compared to surviving fractions of cells cultured in the control medium. With respect to the two-hour exposure to cisplatin alone, that decreased the colony formation capabilities of both cell lines in a dose-dependent manner (Figure 2, open symbols), the incubation of cisplatin-treated cells with 50 µg/ml of MsTI further affected their clonogenic survival (Figure 2, closed symbols).

In the second series of experiments (Figure 3), after the 24-hour pre-treatment with MsTI and after the exposure to cisplatin, the cells were incubated with 50 µg/ml of MsTI for 9-14 days without a medium change, allowing the continuous exposure to MsTI. The survival reduction after the exposure to MsTI alone was similar to that observed after the 24-hour MsTI pre-treatment. The survival reduction in the presence of 50 µg/ml MsTI on the cisplatin-treated cells (Figure 3, closed symbols) was enhanced in both cell lines, in comparison with the exposure to cisplatin alone (Figure 3, open symbols). The dotted lines in Figure 3 report the survival curves of 24-hour MsTI pre-treatment as shown in Figure 2.

Table I reports the D<sub>37</sub> cisplatin values (the cisplatin dose resulting in 37% survival) for both the cell lines tested. After the 24 hours of MsTI pre-treatment, the D<sub>37</sub> cisplatin values were reduced by approximately 40% and 20% in MCF7 and HeLa cells, respectively. The presence of MsTI at a concentration of 50 µg/ml during the entire incubation period did not further affect the clonogenic survival of MCF7 cells, but enhanced the cisplatin-induced cytotoxicity of HeLa cells reducing the D<sub>37</sub> by 40%.

Table II. Cell cycle analysis of MCF7-treated cells.

	Phase of cell cycle (% of cells)		
	G <sub>1</sub> /G <sub>1</sub>	S	G <sub>2</sub> /M
Control	69.7	15.5	14.2
Cisplatin, 4 µg/ml	72.5	14.9	12.5
BBI, 100 µg/ml	70.7	14.5	14.2
BBI, 100 µg/ml + Cisplatin, 4 µg/ml	69.1	16.8	13.7
MsTI, 50 µg/ml	75.3	14.6	9.4
MsTI, 50 µg/ml + Cisplatin, 4 µg/ml	73.2	14.7	11.5

To analyse the MsTI effect on cell cycle progression in comparison with BBI, the cell cycle phase distribution in MCF7 cultures was measured by flow cytometric analysis of propidium iodide-stained nuclei (Table II). MCF7 cells exposed for 24 hours to 50 µg/ml MsTI showed an increase in the population in G<sub>1</sub>-phase (5.6%) with a concomitant decrease in the percentage of cells in G<sub>2</sub>/M-phase (5%) with respect to the control cells. No differences in the percentage of cells in the different phases were observed after exposure to BBI. Cisplatin, at the tested concentration, did not seem to affect the cell cycle progression of MCF7 cells, in the presence or in the absence of MsTI or BBI.

## Discussion

Despite the low sequence identity with other trypsin inhibitors from the *Leguminosae* family, it was interesting to investigate the anticellular activity of MsTI on the basis of the antitumoral properties described for the soybean inhibitor BBI. The potential of BBI to affect radiation and cisplatin-induced cell killing in different cancer cell lines has already been assessed (16, 17).

In order to determine whether MsTI potentiates cisplatin-induced cell killing, the highly purified protein was considered and tested on the well known and characterised human breast adenocarcinoma (MCF7) and cervical carcinoma (HeLa) cell lines. When MsTI was tested in the range 10-80 µg/ml, the same cytotoxic activity on both the cellular lines was observed and, more importantly, higher cytotoxic activity than BBI on HeLa cancer cells was registered (Figure 1).

In order to verify if the presence of MsTI enhances the cisplatin-induced cytotoxicity, different experiments were performed by using a colony formation assay. The exposure to MsTI (50 µg/ml) was by itself not cytotoxic, but the presence of 50 µg/ml MsTI in the culture medium before, during and after the treatment with cisplatin caused a reduction of the percentage of cells that were

able to form clones (Figures 2 and 3). This finding was also confirmed by the 40% reduction of  $D_{37}$  cisplatin values as reported in Table I. Treatment of MCF7 and HeLa cells with cisplatin in combination with MsTI decreased the clonogenic survival of these cells to a much greater extent than treatment with cisplatin in the absence of MsTI. Among the two cell lines analysed, the effects of MsTI on clonogenic capacity were relatively higher in MCF7 cells as compared to HeLa cells.

A preliminary study to examine the effect of MsTI on the progression of cell cycle of MCF7 cells was carried out. Under our experimental conditions, the cell cycle phase distribution in MCF7 cultures was not affected by soybean BBI or cisplatin alone. On the contrary, a moderate although not significant increase in the population in  $G_0/G_1$ -phase (5.6%) with a concomitant decrease in the percentage of cells in  $G_2/M$ -phase (5%) was observed in cells treated with MsTI for 24 hours (Table II). This observed increase in  $G_0/G_1$ -population can suggest a possible block of the cells in  $G_1$ -phase. Although the cisplatin is not strictly cell-cycle-specific, cells are more sensitive to cisplatin actions in late  $G_1$ - and S-phases. The enhancement of cisplatin cytotoxicity could be related to  $G_1$ -phase elongation. Further investigations are needed to clarify this hypothesis.

Although the mechanism for the cytotoxicity induced by protease inhibitors is unknown, several possible mechanisms can be postulated, as previously mentioned. Among these is that the clonogenic survival could be reduced by the inhibition of some proteases involved in vital functions of the cancer cells. In this respect, the cytotoxic properties of MsTI here reported indicate that antichymotryptic activity is not a strict requirement for antitumoral activity, as often reported for other protease inhibitors (27).

Further studies are warranted to evaluate the precise mechanisms for the enhancing effects of MsTI on cisplatin-induced killing, including: i) structural studies for elucidation of the interaction of MsTI with other enzymes or macromolecules which could be important in antitumoral activity, and ii) association with other substances, such as saponins, able to increase cell permeability (28, 29) and facilitate the uptake of the protease inhibitor and cisplatin into the cells.

At present, our finding indicate that MsTI could be considered as a part of a therapeutic regimen for the potentiation of cisplatin-mediated cancer treatment.

#### Acknowledgements

This work was supported by the Italian National Research Council "Progetto Agenzia 2000". The authors thank Dr. Carlo Zibera for assistance in performing the flow cytometry study and Dr. Roberta Riboni for her helpful comments.

#### References

- 1 Liener IE and Kakade ML: Protease inhibitors. In: Toxic Constituents of Plant Foodstuffs (Liener IE, ed.). New York, Academic Press 1980. pp 7-71.
- 2 Birk Y: In: Bioactive Substances in Food of Plant Origin (Kozłowska H, Formal J and Zdunczyk Z, eds.) 1994. pp 202.
- 3 Richardson M: Seed storage proteins: the enzyme inhibitors. In: Methods in Plant Biochemistry. New York, Academic Press 1991. pp 259-305.
- 4 Ryan AC: Protease inhibitors in plants: genes for improving defences against insects and pathogens. *Ann Rev Phytopathol* 28: 425-449, 1990.
- 5 Papastoitis G and Wilson KA: Initiation of the degradation of the soybean Kuntz and Bowman-Birk trypsin inhibitors by a cysteine protease. *Plant Physiol* 96: 1086-1092, 1991.
- 6 Ceciliani F, Tava A, Iori R, Mortarino M, Odoardi M and Ronchi S: A trypsin inhibitor from snail medic seed active against pest proteases. *Phytochem* 44: 393-398, 1997.
- 7 Catalano M, Ragona L, Molinari H, Tava A and Zetta L: Anticarcinogenic Bowman-Birk inhibitor isolated from snail medic (*Medicago scutellata* L.) seeds: solution structure and analysis of self association behavior. *Biochemistry* 42: 2836-2846, 2003.
- 8 Kennedy AR: Anticarcinogenic activity of protease inhibitors: overview. In: Protease Inhibitors as Cancer Chemopreventive Agents (Troll W and Kennedy AR, eds.) 1993. pp 9-64.
- 9 Clark DA, Day R, Seidah N, Moody TW and Cuttitta F: Protease inhibitors suppresses *in vitro* growth of human small cell lung cancer. *Peptides* 14: 1021-1028, 1993.
- 10 Clemente A and Domoney C: Anticarcinogenic activity of protease inhibitors in legumes. in IVth European Conference on Grain Legumes. 2001. Cracow, Poland.
- 11 Kennedy AR: *In vitro* studies of anticarcinogenic protease inhibitors. In: Protease Inhibitors as Cancer Chemopreventive Agents (Troll W and Kennedy AR, eds.) 1993. pp 65-91.
- 12 Kennedy AR: Prevention of carcinogenesis by protease inhibitors. *Cancer Res* 54: 1999s-2005s, 1994.
- 13 Kennedy AR, Beazer-Barclay Y, Kinzler KW and Newberne PM: Suppression of carcinogenesis in the intestines of Min mice by the soybean-derived Bowman-Birk inhibitor. *Cancer Res* 56: 679-682, 1996.
- 14 Kennedy AR and Wan XS: Effects of the Bowman-Birk inhibitor on growth, invasion, and clonogenic survival of human prostate epithelial cells and prostate cancer cells. *Prostate* 50: 125-133, 2002.
- 15 Kennedy CW, Donahue JJ and Wan XS: Effects of the Bowman-Birk protease inhibitor on survival of fibroblasts and cancer cells exposed to radiation and cis-platinum. *Nutr Cancer* 26: 209-217, 1996.
- 16 Wan XS, Hamilton TC, Ware JH, Donahue JJ and Kennedy AR: Growth inhibition and cytotoxicity induced by Bowman-Birk inhibitor concentrate in cisplatin-resistant human ovarian cancer cells. *Nutr Cancer* 31: 8-17, 1998.
- 17 Zhang L, Wan XS, Donahue JJ, Ware JH and Kennedy AR: Effects of the Bowman-Birk inhibitor on clonogenic survival and cisplatin or radiation-induced cytotoxicity in human breast, cervical, and head and neck cancer cells. *Nutr Cancer* 33: 165-173, 1999.

- 18 Dittmann K, Dikomey E, Mayer C and Rodemann HP: The Bowman-Birk protease inhibitor enhances clonogenic cell survival of ionizing radiation-treated nucleotide excision repair-competent cells but not of xeroderma pigmentosum cells. *Int J Radiat Biol* 76: 223-229, 2000.
- 19 Chu SC, Chou FP, Liu LJ and Hsieh YS: Proteolytic activities of mouse sarcoma 180 cells that are inhibited by Bowman-Birk and Kunitz protease inhibitors. *Biochem Mol Biol Int* 42: 965-975, 1997.
- 20 Gueven N, Dittmann K, Mayer C and Rodemann HP: Bowman-Birk protease inhibitor reduce the radiation-induced activation of EGF receptor and induces tyrosine phosphatase activity. *Radiat Biol* 73: 157-162, 1998.
- 21 Dittmann K, Löffler H, Bamberg M and Rodemann HP: Bowman-Birk protease inhibitor (BBI) modulates radiosensitivity and radiation-induced differentiation of human fibroblasts in culture. *Radiother Oncol* 34: 137-143, 1995.
- 22 Caggana M and Kennedy AR: *c-fos* m-RNA levels are reduced in the presence of antipain and Bowman-Birk inhibitor. *Carcinogenesis* 10: 2145-2148, 1989.
- 23 St. Clair WH and St. Clair DK: Effect of the protease inhibitor on the expression of oncogenes in the irradiated rat colon. *Cancer Res* 51: 4539-4543, 1991.
- 24 Chang JD and Kennedy AR: Suppression of *c-myc* by anticarcinogenic protease inhibitors. *In: Protease Inhibitors as Cancer Chemopreventive Agents* (Troll W and Kennedy AR, eds.) 1993. pp 265-280.
- 25 Lanza A, Tava A, Zetta L, Di Francesco S, Robustelli della Cuna FS and Robustelli della Cuna G: Effetto anticellulare di un nuovo inibitore di proteasi su linee tumorali umane (MCF7 e HeLa). *In: XXVII Congresso Nazionale di Oncologia, Il cammino della ricerca in Oncologia*. 2001. Roma, Italy: Pubblicità & Stampa s.n.c.
- 26 Palavicini MG: *In: Techniques in Cell Cycle Analysis* (Gray JW and Darzynkiewicz Z, eds.). Clifton, NJ, Humana Press 1987. pp 139-162.
- 27 Kennedy AR: The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. *Am Clin Nutr* 68: 1406S-1412S, 1998.
- 28 Johnson IT, Gee JM, Price K, Curl C and Fenwick GR: Influence of saponins on gut permeability and active nutrient transport *in vitro*. *J Nutr* 116: 2270-2277, 1986.
- 29 Gee JM, Wal JM, Miller K, Atkinson H, Grigoriadou F, Wijnands MV, Penninks AH, Wortley G and Johnson IT: Effect of saponins on the transmucosal passage of beta-lactoglobulin across the proximal small intestine of normal and beta-lactoglobulin-sensitized rats. *Toxicology* 117: 219-228, 1997.

Received July 31, 2003

Revised October 7, 2003

Accepted December 12, 2003

# NMR Solution Structure of Viscotoxin C1 from *Viscum Album* Species *Coloratum ohwi*: Toward a Structure–Function Analysis of Viscotoxins<sup>†</sup>

Silvia Romagnoli,<sup>‡</sup> Federico Fogolari,<sup>‡</sup> Maddalena Catalano,<sup>§</sup> Lucia Zetta,<sup>§</sup> Gerhard Schaller,<sup>||</sup> Konrad Urech,<sup>||</sup> Matteo Giannattasio,<sup>⊥</sup> Laura Ragona,<sup>§</sup> and Henriette Molinari<sup>\*‡</sup>

Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, Strada le Grazie 15, 37134 Verona, Italy, Laboratorio NMR, ISMAC-CNR, Via Bassini 15, 20133 Milano, Italy, Verein für Krebsforschung, Hiscia Institute, 4144 Arlesheim, Switzerland, and Dipartimento di Arboricoltura, Botanica e Patologia Vegetale, Facoltà di Agraria, Università di Napoli, Portici, 80100 Napoli, Italy

Received May 9, 2003; Revised Manuscript Received September 5, 2003

**ABSTRACT:** The high resolution three-dimensional structure of the newly discovered plant viscotoxin C1, from the Asiatic *Viscum album* ssp. *Coloratum ohwi*, has been determined in solution by <sup>1</sup>H NMR spectroscopy at pH 3.6 and 285 K. The viscotoxin C1-fold, consisting of a helix–turn–helix motif and a short stretch of an antiparallel  $\beta$ -sheet is very similar to that found for the highly similar viscotoxins A2 and A3 and for other related thionins. Different functional properties of members of the thionin family are discussed here in light of the structural and electrostatic properties. Among the very homologous family of  $\alpha$ - and  $\beta$ -thionins, known for their antimicrobial activity, the viscotoxin subfamily differs from the other members because of its high toxicity against tumoral cells. Key residues for the modulation of viscotoxin cytotoxicity have been identified on the basis of sequence and structural alignment.

Viscotoxins are a group of basic low molecular weight proteins (MW approximately 5000 Da) found in the European mistletoe subspecies of *Viscum album*. On the basis of sequence homology, they have been classified as belonging to the  $\alpha$ - and  $\beta$ -thionin family (1, 2). Thionins include several plant proteins from cereals (hordothionins, purothionins, and avenothionins), *Brassicaceae* (crambins), and *Pyrularia pubera* and *Viscaceae* (viscotoxins from European mistletoes and phoratoxins from American mistletoes *Phoradendron* sp.). The main characteristic of all thionins, except crambin, is their toxic effect on different biological systems: evidences of bactericidal and fungicidal properties as well as toxicity in insects were reported (3, 4). Thionin toxicity is exerted through cell membrane destabilization and disruption, but the mechanism of action is not yet fully understood. To account for this generalized toxicity, it was proposed that thionins induce ion channel formation in cell membranes, causing the dissipation of ion concentration gradients (5). An alternative theory suggests that thionins lead to cell membrane breakdown through purely electrostatic interactions with the negatively charged phospholipids (6).

Viscotoxins share a high sequence similarity with thionins (7), and their cysteine residues, all involved in disulfide bridges, are kept in conserved positions within the amino acid sequence defining a structural motif, known as a concentric motif. This kind of disulfide pattern is suggested to be able to stabilize a common structure occurring in

various small proteins able to interact with cell membranes (1). Besides the poisonous effects on the whole organism, viscotoxins have been reported to be cytotoxic against many tumoral cell lines including *Yoshida sarcoma* cells (8, 9), K 562 human myelogenous leukaemia (9), and HeLa cells (10).

Until now, several viscotoxin isoforms have been characterized; namely, viscotoxins A1, A2, A3, B, and 1-PS. These proteins are highly basic, possessing three to four arginines and three to four lysines. Their alignment together with related  $\alpha$ - and  $\beta$ -thionins is reported in Figure 1. For viscotoxin A3, displaying the highest toxicity against *Y. sarcoma* cells, the three-dimensional structure has been previously reported by our group (7), and viscotoxin A2 has been recently deposited (PDB 1JMN) (11). Here, we report the discovery and the NMR<sup>1</sup> structural and electrostatic characterization of a new isoform, viscotoxin C1, from the Asiatic *V. album* ssp. *C. ohwi*. Structure and electrostatic properties of the other isoforms A1, B, and 1-PS were obtained on the basis of homology modeling and discussed as a function of the measured cytotoxicity against *Y. sarcoma* cell lines.

## MATERIALS AND METHODS

**Sample Preparation.** Viscotoxin C1 was detected by bioactivity guided fractionation of acetic acid (0.2 M) extracts from leaves of *V. album* ssp. *C. ohwi* using the viscotoxin

<sup>†</sup> S.R. acknowledges CNR for Fellowship 201.22/04.

\* To whom correspondence should be addressed. Tel.: +39 45 8027097. Fax: +39 45 8027929. E-mail: molinari@sci.univr.it.

<sup>‡</sup> Università degli Studi di Verona.

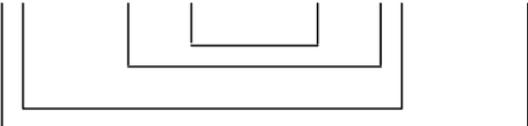
<sup>§</sup> ISMAC-CNR.

<sup>||</sup> Hiscia Institute.

<sup>⊥</sup> Università di Napoli.

<sup>1</sup> Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; DQF–COSY, double quantum filtered–correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; HPLC, high performance liquid chromatography; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; RMSD, root mean square deviation.

PROTEIN	SEQUENCE	IDENTITY %	ED <sub>50</sub> μM
	 VVVVVVVVVV      VVVVVVVV		
viscotoxin A3	KSCCPNTTGRNIYNACRLTGAPRPTCAKLSGCKIISGSTCPSDYPK	100.0	0.068
viscotoxin 1PS	KSCCPNTTGRNIYNTCRFGGGSREVCAKLSGCKIISASTCPSDYPK	78.3	0.089
viscotoxin A1	KSCCPNTTGRNIYNTCRLTGSSRETCAKLSGCKIISASTCPSNYPK	87.0	0.177
viscotoxin C1	KSCCPNTTGRNIYNTCRFAGGSRERCAKLSGCKIISASTCPSDYPK	82.6	0.198
viscotoxin A2	KSCCPNTTGRNIYNTCRFGGGSQVCAKLSGCKIISASTCPSDYPK	80.4	0.223
viscotoxin B	KSCCPNTTGRNIYNTCRLGGGSRERCAKLSGCKIISASTCPSDYPK	82.6	0.942
phoratoxin-A	KSCCPTTARNIYNTCRFGGGSRPVCAKLSGCKIISGKCDNSNGH	67.4	
α-purothionin	KSCCRSTLGRNCYNLCRARGAQK-LCAGVCRCKISSGLSCPFGFPK	56.5	
β-purothionin	KSCCKSTLGRNCYNLCRARGAQK-LCANVCRCKLTSGLSCPFDKFPK	56.5	
β-hordothionin	KSCCRSTLGRNCYNLCRVRGAQK-LCANACRCKLTSGLSCPFGFPK	56.5	
α-hordothionin	KSCCRSTLGRNCYNLCRVRGAQK-LCAGVCRCKLTSGLSCPFGFPK	52.2	
crambin	TTCCPSIVARSNFVNCRLPGTPEALCATYTGCIIPGATCPGDYAN	47.8	



CONCENTRIC MOTIF

FIGURE 1: Alignment of all reported viscotoxin isoforms, listed in order of decreasing cytotoxicity (ED<sub>50</sub>) together with related α- and β-thionins of known structure. Conserved residues in the viscotoxin sequences are shaded. Sequence identity and biological activity data (ED<sub>50</sub>) are reported in the last two columns. Secondary structural elements of viscotoxin A3 are indicated. The disulfide bridge pattern, known as a concentric motif, is shown.

sensitive *Y. sarcoma* cells (9). The purification procedure was the same as described for viscotoxin A3 and B (8). Samples for NMR studies were 1 mM in 50 mM H<sub>3</sub>PO<sub>4</sub>/NaOH buffer at pH = 3.6.

**Viscotoxin C1 Primary Structure Characterization.** Viscotoxin C1 primary structure was determined by Edman degradation coupled to HPLC, NMR proton assignment, and mass spectrometry. A molecular weight of 4945.5 Da over an expected value of 4946.6 was measured.

**NMR Spectroscopy.** NMR spectra were acquired on a Bruker DRX spectrometer operating at 500.13 MHz. Standard homonuclear <sup>1</sup>H DQF-COSY (12), TOCSY (13), and NOESY (14) experiments were recorded, employing mixing times in the range of 20–100 ms for TOCSY and 60–150 ms for NOESY. Data matrixes of 4096 × 512 points were acquired in *f*<sub>2</sub> and *f*<sub>1</sub>, respectively, and 120 scans were accumulated with a sweep width of 5838 Hz in both dimensions. Water suppression was achieved with gradients by a 3-9-19 pulse sequence (14) or using excitation sculpting (15). All NMR experiments were performed at variable temperatures in the range of 280–295 K. Spectra were processed using XWINNMR. Analysis of spectra and cross-peak volumes was performed using XEASY software (16). All chemical shifts were referenced to the methyl resonance of 3-(trimethylsilyl)-propionic acid-*d*<sub>4</sub> sodium salt (TSP) dissolved in the same buffer solution. Secondary chemical shift maps were obtained using as a reference the random coil chemical shifts reported by Wishart et al. (17). Assignment of spin systems to individual amino acids was achieved by using DQF-COSY and TOCSY spectra acquired at 285 K, while complete resonance assignment was obtained by

the combined use of TOCSY and NOESY spectra, following the sequential assignment strategy (18). The obtained assignments were extended at 295 K, and NOESY experiments with different mixing times (60, 100, and 150 ms) and DQF-COSY were performed at all the selected temperatures to check for ambiguous assignments. A single H/D exchange experiment was performed after exchanging the protein in deuterated buffer. Since the protein concentration was quite low (<0.7 mM), acquisition times of 72 h were employed for TOCSY and NOESY experiments. Amide protons stable for at least one week gave us some indication of their involvement in H-bonds.

**Structure Calculations.** Calculated structures were obtained from restrained simulations using DYANA (19) followed by energy minimization using the AMBER force field, as implemented by the program DISCOVER (Molecular Simulations, San Diego, CA). The calibration of the NOESY peak volume was performed on the NOESY spectrum acquired in aqueous solution at *T* = 285 K, with a mixing time of 100 ms, on the basis of the H<sup>γ3</sup>-H<sup>δ2</sup> distance of P<sub>41</sub>, corresponding to an internuclear separation of 3.04 Å. Calibration was executed using the tools of the DYANA package (20), and the obtained list of distance restraints was used as input for DYANA calculations. In all the cases where severe overlap prevented us from quantifying cross-peak volumes, a distance of 4 Å was assumed between the interacting protons. A total of 845 interproton distances were used as input for structure calculation; standard procedures were used to remove irrelevant restraints, thus obtaining 631 final restraints composed of 224 intrasidue, 159 short range, 125 medium range, and 123 long range distances. The NOE

restraints were supplemented by three types of restraints: (i) 24 distance restraints for 12 backbone hydrogen bonds. Hydrogen bond formation or solvent exclusion was assumed to account for the slow exchanging amide protons; for each hydrogen bond, the N–O and NH–O distances were assigned an upper limit value of 3.00 and 2.00 Å, respectively. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints; (ii) 18 distance restraints for the three disulfide bonds; for each disulfide bridge an upper and lower limit value was imposed to the three distance restraints that define an S–S bond:  $S_i-S_j$  (2.00–2.10 Å)  $S_i-C_{\beta j}$  (3.00–3.10 Å), and  $S_j-C_{\beta i}$  (3.00–3.10 Å); and (iii) 18  $\phi$  backbone torsion angle restraints derived from  $^3J_{\text{HN}\alpha}$  coupling constants whose values were estimated by the separation of extrema in the dispersive and absorptive peaks of DQF–COSY spectra (12).  $\phi$  angle restraints of  $-133 \pm 30^\circ$  for  $^3J_{\text{HN}\alpha}$  coupling constants greater than 8.0 Hz and  $-55 \pm 30^\circ$  for  $^3J_{\text{HN}\alpha}$  coupling constants smaller than 4.0 Hz were used for calculations. Two-hundred calculations were initially performed with the program DYANA, and the 20 resulting conformers with the lowest residual target function values (resulting from van der Waals and restraint violations) were analyzed. The restraints were reexamined in view of consistent violations and relaxed where necessary. This procedure was repeated until no consistent violations were found in half or more of the structures. After this step, 600 new calculations were started; the 20 structures selected on the basis of the lowest target function (ranging from 0.34 to 0.52) showed no violations larger than 0.20 Å and were then subjected to energy minimization using the AMBER force field with a distance dependent dielectric constant ( $\epsilon = 4.0$ ) as implemented in the program Discover (Molecular Simulations, San Diego, CA). The 10 structures with the lowest potential energy were then selected for further analysis. INSIGHT II (Molecular Simulations, San Diego, CA) and Swiss-Pdb Viewer (21) programs were used to visually inspect the structures.

**Modeling.** Until now, six viscotoxin isoforms from European mistletoe have been isolated and sequenced, but a structural characterization was previously reported only for viscotoxin A3 (7) and A2 (11). A1, B, and 1-PS three-dimensional structures (showing at least 78.3% identity and 93.5% similarity with isoform A3) were obtained by homology modeling using the software Geno3D (22), employing viscotoxin A3 as template structure (PDB 1ED0).

**Electrostatic Potential Map Calculations.** Electrostatic potential maps for C1, A1, A2, A3, B, and 1-PS viscotoxins were computed using a Poisson–Boltzmann based methodology (23) as implemented in the software package UHBD (24). Partial atomic charges and van der Waals radii were taken from the CHARMM force field (25). Molecular and solvent dielectric constants were set to 4.0 and 80.0, respectively, the ionic strength was set to 100 mM, and the probe radius for generating the dielectric map was set to 1.4 Å. A standard procedure was employed, solving first the Poisson–Boltzmann equation on a coarse grid of  $60 \times 60 \times 60$  points with 2.5 Å spacing imposing the sum of the atomic Debye–Hückel potential as the boundary condition and then focusing on the molecule using a finer grid of  $65 \times 65 \times 65$  points with a spacing of 0.8 Å. The GRASP

program (26) was used to visualize the electrostatic potential surfaces.

**Cytotoxicity Assay.** Cytotoxicity of purified viscotoxin C1 was detected by measuring the inhibition of  $^3\text{H}$ -thymidine incorporation into *Y. Sarcoma* cells as previously described for viscotoxin A1, A2, A3, B, and 1-PS (8). Each assay was done at least in triplicate. The concentration of substance inhibiting 50% of  $^3\text{H}$ -thymidine corresponds to the ED<sub>50</sub> and was defined as 1 inhibitory unit.

## RESULTS

**Primary Structure Determination.** The sequence of viscotoxin C1, as obtained by Edman degradation coupled to HPLC, showed few ambiguities relative to four residues: residue 38 was identified with the same probability as an alanine or serine, and the last three residues were only supposed to be Y<sub>44</sub>, P<sub>45</sub>, and K<sub>46</sub> due to a low signal. The sequence, completely determined through the NMR sequential assignment and confirmed by mass spectrometry, is the following: KSCCPNTTGR<sub>10</sub>NIYNTCRFAG<sub>20</sub>GSRERCAKLS<sub>30</sub>GCKIISASTC<sub>40</sub>PSDYPK (Swiss-Prot accession number P83554).

**Assignment of Spin Systems.** Complete proton assignment of viscotoxin C1 was achieved following the sequential resonance assignment procedure (18) by the combined use of two-dimensional  $^1\text{H}$  DQF–COSY, TOCSY, and NOESY experiments. The list of proton chemical shifts, as obtained from spectra acquired at pH = 3.6 and 285 K, is reported in Table 1. Figure 2 shows a portion of a TOCSY spectrum, illustrating through-bond connectivities between the amide and the aliphatic protons.

**Structure Calculations.** A summary of NOE connectivities is reported in Figure 3. The presence of H-bonds was exploited for structure calculations. The partners for all hydrogen bonds were assigned on the basis of preliminary structure calculations, as obtained by imposing only NOE restraints in the procedure of DYANA molecular dynamics, as described in the Materials and Methods.

Two  $\alpha$ -helices, helix I (7–19) and helix II (23–30), have been identified on the basis of short  $d_{\text{NN}}$  ( $i, i + 1$ ),  $d_{\alpha\text{N}}$  ( $i, i + 3$ ), and  $d_{\alpha\beta}$  ( $i, i + 3$ ) distances and by slow exchanging amide protons (Figure 3). The two short antiparallel  $\beta$ -strands (1–3, 33–35) were identified on the basis of typical  $\text{H}^{\alpha}_i\text{--}\text{H}^{\alpha}_j$  NOEs between residues S<sub>2</sub> and I<sub>34</sub> and C<sub>4</sub> and C<sub>32</sub>; by  $\text{H}^{\alpha}_i\text{--}\text{H}^{\text{N}}_j$  NOEs between S<sub>2</sub> and I<sub>35</sub>, C<sub>4</sub> and K<sub>33</sub>, and I<sub>34</sub> and C<sub>3</sub>; and by  $\text{H}^{\alpha}\text{C}_{32}$  and  $\text{H}^{\delta}\text{P}_5$ .

The analysis of the spectra acquired at 285 K in  $^2\text{H}_2\text{O}$  led to the identification of a few long range interactions ( $\text{H}^{\delta}\text{P}_5\text{--}\text{H}^{\epsilon}\text{Y}_{44}$ ,  $\text{H}^{\delta}\text{P}_5\text{--}\text{H}^{\delta}\text{Y}_{44}$ ,  $\text{H}^{\alpha}\text{C}_{32}\text{--}\text{H}^{\delta}\text{Y}_{44}$ , and  $\text{H}^{\alpha}\text{C}_{32}\text{--}\text{H}^{\epsilon}\text{Y}_{44}$ ). These NOEs suggest that the C-terminal end of the molecule is anchored to the rest of the structure, in agreement with the presence of the C<sub>4</sub>–C<sub>32</sub> disulfide bridge.

The superposition of the 10 final structures of viscotoxin C1 (PDB 1ORL) is shown in Figure 4. For a survey of the conformational restraints and the quality of the structure determination, see Table 2. The average RMSD values in the region of 2–45 are  $0.57 \pm 0.16$  and  $1.22 \pm 0.17$  Å for the backbone and all heavy atoms, respectively (Table 2). The C1 structure resembles a capital letter L, in which the long arm of the L is formed by the two  $\alpha$ -helices running in opposite directions and connected by a short turn (residues

Table 1: Proton Resonance Assignments ( $\delta$ , ppm) for Viscotoxin C1 at 285 K and pH 3.6

residue	HN	H $^{\alpha}$	H $^{\beta}$	H $^{\gamma}$	H $^{\delta}$	others
Lys <sup>1</sup>		4.18	1.83–1.94	1.31–1.40	1.52–1.58	H $^{\epsilon}$ = 3.07 H $^{\xi}$ = 7.56
Ser <sup>2</sup>	8.91	4.98	2.84–3.77			
Cys <sup>3</sup>	9.09	5.01	2.24–4.42			
Cys <sup>4</sup>	9.42	5.27	2.67–2.70			
Pro <sup>5</sup>		4.42	2.02–2.07	2.26	3.81–3.99	
Asn <sup>6</sup>	6.94	4.55	3.30		6.99–7.87	
Thr <sup>7</sup>	9.03	3.89	4.17	1.30		
Thr <sup>8</sup>	8.20	3.98	4.09	1.29		
Gly <sup>9</sup>	8.67	3.59–4.05				
Arg <sup>10</sup>	7.68	4.58	2.14	1.70–1.93	3.30–3.48	H $^{\epsilon}$ = 8.81
Asn <sup>11</sup>	8.46	4.61	2.97–3.06		6.95–7.69	
Ile <sup>12</sup>	8.51	3.69	1.86	CH <sub>2</sub> = 1.11–1.95 CH <sub>3</sub> = 0.86	0.92	
Tyr <sup>13</sup>	8.98	3.67	3.13			H <sup>2,6</sup> = 6.89 H <sup>3,5</sup> = 6.83
Asn <sup>14</sup>	9.01	4.19	2.74–3.08		6.98–7.80	
Thr <sup>15</sup>	8.36	3.95	4.25	1.28		
Cys <sup>16</sup>	8.04	4.20	3.02–3.29			
Arg <sup>17</sup>	8.57	4.01	1.63–1.79	0.89–1.18	2.47–3.03	H $^{\epsilon}$ = 7.23 H $^{\xi}$ = 6.60–7.04
Phe <sup>18</sup>	8.72	4.39	3.26			H <sup>2,6</sup> = 7.29 H <sup>3,5</sup> = 7.34
Ala <sup>19</sup>	7.42	4.38	1.57			
Gly <sup>20</sup>	7.80	3.70–4.37				
Gly <sup>21</sup>	8.25	3.44–4.11				
Ser <sup>22</sup>	8.57	4.35	4.17–4.41			
Arg <sup>23</sup>	9.11	3.83	1.72–1.76	1.62–1.87	3.05–3.26	H $^{\epsilon}$ = 7.74
Glu <sup>24</sup>	8.69	3.98	1.95–2.07	2.35–2.45		
Arg <sup>25</sup>	8.04	4.03	1.83–1.88	1.47–1.64	3.14–3.26	H $^{\epsilon}$ = 7.39
Cys <sup>26</sup>	8.75	4.65	2.36–2.71			
Ala <sup>27</sup>	9.38	3.95	1.52			
Lys <sup>28</sup>	7.53	4.10	2.03–2.05	1.50–1.59	1.72	H $^{\epsilon}$ = 3.00
Leu <sup>29</sup>	7.91	4.19	1.70–1.73	1.68	0.88–0.92	
Ser <sup>30</sup>	7.65	4.30	3.76–4.74	5.85		
Gly <sup>31</sup>	8.03	4.39–3.95				
Cys <sup>32</sup>	8.04	4.96	2.38–2.98			
Lys <sup>33</sup>	9.00	4.50	1.05	0.70–1.28	0.94–1.10	H $^{\epsilon}$ = 2.42–2.49
Ile <sup>34</sup>	8.51	4.55	1.94	CH <sub>2</sub> = 1.09–1.33 CH <sub>3</sub> = 0.72	0.62	
Ile <sup>35</sup>	8.81	4.68	1.95	CH <sub>2</sub> = 1.00–1.32 CH <sub>3</sub> = 0.77	0.71	
Ser <sup>36</sup>	8.77	4.60	3.87			
Ala <sup>37</sup>	7.24	4.51	1.53			
Ser <sup>38</sup>	8.44	4.37	3.93–3.96			
Thr <sup>39</sup>	7.25	4.45	3.94	1.16		
Cys <sup>40</sup>	8.93	4.77	2.57–3.80			
Pro <sup>41</sup>		4.66	2.35–2.39	2.02–2.19	3.66–3.80	
Ser <sup>42</sup>	8.81	4.02	3.91			
Asp <sup>43</sup>	8.74	4.47	2.67			
Tyr <sup>44</sup>	7.66	4.36	2.36–2.45			H <sup>2,6</sup> = 6.76 H <sup>3,5</sup> = 6.88
Pro <sup>45</sup>		4.46	2.00	1.62–2.02	3.35	
Lys <sup>46</sup>	8.31	4.34	1.15	1.77	1.47–1.60	H $^{\epsilon}$ = 2.91–2.98 H $^{\xi}$ = 7.61

20–22), while the short arm is formed by the antiparallel  $\beta$ -sheet together with the C-terminal region. The overall topology of viscotoxin C1 is very similar to viscotoxin A3 (7) (RMSD<sub>bb 2–45</sub> = 1.20 Å) and A2 (11) (RMSD<sub>bb 2–45</sub> = 1.38 Å) and to those of other thionins previously reported (27–30).

**Electrostatic Properties.** Electrostatic potential maps for viscotoxins A3, C1, A1, A2, B, and 1-PS have been computed. Viscotoxins A3, A1, and C1 bear a 6.0 unit net charge, viscotoxins B and 1-PS bear a 5.0 unit net charge, and the A2 isoform bears a 4 unit net charge. The six viscotoxins share a common pattern of electrostatic potential at the surface: a net positive potential is differently distributed on the proteins surface, defining two molecular

faces. One face of the molecule is characterized by a high positive electrostatic potential that corresponds to the surface enclosed by the two  $\alpha$ -helices, while the other face displays a less extended positive electrostatic potential. The electrostatic potential at the surface is shown only for the A3 and C1 isoform (Figure 5) since the other structures show a similar pattern of charge.

**Cytotoxicity Assay.** To correlate structural and electrostatic properties of the different viscotoxin isoforms with their biological activity against tumors, their cytotoxicity was analyzed by measuring the inhibition of <sup>3</sup>H-thymidine incorporation into *Y. sarcoma* cells, as previously described (8, 9). The toxicity values obtained were the following: viscotoxin A3, ED<sub>50</sub> = 0.068  $\mu$ M; 1-PS, ED<sub>50</sub> = 0.089  $\mu$ M;

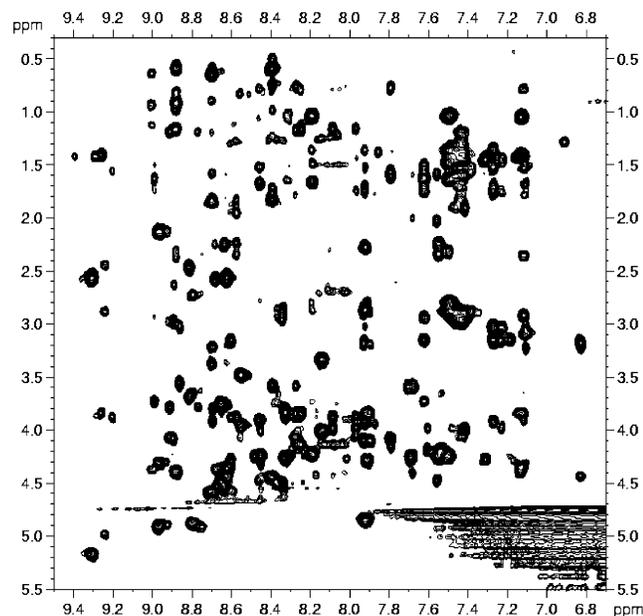


FIGURE 2: Fingerprint region of 500 MHz  $^1\text{H}$  TOCSY spectrum (isotropic mixing 50 ms) of viscotoxin C1 (1 mM) in aqueous 50 mM  $\text{H}_3\text{PO}_4/\text{NaOH}$  buffer ( $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ , 90:10, v/v), pH 3.6, 285 K.

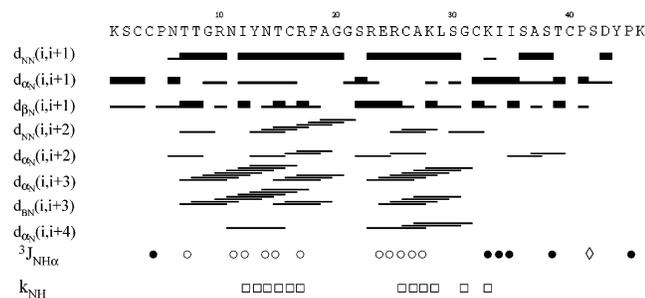


FIGURE 3: Summary of NMR data used to deduce viscotoxin C1 secondary structure. The notation  $d_{NN}$ ,  $d_{\alpha N}$ ,  $d_{\beta N}$ ,  $d_{\alpha\beta}$  ( $i, j$ ) indicates the NOE intensity observed for the protons  $\text{H}^N$ ,  $\text{H}^\alpha$ , and  $\text{H}^\beta$  of residue  $i$  with  $\text{H}^N$  and  $\text{H}^\beta$  of residue  $j$ ; the thickness of the bar indicates the relative normalized intensity of the NOE. The  $^3J_{\text{NH}\alpha}$  are three-bond coupling constants between  $\text{H}^N$  and  $\text{H}^\alpha$ , where the symbols represent the following:  $\circ$  helical values  $< 4.0$  Hz,  $\bullet$   $\beta$ -sheet-like values  $> 8$  Hz, and  $\diamond$  values in the range of 4–8 Hz. Exchange rates,  $k_{\text{NH}}$ , are reported in the last line to identify residues with sufficiently slow amide exchange rates to enable observation of the  $\text{H}^N$ - $\text{H}^\alpha$  cross-peaks in a TOCSY spectrum recorded 144 h after exchanging the buffer against a deuterated solution at pH = 3.6.

A1,  $\text{ED}_{50} = 0.177 \mu\text{M}$ ; C1,  $\text{ED}_{50} = 0.198 \mu\text{M}$ ; A2,  $\text{ED}_{50} = 0.223 \mu\text{M}$ ; and B,  $\text{ED}_{50} = 0.942 \mu\text{M}$  (Figure 1). The data show that the newly identified C1 isoform displays an intermediate toxicity with respect to the most toxic A3 and IPS isoforms.

## DISCUSSION

The structural comparison of viscotoxins C1, A3, and A2 allows for the identification of a conserved hydrophobic cluster involving residues 12 and 13 (helix I), 27 and 29 (helix II), and 34 (C-terminal  $\beta$ -strand). These residues are conserved in all viscotoxin sequences, thus suggesting that this hydrophobic cluster bears some relevance for viscotoxin topology, stabilizing the packing of the two helices and of the  $\beta$ -sheet. The relevant role of this cluster is further underlined by its conservation in the other thionins of known structure.

Table 2: Conformational Restraints and Structural Parameters Determined Using DYANA

Restrains		
upper distance restraints		631
hydrogen bond restraints		24
torsion angle ( $\phi$ ) restraints		18
DYANA (20 Best Structures)		
target function ( $\text{\AA}^2$ )		$0.45 \pm 0.05$
average number of upper restraints violations $> 0.2 \text{\AA}/\text{structure}$		0.0
maximum violation ( $\text{\AA}$ )		0.2
average number of angle restraints violations $> 5^\circ/\text{structure}$		0.0
$\langle \text{RMSD} \rangle$ (2–45, backbone atoms)		$0.7 \pm 0.2$
$\langle \text{RMSD} \rangle$ (2–45, heavy atoms)		$1.3 \pm 0.2$
Discover (AMBER Force Field) (10 Best Structures)		
total energy (kcal/mol)		$-202 \pm 6$
bond energy (kcal/mol)		$5.7 \pm 0.2$
angle energy (kcal/mol)		$38 \pm 1$
torsion energy (kcal/mol)		$49 \pm 2$
out of plane energy (kcal/mol)		$0.9 \pm 0.1$
hydrogen bond energy (kcal/mol)		$23 \pm 1$
lennard-Jones energy (kcal/mol)		$-157 \pm 2$
Coulomb energy (kcal/mol)		$-115 \pm 3$
restraining potential energy (kcal/mol)		$8 \pm 1$
average number of upper restraints violations $> 0.2 \text{\AA}/\text{structure}$		$5 \pm 3$
maximum violation ( $\text{\AA}$ )		0.4
average number of angle restraints violations $> 5^\circ/\text{structure}$		0
maximum violation (degree)		0
$\langle \text{RMSD} \rangle$ (2–45, backbone atoms) ( $\text{\AA}$ )		$0.57 \pm 0.16$
$\langle \text{RMSD} \rangle$ (2–45, heavy atoms) ( $\text{\AA}$ )		$1.22 \pm 0.17$

It is worth mentioning that the two helices in viscotoxins are not amphipathic, and like other peptides with this same characteristic (such as, e.g., magainin1), display no haemolytic activity (31). It has been shown that while the N-terminal amphipathic helical structure is not required for the cytolytic activity toward negatively charged membranes and bacterial cells, it appears to be crucial for binding and insertion into zwitterionic membranes and for haemolytic activity (32). Haemolytic activity is clearly not a desirable feature in therapeutic peptides.

The presence of extended high positive potential areas at the molecule surface of all the studied viscotoxins hints at possible interaction sites with molecules bearing negative surface potential, such as the heads of cell membrane phospholipids. Indeed crambin, the only nontoxic thionin, bears a net unit charge equal to zero. The high isoelectric point of viscotoxins ( $\text{pI} > 9.0$ ) ensures that the proteins remain charged at neutral pH and therefore will interact more strongly with negatively charged membrane phospholipids, such as phosphatidylserine membranes, rather than zwitterionic or neutral membranes.

We have previously suggested, from the comparison of the electrostatic properties of highly active viscotoxin A3 and the inactive crambin, that the loss of a positively charged residue at position 23 in crambin, together with the lack of other positively charged residues in the C-terminal region (namely, lysines 28, 33, and 46), are responsible for the absence of antimicrobial activity (7). A possible role in modulating viscotoxin biological activity may be played by residue 28, which is the only one among the four charged side chains exhibiting charge mutation in the different viscotoxin isoforms. However, even if the electrostatic

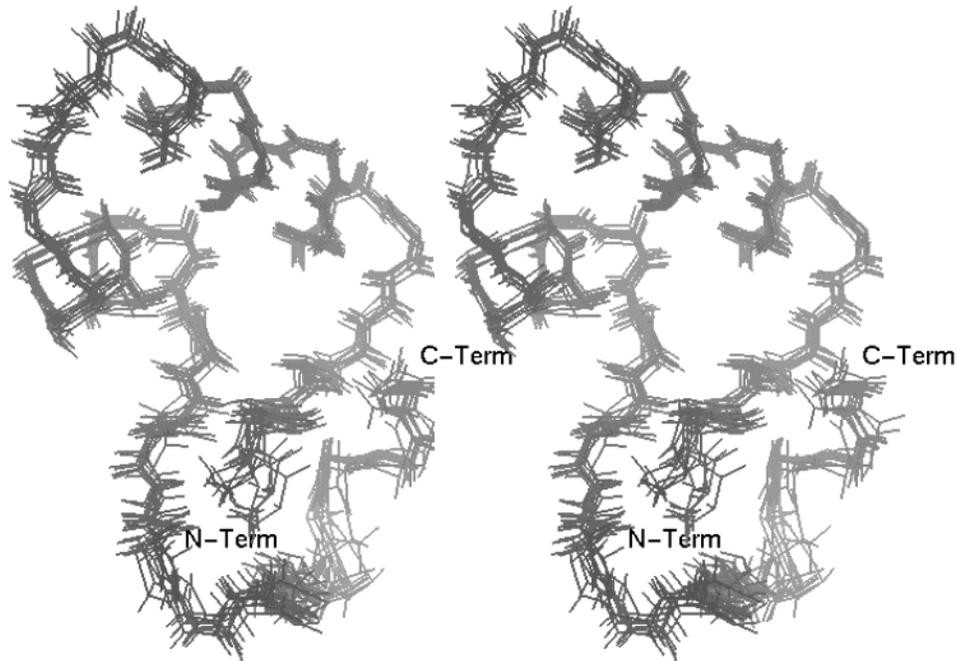


FIGURE 4: Superposition (stereoview) of the 10 final structures of viscotoxin C1 (PDB code: 1ORL),  $\text{RMSD}_{\text{bb } 2-45} = 0.57 \pm 0.16 \text{ \AA}$ .

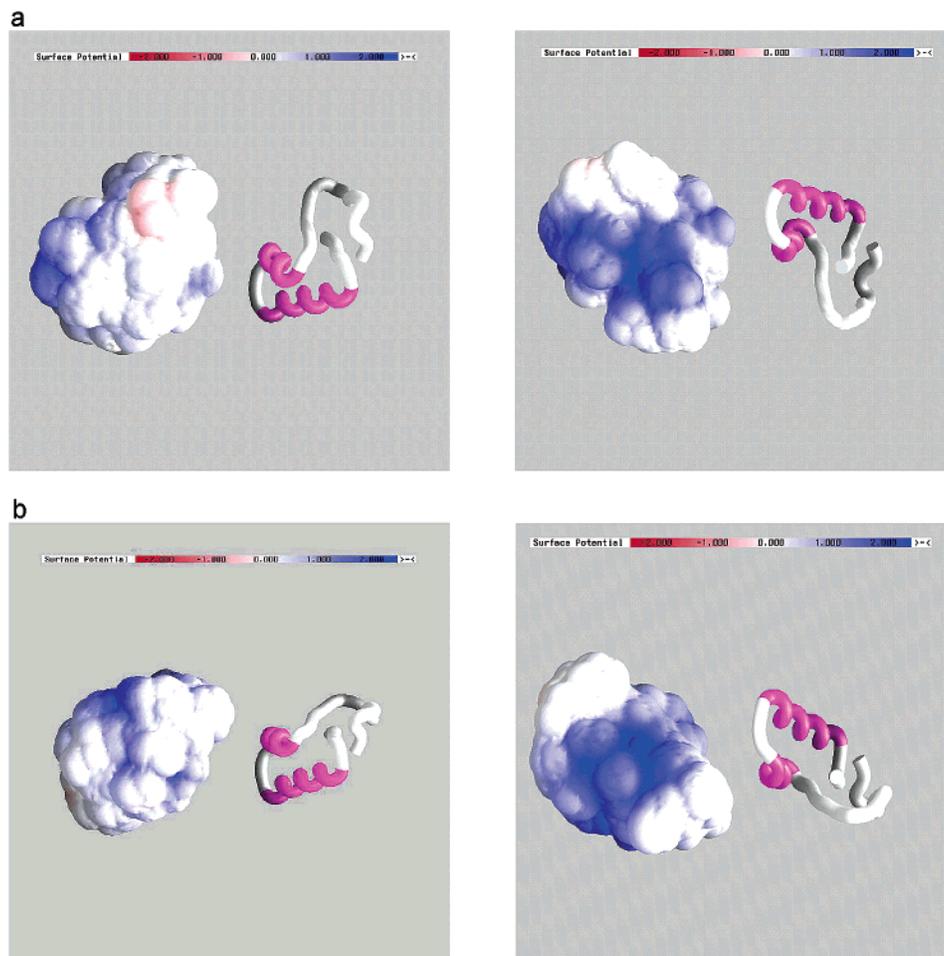


FIGURE 5: Electrostatic potential maps calculated for viscotoxins C1 (a) and A3 (b). Two views of the molecule are represented in the left and right panels.

surface properties of viscotoxins may provide structural evidence for an electrostatically driven interaction with negatively charged molecular targets, they are not sufficient to be assessed for the different levels of toxicity displayed,

and a possible explanation should be searched in single residue mutations. Comparison of viscotoxin sequences (Figure 1) reveals that only 11 positions exhibit mutations (positions 15, 18, 19, 21, 22, 24, 25, 28, 29, 37, and 43),

mostly localized in the region of 15–29, including the C-terminal end of the first  $\alpha$ -helix, the second  $\alpha$ -helix, and the connecting loop. No mutation ever occurs in the regions involved in the formation of the short  $\beta$ -strands (1–3, 33–35). Residues at positions 18, 29, and 43, exhibiting conservative mutations, were not considered relevant for the modulation of cytotoxicity. In view of the similar toxicity ( $ED_{50}$  in Figure 1) exhibited by the A3 and 1PS isoforms, residues 15, 22, and 37 (mutated only in viscotoxin A3) should not play a key role for the activity. The analysis of the role of specific mutations will therefore include only residues 19, 21, 24, 25, and 28. Interestingly, when pairs of sequences exhibiting the lowest number of mutations (three) are considered (i.e., the pairs C1/B, 1PS/A2, and A2/B), the relative role of every mutation can be discussed. Only one nonconservative mutation (K28S) is observed for the pair C1/B ( $ED_{50}C1 \sim 5 ED_{50}B$ ), underlying the relevance of a positively charged residue at position 28. Two nonconservative mutations (E24Q, R28S) are observed for the pair 1PS/A2, where only a 2.5-fold decrease in activity is measured, suggesting that the lack of a negatively charged residue at position 24 may enhance the toxicity. The 4-fold decrease in activity on going from the A2 to the B isoform can be ascribed to the two nonconservative mutations Q24E and V25R, where a further decrease in activity is possibly introduced by a positively charged residue at position 25. When the comparison is extended to the pair 1PS/B, where a 10-fold decrease in activity was measured, only two (V25R and R28S) out of the observed four mutations (F18L, V25R, R28S, and I29L) are nonconservative, strengthening our observation on the relevance of a noncharged residue at position 25 and of a positively charged residue at position 28. It is more difficult to account for the similar activity displayed by the A1/C1 pair where four out of five mutations are nonconservative. It would be, however, interesting to determine the A1 structure to clarify the relevance of the type and orientation of the side chain at position 25.

In a recent study, NMR and Fourier transform infrared spectroscopy have been used to show that  $\beta$ -purothionins, structurally similar to viscotoxins, strongly modify the lipid packing at the surface of the bilayer through electrostatic interactions between the cationic protein and the anionic phospholipid (33). Indeed, when the alignment analysis of viscotoxin isoforms is extended to related  $\alpha$ - and  $\beta$ -thionins (Figure 1), charged residues are always located at positions 1, 10, 17, 23, 33, and 46, with the only exception being the inactive crambin. Although few studies have compared the in vitro activity of viscotoxins with that of other thionins, viscotoxins are known to be more toxic to eukaryotic cells (34). For example,  $ED_{50}$  values measured on human tumor HeLa cells were in the range of  $0.2\text{--}1.7 \mu\text{g mL}^{-1}$  for viscotoxins and  $17 \mu\text{g mL}^{-1}$  for *Pyricularia* thionin. The higher viscotoxin toxicity nicely correlates with the presence of a charged residue at position 28, as discussed before.

The discussed key residues 19, 21, 24, 25, and 28 are highlighted on the superimposed C1, A3, and A2 viscotoxin structures (Figure 6a). It is clear from the figure that the most relevant residues 24, 25, and 28 are localized on the solvent exposed face of the second helix, which could behave as a recognition site for membrane interaction. In the three proteins, basic amino acids lie on one side of a plane roughly defined by the two helices (Figure 6b). The only exception

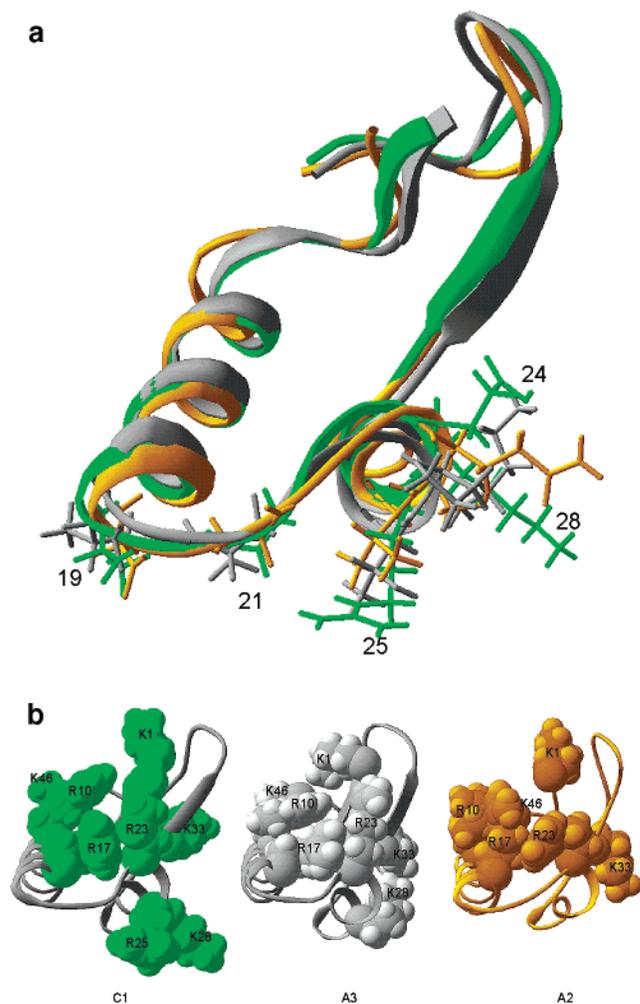


FIGURE 6: Comparison of C1 (green), A3 (dark gray), and A2 (orange) viscotoxin structures: (a) side chains of 19, 21, 24, 25, and 28 residues are shown and (b) basic amino acid side chains are shown as CPK.

is represented by R25 in viscotoxin C1, which points to the opposite side of the plane. Some have recently suggested that the orientation of the R25 side chain, protruding outside the plane formed by the two  $\alpha$ -helices of viscotoxin B, probably acts as an arm that keeps the toxin far from the lipidic interface, thus preventing its insertion into the membrane bilayer (11), in agreement with our observation on the requirement of a noncharged residue at this position.

To further correlate the observed mutations with biological activity, the following considerations may be added. It has been recently suggested (34), on the basis of fluorescence polarization measurements, that the most active viscotoxin A3 isoform has a high affinity for the most acidic phosphatidylserine (PS) membranes and binds parallel to the surface of the membrane in a so-called carpet-like fashion, perturbing the PS structure through a stiffening effect (35). It is worth mentioning that several types of cancer cells, as well as cells undergoing apoptosis, generally have a significant proportion of PS headgroups on the outer leaflet of their membranes. This is a rare occurrence in healthy eukaryotic cells, where lipid asymmetry is actively maintained, and amino phospholipids, such as PS, occur only on the inner leaflet of the membrane (36). PS headgroups bear a net negative charge, thus offering a partial explanation for the

observation that cationic antimicrobial peptides, such as viscotoxins, also have antitumoral properties.

The results reported here provide an overall coherent framework for clarifying the mechanism of the action of viscotoxins and open the way to a rational site-directed mutagenesis approach.

## ACKNOWLEDGMENT

We gratefully acknowledge the technical support offered by the Large Scale Facilities of Florence (CERM) and the Frankfurt Biophysik Centre.

## REFERENCES

- Orrù, S., Scaloni, A., Giannattasio, M., Urech, K., Pucci, P., and Schaller, G. (1997) Amino acid sequence, S–S bridge arrangement and distribution in plant tissues of thionins from *Viscum album*, *Biol. Chem.* 378, 986–996.
- Samuelsson, G., and Jayawardene, A. L. (1974) Isolation and characterisation of viscotoxin 1-PS from *Viscum album*, *Acta Pharm. Suec.* 11, 175–184.
- Bohlmann, H., and Apel, K. (1991) Thionins, *Annu. Rev. Plant Mol. Biol.* 42, 227–240.
- Bohlmann, H., Clausen, S., Behnke, S., Giese, H., Hiller, C., Ulrich, R. P., Shrader, G., Vibeke, B., and Apel, K. (1988) Leaf-specific thionins of barley: a novel class of cell wall proteins toxic to plant pathogenic fungi and possibly involved in the defense mechanism of plants, *EMBO J.* 7(6), 1559–1565.
- Huges, P., Dennis, E., Whitcross, M., Llewellyn, D., and Gage, P. (2000) The cytotoxic plant protein,  $\beta$ -purothionin, forms ion channels in lipid membranes, *J. Biol. Chem.* 275, 823–827.
- Thevissen, K., Ghazi, A., De Samblanx, G. W., Brownlee, C., Osborn, R. W., and Broekaert, W. F. (1996) Fungal membrane responses induced by plant defensins and thionins, *J. Biol. Chem.* 271, 15018–15025.
- Romagnoli, S., Ugolini, R., Fogolari, F., Schaller, G., Urech, K., Giannattasio, M., Ragona, L., and Molinari, H. (2000) NMR structural determination of viscotoxin A3 from *Viscum album L.*, *Biochem. J.* 350, 569–577.
- Schaller, G., Urech, K., and Giannattasio, M. (1996) Cytotoxicity of different viscotoxins and extracts from the European subspecies of *Viscum album L.*, *Phytother. Res.* 10, 473–477.
- Urech, K., Schaller, G., Ziska, P., and Giannattasio, M. (1995) Comparative study on the cytotoxic effect of viscotoxin and mistletoe lectin on tumor cells in culture, *Phytother. Res.* 9, 49–55.
- Konopa, J., Woynarowski, J. M., and Lewandowska-Gumieniak, M. (1980) Isolation of viscotoxins, cytotoxic basic polypeptides from *Viscum album L.*, *Hoppe-Seyler's Z. Physiol. Chem.* 361, 1525–1533.
- Coulon, A., Mosbah, A., Lopez, A., Sautereau, A., Schaller, G., Urech, K., Rougè, P., and Darbon, H. (2003) Comparative membrane interaction study of Viscotoxins A3, A2, and B from mistletoe (*Viscum album*) and connections with their structures, *Biochem. J.* 374, 71–78.
- Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) Multiple quantum filters for elucidating NMR coupling networks, *J. Am. Chem. Soc.* 104, 6800–6801.
- Bax, A., and Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, *J. Magn. Reson.* 65, 355–360.
- Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient tailored water suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments optimized to retain full sensitivity, *J. Magn. Reson., Ser. A* 102, 241–245.
- Hwang, T. L., and Shaka, A. J. (1995) Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients, *J. Magn. Reson., Ser. A* 112, 275–279.
- Bartels, C., Xia, T., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectra analysis of biological macromolecules, *J. Biomol. NMR* 5, 1–10.
- Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995)  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  Random Coil Chemical Shifts of the Common Amino Acids: I. Investigations of Nearest Neighbor Effects, *J. Biomol. NMR* 5, 67–81.
- Wüthrich, K. (1986) Three-dimensional protein structure by NMR, in *NMR of proteins and nucleic acids*, pp 176–199, John Wiley & Sons, Inc., New York.
- Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculations with the new program DYANA, *J. Mol. Biol.* 273, 283–298.
- Güntert, P. (1998) Structure calculation of biological macromolecules from NMR data, *Q. Rev. Biophys.* 31, 145–237.
- Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling, *Electrophoresis* 18, 2714–2723.
- Combet, C., Jambon, M., Deléage, G., and Geourjon, C. (2002) Geno3D: Automatic comparative molecular modeling of protein, *Bioinformatics* 18, 213–214.
- Fogolari, F., Brigo, A., and Molinari, H. (2002) The Piosson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology, *J. Mol. Recog.* 15, 377–392.
- Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Llin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., and McCammon, J. A. (1995) Electrostatic and diffusion of molecules in solution: simulations with the University of Houston Dynamics program, *Comput. Phys. Comm.* 91, 57–95.
- MacKerell, A. D., Jr., Bashford, D., Bellot, M., Dunbrak, R. L., Jr., Field, M. J., Fisher, S., Gao, J., Guo, H., Ha, S., and Joseph, D. et al. (1992) Self-consistent parametrization of biomolecules for molecular modeling and condensed phase simulations, *FASEB J.* 6, A143.
- Nicholis, A. (1993) *GRASP: Graphical representation and analysis of surface properties*, Columbia University, New York.
- Brünger, A. T., Clore, G. M., Gronenborn, A. M., and Karplus, M. et al. (1986) Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin, *Proc. Natl. Acad. Sci. U.S.A.* 83, 3801–3805.
- Clore, G. M., Sukumaran, D. K., Gronenborn, A. M., Teeter, M. M., Whitlow, M., and Jones, B. L. (1987) Nuclear magnetic resonance study of the solution structure of  $\alpha$ -purothionin. Sequential assignment, secondary structure, and low resolution tertiary structure, *J. Mol. Biol.* 193, 571–578.
- Lecomte, J. T. J., Jones, B. L., and Llinas, M. (1982) Proton magnetic resonance studies of barley and wheat thionins: structural homology with crambin, *Biochemistry* 21, 4843–4849.
- Lecomte, J. T. J., Kaplan, D., Thunberg, E., and Samuelsson, G. (1987) Proton magnetic resonance characterization of phoratoxins and homologous proteins related to crambin, *Biochemistry* 26, 1187–1194.
- Lankisch, P. G., and Vogt, W. (1971) Potentiation of hemolysis by the combined action of phospholipase A and a basic polypeptide containing an S–S bond, *Experientia* 2, 122–123.
- Corzo, G., Escoubas, P., Villegas, E., Barnham, K. J., He, W., and Norton, R. (2001) Characterization of unique amphipathic antimicrobial peptides from venom of the scorpion *Pandinus imperator*, *Biochem. J.* 359, 35–45.
- Richard, J. A., Kelly, I., Marion, D., Pezolet, M., and Auger, M. (2002) Interaction between  $\beta$ -Purothionin and Dimyristoylphosphatidylglycerol: a  $^{31}\text{P}$  NMR and Infrared Spectroscopic Study, *Biophys. J.* 83, 2074–2083.
- Coulon, A., Berkane, E., Sautereau, A. M., Urech, K., Rougè, P., and Lopez, A. (2002) Modes of membrane interaction of a natural cysteine-rich peptide: viscotoxin A3, *Biochim. Biophys. Acta* 1559(2), 145–159.
- Shai, Y. (1999) Mechanism of the binding, insertion, and destabilization of phospholipid bilayer membranes by  $\alpha$ -helical antimicrobial and cell nonselective membrane-lytic peptides, *Biochim. Biophys. Acta* 1462, 55–70.
- Vogel, H. J., Schibli, D. J., Jing, W., Lohmeier-Vogel, E. M., Epand, R. F., and Epand, R. M. (2002) Toward a structure–function analysis of bovine lactoferricin and related tryptophan- and arginine-containing peptides, *Biochem. Cell Biol.* 80, 49–63.

BI034762T

---

# Competitive binding of fatty acids and the fluorescent probe 1-8-anilino-naphthalene sulfonate to bovine $\beta$ -lactoglobulin

---

MADDALENA COLLINI,<sup>1,4</sup> LAURA D'ALFONSO,<sup>1,4</sup> HENRIETTE MOLINARI,<sup>2,3,4</sup>  
LAURA RAGONA,<sup>2,4</sup> MADDALENA CATALANO,<sup>2,4</sup> AND GIANCARLO BALDINI<sup>1,4</sup>

<sup>1</sup>Dipartimento di Fisica, Università degli Studi di Milano-Bicocca, 20126 Milano, Italy

<sup>2</sup>Istituto per lo Studio delle Macromolecole, Consiglio Nazionale delle Ricerche (CNR), 20131 Milano, Italy

<sup>3</sup>Dipartimento Scientifico e Tecnologico, Università di Verona, Ca' Vignal 1, 37134 Verona, Italy

<sup>4</sup>Istituto Nazionale per la Fisica della Materia, INFN, UdR Milano-Bicocca, 20126 Milano, Italy

(RECEIVED January 24, 2003; FINAL REVISION April 14, 2003; ACCEPTED May 16, 2003)

## Abstract

The use of spectroscopy in the study of fatty acids binding to bovine  $\beta$ -lactoglobulin (BLG) appears to be a difficult task, as these acid compounds, assumed as the protein natural ligands, do not exhibit favorable optical response such as, for example, absorption or fluorescence. Therefore, the BLG fatty-acid equilibrium has been tackled by exploiting the competition between fatty acids and ANS, a widely used fluorescent hydrophobic probe, whose binding sites on the protein have been characterized recently. Two lifetime decays of the ANS–BLG complex have been found; the longer one has been attributed to the internal binding site and the shorter one to the external site. At increasing fatty acids concentration, the fractional weight associated with ANS bound to the internal site drops, in agreement with a model describing the competition of the dye with fatty acids, whereas the external site occupancy appears to be unaffected by the fatty acids binding to BLG. This model is supported by docking studies. An estimate of the acid-binding affinities for BLG has been obtained by implementing the fitting of the bound ANS intensities with a competitive binding model. A relevant dependence has been found upon the solution pH, in the range from 6 to 8, which correlates with the calyx accessibility modulated by the conformation of the EF loop. Fatty acids with longer aliphatic chains (palmitate and laurate) are found to display larger affinities for the protein and the interaction free energy nicely correlates with the number of contacts inside the protein calyx, in agreement with docking simulations.

**Keywords:** Fatty acid;  $\beta$ -lactoglobulin; ANS; time resolved fluorescence; binding; docking

Bovine  $\beta$ -lactoglobulin (BLG), the major component of ruminant milk whey, is a 162 residues soluble protein belonging to the lipocalin family. BLG proper functionality has not been fully clarified yet, despite the great number of studies performed. However, its presence in the milk of several

mammals has suggested a nutritional role, supported also by its ability to interact with a great variety of hydrophobic ligands, such as retinol (Futterman and Heller 1972; Dufour et al. 1990; Dufour and Haertlé 1991; Cho et al. 1994; Narayan and Berliner 1997; Lange et al. 1998), fatty acids and triglycerides (Dufour et al. 1990; Frapin et al. 1993; Narayan and Berliner 1997; Wang et al. 1997a,b; Qin et al. 1998b; Wu et al. 1999; Zsila et al. 2002). The lipocalin family includes, among others, transport proteins such as the retinol-binding protein, the odorant-binding protein, and the major urinary protein, which all share the common structural feature of a  $\beta$ -barrel calyx, built from eight anti-parallel  $\beta$  sheets, arranged as an ideal site for hydrophobic

---

Reprint requests to: Giancarlo Baldini, Università degli Studi di Milano-Bicocca, Dipartimento di Fisica, Piazza della Scienza 3, 20126 Milano, Italy; e-mail: giancarlo.baldini@mib.infn.it; fax: 39-0264482894.

*Abbreviations:* BLG, bovine  $\beta$ -lactoglobulin; NMR, nuclear magnetic resonance; ANS, 1-8-anilino-naphthalene sulfonate; DMSO, dimethylsulfoxide; MIF, molecular interaction field; PDB, Protein Data Bank.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0304403>.

ligands (Brownlow et al. 1997). Among the above-cited ligands, fatty acids are the most abundant endogenous ligands of BLG, thus suggesting the importance of the study of their interaction with the protein. Some obstacles, however, are encountered in the direct study of their binding to the protein, as these acid compounds do not exhibit a convenient spectroscopic response (such as absorption, circular dichroism, or fluorescence). Measurements aimed at the determination of the fatty-acid-binding affinities have been performed by equilibrium dialysis or by an equilibrium partition method using labeled fatty acids (Spector and Fletcher 1970; Spector 1975; Anel et al. 1989; Pérez et al. 1992). Otherwise indirect procedures could be used following, for instance, how the response of other chromophores, intrinsic or bound to the protein, is affected by the presence of fatty acids. Several studies have been reported on changes of intrinsic (Tryptophan) fluorescence (Narayan and Berliner 1997, 1998; Wang et al. 1997b; Lange et al. 1998) or spectral shifts in the fluorescence emission of covalently bound probes (Richieri et al. 1992). However, the variations are often as small as 10% at most, thereby making difficult quantitative estimates of the binding constant. Structural studies of protein fatty-acid complexes, performed with X rays (Qin et al. 1998b; Wu et al. 1999) and NMR spectroscopy (Ragona et al. 2000) have given indications of the location of binding sites and of the protein residues that are involved in the interaction.

The present work deals with a novel investigation of the BLG fatty-acid binding, which exploits the competition between fatty acids and ANS, a widely used fluorescent hydrophobic probe whose binding sites to BLG have been characterized recently (Collini et al. 2000). We have chosen to follow the modification of the ANS fluorescence upon fatty-acid addition to a BLG-ANS solution in order to obtain information on (1) the preferred site for fatty-acid binding, once the locations of the ANS sites are known, and (2) the strength of the fatty-acid-BLG interaction. When performing a competition study, the probe ligand must be chosen in such a way that its binding affinity for the protein is large enough to ensure the observation of a meaningful fluorescent response of the bound dye, but also sufficiently low to allow dye displacement by the fatty acids. ANS has been chosen, as its binding characteristics fulfill the above requirements, its interaction constant lying around  $10^3 \text{ M}^{-1}$ , depending upon solution pH, but nevertheless, lower than the values reported for the endogenous ligands of BLG.

The binding parameters were derived under different solution conditions, in the pH range from 6–8. This neutral-to-alkaline region is particularly interesting, as the charges on the protein modulate the accessibility to its hydrophobic calyx, an ideal site for small hydrophobic ligands. In fact, the EF loop, sensitive to pH, bends over the calyx entrance in a closed conformation when in acid solutions, whereas, at

pH >7.5, it moves away from the calyx (Tanford transition), assuming an open conformation that favors ligand binding (Tanford et al. 1959; Qin et al. 1998a).

To investigate the nature of the fatty acid interaction with BLG, we have studied saturated fatty acids with different chain lengths: palmitic acid, lauric acid, and caprylic acid, which share a similar chemical structure,  $\text{CH}_3(\text{CH}_2)_n\text{COOH}$ , but possess aliphatic chains of different lengths ( $n = 14, 10, 6$ ; see Fig. 1).

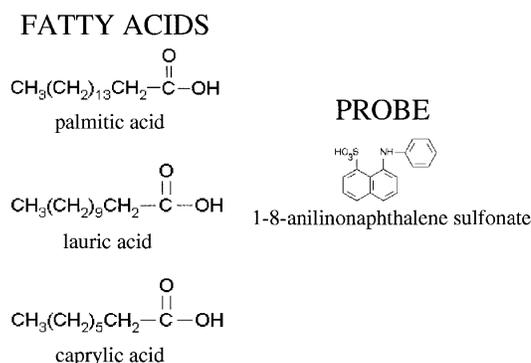
Time-resolved fluorescence measurements yielded ANS fluorescence decay parameters, lifetimes, and fractional intensities values, which have been fitted using a competition model in order to obtain the association constants for fatty-acid BLG interactions.

Fluorescence results have been accompanied by docking simulations performed using the program GRID.

## Results and Discussion

Steady-state fluorescence measurements of the ANS-BLG complex in the presence of fatty acids have been performed in order to identify the suitable range of protein and probe concentration causing the largest fluorescence change upon fatty acid addition. The useful concentrations ranged from 5 to 10  $\mu\text{M}$  for both ANS and BLG, depending upon the pH of the solution, which in turn affects the ANS-binding constant to BLG (Collini et al. 2000). To obtain efficient competition, the concentration of fatty acids was varied from 0 to 12  $\mu\text{M}$ .

Fluorescence lifetime measurements performed with the three different fatty acids in the pH range described above have been fitted to a three-exponential decay, in which the lifetime of free ANS in solutions has been fixed to  $\tau_F = 0.26 \text{ nsec}$ . ANS bound to BLG displays two different lifetimes ( $\sim 14 \text{ nsec}$  and  $\sim 3 \text{ nsec}$ ) whose values are only slightly dependent on pH and ionic strength, in agreement with previous results (D'Alfonso et al. 1999; Collini et al. 2000). These lifetimes have been attributed to two different BLG-binding sites; the shorter lifetime corresponds to a site partially exposed to the solvent and located on a surface



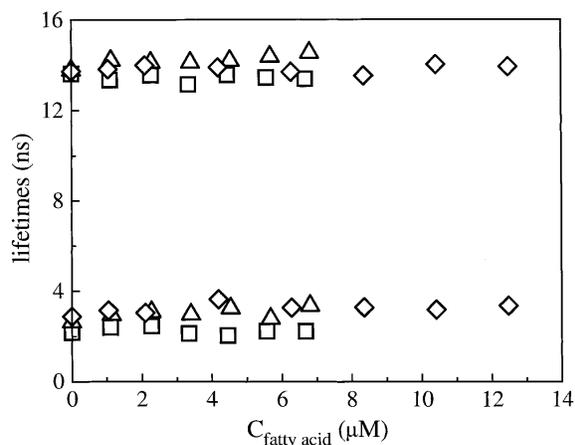
**Figure 1.** Chemical structure of the investigated ligands; fatty acids and ANS (1-8-anilinonaphthalene sulfonate).

hydrophobic patch (Ragona et al. 1997; Fogolari et al. 1998). The longer one corresponds to a binding site shielded from the solvent and located inside protein calyx, in which the probe feels a nonpolar environment, thus decaying with a longer lifetime value.

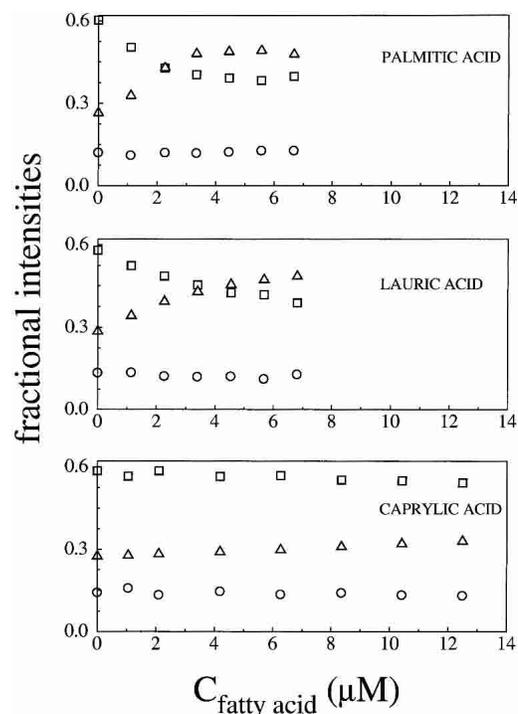
Fluorescence lifetime values of bound ANS are essentially constant, within the experimental error affecting the two exponential analysis, under all of the examined conditions, and they are independent of fatty-acid concentration. In Figure 2, for example, are shown the fluorescence lifetimes of bound ANS versus fatty-acid concentration for the titrations performed at pH 8.3 with palmitic, lauric, and caprylic acid, whereas Figure 3 reports the corresponding fractional intensities. The observation that the lifetimes of bound ANS are unaffected by fatty acids suggests that the presence of fatty acid does not appreciably alter the polarity of ANS-binding sites. It is clear from Figure 3 that the fractional intensities associated with the bound ANS longer lifetime,  $f_1$ , and with the free ANS,  $f_3$ , change at increasing fatty-acid concentration, whereas the fractional intensity associated with the bound ANS shorter lifetime,  $f_2$ , remains essentially constant. The extent of these changes appear to be related to the specific fatty acid under study, being largest for palmitic acid, and also to the solution pH, being larger at the higher pH values.

To derive from the fractional intensities quantitative information on how the ANS bound concentration changes upon fatty-acid addition, one can introduce the ratio of the two bound ANS fractional intensities,  $f_1/f_2$ . In fact, as reported in Materials and Methods (Equation 6), the fractional intensities depend upon the lifetime, the corresponding dye concentration, and an unknown normalizing factor, which can cancel out when taking the ratio  $f_1/f_2$ :

$$\frac{f_1}{f_2} = \frac{\varepsilon_1 C_{b1} \tau_1 k_1^R}{\varepsilon_2 C_{b2} \tau_2 k_2^R} \quad (1)$$



**Figure 2.** Fluorescence lifetimes of ANS bound to BLG versus fatty-acid concentration at pH 8.3. (□) Palmitic acid; (△) lauric acid; (◇) caprylic acid.



**Figure 3.** Fractional intensities of ANS decay in the presence of BLG versus fatty-acid concentration at pH 8.3. Different symbols refer to different fractional intensities. (□)  $f_1$ , corresponding to the longer bound ANS lifetime; (○)  $f_2$ , corresponding to the shorter bound ANS lifetime; (△)  $f_3$ , corresponding to the free ANS lifetime.

The indexes 1 and 2 refer to BLG internal and external sites for ANS, respectively, as stated above. In view of the fact that the extinction coefficient of bound ANS was reported to be almost the same for the two sites (D'Alfonso et al. 1999), and the ANS lifetimes are found to be constant, it is reasonable to assume that the radiative lifetimes do not change (Robinson et al. 1978; Lakowicz 1999). The fractional intensity  $f_2$  is constant as well, thus suggesting that the concentration of ANS bound to the external site does not appreciably change upon fatty-acid binding.

Equation 1 can be rewritten by including all the constant terms in a proportionality factor as:

$$f_1 = (\text{const}) \times C_{b1} \quad (2)$$

In this way, it is possible to relate the experimental fractional intensity to the concentration of ANS bound to the internal site. It is worth noting that the decrease of  $f_1$  at increasing fatty-acid concentration is a strong indication that fatty acids compete with ANS only for the internal binding site. This finding is in good agreement with X-ray and NMR data on the complexes of BLG with dodecanoic and palmitic acids (Qin et al. 1998b; Wu et al. 1999; Ragona et al. 2000).

The aim of the following analysis is the determination of the fatty-acid-binding constant to BLG in terms of the

known ANS-binding constant (Collini et al. 2000), by using a competitive model only for the BLG internal site. The concentration of ANS bound to the external site can be taken into account in the analysis as a renormalization factor of the total ANS concentration; nonetheless, this correction affects the value of the free ANS concentration within the experimental precision of its determination.

When both ANS and fatty acids are present in BLG solution, the following equilibria are present,



in which P, A, and F represent the free protein, ANS, and fatty-acid concentrations, P<sub>A</sub> and P<sub>F</sub> are the concentrations of protein with bound ANS or fatty acid, and K<sub>A</sub> and K<sub>F</sub> are the ANS and fatty-acid-binding constants, respectively. Then P<sub>0</sub>, A<sub>0</sub>, and F<sub>0</sub> represent the total protein, ANS, and fatty-acid concentrations, respectively, and can be expressed as follows:

$$P_0 = P + P_A + P_F$$

$$A_0 = A + P_A$$

$$F_0 = F + P_F$$

The concentration of ANS bound to the protein, P<sub>A</sub>, is the sum of ANS bound to both the external and the internal site of BLG:

$$P_A = C_{b1} + C_{b2}$$

in which the indexes 1 and 2 refer to the internal and external BLG-binding sites, respectively. Because the two ANS-binding affinities (K<sub>1</sub> and K<sub>2</sub>) are known, and C<sub>b2</sub> is constant, it is possible to define the new total dye concentration, A<sub>0</sub><sup>\*</sup> as:

$$A_0^* = A_0 - C_{b2}$$

In this way, an equation relating the unknown fatty-acid-binding constant, K<sub>F</sub>, to the concentration of ANS bound to the internal site, C<sub>b1</sub>, can be derived:

$$C_{b1} = K_A \cdot A \left[ P_0 - A_0^* + A - \frac{F_0 \cdot K_F (A_0^* - A)}{K_A \cdot A + K_F (A_0^* - A)} \right] \quad (3)$$

$$= K_A \cdot (A_0^* - C_{b1}) \left[ P_0 - C_{b1} - \frac{F_0 \cdot K_F C_{b1}}{K_A \cdot (A_0^* - C_{b1}) + K_F C_{b1}} \right]$$

By solving Equation 3 for C<sub>b1</sub>, a cubic expression is obtained in C<sub>b1</sub>, according to:

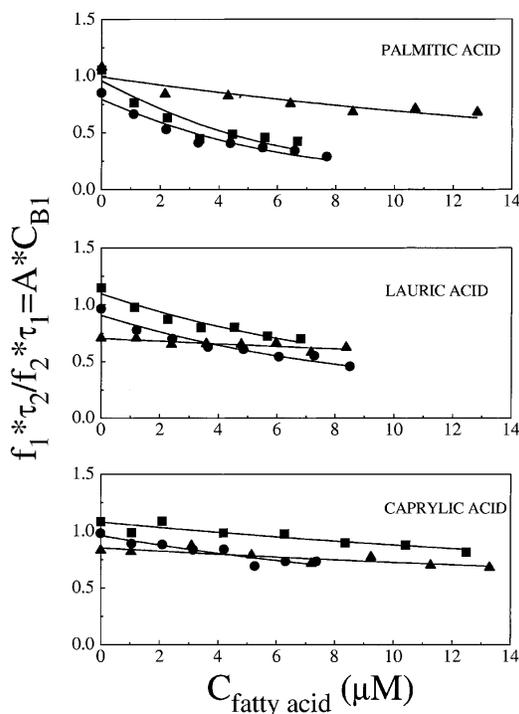
$$C_{b1}^3 \cdot K_A (K_F - K_A) + C_{b1}^2 [(K_A - K_F) + A_0^* \cdot K_A (K_A - K_F) + P_0 \cdot K_A (K_A - K_F) + K_A (A_0^* \cdot K_A + F_0 \cdot K_F)] + C_{b1} [P_0 \cdot A_0^* \cdot K_A (K_F - K_A) - A_0^* \cdot K_A (A_0^* \cdot K_A + F_0 \cdot K_F) - A_0^* \cdot K_A (1 + P_0 \cdot K_A)] + P_0 \cdot A_0^2 \cdot K_A^2 = 0 \quad (4)$$

The amount of ANS initially bound to the external site (C<sub>b2</sub>) can be neglected in the expression of the total protein concentration P<sub>0</sub>, as it induces a change in P<sub>0</sub> of ~0.5%, at most. Similarly, the variation of A<sub>0</sub> due to the free ANS initially bound in the internal site (at most equal to C<sub>b1</sub>) can be neglected in the computation of both C<sub>b1</sub> and C<sub>b2</sub>, because the maximum induced A<sub>0</sub> variation would be of about ≈0.0002 μM, corresponding to a negligible Δf<sub>1</sub>≈0.7%. Therefore, the fractional intensity associated to the longer lifetime component can be fitted by Equation 2. However, it has been often observed that fit parameters, in a three-exponential decomposition of ANS lifetime decay, are not completely independent from each other. For this reason, instead of using the values of f<sub>1</sub> obtained from the fit, we use the ratio f<sub>1</sub>τ<sub>2</sub>/f<sub>2</sub>τ<sub>1</sub> = const' × C<sub>b1</sub>, thereby greatly reducing the noise on the derived parameters. The values of the fatty-acid association constant, K<sub>F</sub>, are shown in Table 1. The good agreement between the data and the fitting model is shown in Figure 4, in which f<sub>1</sub>τ<sub>2</sub>/f<sub>2</sub>τ<sub>1</sub> values and the fitting functions are shown versus fatty-acid concentration. It was not possible to determine the binding constant for caprylic acid, at pH 6.2 and pH 7.2, as the change in the fractional intensity of ANS bound to the internal site is too small; in this case, only an upper limit estimate can be given.

A few values for the association constant of palmitic acid to BLG at different pHs and ionic strengths are reported in the literature. Association constants of palmitic and lauric acids, reported in an earlier work (Spector and Fletcher 1970) are in good agreement with our data and display a similar pH dependence. A larger (about fourfold) association constant was reported recently by Narayan and Berliner (1998). Binding constants were also reported for different fatty acids, such as oleic (K<sub>F</sub> ≈ 0.4 × 10<sup>5</sup> M<sup>-1</sup>) and stearic (K<sub>F</sub> ≈ 4 × 10<sup>5</sup> M<sup>-1</sup>; Spector and Fletcher 1970) and for cholesterol (K<sub>F</sub> ≈ 3 × 10<sup>7</sup> M<sup>-1</sup>; Wang et al. 1997b) and have been summarized in Table 2 of the review by Sawyer and Kontopidis (2000).

**Table 1.** Binding affinities of the fatty acid studied in the different experimental conditions examined

K <sub>F</sub> (*10 <sup>2</sup> ) M <sup>-1</sup>	pH 6	pH 7	pH 8
Palmitic acid	660 ± 30	4700 ± 100	5050 ± 200
Lauric acid	235 ± 20	1630 ± 160	1350 ± 140
Caprylic acid	<3	<7	255 ± 50



**Figure 4.** Ratio  $f_1\tau_2/f_2\tau_1 = \text{const}' \times C_{b1}$  versus fatty-acid concentration at the different investigated pH values: (■) pH 8.3; (●) pH 7.2; (▲) pH 6.2. Broken lines refer to the fitting functions as defined in the text, Equations 4–6.

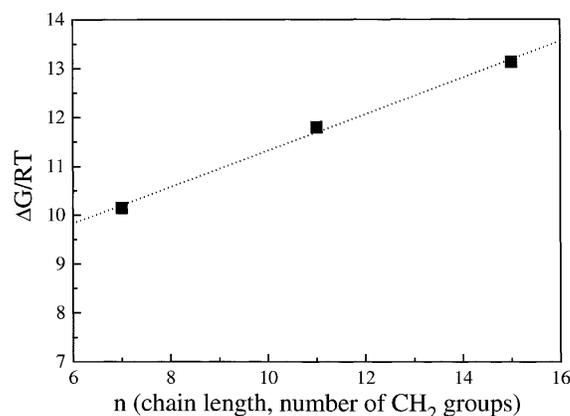
From the analysis of the association constants reported in Table 1, it is clear that near the Tanford transition, in the pH range 6.2–8.3, the strength of the interaction increases with pH, that is, it depends on the opening of the EF loop, regardless of the length of the fatty-acid chain. A strong dependence of the binding constant upon the fatty-acid chain length is also observed, with larger binding affinities for the longer palmitic acid. When the free energies ( $\Delta G$ ) for the binding of fatty acids to BLG are derived from the equilibrium constants at pH 8.3, in which the largest fractional intensity changes allowed for the determination of all the interaction constants, a linear relation of  $\Delta G$  versus the number of methylene groups of the alkylic fatty-acid chain is found, as shown in Figure 5. This free energy can be decomposed in a leading term ( $\Delta G_0$ ), which might depend upon the electrostatic interaction between the fatty-acid carboxylate and BLG lysines located at the calyx entrance (Wu et al. 1999), plus a contribution arising from the number of hydrophobic contacts between fatty-acid methyl and methylene groups ( $n_{\text{met}}$ ) and BLG residues within the calyx  $\Delta G = \Delta G_0 + \xi n_{\text{met}}$ . Data fitting yields  $\Delta G_0 = 7.6 \text{ RT}$  and  $\xi = 0.37 \text{ RT}$  per methylene with  $n_{\text{met}} = 7, 11, 15$  for caprylic, lauric, and palmitic acid, respectively.

To complement experimental fluorescence data with an analysis, at a molecular level, of the binding specificity, the three fatty acids and the ANS probe were docked within

BLG by use of the GRID-docking program (Goodford 1985, Kastenholz et al. 2000). The GRID method has been developed for determining energetically favorable binding sites for small chemical groups (probes) on a target molecule (protein). The probe groups are small chemical entities, such as carboxy oxygens, water molecules, and hydrophobic probes, which are moved through a regular grid of points around the target molecule in order to calculate, at each point of the grid, an interaction energy, thus generating a molecular interaction field (MIF). The ligand molecule is represented as a collection of GRID probes, and MIFs calculated for each probe are used to define the ligand position with respect to the target and to estimate the binding energy of the ligand at its binding site. The docked solutions are ranked on the basis of the total energy of interaction with the target, evaluated by GRID forcefield, to estimate the binding affinity of each solution.

Palmitic, lauric, and caprylic acids have been docked to the BLG structure obtained for the apo-protein at pH 7.1 (PDB code 1bsy), with the EF loop in the open conformation, as it was shown (Ragona et al. 2000) that this is a prerequisite for fatty-acid binding to BLG. At low pH, when the EF loop is in the closed conformation, the protein is unable to bind fatty acids.

The X-ray structure of BLG complexed with palmitic acid (Wu et al. 1999) indicated, in agreement with NMR data (Ragona et al. 2000), that the fatty acid lies within the central cavity of the protein, with the methyl end buried deeply within the protein and the carboxyl end, protruding outside BLG open end. The GRID docking program generates a single solution for palmitic acid, with an interaction energy of  $-14.8 \text{ Kcal/mole}$ . To assess the quality of the docking procedure, we compared the coordinates of the docked palmitic acid with the ones of the crystal structure (PDB code 1b0o; Fig. 6A); a very good agreement is observed for the positioning of both the aliphatic chain and the carboxyl tail. The inspection of MIFs, generated for hydro-



**Figure 5.** Free energy difference for fatty-acid binding to BLG at pH 8.3 versus the number of methylene groups in the fatty-acid chain.

phobic and carboxyl oxygen probe, graphically represented as three-dimensional contours around the target molecule, describing the probe attractive interaction regions, indicates that palmitic-acid docking solution maximizes both electrostatic and hydrophobic interactions (Fig. 6B). This indicates that both of these interactions are relevant for selectivity toward this ligand.

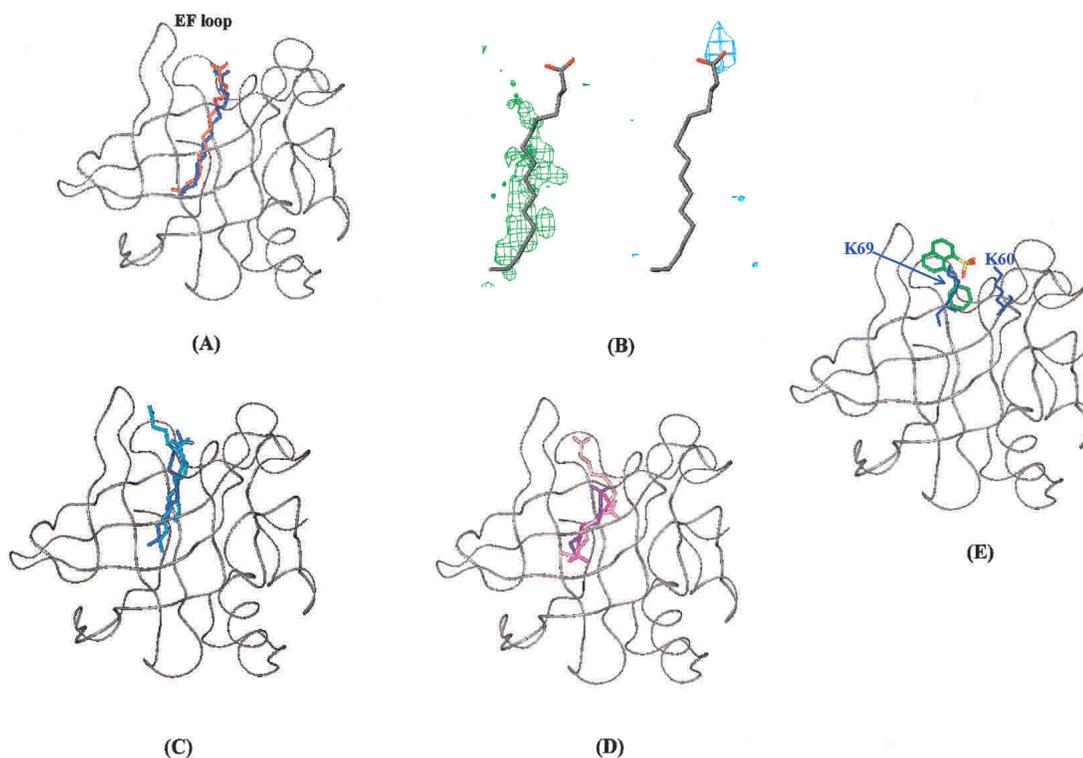
Three different solutions have been generated by GRID for lauric acid, the ligand is always located in the internal cavity with the carboxyl tail toward BLG open end (Fig. 6C). A careful analysis of the three predicted positions with respect to MIFs, indicated that the aliphatic chain always lies along BLG "hydrophobic spine," but only in the second docking solution, the carboxyl group is close to a favorable electrostatic interaction region. Moreover, this solution displays the highest similarity with the previously determined X-ray structure of 12-bromododecanoic acid complexed to BLG (Qin et al. 1998b).

The comparison of palmitic acid-binding energy ( $-14.8$  Kcal/mole), with the most reliable docked solution obtained for lauric acid ( $-11.9$  Kcal/mole), points to a lower affinity for this shorter fatty acid, in agreement with the fluores-

cence data showing a decrease of binding affinity with the shortening of the fatty-acid chain.

Finally, the shorter caprylic acid, docked to BLG, afforded four solutions (Fig. 6D) in which the aliphatic chains fit, for different extensions, the favorable hydrophobic interaction regions. The energy values obtained for the four solutions range from  $-13$  to  $-9$  Kcal/mole; however, the location of the carboxyl tail never fits the favorable electrostatic interaction regions. This observation suggests a very low-binding specificity of caprylic acid to BLG, possibly due to its minor sterical hindrance, conferring a large conformational freedom within a cavity that preferentially hosts the endogenous longer-chain fatty acids (Perez et al. 1989).

Concerning the GRID docking simulations relative to the BLG-ANS complex, three solutions were generated with energy values ranging from  $-6.2$  to  $-5.5$  Kcal/mole. The best docking solution, selected on the basis of MIFs analysis, presents ANS laying at the entrance of the cavity, with the aniline aromatic group pointing within the internal cavity and the negatively charged sulphonate group, in close contact with K60 and K69, nicely fitting the favorable elec-



**Figure 6.** Docking solutions obtained from GRID program for interaction of BLG with palmitic, lauric, caprylic acids, and ANS. (A) Superposition of predicted (blue) and experimentally observed (red) positions of palmitic acid within BLG. BLG backbone (PDB code 1bsy) is shown in gray in a ribbon representation. (B) Graphical representation of MIF in the region of palmitic acid binding site. Palmitic acid docking solution is shown in gray. The green and blue contours indicate regions of favorable interaction energies for hydrophobic (energy level =  $-0.5$  Kcal/mole) and carboxyl oxygen (energy level =  $-6.2$  Kcal/mole) probes, respectively. (C) Docking solutions obtained for lauric acid. (D) Docking solutions obtained for caprylic acid. Darker colors correspond to lower energy solutions. (E) Docking solution obtained for ANS. ANS molecule is colored by atom type. K60 and K69 side-chains are displayed in blue.

trostatic interaction field generated by GRID (Fig. 6E). The obtained energy values suggest a lower affinity of ANS with respect to fatty acids.

### Conclusions

The competition of ANS with fatty acids for a site on BLG, as detected here by time-resolved fluorescence of the dye during titration, has led to establish the strength of the protein to acids affinity. In particular, the fatty-acid-binding affinities values, which were shown to increase with pH in the range 6.2–8.3, independent of the acid-chain length, confirm the major role of the binding played by the EF loop conformation. Furthermore, palmitic acid, exhibiting the longest aliphatic chain, was shown to display the highest affinity for BLG, and a linear dependence of the binding-free energy versus fatty-acid chain length was observed, in agreement with the results of docking simulations performed using GRID. Both fluorescence data and docking simulations thus suggest the leading role of the electrostatic interactions in modulating the binding interaction energies. Docking solutions for the short caprylic acid, consistent with the very low affinity measured at all the tested pHs, were characterized by a large variability of carboxylate locations, and did not fit the favorable electrostatic interaction regions.

The docking solution obtained for ANS is in nice agreement with fluorescence data, indicating that the ANS probe lays within the protein (Collini et al. 2000). The comparison of the GRID energy figures of the investigated complexes indicate that the ANS-binding energy values are lower than those obtained for fatty acids, in agreement with fluorescence lifetime competition data, which suggest that fatty acids displace ANS.

It is worth mentioning that the higher specificity displayed by BLG for palmitic acid correlates well with biochemical data on the composition of fatty acids bound to BLG isolated from milk, in which palmitic acid is 30 times more abundant than lauric acid (Perez et al. 1989). This figure should be compared with that reported for cow milk fatty-acid content showing only an eightfold amount of palmitic acid, with respect to lauric acid (Belitz and Gosch 1999). The larger amount of palmitic acid complexed to BLG can be justified confidently by the stronger affinity of the protein for this ligand.

### Materials and methods

#### Sample preparation

Lyophilized bovine  $\beta$ -lactoglobulin, genetic variant B, (lot. 11K7032, Sigma-Aldrich Inc.) was dissolved in the proper buffer at the desired concentration determined photometrically using the molar extinction coefficient at the absorption peak,  $\epsilon(280) = 17600 \text{ cm}^{-1} \text{ M}^{-1}$ . Palmitic acid, lauric acid, and caprylic acid have been purchased from Sigma Chemical Co. and dissolved

(or diluted in the case of caprylic acid) in ethanol at a concentration of  $\sim 1 \text{ mM}$  (stock solutions). ANS (8-anilino-1-naphthalene-sulfonic acid) ammonium salt has been purchased from Fluka Chemical Co. Dye concentrations were determined photometrically using  $\epsilon(350) = 5000 \text{ cm}^{-1} \text{ M}^{-1}$ . Fresh BLG-ANS solutions were prepared before each measurement by adding to a BLG solution proper aliquots of a stock solution of ANS dissolved in the desired buffer, obtaining final probe and protein concentrations of 5–10  $\mu\text{M}$ , depending upon the particular experiment, as explained in the Results section.

Phosphates buffers used in the titration experiments were as follows: (1) 0.010 M  $\text{KH}_2\text{PO}_4\text{-Na}_2\text{HPO}_4$  at pH 6.2; (2) 0.010 M  $\text{KH}_2\text{PO}_4\text{-Na}_2\text{HPO}_4$  at pH 7.2; (3) 0.010 M  $\text{KH}_2\text{PO}_4\text{-Na}_2\text{HPO}_4\text{-NaOH}$  at pH 8.3. All of the reagents used in sample preparation were of analytical grade.

#### Fluorescence measurements

Fluorescence titration experiments with fatty acids were performed at  $25^\circ\text{C}$  by adding small aliquots (typically  $2\mu\text{L}$ ) of the fatty-acid stock solutions to a BLG-ANS solution, reaching a concentration ratio fatty acid:BLG  $>1:1$  in the presence of, at most, 1% w/w ethanol. Such a small fatty-acid total added volume allows us to consider the protein and ANS concentrations as constants during the titration; the concentration values used were 5 or 10  $\mu\text{M}$ , according to the solution pH value. Each titration has been performed at least twice.

Steady-state spectra were acquired on a Cary Eclipse (Varian Inc.) spectrofluorometer, recording the fluorescence signal between 380 and 650 nm after excitation at 363 nm.

Dynamic fluorescence measurements were performed with a frequency-modulated phase fluorometer (Digital K2, I.S.S.). The excitation was accomplished by the 363.7-nm line of an Argon ion laser at 30 mW power (2025, Spectra Physics). For further details, see Collini et al. 1995. Digital data acquisition and storage was provided by the ISS-A2D ACD card inserted in a personal computer. For each data set, at least 15 logarithmically spaced frequencies were used in the range 2–220 MHz with a cross-correlation frequency of 400 Hz. Each lifetime measurement has been repeated at least twice in order to obtain an estimate of the errors affecting the results. Phase angles and modulation ratios accuracy were of  $0.2^\circ$  and 0.004, respectively. Lifetime measurements have been performed under the magic angle conditions and a long pass filter at 435 nm (Andover Co.) was used to cut Rayleigh and Raman scattering. A solution of dimethyl-popop [1,4-bis(4-methyl-5-phenyloxazol-2-yl)benzene] in ethanol was used as a reference sample of known lifetime ( $\tau = 1.45 \text{ nsec}$ ). Data fitting was accomplished by means of a least square minimization procedures based on the Marquardt algorithm. Fluorescence lifetimes were analyzed in terms of sums of discrete exponential components, with the lifetimes values  $\tau_i$  and their corresponding fractional intensities  $f_i$  as unknown parameters, according to the equations:

$$I(t) = \sum_i x_i e^{-\frac{t}{\tau_i}} \quad \text{and} \quad f_i = \frac{x_i \cdot \tau_i}{\sum_j x_j \cdot \tau_j} \quad (5)$$

in which  $x_i$  represent the pre-exponential factors, which are proportional to the concentration ( $C$ ), to the molar extinction coefficient ( $\epsilon$ ) and to the radiative constant ( $k^R$ ) of the corresponding emitting chromophore, leading to:

$$f_i = \frac{\epsilon_i \cdot C_i \cdot \tau_i \cdot k_i^R}{\sum_j \epsilon_j \cdot C_j \cdot \tau_j \cdot k_j^R} \quad (6)$$

### Circular dichroism measurements

To test whether the BLG conformation was affected by the addition of ethanol or palmitic acid, circular dichroism (CD) experiments were performed with a Jasco (Easton) J-500A spectropolarimeter both in the near (350–250 nm) and in the far (250–190 nm) UV, by use of 1- and 0.2-cm path length cuvettes, 40- and 4- $\mu$ M protein concentrations, ethanol concentration about 1% w/w, and 12- $\mu$ M palmitic acid samples. In all of the tested experiments, the presence of a small quantity of ethanol or of the fatty acids used did not induce any modification in BLG ellipticity.

### GRID calculations

BLG X-ray coordinates, obtained for the apo-protein at pH 7.1 (PDB code 1bsy), were used for the docking program. Crystallographic water molecules were removed. The coordinates of the palmitic, lauric, and caprylic acids were derived from the coordinates of palmitic acid bound to BLG (PDB code 1b0o). The fatty acid carboxyl group was considered deprotonated, bearing a net negative charge of  $-1$ . ANS molecule was built with InsightII (Accelrys), and minimized with CVFF forcefield (200 steps steep-descent and 2000 steps of conjugate gradient).

The calculation was performed with version 20 of the GRID software (Molecular Discovery Ltd). The protein was considered rigid, and hydrogens were added with the program GRIN (part of GRID package). The docking search was performed on the whole protein. All GRID input parameters retained their default values. The GRID probes used were DRY (for hydrophobic interactions), carboxy oxygen, water, and neutral hydrogen atom. The calculated molecular interaction fields (MIF) were inspected with Gview (part of GRID package). The docking results were visualized with Gview and InsightII (Accelrys).

### Acknowledgments

We thank Valentina Guatteo for her help during measurements and Silvio Mecucci for his support with the GRID program. This work has been partly supported by the National Institute for the Physics of Matter (INFM), PAIS-FOLGIN. H.M., L.R. and M.C. acknowledge Ministero dell'Istruzione, delle Università e della Ricerca (M.I.U.R.) ex 40% 2002 and Fondo per gli Investimenti della ricerca di Base (F.I.R.B.) 2001 grants.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

### References

Anel, A., Calvo, M., Naval, J., Iturralde, M., Alava, M.A., and Piñeiro, A. 1989. Interaction of rat  $\alpha$ -fetoprotein and albumin with polyunsaturated and other fatty acids: Determination of apparent association. *FEBS Lett.* **250**: 22–24.

Belitz, H.D. and Gosch, W. 1999. *Food chemistry*, p. 387. ed. Springer-Verlag, Heidelberg, Germany.

Brownlow, S., Cabral, J.H.M., Cooper, R., Flower, D.R., Yewdall, S.J., Polikarpov, I., North, A.C.T., and Sawyer, L. 1997. Bovine  $\beta$ -lactoglobulin at 1.8 Å resolution—Still an enigmatic lipocalin. *Structure* **5**: 481–495.

Cho, Y., Batt, C.A., and Sawyer, L. 1994. Probing the retinol-binding site of bovine  $\beta$ -lactoglobulin. *J. Biol. Chem.* **269**: 11102–11107.

Collini, M., Chirico, G., Baldini, G., and Bianchi, M.E. 1995. Conformation of short DNA fragments by modulated fluorescence polarization anisotropy. *Biopolymers* **36**: 211–225.

Collini, M., D'Alfonso, L., and Baldini, G. 2000. New insight on  $\beta$ -lactoglobulin binding sites by 1-anilinonaphthalene-8-sulfonate fluorescence decay. *Protein Sci.* **9**: 1968–1974.

D'Alfonso, L., Collini, M., and Baldini, G. 1999. Evidence of heterogeneous 1-anilinonaphthalene-8-sulfonate binding to  $\beta$ -lactoglobulin from fluorescence spectroscopy. *Biochim. Biophys. Acta* **1432**: 194–202.

Dufour, E. and Haertlé, T. 1991. Binding of retinoids and  $\beta$ -carotene to  $\beta$ -lactoglobulin. Influence of protein modifications. *Biochim. Biophys. Acta* **1079**: 316–320.

Dufour, E., Marden, M.C., and Haertlé, T. 1990.  $\beta$ -Lactoglobulin binds retinol and protoporphyrin IX at two different binding sites. *FEBS Lett.* **277**: 223–226.

Fogolari, F., Ragona, L., Zetta, L., Romagnoli, S., De Kruijff, K.G., and Molinari, H. 1998. Monomeric bovine  $\beta$ -lactoglobulin adopts a  $\beta$ -barrel fold at pH 2. *FEBS Lett.* **436**: 149–154.

Frapin, D., Dufour, E., and Haertlé, T. 1993. Probing the fatty acid binding site of  $\beta$ -lactoglobulins. *J. Protein. Chem.* **12**: 443–449.

Futterman, S. and Heller, J. 1972. The enhancement of fluorescence and the decreased susceptibility to enzymatic oxidation of retinol complexed with bovine serum albumin,  $\beta$ -lactoglobulin, and the retinol-binding protein of human plasma. *J. Biol. Chem.* **247**: 5168–5172.

Goodford, P.J. 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**: 849–857.

Kastenholz, M.A., Pastor, M., Cruciani, G., Haaksma, E.E., and Fox, T. 2000. GRID/CPCA: A new computational tool to design selective ligands. *J. Med. Chem.* **43**: 3033–3044.

Lakowicz, J.R. 1999. *Principles of fluorescence spectroscopy*, 2nd ed. Kluwer Academic/Plenum Publishers, New York.

Lange, D.C., Kothari, R., Patel, R.C., and Patel, S.C. 1998. Retinol and retinoic acid bind to a surface cleft in bovine  $\beta$ -lactoglobulin: A method of binding site determination using fluorescence resonance energy transfer. *Biophys. Chem.* **74**: 45–51.

Narayan, M. and Berliner, L.J. 1997. Fatty acids and retinoids bind independently and simultaneously to  $\beta$ -lactoglobulin. *Biochemistry* **36**: 1906–1911.

———. 1998. Mapping fatty acid binding to  $\beta$ -lactoglobulin: Ligand binding is restricted by modification of Cys 121. *Protein Sci.* **7**: 150–157.

Perez, M.D., Diaz de Villegas, C., Sanchez, L., Aranda, P., Ena, J.M., and Calvo, M. 1989. Interaction of fatty acids with  $\beta$ -lactoglobulin and albumin from ruminant milk. *J. Biochem.* **106**: 1094–1097.

Perez, M.D., Sanchez, L., Aranda, P., Ena, J.M., Oriá, R., and Calvo, M. 1992. Effect of  $\beta$ -lactoglobulin on the activity of pregastric lipase. A possible role for this protein in ruminant milk. *Biochim. Biophys. Acta* **1123**: 151–155.

Qin, B.Y., Bewley, M.C., Creamer, L.K., Baker, H.M., Baker, E.N., and Bewley, G.B. 1998a. Structural basis of the Tanford transition of bovine  $\beta$ -lactoglobulin. *Biochemistry* **37**: 14014–14023.

Qin, B.Y., Creamer, L.K., Baker, E.N., and Jameson, G.B. 1998b. 12-bromododecanoic acid binds inside the calyx of bovine  $\beta$ -lactoglobulin. *FEBS Lett.* **438**: 272–278.

Ragona, L., Pusterla, F., Zetta, L., Monaco, H.L., and Molinari, H. 1997. Identification of a conserved hydrophobic cluster in partially folded bovine  $\beta$ -lactoglobulin at pH 2. *Fold. Des.* **2**: 281–290.

Ragona, L., Fogolari, F., Zetta, L., Pérez, D.M., Puyol, P., De Kruijff, K., Löhr, F., Rüterjans, H., and Molinari, H. 2000. Bovine  $\beta$ -lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* **9**: 1347–1356.

Richieri, G.V., Ogata, R.T., and Kleinfeld, A.M. 1992. A fluorescently labeled intestinal fatty acid binding protein. Interactions with fatty acids and its use in monitoring free fatty acids. *J. Biol. Chem.* **267**: 23495–23501.

Robinson, G.W., Robbins, R.J., Fleming, G.R., Morris, J.M., Knight, A.E.W., and Morrison, R.J.S. 1978. Picosecond studies of the fluorescence probe molecule 8-anilino-1-naphthalenesulfonic acid. *J. Am. Chem. Soc.* **100**: 7145–7150.

Sawyer, L. and Kontopidis, G. 2000. The core lipocalin, bovine  $\beta$ -lactoglobulin. *Biochim. Biophys. Acta* **1482**: 136–148.

Spector, A.A. 1975. Fatty acid binding to plasma albumin. *J. Lipid Res.* **16**: 165–179.

Spector, A.A. and Fletcher, J.E. 1970. Binding of long chain fatty acids to  $\beta$ -lactoglobulin. *Lipids* **5**: 403–411.

Tanford, C., Bunville, L.G., and Nozaki, Y. 1959. The reversible transformation of  $\beta$ -lactoglobulin at pH 7.5. *J. Am. Chem. Soc.* **81**: 4032–4035.

Wang, Q., Allen, J.C., and Swaisgood, H.E. 1997a. Binding of retinoids to  $\beta$ -lactoglobulin isolated by bioselective adsorption. *J. Dairy Sci.* **80**: 1047–1053.

———. 1997b. Binding of vitamin D and cholesterol to  $\beta$ -lactoglobulin. *J. Dairy Sci.* **80**: 1054–1059.

Wu, S., Perez, M.D., Puyol, P., and Sawyer, L. 1999.  $\beta$ -Lactoglobulin binds palmitate within its central cavity. *J. Biol. Chem.* **274**: 170–174.

Zsila, F., Imre, T., Szabo, P.T., Bikadi, Z., and Simonyi, M. 2002. Induced chirality upon binding of *cis*-parinaric acid to bovine  $\beta$ -lactoglobulin: Spectroscopic characterization of the complex. *FEBS Lett.* **520**: 81–87.

## EF Loop Conformational Change Triggers Ligand Binding in $\beta$ -Lactoglobulins\*

Received for publication, June 13, 2003, and in revised form, July 11, 2003  
Published, JBC Papers in Press, July 11, 2003, DOI 10.1074/jbc.M306269200

Laura Ragona‡, Federico Fogolari§, Maddalena Catalano‡§, Raffaella Ugolini§, Lucia Zetta‡, and Henriette Molinari§¶

From the ‡Laboratorio Risonanza Magnetica Nucleare, Istituto Macromolecole, Consiglio Nazionale delle Ricerche, via Bassini 15, 20133 Milano, Italy and §Dipartimento Scientifico e Tecnologico, Strada Le Grazie 15, 37134 Verona, Italy

$\beta$ -Lactoglobulins, belonging to the lipocalin family, are a widely studied group of proteins, characterized by the ability to solubilize and transport hydrophobic ligands, especially fatty acids. Despite many reports, the mechanism of ligand binding and the functional role of these proteins is still unclear, and many contradicting concepts are often encountered in the literature. In the present paper the comparative analysis of the binding properties of  $\beta$ -lactoglobulins has been performed using sequence-derived information, structure-based electrostatic calculations, docking simulations, and NMR experiments. Our results reveal for the first time the mechanism of  $\beta$ -lactoglobulin ligand binding, which is completely determined by the opening-closing of EF loop, triggered by Glu<sup>89</sup> protonation. The alkaline shift observed for Glu<sup>89</sup>  $pK_a$  in porcine  $\beta$ -lactoglobulin ( $pK_a$  9.7) with respect to the bovine species ( $pK_a$  5.5) depends upon the interplay of electrostatic effects of few nearby key residues. Porcine protein is therefore able to bind fatty acids provided that the appropriate pH solution conditions are met (pH > 8.6), where the EF loop conformational change can take place. The unusually high pH of binding detected for porcine  $\beta$ -lactoglobulin seems to be functional to lipases activity. Theoretical  $pK_a$  calculations extended to representative  $\beta$ -lactoglobulins allowed the identification of key residues involved in structurally and functionally important electrostatic interactions. The results presented here provide a strong indication that the described conformational change is a common feature of all  $\beta$ -lactoglobulins.

The physicochemical and biological characteristics of  $\beta$ -lactoglobulins, which belong to the lipocalin family, have been extensively studied in the last 30 years, but despite the wealth of data, the biological function of these extracellular proteins is still undefined (Ref. 1 and references therein).  $\beta$ -Lactoglobulins isolated from cow, goat, and sheep milk samples, under non-denaturing conditions, showed endogenously bound fatty acids (2). Many authors have suggested that bovine  $\beta$ -lactoglobulin (BLG)<sup>1</sup> has a transport and/or protective role toward bound ligands in the stomach (3). However, we have previously shown

(4) that BLG, despite its high stability at acidic pH, is unable to bind fatty acids at low pH, thus indicating that it could not be employed "as is" as a transporter through the human gastric tract.

Retinoids and fatty acids have been reported to bind to BLG *in vitro* in the pH range of 6.5–8.5, with dissociation constants on the order of 65 nM and 0.6  $\mu$ M, respectively (5). Specifically, titration experiments of BLG with palmitic acid (PA) (4) have clearly shown that: (i) at neutral pH the primary site for palmitic acid binding is within the protein calyx; (ii) the amount of bound PA is drastically reduced upon decreasing pH and the ligand is completely released at pH 2; (iii) in the pH range 7.3–6.4, a conformational equilibrium was observed for the bound ligand reflecting the dynamics of EF loop (region 85–90) (Fig. 1), triggered by the titration of Glu<sup>89</sup> at anomalously high  $pK_a$  (~6.5) (6). On the contrary, it was shown that the highly similar porcine (PLG) (62% identity, 83% similarity) and equine (58% identity, 74% similarity)  $\beta$ -lactoglobulins had neither fatty acids physiologically bound nor the ability to bind them *in vitro* at neutral pH (7). We have previously reported PLG interaction studies with palmitic acid in the pH range 2–8, aimed at clarifying whether dimer formation could possibly have a role in binding (8). Indeed PLG exhibits a pH dependence of the monomer-dimer equilibrium opposite to that observed for BLG. NMR interaction studies demonstrated that PLG is unable to bind palmitic acid in this pH range. These results were rationalized by us and others (8, 9) hypothesizing that Lys<sup>60</sup> and Lys<sup>69</sup>, forming a superficial positively charged patch at the open end of the BLG calyx, could be responsible for electrostatic interactions with palmitic acid carboxylate, thus driving the binding. Both PLG and equine  $\beta$ -lactoglobulin, which did not show endogenously bound fatty acids, exhibited the K69E mutation. However, tear lipocalin, a protein closely related to BLG, was shown to bind 16-doxyl stearic acid despite the absence of any charged side chain at positions 60 (Met) and 69 (Val) (BLG numbering) (10).

The importance of clarifying the binding mechanism and, hence, the functional role of  $\beta$ -lactoglobulins has prompted us to extend our previous investigations to all members of this family. In the present paper the comparative analysis of  $\beta$ -lactoglobulin binding properties has been performed using sequence-derived information, structure-based electrostatic calculations, docking simulations, and NMR experiments. Our results reveal for the first time the mechanism of  $\beta$ -lactoglobulin ligand binding, which is fully determined by the conformational change involving the opening-closing of the EF loop (11).

mitic acid; <sup>13</sup>C PA, fully enriched palmitic acid; HSQC, heteronuclear single-quantum coherence; MIF, molecular interaction field; PDB, Protein Data Bank.

\* This project was supported by grants from Ministero Istruzione Università e Ricerca (MIUR) 2002, Fondo per Investimenti della Ricerca di Base (FIRB) 2001, Progetti Avanzati ed Iniziative di Sessione, and Fondazione Antonio de Marco. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 390-45-8027906; Fax: 390-45-8027929; E-mail: molinari@sci.univr.it.

<sup>1</sup> The abbreviations used are: BLG, bovine  $\beta$ -lactoglobulin; PLG, porcine  $\beta$ -lactoglobulin; PA, palmitic acid; <sup>13</sup>C<sub>1</sub> PA, singly enriched pal-

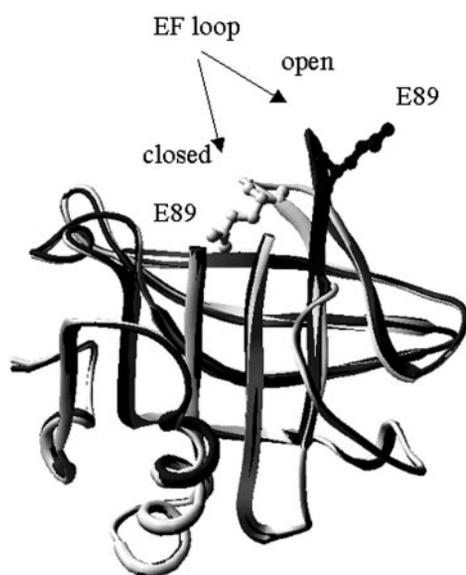


FIG. 1. Superposition of BLG x-ray structures with EF loop in the closed (PDB code 3blg; light gray) and open (PDB code 2blg; dark gray) conformations. Region 109–116 of both proteins has been removed from the ribbon representation in order to better visualize the EF loop region-(85–90). Asp<sup>89</sup> side chain is shown.

#### EXPERIMENTAL PROCEDURES

**Materials**—PLG was purified from milk at NIZO Food Research (The Netherlands) as previously described (8). Palmitic acid was purchased from Sigma. Palmitic acid-protein complexes were prepared as described previously (4). 1.5 mM protein solutions were prepared in 10 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> solution at pH 7.0, and a molar excess of PA (4:1 molar ratio) was employed for complex preparation. NMR titrations at different pH levels were performed adding a few microliters of 0.25 N H<sub>3</sub>PO<sub>4</sub> or 1 M NaOH to the PA-protein complex prepared at pH 7.0.

10 mM phosphate buffer solution, at different pH levels (7–10), was added to the tube containing PA, and NMR experiments were acquired to check the solubility of the fatty acid. A solution of 50 mM KOH at pH 12.5 was necessary to dissolve a uniformly <sup>13</sup>C-labeled PA, 5 mM, for NMR analysis.

**NMR Characterization**—NMR spectra were acquired on an Avance Bruker 500 MHz spectrometers at 27 °C. <sup>1</sup>H chemical shifts were referred to 3-trimethylsilylpropionate, and <sup>13</sup>C chemical shifts were referenced as described previously (4). <sup>1</sup>H one-dimensional NMR spectra of apo- and holo-PLG were run at each pH to check protein stability. One-dimensional proton decoupled <sup>13</sup>C spectra were recorded on the complex prepared with carboxyl-enriched PA (<sup>13</sup>C<sub>1</sub> PA) in the pH range 7.0–10.0. The carboxyl region was acquired with a sweep width of 2,520 Hz, 2,000 time domain points, and 20,000 scans. The aliphatic PA signals were observed through one-dimensional proton decoupled <sup>13</sup>C spectra of completely enriched PA (<sup>13</sup>C-PA) in complex with PLG, in the pH range 7.0–10.0 and with BLG, in the pH range 2.5–7.3. A sweep width of 5,040 Hz, 16,000 time domain points, and 3,584 scans were employed. Two-dimensional <sup>1</sup>H-<sup>13</sup>C HSQC experiments were recorded on <sup>13</sup>C-PA and on PLG-<sup>13</sup>C-PA complex with 512 and 2048 data points in the *t*<sub>1</sub> and *t*<sub>2</sub> dimensions, respectively, and a spectral width of 7002 (*t*<sub>2</sub>) and 5040 (*t*<sub>1</sub>) Hz. The spectra were processed and analyzed with the programs XWINNMR (Bruker) and XEASY (12).

**Docking Simulations**—Docking simulations were performed using the program GRID, version 21 (Molecular Discovery Ltd.) (13, 14). All available BLG structures deposited in the PDB and the PLG x-ray structure (PDB code 1exs) were employed as targets. Crystallographic water molecules were removed. The coordinates of palmitic (C16) and caprylic (C8) acids were derived from the coordinates of palmitic acid bound to BLG (PDB code 1b0o). The fatty acid carboxyl group was considered deprotonated, bearing a net negative charge of -1. The target proteins were considered rigid, and hydrogens were added with the program GRIN (part of the GRID package). The docking search was performed on the whole protein. All GRID input parameters were employed with their default values. The calculated molecular interaction fields (MIF) were inspected with Gview (part of the GRID package). The docking results were visualized with Gview, InsightII (Accelrys, San Diego, CA), and Swiss-PdbViewer (15).

**Modeling and pK<sub>a</sub> Calculations**—Models for lactoglobulins have been built based on the x-ray structures of BLG with the EF loop in the open (PDB code 2blg) and closed (PDB code 3blg) conformations (11) using the Swiss-PdbViewer program (15). Gaps and insertions were shifted out of secondary structure elements, where this was possible, and built using the scan loop data-base module of Swiss-PdbViewer. The anchor residues were chosen such as to make possible the closure of the loop. Hydrogens have been added using the program Gromacs (17).

The University of Houston Brownian Dynamics (UHBD) program, version 5.1 (18), was used for the electrostatic and pK<sub>a</sub> calculations. We employed the single-site titration model (19), which includes partial atomic charges of each ionizable group and models the ionization by adding a +1 or -1 point charge to one central atom of the ionizing group. We used a dielectric constant of 20 to represent protein interior. Partial atomic charges and radii for the protein were taken from the pK<sub>a</sub> files provided by UHBD. The electrostatic potential computed by UHBD has been used for generating an ensemble of protonation states at different pHs using a Monte Carlo procedure as described previously (20). A straightforward application of this procedure led to few but very large pK<sub>a</sub> shifts, which were associated with atoms (and therefore with charges) too close to each other. The bumps were then removed using the program WHATIF (21), and pK<sub>a</sub> values were computed again. The solvent was assigned a dielectric constant of 80 and ionic strength of 100 mM, and the ion exclusion radius was set to 2.0 Å. The ionic strength was chosen based on the observation that the pK<sub>a</sub> values are fairly insensitive to ionic concentrations over 100 mM. A probe with a radius of 1.4 Å was used to define the boundary between the protein and solvent dielectric regions. Electrostatic potentials were calculated by focusing on one titrable residue at a time. The last focusing run employed a grid spacing of 0.25 Å. The temperature was set at 25 °C.

#### RESULTS

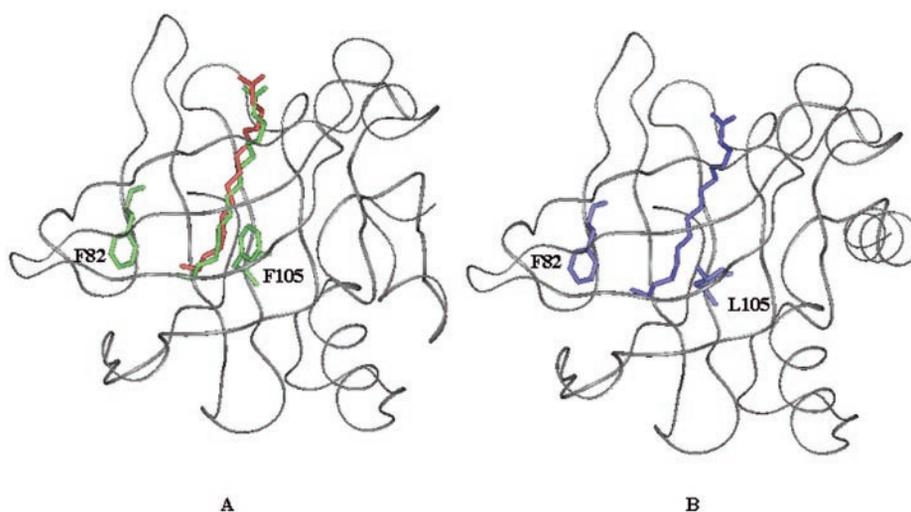
**Docking Experiments**—To evaluate the role of the EF loop conformational change on binding, different chain length fatty acids were docked within BLG and PLG employing the GRID docking program. The GRID method has been developed for determining energetically favorable binding sites for small chemical groups (probes) on a target molecule (protein) (13, 14). The probe groups are small chemical entities that are moved through a regular grid of points around the target molecule in order to calculate, at each point of the grid, an interaction energy, thus generating MIFs. The ligand molecule is represented as a collection of GRID probes, and MIFs calculated for each probe are used to define the ligand position with respect to the target and to estimate the binding energy of the ligand at its binding site.

Palmitic acid, the most abundant BLG endogenous ligand, was docked to all available BLG structures. GRID generated a docking solution for PA within the protein calyx only when the target BLG presented the EF loop in the open conformation. A very good agreement was observed for the positioning of both the aliphatic chain and the carboxyl tail of the docked and experimentally determined PA (PDB code 1b0o) (Fig. 2A). The inspection of MIFs, generated for hydrophobic and carboxyl oxygen probes, indicates that PA docking solution maximizes both electrostatic and hydrophobic interactions.

Docking experiments performed with the available PLG x-ray structure (PDB code 1exs), obtained at pH 3.2 with the EF loop in the closed conformation, were never successful in locating PA within the protein cavity. A model of PLG structure with the EF loop in the open conformation was obtained with the Swiss-PdbViewer program, and GRID simulations showed fatty acids located within the hydrophobic pocket of the protein (Fig. 2B). Interestingly the opening of the EF loop appeared to be a requisite for binding even when the shorter caprylic acid (C8) was used in docking experiments.

**Electrostatics and pK<sub>a</sub> Calculations on  $\beta$ -Lactoglobulins**—The opening of the EF loop-(85–90) in BLG has been shown to be triggered by the titration of Glu<sup>89</sup> (11). EF loop closed conformation is stabilized in BLG by a pattern of hydrogen bonds involving Asp<sup>88</sup>, Glu<sup>89</sup>, Asn<sup>90</sup>, Asn<sup>109</sup>, Ser<sup>110</sup>, and Ser<sup>116</sup>

FIG. 2. GRID docking solutions obtained for BLG (A) and PLG (B). A, superposition of predicted (green) and experimentally observed (red) positions of PA. B, docked solution obtained for PLG (blue). The side chains of residues 82 and 105 are shown.



residues. The titration of the Glu<sup>89</sup> side chain at unusually high pH (~6.5) is due to the loss of this H-bond pattern and causes the fold-back of the EF loop with a consequent solvent exposure of the glutamic acid side chain (Fig. 1).

If the conformational change of the EF loop determines the binding ability, it is clear that all factors affecting Glu<sup>89</sup> p*K*<sub>a</sub>, such as the nature and charge of close residues, will modulate the pH of binding.

A comparison of the amino acid sequences of the EF loop region of all  $\beta$ -lactoglobulins, (Fig. 3) revealed that although Glu<sup>89</sup> is always conserved, its flanking residues (88 and 90) differ both in side chain length and charge. Charged residues close in space to Glu<sup>89</sup> influence its titration as well. To study the pH dependence of the EF loop conformational change and correlate the sequence mutations with the binding ability, p*K*<sub>a</sub> calculations were performed on the BLG and PLG experimentally determined structures and on the modeled structures of a few lactoglobulins (*lacc\_felca*, *laca\_canfa*, *lacb\_horse*, *laca\_horse*, *lacb\_macgi*) selected on the basis of their clustering in the phylogenetic tree (built with the ClustalW program (22)) (Fig. 4).

Two models were built for each  $\beta$ -lactoglobulin, using as a template BLG structures with the EF loop in the open (PDB code 2blg) and closed conformations (PDB code 3blg), respectively. p*K*<sub>a</sub> calculations were performed using the University of Houston Brownian Dynamics program, which solves the Poisson-Boltzmann equation and provides electrostatic free energies, generating a statistical ensemble of protonation states by a Monte Carlo procedure (18). The obtained Glu<sup>89</sup> p*K*<sub>a</sub> values are reported in Table I.

The inspection of all  $\beta$ -lactoglobulin structures allowed the identification of those side chains close to the Glu<sup>89</sup> carboxylic group. Most of these residues never display any charge in different  $\beta$ -lactoglobulins; however, a small subset, consisting of residues 88, 90, 107, 108, and 116, shows a charge change along the different species; *lacb\_macgi* was excluded from this analysis because its low sequence identity with BLG (34%), localized mainly at the level of the EF loop, made its model less reliable. Moreover, residue 116 is close to an insertion making it impossible to reliably discuss any result on Glu<sup>89</sup> p*K*<sub>a</sub>. It is worth noting that on going from BLG to PLG, Glu<sup>89</sup> p*K*<sub>a</sub> increases by nearly two pH units, thus suggesting that PLG binding should indeed occur but at pH higher than 8.0.

**NMR Titration Experiments**—PLG interaction studies with <sup>13</sup>C<sub>1</sub> and <sup>13</sup>C PA were performed in the pH range 8.0–10.0, *i.e.* at pH values higher than the calculated Glu<sup>89</sup> p*K*<sub>a</sub>. The increase in PA resonance intensities was used to monitor the

uptake of ligand by the protein, because PA is solubilized only through protein binding. Fig. 5 shows that <sup>13</sup>C<sub>1</sub> resonance becomes measurable at pH 8.6, and its intensity increases with pH, showing that PLG can bind PA upon EF loop opening. A line width increase of carboxyl resonance is observed raising the pH from 8.6 ( $\Delta\nu_{1/2} = 11$  Hz) to 9.7 ( $\Delta\nu_{1/2} = 17$  Hz). The amount of bound PA as a function of pH, as derived from the integration of C<sub>16</sub> methyl resonance, is shown in Fig. 6. The transition midpoint occurs at pH 9.7, thus affording a clear indirect estimate of Glu<sup>89</sup> p*K*<sub>a</sub> of PLG.

We observed, from the analysis of <sup>1</sup>H NMR spectra, that PLG maintains its conformation up to pH 10, whereas at higher pH the protein starts to precipitate. <sup>1</sup>H and <sup>13</sup>C chemical shifts of PA complexed with PLG are reported in Table II.

#### DISCUSSION

The influence that pH exerts on protein structure is widely believed to be electrostatic in nature via changes in the protonation state of titrable groups, which in turn influence processes like ligand uptake or release, partial or global unfolding, and protein-protein association (23). Our results clearly show for the first time that the protonation state of Glu<sup>89</sup> influences ligand binding in  $\beta$ -lactoglobulins, forcing the EF loop to act as a mobile lid, hindering the access to the protein cavity when it is in the closed conformation. NMR titration experiments have shown that PLG is able to bind PA, provided that the appropriate pH solution conditions (pH > 8.6) are met. The dependence of the amount of bound PA on pH can be fitted with a sigmoid curve, both for BLG and PLG, thus indicating that the same binding mechanism is at work for the two proteins. In PLG the observed transition midpoint is highly shifted toward alkaline pH (pH 9.7) with respect to BLG (pH 5.5) (Fig. 6), reflecting the different p*K*<sub>a</sub> of Glu<sup>89</sup> side chains in the two proteins. These data strongly suggests that, contrary to previous reports (24), the EF loop conformational change also occurs in PLG and is likely to be a common feature of all  $\beta$ -lactoglobulins. To further investigate this phenomenon, p*K*<sub>a</sub> calculations were performed for representative  $\beta$ -lactoglobulins from different species. It has been shown by us and others that theoretically determined p*K*<sub>a</sub> values are in good agreement with experimentally determined ones (25), and we feel confident that the p*K*<sub>a</sub> shifts calculated for the different models reflect a reliable trend, which should be considered as such, rather than as an assessment of the exact p*K*<sub>a</sub> value of each titrating group. It is generally agreed that large p*K*<sub>a</sub> shifts can be ascribed to the electrostatic effects of desolvation, nearby charges, and/or to the disruption of hydrogen bonds (25). To rationalize Glu<sup>89</sup> p*K*<sub>a</sub>

P02754 LACB_BOVIN	LIVTQTMKGL	DIQKVAGTWY	SLAMAASDIS	LLDAQSAPLR	
P02755 LACB_BUBBU	IIVTQTMKGL	DIQKVAGTWY	SLAMAASDIS	LLDAQSAPLR	
P02756 LACB_CAPHI	IIVTQTMKGL	DIQKVAGTWY	SLAMAASDIS	LLDAQSAPLR	
P02757 LACB_SHEEP	IIVTQTMKGL	DIQKVAGTWH	SLAMAASDIS	LLDAQSAPLR	
P33685 LACA_CANFA	IVVPRTMEDL	DLQKVAGTWH	SMAMAASDIS	LLDSEAPLR	
P33688 LACC_FELCA	ATVPLTMDGL	DLQKVAGTWH	SMAMAASDIS	LLDSEAPLR	
P33687 LACB_FELCA	ATVPLTMDGL	DLQKVAGTWH	SMAMAASDIS	LLDSEAPLR	
P33686 LACC_CANFA	IVIPRTMEDL	DLQKVAGTWH	SMAMAASDIS	LLDSEAPLR	
P04119 LACB_PIG	VEVTPIMTEL	DTQKVAGTWH	TVAMAVSDVS	LLDAKSSPLK	
P02758 LACB_HORSE	TNIPQTMQDL	DLQEVAGKWH	SVAMAASDIS	LLDSESAPLR	
P13613 LACB_EQUAS	TNIPQTMQDL	DLQEVAGKWH	SVAMAASDIS	LLDSEAPLR	
P21664 LACA_FELCA	ATLPPTMEDL	DIRQVAGTWH	SMAMAASDIS	LLDSEAPLR	
P19647 LACA_EQUAS	TDIPQTMQDL	DLQEVAGRWH	SVAMVASDIS	LLDSESAPLR	
P07380 LACA_HORSE	TDIPQTMQDL	DLQEVAGRWH	SVAMVASDIS	LLDSEAPLR	
P11944 LACB_MACGI	VENIRSKNDL	GVEKFGVSWY	LREAAK--T-	-MEFSIPLFD	
Q29614 LACB_MACEU	VENIRSKNDL	GVEKFGVSWY	LREAAK--T-	-MEFSIPLFD	
Q29146 LACB_TRIVU	IENIHSKEEL	VVEKLIGPWY	RVEEAK--A-	-MEFSIPLFD	
P02754 LACB_BOVIN	VYVEELKPTP	EGDLEILLQK	WENGCEAQQK	IIAEKTKIPA	
P02755 LACB_BUBBU	VYVEELKPTP	EGDLEILLQK	WENGCEAQQK	IIAEKTKIPA	
P02756 LACB_CAPHI	VYVEELKPTP	EGNLEILLQK	WENGCEAQQK	IIAEKTKIPA	
P02757 LACB_SHEEP	VYVEELKPTP	EGNLEILLQK	WENGCEAQQK	IIAEKTKIPA	
P33685 LACA_CANFA	VYIQLRPTP	QDNLEIVLRK	WEDGRCAEQK	VLAEKTEVPA	
P33688 LACC_FELCA	VYVQELRPTP	RDNLEIILRK	WEQKRCVQKK	LLAQKTELPA	
P33687 LACB_FELCA	VYVQELRPTP	RDNLEIILRK	WEDNRCVEKK	VLAEKTECAA	
P33686 LACC_CANFA	VYIQLRPTP	QDNLEIVLRK	WEDNRCVEKK	VFAEKTECAA	
P04119 LACB_PIG	AYVEGLKPTP	EGDLEILLQK	RENKCAQEV	LLAKKTDIPA	
P02758 LACB_HORSE	VYIEKLRPTP	EDNLEIILRE	GENKGCQAEK	IFAEKTESPA	
P13613 LACB_EQUAS	VYIEKLRPTP	EDNLEIILRE	GENKGCQAEK	IFAEKTESPA	
P21664 LACA_FELCA	VYVQELRPTP	RDNLEIILRK	RENHACIEGN	IMAQRTEDEPA	
P19647 LACA_EQUAS	VYVEELRPTP	EGNLEIILRE	GANHVCCVERN	IVAQRTEDEPA	
P07380 LACA_HORSE	VYVEELRPTP	EGNLEIILRE	GANHACVERN	IVAQRTEDEPA	
P11944 LACB_MACGI	MDIKEVNLTP	EGNLELVLE	KTDR-CVEKK	LLLKKTQKPT	
Q29614 LACB_MACEU	MDIKEVNLTP	EGNLELVLE	KADR-CVEKK	LLLKKTQKPT	
Q29146 LACB_TRIVU	MNIKEVNRTP	EGNLELVLE	QTDS-CVEKK	FLLKKTKEPA	
P02754 LACB_BOVIN	VFKIDALNEN	---KVLVLD	DYKYYLLFCM	ENSAEPEQS-	
P02755 LACB_BUBBU	VFKIDALNEN	---KVLVLD	DYKYYLLFCM	ENSAEPEQS-	
P02756 LACB_CAPHI	VFKIDALNEN	---KVLVLD	DYKYYLLFCM	ENSAEPEQS-	
P02757 LACB_SHEEP	VFKIDALNEN	---KVLVLD	DYKYYLLFCM	ENSAEPEQS-	
P33685 LACA_CANFA	EFKINYVEEN	---QIFLLD	DYDNYLFFCE	MNADAPQOS-	
P33688 LACC_FELCA	EFKISYLDEN	---ELIVLD	DYENYLFFCL	ENADAPQON-	
P33687 LACB_FELCA	KFNINYLDE	---ELIVLD	DYENYLFFCL	ENADAPQON-	
P33686 LACC_CANFA	XFSINYVEEN	---QIFLLD	DYDNYLFFCM	ENANAPQOS-	
P04119 LACB_PIG	VFKINALDEN	---QLFLLD	DYDSHLLLCM	ENASPEHS-	
P02758 LACB_HORSE	EFKINYLDED	---TVFALD	DYKMYLFLCM	KNAATPGQS-	
P13613 LACB_EQUAS	EFKINYLDED	---TVFALD	DYKMYLFLCM	KNAATPGQS-	
P21664 LACA_FELCA	VFMVDYQGEK	---KISVLD	DYTHYMFPCM	EAPAPGTENG	
P19647 LACA_EQUAS	VFTVNYQGER	---KISVLD	DYAHYMFPCV	GPCLPSAEHG	
P07380 LACA_HORSE	VFTVNYQGER	---KISVLD	DYAHYMFPCV	GPPLPSAEHG	
P11944 LACB_MACGI	EFEIYISSES	-SYTFCVME	DYDSYFLFCL	YNISDREK--	
Q29614 LACB_MACEU	EFEIYISSES	ASYTFSVME	DYDSYFLFCL	YNISDREK--	
Q29146 LACB_TRIVU	EFEIYIPSES	ASYTFSVME	DYDNYILGCL	ENWNYREK--	
P02754 LACB_BOVIN	LACQCLVRTP	EVDDEALEKF	DKALKALPMH	IRLSFNPTQL	EEQCHI
P02755 LACB_BUBBU	LACQCLVRTP	EVDDEALEKF	DKALKALPMH	IRLSFNPTQL	EEQCHV
P02756 LACB_CAPHI	LACQCLVRTP	EVDKEALEKF	DKALKALPMH	IRLAFNPTQL	EGQCHV
P02757 LACB_SHEEP	LACQCLVRTP	EVDNEALEKF	DKALKALPMH	IRLAFNPTQL	EGQCHV
P33685 LACA_CANFA	LMCQCLARTL	EVDNEVMEKF	NRALKTLFVH	MQLLN-PTQA	EEQCLI
P33688 LACC_FELCA	LVCQCLTRTL	KADNEVMEKF	DRALQTLFVH	VRLFFDPTQV	AEQCRI
P33687 LACB_FELCA	LVCQCLTRTL	KADNEVMEKF	DRALQTLFVH	VRLFFDPTQV	AEQCRI
P33686 LACC_CANFA	LMCQCLARTL	EVNNEVIGKF	NRALKTLFVH	MQLLN-PTQV	EEQCLV
P04119 LACB_PIG	LVCQCLARTL	EVDDQIREKF	EDALKTLFVH	MR-ILPAQL	EEQCRV
P02758 LACB_HORSE	LVCQCLARTQ	MVDEEIMEKF	RRALQPLPGR	VQIVPDLTRM	AERCRI
P13613 LACB_EQUAS	LVCQCLARTQ	MVDEEIMEKF	RRALQPLPGR	VQIVPDLTRM	AERCRI
P21664 LACA_FELCA	MMCQCLARTL	KADNEVMEKF	DRALQTLFVH	IRIILDLTQG	KEQCRI
P19647 LACA_EQUAS	MVCQCLARTQ	KVDEEVMKEF	SRALQPLPGR	VQIIQDPSGG	QERCGR
P07380 LACA_HORSE	MVCQCLARTQ	KVDEEVMKEF	SRALQPLPGR	VQIVQDPSGG	QERCGR
P11944 LACB_MACGI	MACHYVRRRI	E-ENKGMNEF	KKILRTLAMP	YTVIEVTRTR-	-DMCHV
Q29614 LACB_MACEU	MACHYVRRRI	E-ENKGMNEF	KKILRTLAMP	YTVIEVTRTR-	-DMCHV
Q29146 LACB_TRIVU	MACHYERRI	E-ENKGMNEF	KKIVRTLAMP	YTHIEAQTTR-	-EMCRV

FIG. 3. ClustalW (22) alignment of  $\beta$ -lactoglobulins. The following species are reported: BOVIN (cow), BUBBU (domestic water buffalo), CAPHI (goat), SHEEP, CANFA (dog), FELCA (cat), PIG, HORSE, EQUAS (donkey), MACGI (eastern gray kangaroo), MACEU (tammar wallaby), and TRIVU (brush-tailed possum).

shifts in the analyzed models, contributions due to mutations at the level of residues either involved in H-bonds or close in space to the Glu<sup>89</sup> side chain were therefore considered. As shown in Table I, Glu<sup>89</sup>  $pK_a$  computed for lacc\_pig, lacc\_felca, and lacc\_horse are higher than that computed for lacc\_bovin, reflecting the mutation of the flanking Asn<sup>88</sup> residue to a neg-

atively charged aspartic acid. The strongest validation of these observations comes from NMR titration experiments demonstrating that PLG binding occurs at higher pH with respect to BLG. The agreement between experimental and theoretical  $pK_a$  gave us confidence in the interpretation of the  $pK_a$  trend calculated for all  $\beta$ -lactoglobulins. It is important to stress that

FIG. 4. **Phylogenetic tree of  $\beta$ -lactoglobulin family.** This diagram, produced by the program ClustalW (22), is drawn only approximately to scale, but the branch points are in correct order.

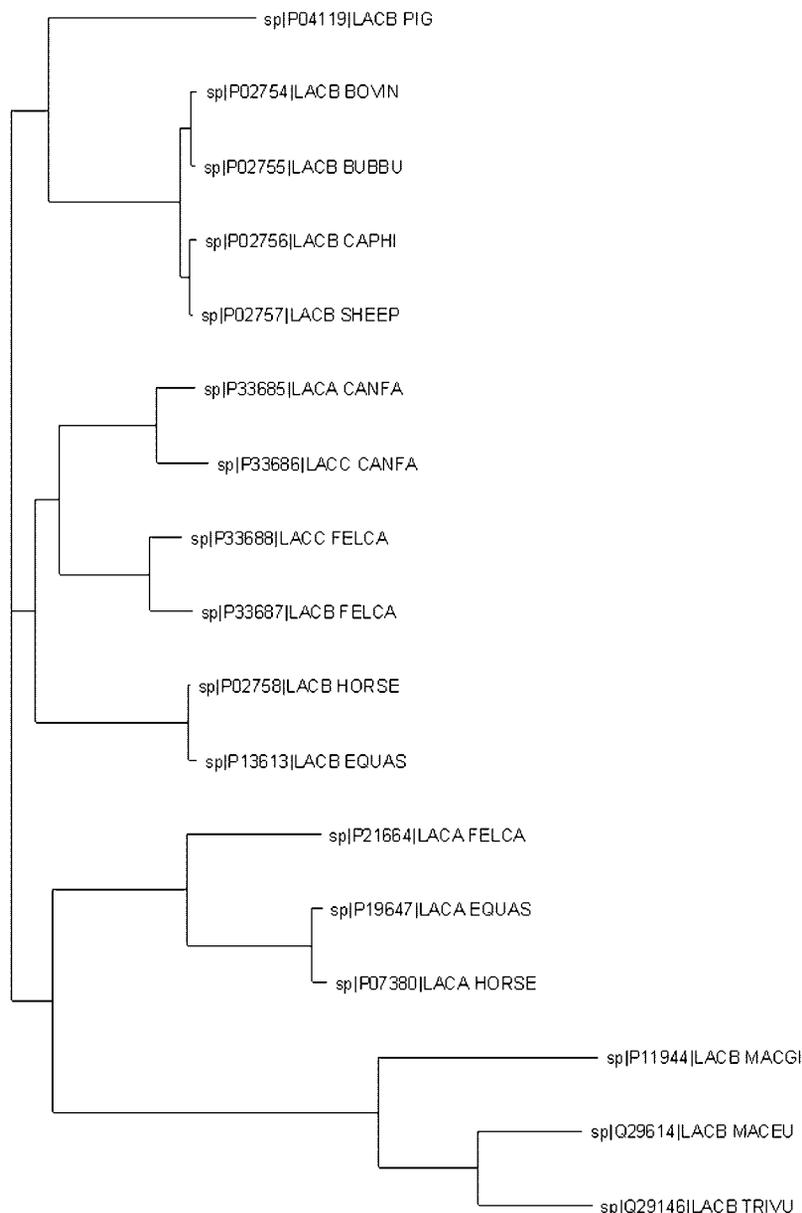


TABLE I  
Calculated  $pK_a$  values of  $Glu^{89}$  in  $\beta$ -lactoglobulins from different species

$\beta$ -Lactoglobulin	Key residues				$Glu^{89}$ $pK_a$	
	88	89	90	116	Closed	Open
Lacb_bovin	Asn	Glu	Asn	Ser	5.59	3.75
Lacb_pig	Asp	Glu	Asn	Ser	7.40	4.42
Lacc_felca	Asp	Glu	Asn	Asn	6.51	4.36
Laca_canfa	Glu	Glu	Asn	Ser	5.64	3.31
Lacb_horse	Asp	Glu	Asp	Ser	6.24	3.96
Laca_horse	Gly	Glu	Arg	His	3.92	3.96
Lacb_bovin mutant 1	Gly	Glu	Arg	Ser	5.10	
Lacb_bovin mutant 2	Gly	Glu	Arg	His	3.30	

laca\_canfa, showing a negative charged residue at position 88, similar to lacb\_pig, lacc\_felca, and lacb\_horse, exhibits a calculated  $pK_a$  of 5.64, similar to that obtained for lacb\_bovin (5.61). The analysis of the model showed that  $Glu^{88}$  side chain in laca\_canfa points in the opposite direction with respect to  $Glu^{89}$ , with the distance between the two carboxylic groups being the highest (8.5 Å) among all the analyzed structures (in

the range 3.5–6.0 Å), thus reducing the electrostatic effects.

The low  $pK_a$  value calculated for laca\_horse correlates well with the presence of two close, positively charged residues ( $Arg^{90}$  and  $His^{116}$ ). To single out whether the main role is played by the flanking  $Arg^{90}$  or by the close  $His^{116}$ , we modeled the following multiple mutations in BLG, N88G,N90R (mutant 1) and N88G,N90R,S116H (mutant 2), and calculated the corresponding  $Glu^{89}$   $pK_a$  values. Interestingly a  $pK_a$  of 5.10 was calculated for mutant 1 and a  $pK_a$  of 3.30 was obtained for mutant 2, thus indicating that the close, positively charged residue 116 plays the determining role in influencing the  $pK_a$ .

It is worth noting that when  $Glu^{89}$   $pK_a$  was calculated for all of the models with the EF loop in the open conformation, very similar values were obtained, in the range 3.75–4.36 (Table I), as expected for accessible, solvent-exposed carboxylates.

A comment could be added that further supports the reliability of  $pK_a$  calculations. The low  $pK_a$  value (1.05) estimated for  $Asp^{98}$  in lacb\_bovin is in good agreement with the value of 2 that was determined experimentally by NMR (20).

Docking experiments run on BLG and PLG indicated that the opening of the EF loop plays a key role in modulating the binding of both proteins. The GRID solution obtained for the

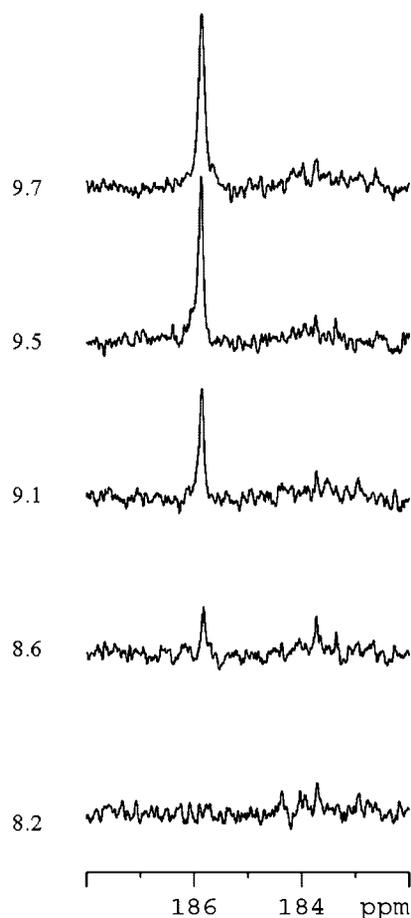


FIG. 5. Carboxylic region of 125.7 MHz one-dimensional proton decoupled  $^{13}\text{C}$  spectra acquired in pH range 8.2–9.7 on the complex of  $^{13}\text{C}_1$  PA with PLG at 27 °C.

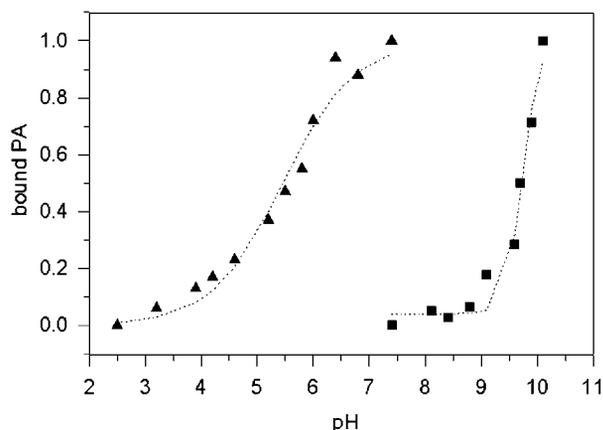


FIG. 6. Plot of  $^{13}\text{C}$  PA methyl resonance areas versus pH for BLG (triangle) and PLG (square). Intensities are normalized according to the spectrum obtained at higher pH. The dotted lines represent the curve fits obtained with the program Sigma Plot.

PLG-PA complex clearly showed that fatty acid is located within the protein calyx adopting a conformation similar to that observed experimentally in BLG (Fig. 2B). PLG residues showing short contacts with bound PA are substantially the same as observed in holo-BLG (26). However the comparison of  $^1\text{H}$  and  $^{13}\text{C}$  one-dimensional and two-dimensional HSQC NMR spectra of holo-PLG and -BLG revealed that PA H-16 and H-15 resonances exhibited less marked up-field shifts when bound to PLG ( $\delta\text{H}16 = 0.69$  ppm,  $\delta\text{H}15 = 1.00$ ) rather than to BLG ( $\delta\text{H}16 = 0.20$  ppm,  $\delta\text{H}15 = 0.54$ – $0.43$ ). This behavior can be

TABLE II  
 $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of PA complexed with PLG, pH 9.7, 27 °C

Position number	$^1\text{H}$	$^{13}\text{C}$
C-1		185.85
C-2	2.08	40.99
C-3	1.43	29.29
Other	1.05	32.17
C-15	1.00	25.14
C-16	0.69	17.66

attributed to a different distribution of aromatic residues in the binding cavity; in BLG, PA methyl protons feel the shielding effect of two aromatic residues (Phe<sup>82</sup> and Phe<sup>105</sup>) lying within 5 Å of its center (Fig. 2A), whereas in PLG the mutation F105L accounts for the reduced shielding effect (Fig. 2B). The minor chemical shift dispersion observed in PLG for C-4–C-12 carbons can be ascribed to the same effect. All of these NMR data confirm that PA should have the same spatial arrangement within the calyx of the two proteins. The conformational equilibrium clearly observed for PA bound to BLG, involving the carboxyl tail, at a pH value close to that of the EF loop conformational change (4), has been also observed for holo-PLG, as indicated by the  $^{13}\text{C}$  carboxyl PA line width increase on going from pH 8.6 to 9.7.

Docking simulations provided a lower binding energy for the interaction of PA with PLG (–10.6 kcal/mol) with respect to BLG (–14.81 kcal/mol), suggesting a minor affinity of the porcine protein for fatty acids. We have previously shown that binding energies provided by GRID for BLG were in good agreement with those measured through dynamic fluorescence experiments (27), and we are therefore confident of the reliability of the given energy differences. The mutation K69E, observed on going from BLG to PLG, may be held responsible for the lower interaction energy as deduced from the inspection of MIFs, showing less favorable electrostatic interactions with the ligand in PLG.

Altogether these data indicate that the pH-dependent conformational change of EF loop is a common feature of all  $\beta$ -lactoglobulins. The analysis presented here, while allowing the identification of structurally and functionally important electrostatic interactions in  $\beta$ -lactoglobulins, opens the way to the design of engineered proteins characterized by binding capability in a selected pH range.

Interestingly the binding mechanism identified for  $\beta$ -lactoglobulins is reminiscent of the mechanism of lipase interfacial activation associated with a conformational change, in which a lid, consisting of one  $\alpha$ -helix, opens up by rotating around its hinge regions, thus allowing ligand binding. As it was reported that (i)  $\beta$ -lactoglobulins may play a role in increasing lipases activity by removing free fatty acids (28) and (ii) liver pig triacylglycerol hydrolases exhibit activity at an optimum alkaline pH of 8.5 (16), it is likely that the unusually high pH of binding detected for PLG is functional in lipases activity.

**Acknowledgments**—S. Mecucci and P. Benedetti are gratefully acknowledged for their support with the GRID program. We thank E. Silletti, A. Altling, and R. W. Visschers (NIZO Food Research, the Netherlands) for isolation and purification of PLG.

#### REFERENCES

- Sawyer, L., and Kontopidis, G. (2000) *Biochim. Biophys. Acta* **1482**, 136–148
- Pérez, M., D., de Villegas, C. D., Sanchez, L., Aranda, P., Ena, J. M., and Calvo, M. (1989) *J. Biochem.* **106**, 1094–1097
- Uhrinova, S., Smith, M., Jameson, G. B., Uhrin, D., Sawyer, L., and Barlow, P. N. (2000) *Biochemistry* **39**, 3567–3574
- Ragona, L., Zetta, L., Fogolari, F., Perez, D. M., Pujol, P., De Kruif, K., Lohr, F., Ruterjans, H., and Molinari, H. (2000) *Protein Sci.* **9**, 1347–1356
- Narayan, M., and Berliner, L. J. (1997) *Biochemistry* **36**, 1906–1911
- Tanford, C., and Nozaki, Y. (1959) *J. Biol. Chem.* **234**, 2874–2876
- Pérez M. D., Puyol, P., Ena, J. M., and Calvo, M. (1993) *J. Dairy Res.* **60**, 55–63
- Ugolini, R., Ragona, L., Silletti, E., Fogolari, F., Visschers, R. W., Altling, A. C.,

- and Molinari, H. (2001) *Eur. J. Biochem.* **268**, 4477–4488
9. Frapin, D., Dufour, E., and Haertle, T. (1994) *Biochim. Biophys. Acta* **1205**, 105–112
10. Gasymov, O. K., Abduragimov, A. R., Yusifov, T. N., and Glasgow, B. J. (2001) *Biochemistry* **40**, 14754–14762
11. Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, H. M., Baker, E. N., and Jameson, G. B. (1998) *Biochemistry* **37**, 14014–14023
12. Bartels, C., Xia, T., Billeter, M., Guntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR* **5**, 1–10
13. Goodford, P. J. (1985) *J. Med. Chem.* **28**, 849–857
14. Kastenholz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E., and Fox, T. (2000) *J. Med. Chem.* **43**, 3033–3044
15. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723
16. Verger, R., Rietsch, J., Dam-Mieras, M. C. E., and de Haas, G. H. (1976) *J. Biol. Chem.* **251**, 3128–3133
17. Lindahl, E., Hess, B., and van der Spoel, D. (2001) *J. Mol. Model.* **7**, 306–317
18. Madura, J. D., Briggs, J. M., Wade, R., Davis, M. E., Luty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., and McCammon, J. A. (1995) *Comput. Phys. Commun.* **91**, 57–95
19. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) *J. Mol. Biol.* **238**, 415–436
20. Fogolari, F., Ragona, L., Licciardi, S., Romagnoli, S., Michelutti, R., Ugolini, R., and Molinari, H. (2000) *Proteins Struct. Funct. Genet.* **39**, 317–330
21. Vriend, G. (1990) *J. Mol. Graph.* **8**, 52–56
22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
23. Fogolari, F., Brigo, A., and Molinari, H. (2002) *J. Mol. Recognit.* **15**, 377–392
24. Burova, T. V., Grinberg, N. V., Visschers, R. W., Grinberg, V. Y., and de Kruijff, C. G. (2002) *Eur. J. Biochem.* **269**, 3958–3968
25. Herrgard, S., Gibas, C. J., and Subramaniam, S. (2000) *Biochemistry* **39**, 2921–2930
26. Wu, S. Y., Pérez, M. D., Puyol, P., and Sawyer, L. (1999) *J. Biol. Chem.* **274**, 170–174
27. Collini, M., D'Alfonso, L., Molinari, H., Ragona, L., Catalano, M., and Baldini, G. (2003) *Protein Sci.* **12**, 1596–1603
28. Pérez, M. D., Sanchez, L., Aranda, P., Ena, J. M., Oria, R., and Calvo, M. (1992) *Biochim. Biophys. Acta* **1123**, 151–155



## Heterologous expression of bovine and porcine $\beta$ -lactoglobulins in *Pichia pastoris*: towards a comparative functional characterisation

Gaetano Invernizzi<sup>a</sup>, Laura Ragona<sup>b</sup>, Stefania Brocca<sup>a</sup>,  
Eros Pedrazzoli<sup>a</sup>, Henriette Molinari<sup>d</sup>, Piero Morandini<sup>c</sup>,  
Maddalena Catalano<sup>b,d</sup>, Marina Lotti<sup>a,\*</sup>

<sup>a</sup> Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca, Piazza della Scienza 2, Milan, Italy

<sup>b</sup> Laboratorio NMR, ISMAC, CNR, Milan, Italy

<sup>c</sup> Dipartimento di Biologia, Università degli Studi di Milano, Milan, Italy

<sup>d</sup> Dipartimento Scientifico e Tecnologico, Università di Verona, Verona, Italy

Received 16 November 2002; accepted 14 October 2003

### Abstract

Bovine and porcine  $\beta$ -lactoglobulins were cloned and expressed in host cells with the aim of developing the tools necessary for their structural, functional and conformational characterisation by NMR techniques. Both lipocalins were expressed in *Pichia pastoris*, where the use of a constitutive promoter turned out to allow the highest productivity. The yield of recombinant proteins was further improved through multiple integration of the encoding genes and by increasing aeration of the transformed cultures. Both proteins were obtained in the culture medium at the concentration of 200  $\mu\text{g}/\text{ml}$ . Recombinant lipocalins were purified by ion-exchange chromatography from the culture medium. A preliminary NMR characterisation showed that both proteins were correctly folded.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:**  $\beta$ -Lactoglobulin; Expression; *Pichia pastoris*; NMR

### 1. Introduction

$\beta$ -Lactoglobulins ( $\beta$ -LGs) are whey proteins that belong to the family of lipocalins, a large group of small extracellular proteins varied in amino acid sequence and function but sharing a common an-

tiparallel  $\beta$ -barrel protein fold and the ability to bind small, mostly hydrophobic ligands, such as retinol and steroids (for a review, see Flower et al., 2000). Although lipocalins are the target of several studies addressing their structure and biochemical properties, fundamental issues related to their physiological function and to the structural basis of protein properties still remain largely undisclosed. Bovine (BLG) and porcine (PLG) lactoglobulins are a convenient

\* Corresponding author. Fax: +39-02-6448-3565.

E-mail address: [marina.lotti@unimib.it](mailto:marina.lotti@unimib.it) (M. Lotti).

tool for a systematic and comparative study on structure–function in this protein family. In fact, they share 63% identity and 80% similarity in their primary sequence and yet differ in functional properties. It has been shown that the two proteins display an opposite dependence on the pH for their monomer–dimer equilibrium: bovine  $\beta$ -LG is monomeric at pH 2, while porcine  $\beta$ -LG is monomeric at pH higher than 5 (Ugolini et al., 2001). Moreover, while BLG binds a number of hydrophobic ligands, the porcine protein is unable to bind fatty acids in the pH range 2–8. Besides, other considerations might draw the attention of researchers on functional properties of lactoglobulins. In particular, bovine  $\beta$ -LG is involved in the irreversible denaturation of milk proteins, due to its tendency to form insoluble aggregates with itself and other milk proteins (De Jong, 1997) and it is being exploited as a gelling agent for food processing (Vittayanont et al., 2002). The relevance of lipocalins in biotechnology has been recently stressed by a novel application of engineered lipocalin variants (anticalins) for the molecular recognition of prescribed ligands (Schlehuber and Skerra, 2000).

A comprehensive mutagenesis approach coupled with the structural and conformational characterisation of these two proteins might provide insight into the determinants of features such as stability, tendency to aggregation, ligand binding and mechanism of folding. NMR spectroscopy is a powerful technique for the detailed analysis of structural and dynamic properties of proteins. The monomer–dimer equilibrium should be taken into account when performing the NMR spectra of these proteins: indeed the dimeric proteins are beyond the routine size limit of NMR, and the proteins must be studied in the pH range where they are monomeric. Even for the monomeric form, however, simple  $^1\text{H}$  experiments, even if multi-dimensional, cannot lead, due to severe overlap, to a complete assignment, which depends upon triple resonance NMR techniques, for which isotope labelling is mandatory. Isotope labelling also allows for heteronuclear relaxation measurements, which can give important information on protein dynamics.

The requirement of a complete NMR structure–function correlation study for proteins belonging to the lipocalin family put the highest pressure for developing suitable protocols for the over-expression of  $\beta$ -lactoglobulins from different species. Several hosts

have been exploited for the heterologous expression of bovine  $\beta$ -LG: *Escherichia coli* (Batt et al., 1990), *Saccharomyces cerevisiae* (Totsuka et al., 1990), transgenic animals (Hytinen et al., 1998), with relatively low yields. On the other hand high expression and secretion of the bovine  $\beta$ -LG variant A was obtained in *Pichia pastoris* from an expression vector controlled by a methanol inducible promoter (Kim et al., 1997). On the contrary, heterologous expression of porcine  $\beta$ -LG has not been attempted yet. In this paper, we report about the optimisation of the production of bovine  $\beta$ -LG variant B and porcine  $\beta$ -LG expressed in *P. pastoris* under the control of a constitutive promoter and the purification of NMR grade recombinant proteins.

## 2. Materials and methods

### 2.1. Purified protein samples

Bovine and porcine  $\beta$ -LG were purified from pooled milk at the NIZO Institute (Netherlands Institute for Dairy Research, Ede, The Netherlands) using standard procedures (Fogolari et al., 1998). The purity of the proteins was checked by reverse phase HPLC, electrospray mass spectroscopy and NMR measurements. The bovine  $\beta$ -LG sample contained the pure B variant. The mass spectra of the porcine species revealed the presence of two isoforms, with relative molecular weights of  $17\,857 \pm 1$  and  $17\,852 \pm 1$ , corresponding to A and C isoforms, respectively. The C and A variants differ for the substitutions E9D, Q68H.

For all the experiments, the pH was measured before and after the set of measurements with a conventional micro glass electrode.

### 2.2. Strains, plasmids and media

*E. coli* DH5 $\alpha$  (Stratagene GmbH, Heidelberg, Germany) was used as the host for steps in DNA manipulation and *P. pastoris* strain X-33 (Invitrogen Corporation, San Diego, LA) for the expression of recombinant  $\beta$ -lactoglobulins. Plasmids were pGem7Z (Promega Corporation, Madison, WI), pPIC $\alpha$ B and pGAPZ $\alpha$ B (Invitrogen Corporation, San Diego, LA). The latter two vectors are suited for inducible or constitutive expression of recombinant proteins in *P. pastoris* and both contain a signal sequence for their

secretion in the culture medium. *E. coli* was grown in shaking flasks according to standard procedures. Selection of transformants was on low salt LB containing 25 µg/ml zeocine.

*P. pastoris* was grown in shaking flasks at 30 °C in a rich buffered medium (1% yeast extract, 2% peptone, 1.34% Yeast Nitrogen Base) containing 0.1 M phosphate buffer pH 6.0 and 1% glucose (BMD) or sorbitol (BMS). Transformants were selected in YPD medium (1% yeast extract, 2% peptone, 2% D-glucose) containing 100 µg/ml zeocine (Invitrogen Corporation, San Diego, CA).

### 2.3. Cloning, mutagenesis and expression

The B isoform of the bovine β-lactoglobulin gene was isolated by PCR from a *Bos taurus* cDNA library (Stratagene, Cat. #937717). The oligonucleotides used were as follows: 5'-CCGCTCGAGAAAAGAGAGGCTGAAGCTTACGTWACC-3' (sense) and 5'-CGTC-TAGACCTTRGATGTGGCACTGCTCCTC-3' (antisense) and allowed for insertion into the *Eco*RI restriction site and in-frame fusion with α-factor gene from *S. cerevisiae* in the two expression vectors giving the pPIC-BLG and pGAP-BLG constructs.

A point mutation necessary to restore the wt sequence was introduced in the amplified cDNA by two PCR steps generating two DNA fragments overlapping in the region where the mutation was introduced. The oligonucleotide pair was: 5'-CTCAC-CATTCTCCCATTCTGC-3' (sense)/5'-GAGAATG-ATGAGTGTGCTCAG-3' (antisense). The two overlapping fragments were joined by PCR by the overlap extension method (Ho et al., 1989) giving a 1 kbp sequence carrying the mutation site enclosed in a *Xho*I-*Xba*I cassette (512 bp). This was substituted within the pPIC-BLG expression vector. Porcine β-LG (isoform A) was synthesised using PCR overlapping extension: the sequence was designed according to the protein primary structure and adapted to the codon usage of *P. pastoris*. Restriction sites useful in the following phases of cloning were also introduced. Recombinant gene cassette *Xho*I-*Xba*I was cloned into constitutive expression vector pGAPZαB. *P. pastoris* was transformed by electroporation (Scorer et al., 1994) using a Bio-Rad Gene Pulser with 10 µg of *Rca*I-linearised plasmid and plated onto YPDS plates containing zeocine.

Positive transformants were checked for β-LG production by shaking cultures from single yeast colonies for 5 days at 30 °C in 25 ml of the appropriate medium. Secretion was monitored on 16% SDS-PAGE (Laemmli, 1970). Large scale constitutive expression was carried out in 31 flasks at 30 °C and 340 rpm in 500–1000 ml BMD medium at pH 6.0.

### 2.4. Purification of recombinant proteins

After 120 h fermentation, cells were removed by centrifugation (11 000 rpm, 40 min) and the culture supernatants were subjected to sterile filtration (Millipore Stericup device with 0.22 µm pore membrane). Concentration to 5–10% of the original volume was obtained using a tangential flow device (Minitan or Pellicon Apparatus, Millipore) through cellulose filter (5 kDa cut-off). One litre culture supernatant, obtained and concentrated as previously described, was diafiltered against 50 mM phosphate buffer pH 6.3 and applied to a DEAE cellulose c26/70 Pharmacia column (150–300 ml packed volume), pre-equilibrated with 50 mM phosphate buffer pH 6.3. The column was washed with 3–5 column volumes of the same buffer and rβ-LGs were eluted with a linear gradient of 0–0.3 M NaCl in 50 mM phosphate buffer, pH 6.3. Elution was achieved at flow rate of 1 ml/min. Elution fractions were desalted by diafiltration and subjected to SDS-PAGE. Fractions containing the recombinant purified proteins were pooled, concentrated and diafiltered against 12 mM phosphoric acid, pH 2.4 by Centricon centrifugal concentrator (Amicon, Millipore) and concentrated to 17–20 mg/ml for NMR characterisation.

### 2.5. NMR spectroscopy

<sup>1</sup>H NMR were acquired on a Bruker DRX 500 MHz spectrometer. Homonuclear TOCSY spectra (Bax and Davis, 1985) of protein samples (1.0 mM) were recorded in aqueous buffers at different selected pHs. The employed buffers were as following: (i) pH 2: 12 mM H<sub>3</sub>PO<sub>4</sub>:NaOH; (ii) pH 5 and pH 7: 10 mM Na<sub>2</sub>HPO<sub>4</sub> and 10 mM KH<sub>2</sub>PO<sub>4</sub> solutions mixed in the appropriate ratio. Experiments were carried out in the temperature range 27–37 °C. Each of the 512 spectra contained 2048 data points and 64 scans, and had a sweep width of 7000 Hz in both dimensions. The water

suppression was achieved by low power irradiation during relaxation delay introduced between scans or by 3–9–19 pulse sequence with gradients (Sklenar et al., 1993). For TOCSY spectra various isotropic mixing were employed, ranging from 20 to 80 ms. Spectra were processed with XWINNMR using Gaussian enhancement in both dimensions and zero-filled once in the  $F_1$  dimension. Analysis of the spectra were performed using both XWINNMR and XEASY (Bartels et al., 1995). Chemical shift references for all acquired spectra was obtained using internal capillary with TSP dissolved in the appropriate buffer solution.

### 3. Results and discussion

#### 3.1. Cloning and expression of bovine $\beta$ -lactoglobulin

Primers were designed based on the sequence of  $\beta$ -lactoglobulin isoform B, to amplify by PCR its complete ORF from a female *B. taurus* cDNA library. Amplification produced a single band of about 0.5 kb that was subcloned directly into pPICZ $\alpha$ B. Several clones were sequenced and clone BGL4 was selected for having the highest identity with the native sequence (Accession Number NCBI: Z48305; protein ID: CAA88303.1), with single nucleotide difference leading to the substitution D64G. In order to reconstitute the wild type  $\beta$ -lactoglobulin gene, the *XhoI*–*XbaI* 512 bp DNA cassette containing the mutation was replaced with the corresponding fragment obtained by PCR overlap extension. The oligonucleotides used for ORF amplification allow for its in-frame fusion with the  $\alpha$ -factor gene from *S. cerevisiae* that drives secretion of the recombinant protein to the culture medium. The resulting expression plasmid was transformed in *P. pastoris* strain X-33. Transformants were able to grow on medium added with 100  $\mu$ g/ml zeocine, even after several growth generations in non-selective medium, indicating that plasmid DNA was stably integrated in the yeast genome. Five to 10 transformed clones were inoculated in small-scale cultures (25 ml in 100 ml shaking flask) to assay their ability to produce and to secrete recombinant BLG. Under these preliminary tests about 50  $\mu$ g/ml recombinant product were secreted into the medium. Several different cultivation conditions were tested with a moderate

increase in productivity. Higher yields were achieved by expressing BLG from vector pGAPz $\alpha$ B, which is identical to the previous but for the presence of the constitutive GAPDH promoter. After 5 days fermentation in BMD medium, 100  $\mu$ g/ml recombinant protein were detectable in the culture medium. As referred by other authors, integration of the expression vector in multiple copies increases both zeocine resistance and heterologous protein expression (Scorer et al., 1994). Thus, resistance to high zeocine concentration can be considered as a diagnostic trait of multi-copies integration of pGAP–BLG constructs in the *P. pastoris* genome and could indirectly indicate clones with increased capability to express rBLG. Twenty clones bearing the pGAP–BLG construct, already selected for their resistance to 100  $\mu$ g/ml zeocine, were spotted at controlled density ( $10^3$ ,  $10^4$ ,  $10^5$  cells) onto YEPD medium containing 0.1 mg/ml ( $10\times$ ) and 0.2 mg/ml ( $20\times$ ) zeocine. Clones resistant to the highest zeocine concentration were inoculated in small-scale YEPD cultures and examined for their production of rBLG by SDS-PAGE. In agreement with previous reports and with the expected,  $\beta$ -lactoglobulin was detected in all assayed clones, but its expression was higher in cells resistant to the highest zeocine concentration. As it is well established that one of the main limiting factor in cell growth and heterologous protein production, especially in flask fermentation, is the culture oxygenation, which in turn depends on the culture agitation efficiency, we exploited the use of baffled shaking flasks. The final yield of secreted recombinant protein was 200  $\mu$ g/ml in rich buffered medium containing 0.1 M phosphate pH 6.0. On the basis of these observations, porcine  $\beta$ -lactoglobulin was expressed from the pGAPz $\alpha$ B vector using identical cultural conditions and obtaining similar yields.

#### 3.2. Characterisation of recombinant $\beta$ -lactoglobulins

In SDS-PAGE, recombinant bovine  $\beta$ -lactoglobulin had an apparent molecular weight of about 18 kDa, similar to that of the protein purified from cow milk (Fig. 1A). A slight molecular weight difference can be ascribed to the presence of four extra amino acid residues at the N-terminus of the protein, due to ineffective cleavage of the  $\alpha$ -factor sequence by the Ste13 proteolytic enzyme (Kim et al., 1997). Recom-

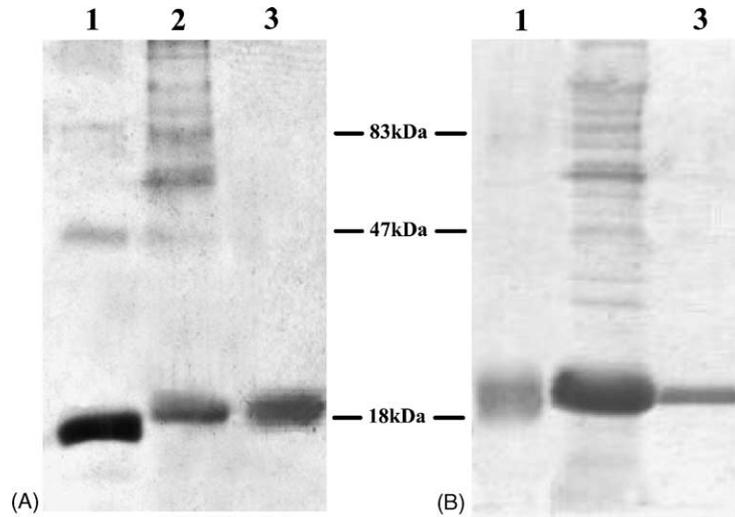


Fig. 1. SDS-PAGE of bovine (A) and porcine (B)  $\beta$ -lactoglobulins. Lane 1: standard 5  $\mu$ g; lane 2: 40  $\mu$ l of supernatants from *P. pastoris* cultures expressing recombinant proteins; lane 3: samples after purification by chromatography.

binant porcine  $\beta$ -LG displayed an apparent molecular mass consistent with the full processing of the leader sequence (Fig. 1B). Recombinant  $\beta$ -LGs were subjected to preparative purification on a DEAE cellulose column. Fractions obtained during the elution in preparative chromatography were assayed for protein content and analysed by SDS-PAGE to assess their purity (Fig. 1). Samples containing lonely pure recombinant BLG were pooled and diafiltrated against 20 mM phosphate buffer pH 5.5.

In order to assess that the recombinant proteins have the correct folded native structure, a comparison of NMR spectra of purified and expressed bovine and porcine  $\beta$ -lactoglobulins is reported.

The 1D spectra of bovine and porcine  $\beta$ -LG (Figs. 2 and 3) are typical of folded proteins, showing widely dispersed chemical shifts ( $\delta$ ) characterised by the presence of downfield shifted alpha protons ( $\delta$  in the range 4.7–6.0 ppm), typical of beta structures, and by high-field shifted resonance ( $\delta$  in the range 0.7–0.5 ppm), indicating the presence of an aromatic tertiary environment inducing ring current shifts in nearby protons.

It is worth mentioning that NMR studies of the bovine protein were previously reported, and the protein spectrum has been completely assigned (Ragona et al., 1997; Uhrinova et al., 1998). As a consequence, given the NMR capability of providing detailed

structural information at the atomic level, it is possible to check the location of any possible conformational change resulting for the recombinant bovine protein.

Fig. 4a and b report an expansion of the  $^1\text{H}$  1D high-field regions of bovine  $\beta$ -LGs where the same pattern of resonances is clearly observed for the well-resolved aliphatic protons, thus giving a first indication that the purified and recombinant proteins have the same conformation.

In view of the high spectral overlap of the 1D spectra, a further analysis was performed, by comparing the finger-print regions of 2D scalar correlated experiments ( $^1\text{H}$  2D TOCSY), where each cross-peak represents the correlation of an amide with its scalarly coupled alpha proton.

Fig. 5 reports the superimposition of the fingerprint regions of 2D TOCSY spectra of purified and recombinant bovine  $\beta$ -LG. Amide and  $\text{H}\alpha$  secondary shifts have been shown to depend on the secondary structure (Wishart et al., 1995), and the strict correspondence of these two regions is the strongest indication that the two proteins share an identical conformation. A minor difference in the secondary shift is observed for the cross-peaks corresponding to the NH- $\text{H}\alpha$  correlations of K<sub>101</sub> and Y<sub>102</sub>, indicated with arrows in the figure. These resonances were shown to be strongly affected by the pH and ionic strength of the solution, exhibiting up to 0.5 ppm shift increasing the ionic strength of

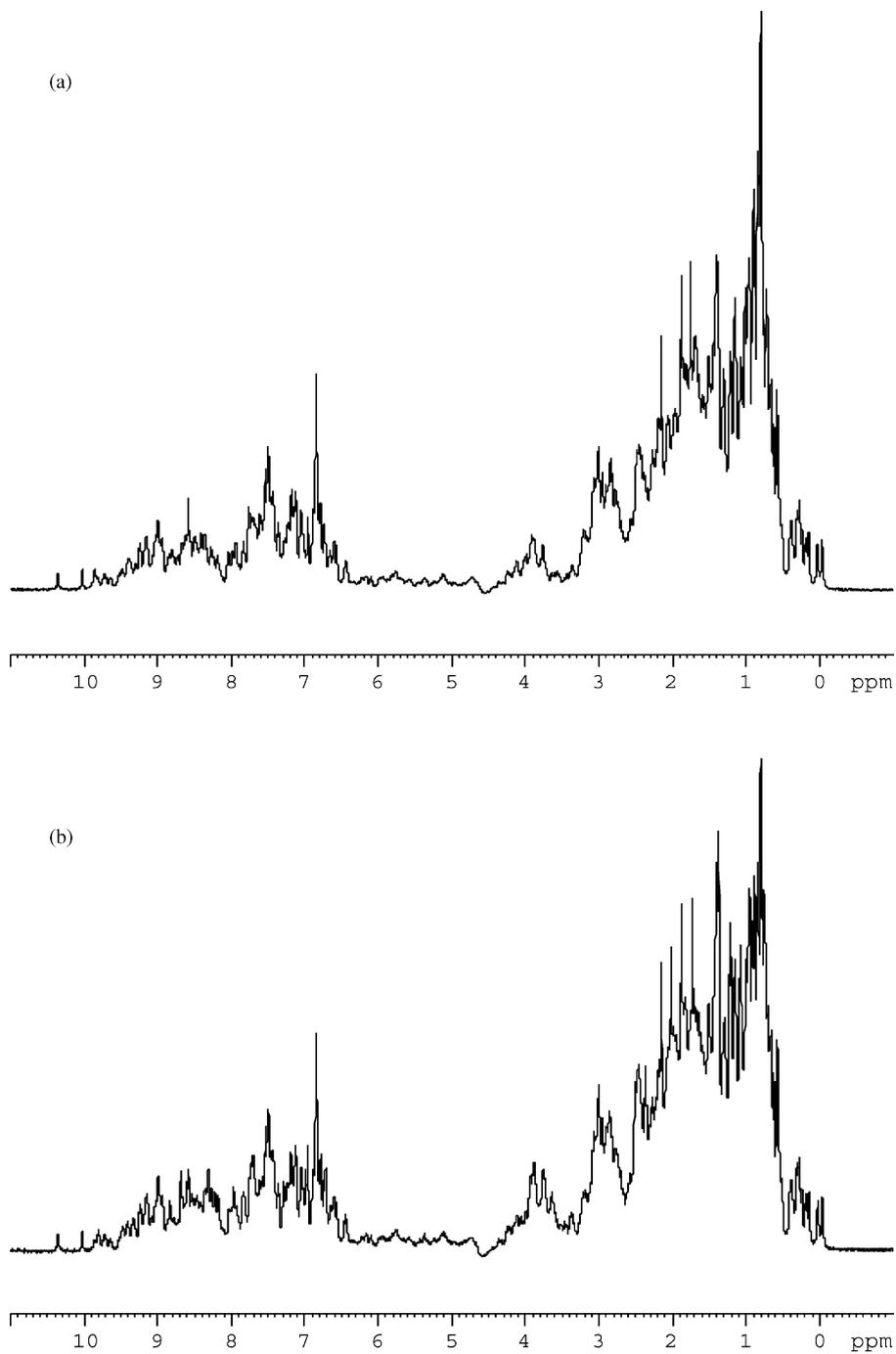


Fig. 2.  $^1\text{H}$  500 MHz 1D NMR spectra of 1 mM bovine  $\beta$ -LGs, recorded at pH 2 and 37 °C: (a) purified; (b) recombinant.

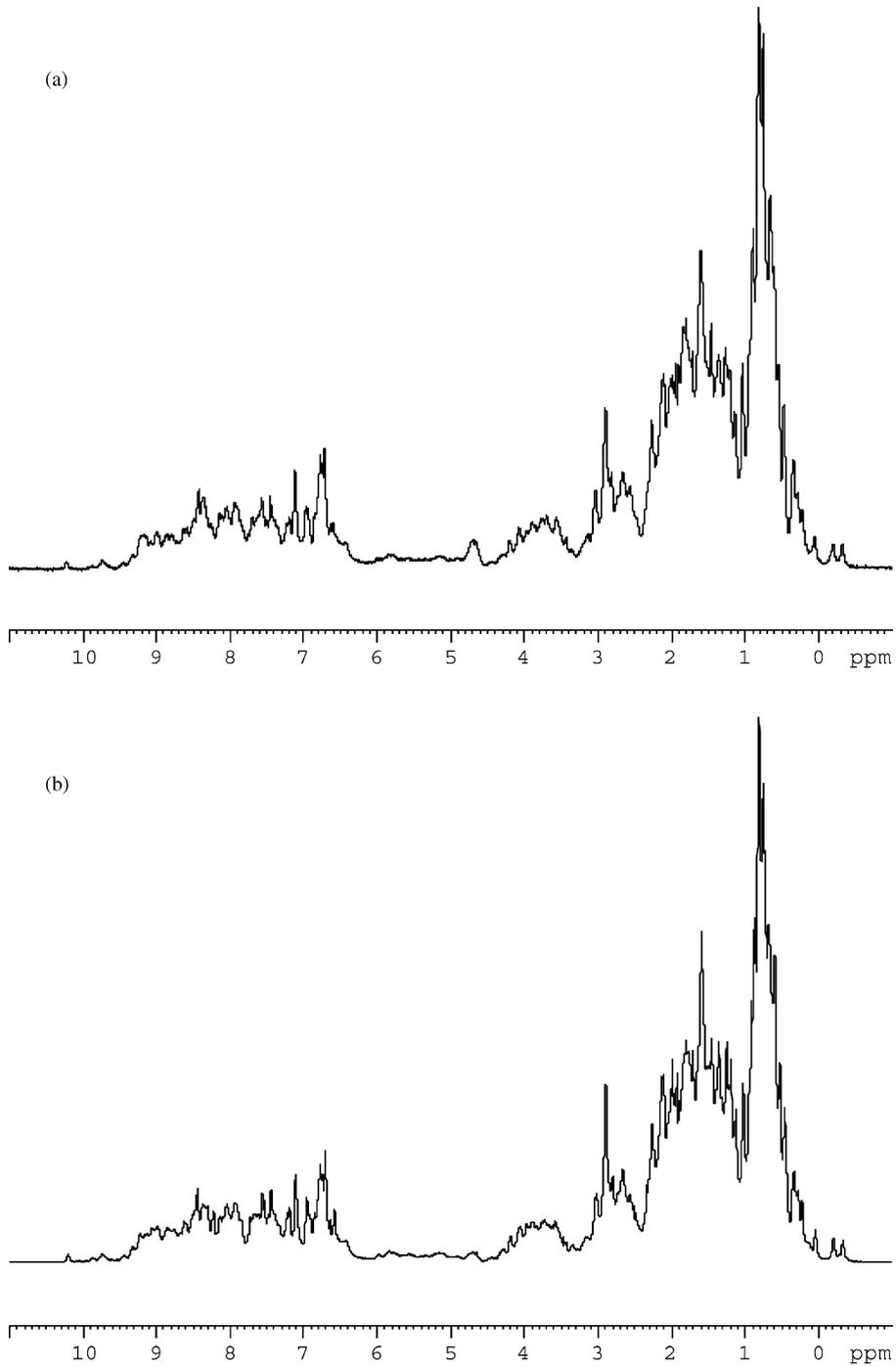


Fig. 3.  $^1\text{H}$  500MHz 1D NMR spectra of 1 mM porcine  $\beta$ -LGs, recorded at pH 5.6 and 27 °C: (a) purified; (b) recombinant.

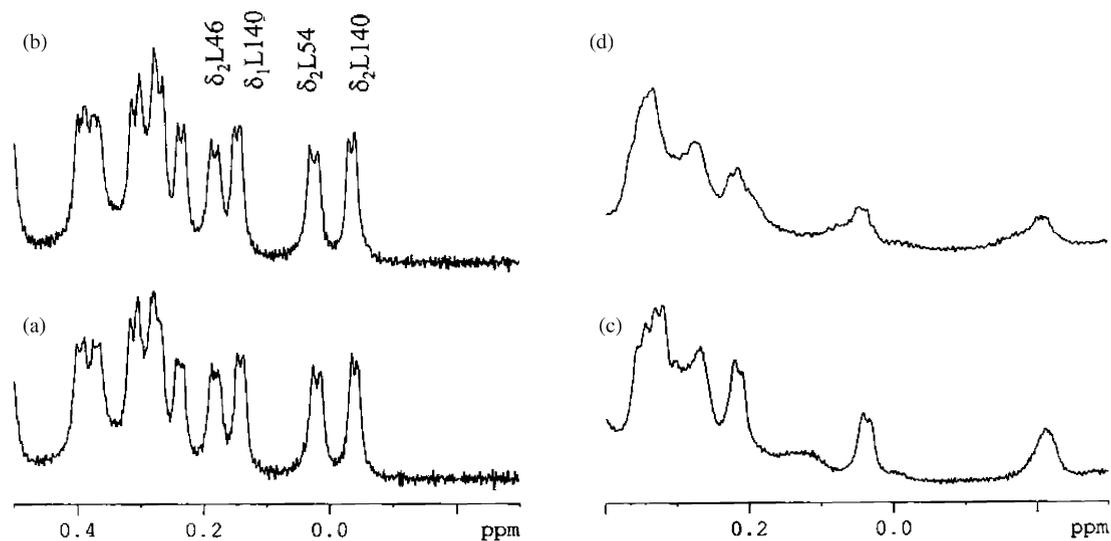


Fig. 4. High-field regions of  $^1\text{H}$  500 MHz 1D NMR spectra of: (a) purified bovine  $\beta$ -LG; (b) recombinant bovine  $\beta$ -LG. The assignment of methyl ( $\delta\text{CH}_3$ ) resonance of L<sub>46</sub>, L<sub>54</sub> and L<sub>140</sub> are indicated; (c) recombinant porcine  $\beta$ -LG; (d) purified porcine  $\beta$ -LGs.

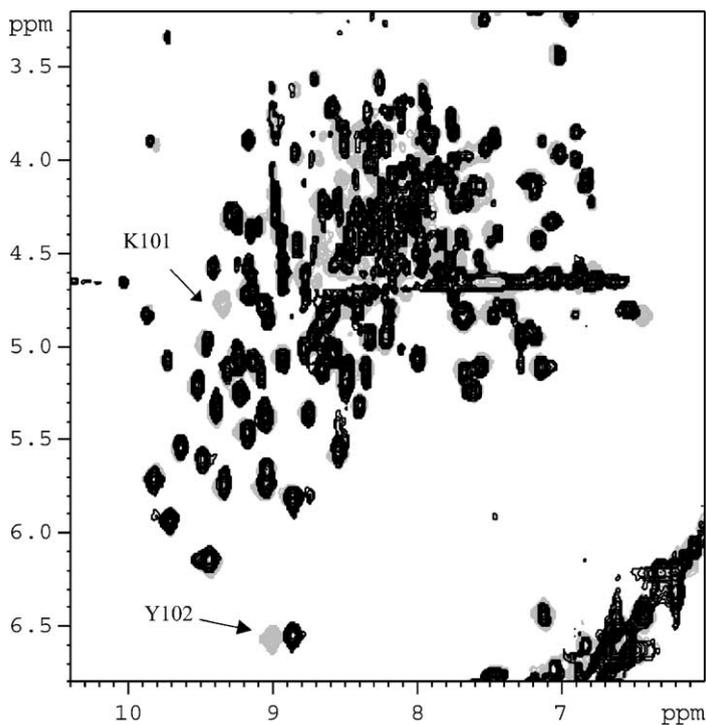


Fig. 5. Superposition of  $^1\text{H}$  500 MHz 2D TOCSY fingerprint regions, showing correlations between amide and  $\text{H}\alpha$  protons, of purified (black) and recombinant (grey) 1 mM samples of bovine  $\beta$ -LG, pH 2, 37°C. The labelled cross-peaks undergo some shift as described in the text.

the phosphate buffer from 12 to 120 mM. Indeed K<sub>101</sub> and Y<sub>102</sub> belong, as defined on the basis of electrostatic calculations, to a high potential region, located in proximity of the FG loop– $\alpha$  helix interface (Fogolari et al., 1998, 2000). These observed differences can be therefore ascribed only to a minor difference in the ionic strength of the two samples. Both spectra also contain some overlapped broad peaks.

We can conclude that the recombinant bovine protein is correctly folded, has the same conformation of the purified protein, thus indicating that the described expression protocol successfully yielded the desired protein.

It is worth mentioning that while a very detailed analysis can be performed for the bovine  $\beta$ -LG spectra, the correctness of the recombinant porcine  $\beta$ -LG fold may be inferred only from its similarity to the spectra of samples purified from milk, due to the fact that we do not still have the complete NMR assignment of the porcine protein. Indeed this assignment was hampered by the fact that milk purification of porcine species lead to a mixture of two isoforms, A and C, which could not be further purified, showing a

double mutation at position 9, glutamic or aspartic and 68, glutamine or histidine, for C and A isoforms, respectively. Therefore, all NMR data reported and obtained for the purified porcine species refer to mixture of proteins, while the recombinant protein is the pure A variant. Some caution is therefore requested when comparing porcine  $\beta$ -LG spectra. It is interesting to observe that the comparison of 1D spectra of Fig. 4c and d, indicates that the linewidth of the recombinant porcine protein are sharper with respect to those of the purified PLG, which refer to an isoform mixture.

From the superposition of the fingerprint regions of 2D TOCSY spectra of purified and recombinant porcine species (Fig. 6) it appears that the linewidth of the milk protein is wider and the resonances less dispersed, possibly due to the heterogeneity of the sample. It is however clear that the displayed pattern of resonances is similar, and it can be concluded that the two proteins share the same conformation, although no further detail could be given. We expect that the higher quality of the spectra and the availability of <sup>15</sup>N isotope labelled samples will allow a prompt structural assignment of the porcine  $\beta$ -LG spectra.

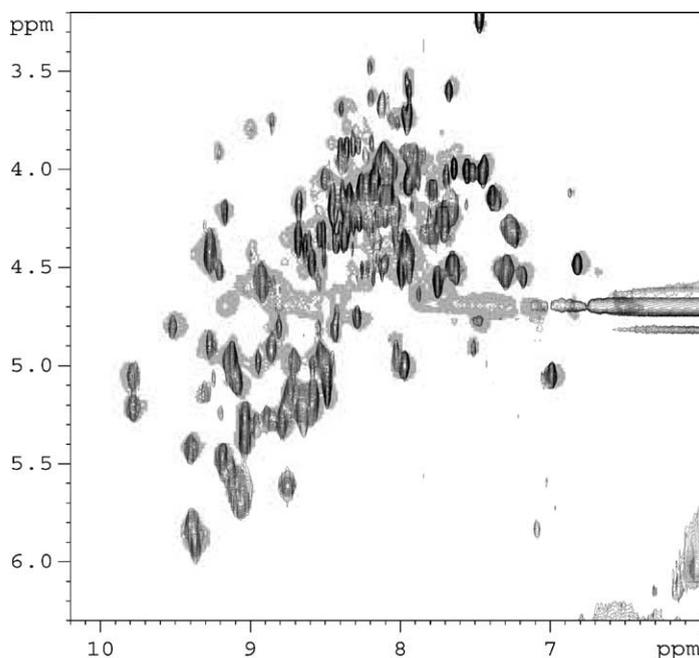


Fig. 6. Superposition of <sup>1</sup>H 500MHz 2D TOCSY fingerprint regions, showing correlations between amide and H $\alpha$  protons, of purified (black) and recombinant (grey) 1 mM samples of porcine  $\beta$ -LG, pH 5.6, 27 °C.

## Acknowledgements

This work was partly supported by a grant M.I.U.R. ex 40% and by a grant F.I.R.B. to HM. The authors wish to thank Monica Galliano for help in purification procedures and Piero Pucci for MS-analysis.

## References

- Bartels, C., Xia, T., Billeter, M., Güntert, P., Wüthrich, K., 1995. The program XEASY for computers-supported NMR spectra analysis of biological macromolecules. *J. Biomol. NMR* 5, 1–10.
- Batt, C.A., Rabson, L.D., Wong, D.W., Kinsella, J.E., 1990. Expression of recombinant bovine beta-lactoglobulin in *Escherichia coli*. *Agric. Biol. Chem.* 54 (4), 949–955.
- Bax, A., Davis, D.G., 1985. MLEV-17 based two-dimensional homonuclear magnetisation transfer spectroscopy. *J. Magn. Reson.* 65, 355–360.
- De Jong, P., 1997.  $\beta$ -Lactoglobulin as a major responsible of heat exchanger inefficiency. *Trends Food Sci. Technol.* 8, 401–405.
- Flower, S.R., North, A.C.T., Sansom, C.E., 2000. The lipocalin protein family: structural and sequence overview. *Biochim. Biophys. Acta* 1482, 9–24.
- Fogolari, F., Ragona, L., Zetta, L., Romagnoli, S., De Kruijff, K.G., Molinari, H., 1998. Monomeric bovine  $\beta$ -lactoglobulin adopts a  $\beta$ -barrel fold at pH. *FEBS Lett.* 436, 149–154.
- Fogolari, F., Licciardi, S., Romagnoli, S., Ragona, L., Michelutti, R., Ugolini, R., Molinari, H., 2000. Electrostatic properties of bovine beta lactoglobulin. *Proteins: Struct. Funct. Genet.* 39, 317–330.
- Ho, S.M., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51–59.
- Hytinen, J.M., Korhonen, V.P., Hiltunen, M.O., Myohanen, S., Janne, J., 1998. High-level expression of bovine beta-lactoglobulin gene in transgenic mice. *J. Biotechnol.* 61 (3), 191–198.
- Kim, T.R., Goto, Y., Hirota, N., Kuwata, K., Denton, H., Wu, S.Y., Sawyer, L., Batt, C.A., 1997. High-level expression of bovine beta-lactoglobulin in *Pichia pastoris* and characterization of its physical properties. *Protein Eng.* 10 (11), 1339–1345.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* 227, 680–685.
- Ragona, L., Pusterla, F., Zetta, L., Monaco, H.L., Molinari, H., 1997. Identification of a conserved hydrophobic cluster in partially folded bovine lactoglobulin at pH 2. *Fold. Des.* 2 (5), 281–290.
- Schlehuber, S., Skerra, A., 2000. Tuning ligand affinity specificity and folding stability of an engineered lipocalin variant—a so-called “anticalin”—using a molecular random approach. *Biophys. Chem.* 96, 213–228.
- Scorer, C.A., Clare, J.J., McCombie, W.R., Romanos, M.A., Sreekrishna, K., 1994. Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *Biotechnology* 12, 181–184.
- Sklenar, V., Piotto, M., Leppik, R., Saudek, V., 1993. Gradient tailored water suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments optimised to retain full sensitivity. *J. Magn. Reson., Ser. A* 102, 241–245.
- Totsuka, M., Katakura, Y., Shimizu, M., Kumagai, I., Miura, K., Kaminogawa, S., 1990. Expression and secretion of bovine beta-lactoglobulin in *Saccharomyces cerevisiae*. *Agric. Biol. Chem.* 54 (12), 3111–3116.
- Ugolini, R., Ragona, L., Silletti, E., Fogolari, F., Visschers, R.W., Alting, A.C., Molinari, H., 2001. Dimerization, stability and electrostatic properties of porcine beta-lactoglobulin. *Eur. J. Biochem.* 268, 4477–4488.
- Uhrinova, S., Uhrin, D., Denton, H., Smith, M., Sawyer, L., Barlow, P.N., 1998. Complete assignment of  $^1\text{H}$ ,  $^{13}\text{C}$   $^{15}\text{N}$  chemical shifts for bovine  $\beta$ -lactoglobulin: secondary structure and topology of the native state is retained in a partially unfolded form. *J. Biomol. NMR* 12, 89–107.
- Vittayanont, M., Steffe, J.F., Fleger, F.L., Smith, D.M., 2002. Gelling properties of heat-denatured beta-lactoglobulin aggregates in high-salt buffer. *J. Agric. Food Chem.* 50 (10), 2987–2992.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., Sykes, B.D., 1995.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 5, 67–81.

1 **Determinants of protein stability and folding: comparative analysis of beta-lactoglobulins**  
2 **and Liver basic fatty acid binding protein**

3  
4 Laura Ragona<sup>1</sup>, Giorgio Colombo<sup>2,\*</sup>, Maddalena Catalano<sup>1</sup> and Henriette Molinari<sup>3,\*</sup>

5  
6 <sup>1</sup>Laboratorio NMR, ISMAC, CNR, via Bassini 15, 20133 Milano, Italy; <sup>2</sup>Istituto di Chimica del  
7 Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131 Milano, Italy; <sup>3</sup>Dipartimento  
8 Scientifico e Tecnologico, Strada Le Grazie 15, 37134 Verona, Italy.

9  
10 \*To whom correspondence should be sent:

11 Prof. Henriette Molinari

17 Dr. Giorgio Colombo

12 Dipartimento Scientifico e Tecnologico,

18 ICRM-CNR

13 Università di Verona, Ca' Vignal 1,

19 Via Mario Bianco 9

14 Strada le Grazie 15, 37134, Verona, Italy.

20 20131 Milano, Italy.

15 Fax:+390223699620

21 Fax.+390228500031

16 E-mail: molinari@sci.univr.it

22 E-mail: g.colombo@icrm.cnr.it

1 Abbreviations: bovine beta-lactoglobulin (BLG), free energy difference ( $\Delta G$ ) , Liver basic fatty

2 Acid Binding Protein (Lb-FABP), porcine beta-lactoglobulin (PLG).

3

4 Key words: lipocalins, fatty acid binding proteins, beta-lactoglobulins, molecular dynamics

5 simulations, folding determinants, sidechain-sidechain conservation

6

**1 Abstract**

2 A new energy decomposition approach, aimed at identifying residues playing a folding key role,  
3 has been applied here to three homologous proteins, belonging to the calycin superfamily,  
4 namely bovine and porcine beta-lactoglobulins and Liver basic fatty acid binding protein,  
5 sharing the same beta-barrel fold and different degree of sequence identities. All-atom, explicit  
6 solvent Molecular Dynamics simulations around the native conformation were used to generate,  
7 for each of the three proteins, energy maps which were further simplified through eigenvalue  
8 decomposition. Analysis of the components of the eigenvector associated with the lowest  
9 eigenvalue singled out those residues (*hot sites*) behaving as strongly interacting and possible  
10 nucleation centers. The results fit well with experimental folding data and, especially, with the  
11 analysis of side-chain/side-chain interaction conservation.

12

13

## 1 **Introduction**

2 A major challenge in the study of folding of protein families is the capability to distinguish  
3 residues conserved for folding from those conserved for function. Indeed conserved sequences in  
4 homologous proteins are recognised to be important for function while contradictory data have  
5 been reported correlating sequence conservation and protein folding. Relevant to this issue, a  
6 conserved signature for folding has been recently proposed for proteins belonging to the  
7 lipocalin family<sup>1</sup>. Indeed, exploratory kinetic studies conducted on human serum Retinol-  
8 Binding Protein, a model lipocalin member, suggested that evolutionary conserved regions fold  
9 on a faster folding timescale than the non conserved regions. To further address whether related  
10 proteins are stabilised by topologically conserved interactions and in light of our long time  
11 interest in the study of calycons<sup>2-6</sup>, we report here the results of a new energy decomposition  
12 strategy, recently proposed by some of us<sup>7</sup>, aimed at obtaining a comparative analysis of the sites  
13 mostly responsible for the stability of the folded state of bovine and porcine beta-lactoglobulins  
14 (BLG, PLG) and of chicken liver-basic fatty acid binding protein (Lb-FABP), belonging to  
15 lipocalins and fatty acid binding protein families, respectively. Both lipocalins and FABPs are  
16 part of the calycin superfamily, sharing the same beta-barrel topology and different degrees of  
17 sequence identities. The approach used here was designed to single out protein sites contributing  
18 mostly to the stability and accessibility of the native state, by means of a simple energetic  
19 analysis of all-atom, explicit solvent Molecular Dynamics (MD) simulation around the native  
20 conformation. The generated energy map (a symmetric matrix) can be simplified through  
21 eigenvalue decomposition, and the analysis of the components of the eigenvector associated with  
22 the lowest eigenvalue proved to be able to identify those sites, referred to as “hot sites”,

1 responsible for protein stability<sup>7</sup>. Hot sites are generally referred to as the sites in proteins which,  
2 if mutated, can cause severe variations in proteins' stability, folding kinetics and mechanisms. In  
3 other words, this analysis is able to single out, in each protein, those residues building clusters of  
4 strong interactions during the simulation. These clusters represent local elementary structures  
5 which may serve as the local native-like nucleation centres, driving the polypeptide to the native  
6 conformation. Indeed this method, tested on a series of proteins representing different folds, was  
7 able to identify between 60 and 80% of the residues playing a key role in folding<sup>7</sup>.

8 Two of the proteins under study, BLG and PLG, are extracellular lipocalins and share: i) the  
9 same beta-barrel topology made up of eight antiparallel beta strands, named from A to H<sup>4,8-10</sup>,  
10 and ii) a high degree of sequence identity and similarity (64% and 82%, respectively). The third  
11 protein, Lb-FABP<sup>6</sup>, is an intracellular lipid binding protein, displaying a low sequence identity  
12 with beta-lactoglobulins (< 20%), but the same beta barrel topology, with the insertion of two  
13 additional beta-strands. With the aim of deriving new information on the folding determinants of  
14 the described proteins, the results of the energetic analysis from 6ns MD simulations in explicit  
15 solvents of the three proteins have been examined in this paper in light of i) experimental folding  
16 data and ii) analysis of side-chain interaction conservation within each protein family.

17

## 18 **Experimental Details**

19 The starting structure for the three all-atom MD simulations were taken from the protein data  
20 bank: cj5 for BLG, 1exs for PLG and 1mvg for Lb-FAB. The mutants of PLG (K35Q and  
21 D88N) were created with the program SWISS-PDBviewer. The proteins were protonated to  
22 obtain a zwitterionic form (with N-terminal NH<sub>3</sub><sup>+</sup> and C-terminal COO<sup>-</sup> groups) with the carboxyl

1 and NH side chain groups in their charged states. The total charges on the proteins resulted  
2 +19 for BLG, -12 for PLG and +2 for Lb-FAB. The proteins were solvated with water in a  
3 octahedral box large enough to contain 1.2 *nm* of solvent around the peptide. The simple point  
4 charge (SPC) water model<sup>11</sup> was used to solvate each protein in the simulation box. In all  
5 simulations, suitably charged counterions were used to yield an electrically neutral system.  
6 Each system was subsequently energy minimized with a steepest descent method for 1000 steps.  
7 The calculation of electrostatic forces utilized the PME implementation of the Ewald summation  
8 method. The LINCS algorithm<sup>12</sup> was used to constrain all bond lengths. For the water molecules  
9 the SETTLE algorithm<sup>13</sup> was used. A dielectric permittivity,  $\epsilon = 1$ , and a time step of 2 fs were  
10 used. All atoms were given an initial velocity obtained from a Maxwellian distribution at the  
11 desired initial temperature of 300K. The density of the system was adjusted performing the  
12 first equilibration runs at NPT condition by weak coupling to a bath of constant pressure  
13 ( $P_0 = 1$  bar, coupling time  $\tau_p = 0.5$  ps)<sup>13,14</sup>. In all simulations the temperature was maintained  
14 close to the intended values by weak coupling to an external temperature bath<sup>14</sup> with a coupling  
15 constant of 0.1 ps. The protein and the rest of the system were coupled separately to the  
16 temperature bath. Each of the five MD simulations was extended to 10 ns. All simulations and  
17 analysis were carried out using the GROMACS package<sup>15</sup> using the GROMOS96 43A1 force  
18 field<sup>15</sup>. All calculations were performed on clusters of PCs, with Linux operating system.

19 The analysis of the energetics of each of the trajectories to identify the residues associated with  
20 low energy intra-protein interactions was carried out on the last 2 ns of each simulation. The  
21 methodology is the same as previously described<sup>7</sup>. In order to summarize it here, the main steps  
22 will be presented. Based on the calculation of the non-bonded (electrostatic, H-bond and van der

1 Waals) interactions among all non consecutive residues in the protein under examination, a  
2 symmetric interaction energy matrix is built (see ref. 7 for more details). This matrix is  
3 simplified through eigenvalue decomposition. The resulting eigenvalues are sorted in order of  
4 increasing values and the differences between the value of the lowest eigenvalue ( $\lambda_1$ ) and the  
5 second lowest eigenvalue ( $\lambda_2$ ) is compared with the average difference between all the other  
6 eigenvalues. If  $\Delta\lambda_{(\lambda_1-2)}$  is larger than the average spacing between the other eigenvalues, then  
7 eigenvector  $\mu^j$  (associated with ( $\lambda_1$ )) can be considered to contain the information on the  
8 spectrum of residues most strongly interacting in the native state of each protein (see the  
9 following Results section). The analysis of the components of this eigenvector is thus able to  
10 identify those sites, referred to as “hot sites”.

## 11 **Results**

12 The structures of lipocalins and fatty acid binding proteins are represented in Figure 1, while  
13 Table 1 reports the structurally aligned regions of BLG and Lb-FABP as obtained through the  
14 VAST server ([www.ncbi.nih.gov/Structure/VAST/vast.html](http://www.ncbi.nih.gov/Structure/VAST/vast.html)).

15 **Hot sites analysis.** Long time-scale (6 ns) MD simulations in explicit solvent were run starting  
16 from the experimental structures of BLG, PLG and Lb-FABP. The relative energy analysis was  
17 run, as described in Experimental details, on the last 2 ns of each simulation. For all the proteins  
18 examined the separation  $\Delta\lambda$  between eigenvalues  $\lambda_1$  and  $\lambda_2$  ( $\Delta\lambda_{(\lambda_1-2)}$ ) is much larger than the  
19 average spacing  $\lambda$  ( $\Delta\bar{\lambda}$ ) between the other eigenvalues.  $\Delta\lambda_{(\lambda_1-2)}$  and  $\Delta\bar{\lambda}$  are 12.3 and 2.1,  
20 respectively, for BLG, 6.3 and 1.4 for PLG and 10.6 and 1.8 for Lb-FABP, showing that the  
21 overall energetic interaction matrix can be approximated by  $M_{ij} = \lambda_1 \mu_i^j \mu_j^j$ , where  $\mu_i^j$  is the  
22 eigenvector associated with  $\lambda_1$ . Eigenvector  $\mu^j$  thus contains the information on the spectrum of

1 residues most strongly interacting in the native state of each protein. In order to show how the  
2 stabilisation energy is distributed among different residues, the components  $\mu_i$  of the normalised  
3 eigenvector with the lowest eigenvalue are plotted against the residue number for BLG and PLG  
4 (Figure 2). The definition of a residue as determinant for the stabilization of protein native state  
5 (hot site) on the basis of the components of the eigenvector was derived from the original  
6 paper<sup>7</sup>. The most probable value of the eigenvector components distribution was calculated, and  
7 all the sites of the protein displaying an eigenvector component larger than the most probable  
8 value (threshold line of Figure 2) were defined as hot sites.

9 In BLG particularly large values of the energetic components are observed in the intervals 18-24  
10 (located in A strand), 40-46 (B strand), 56-61 (C strand), 68-69 (D strand), 102-107 (G strand),  
11 118-124 (H strand), 136, 140 (C-terminal  $\alpha$ -helix), 151, 152, 159.

12 In PLG the highest values are observed at the level of residues 11, 14-15, 19-21 (A strand), 35-  
13 37, 40-41 (AB loop), 42-44 (B strand), 58-60 (C strand), 69 (D strand), 87-89 (EF loop), 93, 95-  
14 99 (F strand), 100-109 (G strand), 120, 122-124 (H strand), 129, 131, 132, 135-136 (C-terminal  
15  $\alpha$ -helix), 159-160 (C terminal tail). For Lb-FABP hot sites were located in the regions 2, 5-19 (A  
16 strand), 35-40 (B strand), 88-91 (G strand), 97-103 (H strand), 105-112 (I strand), 118-124 (J  
17 strand) (Figure 3).

18 **Essential dynamics (ED) analysis.** The characterization of the dynamical properties of the three  
19 proteins was also performed utilizing a principal component analysis of the covariance matrix of  
20 the positional fluctuations (essential dynamics analysis) of the C-alpha atoms, as described  
21 elsewhere<sup>16</sup>. This matrix was built from the equilibrated parts of trajectories (last 2 ns), and its  
22 diagonalization yielded the principal directions (essential eigenvectors) associated to the large-

1 amplitude concerted motions that define the essential subspace of a protein internal dynamics.  
2 Figure 4 shows that the atom displacements along the first two main geometric directions of  
3 motion are, in most cases, minimal ( $< 1\text{\AA}$ ) for those residues corresponding to the highest energy  
4 peaks in the first eigenvalue of the energy decomposition described before. Exceptions are  
5 represented by residue 155 in BLG (Figure 4a) and 11, 14-15, 37 and 87 in PLG (Figure 4b).  
6 From this analysis it appears that hot sites exhibit minimal flexibility. Cross-examination of the  
7 mean displacements along the first two main essential directions in BLG and PLG also suggests  
8 that the latter protein is characterized by wider displacements, indicating increased backbone  
9 flexibility.

10 For Lb\_FABP ED analysis indicated that atom displacement of residues belonging to hot sites is  
11 lower than  $1\text{\AA}$  in almost all cases (exceptions are residues 2 and 103), thus indicating that also  
12 for this protein residues highly contributing to protein stability exhibit low backbone motions  
13 (Figure 4c).

14 ***Analysis of topologically conserved interactions.*** Gierasch and coworkers<sup>17</sup> performed a  
15 sequence and structural analysis of intracellular Lipid binding protein family, including Lb-  
16 FABP, in order to single out the conserved side-chain/side-chain interactions within this protein  
17 family. Following the same approach, 14 sequences, with more than 30% identity with BLG,  
18 were identified in the SWISSPROT sequence database. A CLUSTALW<sup>18</sup> multiple sequence  
19 alignment (Figure 5) was then performed, in order to analyse the conservation of interacting  
20 pairs within the family. Two residues were considered as an “interacting pair” if the interatomic  
21 distance between any side-chain atom of the first residue was within  $4\text{\AA}$  of any side-chain atom  
22 of the second residue. A conservation score was then defined, using a residue classification based

1 on side-chain character hydrophobic/polar, as previously described<sup>17</sup>. A total of 136 long range  
2 hydrophobic-hydrophobic (H-H) and 40 polar-polar (P-P) interacting residues were first  
3 identified in BLG. These figures reduced to 81 H-H and 21 P-P interactions conserved in all the  
4 14 sequences examined (Table II). Out of the 81 H-H conserved interactions, 50 involved  
5 residues located in beta-strands, 6 corresponded to beta barrel/C-terminal  $\alpha$ -helix interactions.  
6 Interestingly only one polar-polar fully conserved interaction is present in the barrel, thus  
7 confirming the relevance of the hydrophobic network in the stabilisation of beta-lactoglobulins,  
8 as previously reported<sup>19</sup>. It is worth noting that most of the conserved P-P interactions lie in loop  
9 regions suggesting a functional role rather than a stabilising function. Indeed we have previously  
10 demonstrated the functional role played in ligand binding by polar-polar interactions among  
11 residues located in the EF loop and residue in position 116 (GH loop)<sup>5</sup>.  
12 The analysis of topologically conserved interactions for intracellular Lipid binding protein  
13 family has been already reported<sup>17</sup>.

14

## 15 **Discussion**

16 Recently a high number of papers have appeared suggesting that folding mechanisms are largely  
17 determined by the native state topology and a significant correlation between folding rate and  
18 measures of the average topological complexity, such as relative contact order, have been  
19 reported<sup>20</sup>. In this line our approach should provide information on the protein sites which  
20 contribute most to the stability of the native state, thus giving information on the interactions  
21 determining the folding pathways and predicting sites where to focus mutagenesis experiments.

1 Even if BLG has not been characterised in depth in terms of the mutations needed to (de)stabilise  
2 the folded state, as is the case for other proteins<sup>21-24</sup>, a wealth of data were reported by us and  
3 others dealing with thermodynamic and kinetic experiments. For a critical assessment of the  
4 results obtained from the application of our computational method to BLG, hot sites distribution  
5 has been compared with many different experimental results, as described in the following  
6 paragraphs.

7 i) BLG stability changes upon the mutations C121S ( $\Delta\Delta G_{N-U} = 5.5$  kcal/mol) and C121V ( $\Delta\Delta G_{N-}$   
8  $U = 4.8$  kcal/mol) indicated a notable destabilising effect depending on steric hindrance and  
9 polarity of the substituent<sup>25</sup>. Interestingly residue 121 (strand H) shows a large value of the  
10 eigenvector component, thus confirming the ability of the computational method to predict the  
11 effect of mutations.

12 ii) Data on BLG amides stability, as derived from H/D exchange from the native state<sup>26</sup> and from  
13 unfolding NMR measurements<sup>2</sup> indicated that highly protected and stable amides are evenly  
14 distributed in all the strands of the barrel and in the terminal helix, while, interestingly, hot sites  
15 are absent in E and F strands (Figure 6 a,b). The stabilisation energy derived from dynamics  
16 simulations is mainly determined by sidechain-sidechain interactions and does not reflect  
17 properly the amides behaviour, possibly explaining this observed discrepancy.

18 iii) High proton occupancy ( $> 0.7$ ), as derived from H/D labelling refolding experiments<sup>27</sup>, was  
19 measured for residues belonging to G and H strands (see Figure 6c), in good agreement with the  
20 high values of  $\mu_1$  in this region of the protein. Further NMR H/D burst phase labelling  
21 experiments suggested the onset of marginally stable hydrogen bonded structures in a more  
22 extended region, including A, F, G, H strands and the major alpha-helix (Figure 6d).

1 iv) It has been recently proposed that evolutionary conserved regions (ECR) within the lipocalin  
2 family may be characterised by a faster folding timescale with respect to the non conserved  
3 regions<sup>1,28</sup>. The five ECR of lipocalins are localised, for beta-lactoglobulins, in the segments: 9-  
4 25 (ECRI, A strand), 80-82 (ECRII, E strand), 94-103 (ECRIII, F and G strand), 121-124  
5 (ECRIV, H strand) and 131-136 (ECRV, terminal helix). The comparison of hot sites distribution  
6 and ECR (Figure 6e), points to a discrepancy at the level of E and F strands, which apparently do  
7 not contribute to the global stabilisation of BLG native state.

8 From the above comparisons it appears that, irrespective of the considered experimental data, the  
9 GH hot site region is always identified as playing a crucial role in protein stability. Indeed we  
10 have shown by NMR<sup>29</sup> that BLG peptide covering residues 102-118 (G and H strands) exhibited  
11 native-like interactions, thus behaving as a folding initiation site.

12 The comparison of the computational results obtained for BLG and PLG offers new insights into  
13 the key determinants of beta-lactoglobulins stability and folding. Interestingly a different  
14 distribution of eigenvector components was observed for the two proteins, sharing the same  $\beta$ -  
15 barrel topology and a relatively high sequence identity (64%). Hot sites in BLG are accumulated  
16 in a contiguous, rather extended, core region, while in PLG only a three-stranded contiguous  
17 core is present (Figure 7). Hot sites are more scattered in PLG, including also AB loop, EF loop  
18 and F strand regions, thus suggesting the absence of a strongly localized stabilized nucleus,  
19 whose absence is reflected in a different flexibility of the protein with respect to BLG (see ED  
20 analysis). AB loop is one of the regions where major structural differences were observed for  
21 BLG and PLG, possibly as a consequence of a different charge distribution due to a charge  
22 inversion for the closed residues 35 and 88 (F strand). Interestingly, the *in silico* mutation of

1 PLG K35 and D88 to Q35 and N88 (to reproduce BLG sequence), resulted in a net variation of  
2 peaks distribution in the first eigenvector. In particular, in the mutated regions, the resulting  
3 pattern was similar to that of BLG, but, at the same time, changes were observed for residues 7-  
4 10, 52-54 and 73-78 (located far apart from the mutated regions) which now become hot sites.  
5 These results indicate that local mutations, at the level of loop regions, are reflected on the global  
6 fold and suggest that polar-polar interactions play an important role in the stabilization of PLG.  
7 This observation correlates also with possible differences in the folding mechanism of the two  
8 proteins, in terms of stabilizing interactions that have to be formed along the folding pathway to  
9 drive the system into the native basin.

10 Side-chains/side-chains interactions have a prominent role on  $\beta$ -barrel protein stability and the  
11 comparison of topologically conserved interactions with hot site distribution further pointed to  
12 differences between the two lactoglobulins. In BLG, out of 46 hot site residues, 22 are involved  
13 in 30 fully conserved H-H interactions (Table I and Figure 8). In PLG only 12 residues (out of  
14 42) are involved in 13 fully conserved H-H interactions. This observation suggests that in  
15 porcine protein the network of conserved hydrophobic interactions is significantly loosened. The  
16 offered picture is in agreement with fluorescence, CD and NMR unfolding data revealing a  
17 minor stability for the porcine protein ( $\Delta G_{\text{NU}} \sim 7$  kcal/mol) with respect to the bovine protein  
18 ( $\Delta G_{\text{NU}} \sim 12$  kcal/mol)<sup>30</sup>. The formation of a non-native  $\alpha$ -helical intermediate has been reported  
19 to be a common feature for BLG and PLG<sup>30</sup> but the thermodynamic stability and, therefore, the  
20 population of this intermediate state is significantly higher for PLG, possibly due to the different  
21 distribution and different kinds of residue-residue interactions stabilising the protein. Indeed  
22 when fully conserved P-P interactions, involving residues belonging to hot sites, are considered,

1 again BLG and PLG behaves differently: in PLG 8 fully conserved P-P interactions are present,  
2 while only three are observed among BLG hot sites. The relevance of P-P interactions in PLG  
3 was also confirmed by the observation that mutations K35Q and D88N, implying the loss of  
4 S36-D88 electrostatic interaction connecting AB and EF loops, determines a significant variation  
5 in the distribution of the energy peaks represented in the first eigenvector.

6 The difference in the stabilizing interactions between the two proteins is also reflected in the  
7 different conformational dynamic behaviour extracted from ED analysis. Essential dynamics  
8 analysis (Figures 4a, b) indicated that both in BLG and PLG atom displacement of residues  
9 belonging to hot sites is almost always lower than 1 Å. However, higher atom displacements  
10 along the two first directions of motion were observed for PLG with respect to BLG, suggesting  
11 an increased backbone dynamics for the porcine protein, possibly related to the observed minor  
12 stability.

13 The energetic analysis has been extended to Lb-FABP, sharing a beta-barrel topology with beta-  
14 lactoglobulins, as shown in Figure 1. Topologically conserved interactions in intracellular Lipid  
15 binding protein family have been investigated by Gunasekaran et al.<sup>17</sup> who analysed 52  
16 sequences from SWISSPROT database, including Lb-FABP, and performed the structural  
17 analysis on the basis of CRABP I interactions. Thirty residue pairs, conserved in 90% of the  
18 examined sequences, were identified using a distance cutoff of 4 Å. A total of 10 conserved H-H  
19 and 2 P-P interactions, involve hot sites residues in Lb-FABP: F2-I40, F2-L89, F2-M107, W6-  
20 I40, W6-V102, W6-M107, Y14-I111, F17-I111, I40-M107, V102-M107, Q11-R120, E109-S122  
21 (Figure 9a). From the analysis of Figure 9a, it is evident that in Lb-FABP most of the residues  
22 highly contributing to global stability are involved in conserved side-chain interactions,

1 suggesting that the intersection of these two categories of residues may highlight residues  
2 playing a key role in folding thus distinguishing amino acids relevant for stability from those  
3 important for protein function. Conserved interactions are mainly located in a beta core involving  
4 strands A, H, I, J and in a region at the interface between helix I and the barrel (Figure 9b).  
5 Residues F2, W6, I40, V102 and M107 belong to a hydrophobic cluster, reported to play an  
6 important role for protein stability<sup>6</sup>. Residues Y14, F17 and I111 belong to the second identified  
7 hydrophobic cluster<sup>6</sup>, located at the entrance of the protein cavity. The regions presenting the  
8 highest number of hot sites and sidechain-sidechain conservation are at the level of A, I and J  
9 strands.

10 The comparison of hot sites distribution in the structurally aligned BLG, PLG and Lb-FABP, is  
11 reported in Figure 10. The three proteins share three hot site regions, namely A, G and H strands  
12 for beta-lactoglobulins and A, I and J strands for Lb-FABP, in very good agreement with the data  
13 reported on the key of G, H<sup>29</sup> (for lipocalins) and I and J (for cellular retinoic acid-binding  
14 protein<sup>31,32</sup>) regions as folding initiation sites.

15 In conclusion, we have shown here that the application of the simplified energetic analysis of all-  
16 atom MD trajectories, applied to homologous proteins belonging to the calycin superfamily, can  
17 be helpful in the identification of key residues, determining the stability and folding properties.  
18 Differences in the hot sites distribution observed for the two highly correlated bovine and  
19 porcine beta-lactoglobulins may in fact be interpreted as an indication of the presence of  
20 different folding pathways.

21 Following these results site-directed mutagenesis experiments are in progress in our laboratory to  
22 further explore determinants of protein folding and stability in calycins.

1

2 **Acknowledgements**

3 This work was financially supported by MIUR ex 40% 2002 and FIRB 2001. The authors are  
4 indebted to Dr. Raffaella Ugolini for helpful discussions.

5

1 **References**

- 2 1. Greene LH, Hamada D, Eyles SJ, Brew K. Conserved signature proposed for folding in  
3 the lipocalin superfamily. *FEBS Lett* 2003;553(1-2):39-44.
- 4 2. Ragona L, Fogolari F, Romagnoli S, Zetta L, Maubois JL, Molinari H. Unfolding and  
5 refolding of bovine beta-lactoglobulin monitored by hydrogen exchange measurements. *J*  
6 *Mol Biol* 1999;293(4):953-969.
- 7 3. Fogolari F, Ragona L, Licciardi S, Romagnoli S, Michelutti R, Ugolini R, Molinari H.  
8 Electrostatic properties of bovine beta-lactoglobulin. *Proteins* 2000;39(4):317-330.
- 9 4. Ugolini R, Ragona L, Silletti E, Fogolari F, Visschers RW, Alting AC, Molinari H.  
10 Dimerization, stability and electrostatic properties of porcine beta-lactoglobulin. *Eur J*  
11 *Biochem* 2001;268(16):4477-4488.
- 12 5. Ragona L, Fogolari F, Catalano M, Ugolini R, Zetta L, Molinari H. EF loop  
13 conformational change triggers ligand binding in beta-lactoglobulins. *J Biol Chem*  
14 2003;278(40):38840-38846.
- 15 6. Vasile F, Ragona L, Catalano M, Zetta L, Perduca M, Monaco H, Molinari H. Solution  
16 structure of chicken liver basic fatty acid binding protein. *J Biomol NMR*  
17 2003;25(2):157-160.
- 18 7. Tiana G, Simona F, De Mori GM, Broglia RA, Colombo G. Understanding the  
19 determinants of stability and folding of small globular proteins from their energetics.  
20 *Protein Sci* 2004;13(1):113-124.

- 1 8. Fogolari F, Ragona L, Zetta L, Romagnoli S, De Kruijff KG, Molinari H. Monomeric  
2 bovine beta-lactoglobulin adopts a beta-barrel fold at pH 2. FEBS Lett 1998;436(2):149-  
3 154.
- 4 9. Brownlow S, Morais Cabral JH, Cooper R, Flower DR, Yewdall SJ, Polikarpov I, North  
5 AC, Sawyer L. Bovine beta-lactoglobulin at 1.8 Å resolution--still an enigmatic lipocalin.  
6 Structure 1997;5(4):481-495.
- 7 10. Kuwata K, Hoshino M, Forge V, Era S, Batt CA, Goto Y. Solution structure and  
8 dynamics of bovine beta-lactoglobulin A. Protein Sci 1999;8(11):2541-2545.
- 9 11. Berendsen H, Grigera J, Straatsma T. The missing term in effective pair potentials. J Phys  
10 Chem 1987;91:6269-6271.
- 11 12. Hess B, Bekker, H., Fraaije, J. G. E. M., and Berendsen, H. J. C. A linear constraint  
12 solver for molecular simulations. JCompChem 1997;18:1463-1472.
- 13 13. Miyamoto S, Kollman P. Settle: An analytical version of the shake and rattle algorithms  
14 for rigid water models. J Comp Chem 1992;13:952-962.
- 15 14. van Gunsteren WF, Berendsen HJ. Computer simulation as a tool for tracing the  
16 conformational differences between proteins in solution and in the crystalline state. J Mol  
17 Biol 1984;176(4):559-564.
- 18 15. van Gunsteren WF, Daura X, Mark AE. Encyclopedia of Computational Chemistry;  
19 1998. 1211-1216 p.
- 20 16. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins  
21 1993;17(4):412-425.

- 1 17. Gunasekaran K, Hagler AT, Gierasch LM. Sequence and structural analysis of cellular  
2 retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions.  
3 *Proteins* 2004;54(2):179-194.
- 4 18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of  
5 progressive multiple sequence alignment through sequence weighting, position-specific  
6 gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22(22):4673-4680.
- 7 19. Ragona L, Pusterla F, Zetta L, Monaco HL, Molinari H. Identification of a conserved  
8 hydrophobic cluster in partially folded bovine beta-lactoglobulin at pH 2. *Fold Des*  
9 1997;2(5):281-290.
- 10 20. Kuznetsov IB, Rackovsky S. Class-specific correlations between protein folding rate,  
11 structure-derived, and sequence-derived descriptors. *Proteins* 2004;54(2):333-341.
- 12 21. Martinez JC, Pisabarro MT, Serrano L. Obligatory steps in protein folding and the  
13 conformational diversity of the transition state. *Nat Struct Biol* 1998;5(8):721-729.
- 14 22. Grantcharova VP, Riddle DS, Santiago JV, Baker D. Important role of hydrogen bonds in  
15 the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct*  
16 *Biol* 1998;5(8):714-720.
- 17 23. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and  
18 theory highlight role of native state topology in SH3 folding. *Nat Struct Biol*  
19 1999;6(11):1016-1024.
- 20 24. McCallister EL, Alm E, Baker D. Critical role of beta-hairpin formation in protein G  
21 folding. *Nat Struct Biol* 2000;7(8):669-673.

- 1 25. Yagi M, Sakurai K, Kalidas C, Batt CA, Goto Y. Reversible unfolding of bovine beta-  
2 lactoglobulin mutants without a free thiol group. *J Biol Chem* 2003;278(47):47009-  
3 47015.
- 4 26. Forge V, Hoshino M, Kuwata K, Arai M, Kuwajima K, Batt CA, Goto Y. Is folding of  
5 beta-lactoglobulin non-hierarchic? Intermediate with native-like beta-sheet and non-  
6 native alpha-helix. *J Mol Biol* 2000;296(4):1039-1051.
- 7 27. Kuwata K, Shastry R, Cheng H, Hoshino M, Batt CA, Goto Y, Roder H. Structural and  
8 kinetic characterization of early folding events in beta-lactoglobulin. *Nat Struct Biol*  
9 2001;8(2):151-155.
- 10 28. Greene LH, Chrysina ED, Irons LI, Papageorgiou AC, Acharya KR, Brew K. Role of  
11 conserved residues in structure and stability: tryptophans of human serum retinol-binding  
12 protein, a model for the lipocalin superfamily. *Protein Sci* 2001;10(11):2301-2316.
- 13 29. Ragona L, Catalano M, Zetta L, Longhi R, Fogolari F, Molinari H. Peptide models of  
14 folding initiation sites of bovine beta-lactoglobulin: identification of nativelylike  
15 hydrophobic interactions involving G and H strands. *Biochemistry* 2002;41(8):2786-  
16 2796.
- 17 30. D'Alfonso L, Collini M, Ragona L, Ugolini R, Baldini G, Molinari H. Porcine beta-  
18 lactoglobulin chemical unfolding: Identification of a non-native alpha-helical  
19 intermediate. *Proteins* 2005;58(1):70-79.
- 20 31. Rotondi KS, Gierasch LM. Role of local sequence in the folding of cellular retinoic  
21 acid-binding protein I: structural propensities of reverse turns. *Biochemistry*  
22 2003;42(26):7976-7985.

- 1 32. Rotondi KS, Gierasch LM. Local sequence information in cellular retinoic acid-binding
- 2 protein I: specific residue roles in beta-turns. *Biopolymers* 2003;71(6):638-651.
- 3

## 1           **Figure Legends**

2   **Figure 1** Schematic calycin evolutionary tree together with the structures of BLG (PDB code  
3 1cj5), PLG (PDB code 1exs) and Lb-FABP (PDB code 1mvg).

4   **Figure 2** The eigenvectors associated with the largest eigenvalue in the all-atom MD simulations  
5 of BLG (black bars) and PLG (grey bars), calculated on the last 2 ns of each simulation. The  
6 horizontal line represents the threshold, calculated as the most probable value of the eigenvector  
7 components distribution. Regions of secondary structure ( $\beta$ -strands ■ and  $\alpha$ -helices ○○○○) are  
8 indicated at the top of the diagram.

9   **Figure 3** The eigenvectors associated with the largest eigenvalue in the all-atom MD simulation  
10 of Lb-FABP, calculated on the last 2 ns of simulation. Other symbols are as in Figure 2.

11   **Figure 4** Mean atom displacements along the first two main essential directions in a) BLG, b)  
12 PLG and c) Lb-FABP. Residues belonging to hot sites are marked with a circle.

13   **Figure 5** Clustal W multiple alignment of beta-lactoglobulins from different species showing  
14 more than 30% identity with BLG.

15   **Figure 6** Comparison of Eigenvectors associated with the largest Eigenvalue in the all atom MD  
16 simulations of BLG with: a) H/D exchange data from the native state<sup>26</sup>: residues **with protection**  
17 **factors (P)** higher than 3000 are marked with a circle; b)  $\Delta G_{\text{unf}}$  of amide protons<sup>2</sup>: residues with  
18  $\Delta G_{\text{unf}}$  higher than 6 kcal/mol are marked with a circle; c) proton occupancy derived from  
19 refolding experiments, detected by H/D exchange<sup>26</sup>: residues with proton occupancies higher  
20 than 0.7 are marked with a circle; d) amide protons protection factors measured by NMR burst  
21 phase labelling (**< 1.8 ms refolding time**)<sup>27</sup>: residues with **P** higher than 2 are marked with a  
22 circle; e) evolutionary conserved regions of lipocalins: residues belonging to ECR are marked

1 with a circle. The meaning of the horizontal line is defined in the caption in Figure 2. Residues  
2 with  $\mu > 0.8$  are considered hot sites.

3 **Figure 7** Graphical representation of the distribution of eigenvector components (hot sites, in  
4 black in the structures) resulting from computation: BLG (left) and PLG (right).

5 **Figure 8** Conserved H-H (filled circle) and P-P (filled triangle) side-chains/side-chains  
6 interactions are indicated on the graph representing the distribution of Eigenvectors associated  
7 with the largest Eigenvalue in the all atom MD simulations of BLG (a) and PLG (b).

8 **Figure 9** Conserved H-H (filled circle) and P-P (filled triangle) side-chains/side-chains  
9 interactions are indicated on the graph representing the distribution of Eigenvectors associated  
10 with the largest Eigenvalue in the all atom MD simulations of Lb-FABP (a). Hot sites involved  
11 in H-H side-chain/side-chain interactions are colored in red while those involved in P-P  
12 interactions are colored in blue (b).

13 **Figure 10** Comparison of Eigenvectors associated with the largest Eigenvalue in the all atom  
14 MD simulations of BLG (continuous line), PLG (stacked line), and Lb\_FABP (dotted line). The  
15 structural alignment was obtained through the VAST server ([www.ncbi.nih.  
16 gov/Structure/VAST/vast.shtml](http://www.ncbi.nih.gov/Structure/VAST/vast.shtml)).

17

18

1 **Tables**2 **Table I. Structurally aligned regions of BLG and Lb-FABP**

BLG	14-21	23-34	42-53	55-61	66-72	74-83	86-97	102-109	115-124
Strand	A	AB loop	B	C	D	DE loop	F	G	H
Lb-FABP	1-8	9-20	36-47	48-54	55-61	82-91	93-104	105-112	116-125
Strand	A	A	B	C	D	F, G	H	I	J

3

4

1 **Table II. Conserved long range interactions**  
 2 **in beta-lactoglobulins<sup>a</sup>**

<b>Hydrophobic-Hydrophobic</b>		<b>Polar-Polar</b>
M7-Y99	V43-L58	D11-D98
V15-L46	V43-F105	Q13-E51
V15-L103	V43-L122	K14-D98
A16-L46	L46-L54	S27-Q115
A16-P48	L46-L56	S27-S116
A16-L54	L54-I71	D28-Q115
W19-V43	L54-I72	S30-Q115
W19-L46	L54-A73	Q35-N63
W19-L103	I56-I71	E44-Q59
W19-L122	L58-A67	E44-Q159
L22-V128	C66-C160	K47-E55
L22-A132	I71-I84	T49-E55
A23-C119	I72-I84	D53-E74
A23-F136	A73-F82	Q59-Q159
M24-L32	A73-I84	D98-R124
M24-L39	A80-V92	K101-R124
M24-V41	A80-V94	K101-T125
M24-A118	F82-V92	K101-E131
M24-C119	F82-V94	K101-K135

<b>Hydrophobic-Hydrophobic</b>		<b>Polar-Polar</b>
A25-L32	I84-V92	T125-E131
A25-A118	V92-F105	K141-R148
A25-C119	V92-C106	
A25-I147	V94-L103	
A26-L32	V94-L104	
A26-L117	V94-F105	
A26-A118	L95-L104	
A26-I147	Y102-C121	
I29-I147	L103-L122	
L31-P38	L104-F136	
L31-L39	L104-A139	
L31-L117	L104-L140	
L31-A118	F105-A118	
L32-L38	F105-C119	
L32-L39	F105-L122	
L32-L118	C106-L117	
L32-I147	C106-A118	
L39-A118	C106-C119	
V41-L58	C119-L140	
Y42-L58	V128-P153	
V43-I56	A132-A139	

<b>Hydrophobic-Hydrophobic</b>	<b>Polar-Polar</b>
V43-L57	

1 <sup>a</sup> Interactions identified in BLG structure 1CJ5  
2 (first model) by using 4 Å distance cutoff and  
3 conserved in the 14 lactoglobulin sequences.  
4

5



## Letter to the Editor: Solution structure of chicken liver basic fatty acid binding protein

Francesca Vasile<sup>a</sup>, Laura Ragona<sup>b</sup>, Maddalena Catalano<sup>b</sup>, Lucia Zetta<sup>b</sup>, Massimiliano Perduca<sup>a</sup>, Hugo Monaco<sup>a</sup> & Henriette Molinari<sup>a,b,\*</sup>

<sup>a</sup>Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy; <sup>b</sup>Laboratorio NMR, ISMAC-CNR, Via Ampere 56, 20131 Milano, Italy

Received 9 October 2002; Accepted 6 November 2002

**Key words:** calycons, chicken liver-basic FABP, NMR structure

### Abstract

Chicken liver basic fatty acid binding protein (Lb-FABP) belongs to the basic-type fatty acid binding proteins, a novel group of proteins isolated from liver of different non mammalian species whose structure is not known. The structure of Lb-FABP has been solved by <sup>1</sup>H NMR. The overall fold of Lb-FABP, common to the other proteins of the family, consists of ten antiparallel  $\beta$ -strands organised in two nearly orthogonal  $\beta$ -sheets with two alpha helices closing the protein cavity where small hydrophobic ligands can be bound. The binding specificity of the protein is not known, however, based on the high sequence and structural similarity with an orthologous protein, ileal lipid binding protein, it is suggested that bile acids may be the putative ligands.

### Biological context

Fatty acid binding proteins (FABP) belong to a large family and bind, with variable stoichiometry, a diverse set of hydrophobic ligands, such as fatty acids, retinoids and bile salts, into a large cavity located in the interior of the protein. The specific function of these proteins still has to be determined, but it is clear that they are involved in fatty acid intracellular metabolism. Fatty acids are converted, within the cell, to many biological active compounds including oxygenated and hydroxylated metabolites, which play an important role in cell proliferation (Bratt, 2000).

More than sixty different FABP amino-acid sequences have been determined and generally classified on the basis of the organ from which they were initially isolated. The highest sequence similarity (up to 70–80%) is observed for FABPs from the same tissue from different species, sometimes quite distant in evolution, thus suggesting that tissue specificity could be correlated to protein function. However several examples are known in which more than one FABP type has been shown to be produced by a single tissue, with the proteins exhibiting variability both in the type, binding mode and stoichiometry of its preferred ligands. Along this line, the structural characterisation and binding

properties of different FABPs purified from the same tissue is of relevance for the understanding of the molecular properties and function of this family.

Liver FABPs (L-FABP) identify a widely studied FABP sub-family (Börchers and Spener, 1993). Besides mammalian FABPs (human, mouse, cow, pig) sharing 79–90% sequence identity, a new basic protein ( $pI = 9$ ), showing a very low sequence similarity with the mammalian species, was purified from chicken liver (Scapin et al., 1988) and hence called liver basic FABP (Lb-FABP). Later, other basic-type FABPs, exhibiting high sequence similarity, were isolated from the liver of other species, such as axolotl, catfish, lungfish, shark, frog, iguana (Di Pietro et al., 1999), though their structures are still unknown.

Here we report the first NMR structure of a basic-type FABP, the chicken Lb-FABP, and suggest, on the basis of comparison with the structures of other proteins of the family, its binding specificity. The study of the subtle structural changes brought about by evolution within the same protein family is relevant for the understanding of the sequence-function specificity relationships.

### Methods and results

Chicken Lb-FABP (125 a.a.) was purified from natural sources (Scapin et al., 1988). The structure determina-

To whom correspondence should be addressed. E-mail: molinari@sci.univr.it

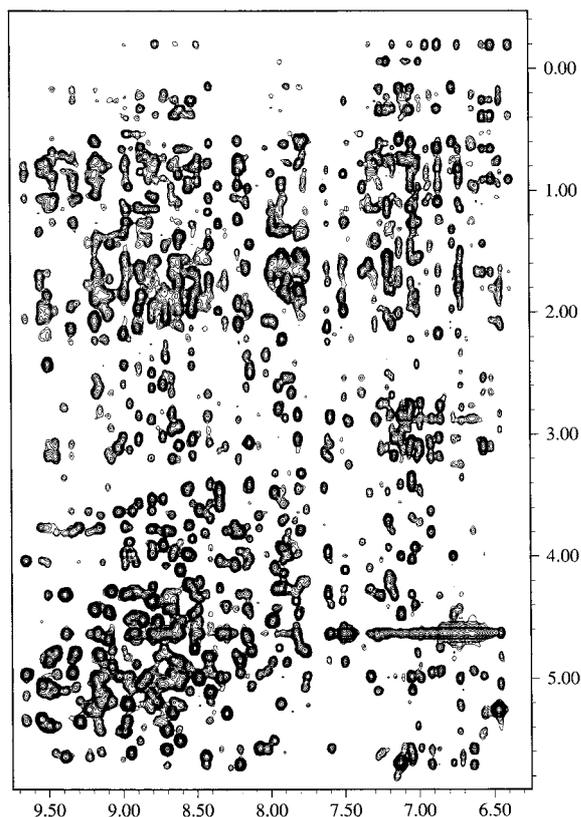


Figure 1. NH-aliphatic region of the 500 MHz  $^1\text{H}$  NOESY spectrum of chicken Lb-FABP (1 mM in 20 mM phosphate buffer, pH = 5.6, 37 °C).

tion was performed at pH 5.6 and 37 °C. Homonuclear  $^1\text{H}$  NMR experiments were recorded at 500 MHz on a Bruker DMX spectrometer. Spectra were analysed with the program XEASY (Bartels et al., 1995), while the programs DYANA (Güntert et al., 1997) and DISCOVER (Molecular Simulations, San Diego, CA, U.S.A.) were used for structure calculation and refinement, respectively.

The quality of the data can be estimated by the NH-aliphatic protons correlation region of the NOESY spectrum reported in Figure 1. The complete sequence-specific  $^1\text{H}$  resonance assignment was obtained from standard  $^1\text{H}$  2D experiments and deposited in the BioMagResBank (accession code BMRB-5512). A partial assignment of this protein was previously reported (Schievano et al., 1999).

The large number of strong  $d_{\alpha\text{N}}$  and  $d_{\alpha\alpha}$  NOE connectivities indicated a high content of antiparallel beta-sheet in the protein. A survey of sequential short and medium range NOEs is reported in Figure 2. The reported secondary structures were also clearly con-

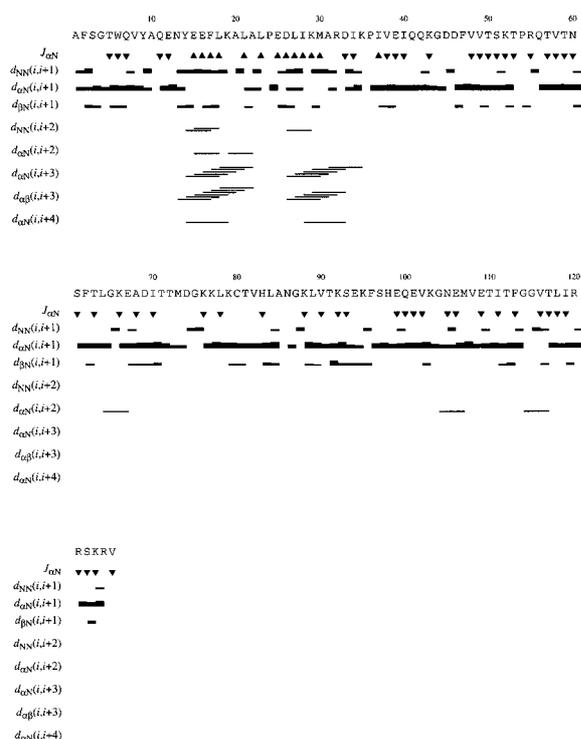
sistent with the  $\text{H}^\alpha$  and  $\text{H}^\text{N}$  shifts (data not shown). The ten antiparallel  $\beta$ -strands,  $\beta_\text{A}$  (3–12),  $\beta_\text{B}$  (37–43),  $\beta_\text{C}$  (46–53),  $\beta_\text{D}$  (56–63),  $\beta_\text{E}$  (67–72),  $\beta_\text{F}$  (76–85),  $\beta_\text{G}$  (88–93),  $\beta_\text{H}$  (96–103),  $\beta_\text{I}$  (106–113) and  $\beta_\text{J}$  (116–125) are connected by short type II turns except for  $\beta_\text{A}$  and  $\beta_\text{B}$ , which are connected by two short helical segments,  $\alpha_1$  (14–21) and  $\alpha_2$  (26–32). The two helices are antiparallel to each other, as demonstrated by the presence of NOEs between residues 14, 15, 18, 19 ( $\alpha_1$  helix) and 28, 31, 32 ( $\alpha_2$  helix). A  $\beta$ -bulge is present at positions 9–10 of  $\beta_\text{A}$ . A gap is observed between  $\beta_\text{D}$  and  $\beta_\text{E}$ : Only seven interstrand side-chain/side-chain NOEs have been identified between these two strands, thus limiting the size of the gap. This feature is common to all the FABPs structures determined to date.

Resonance signals of backbone amide protons of Lb-FABP showed partial protection against hydrogen/deuterium (H/D) exchange. After 14 h at 37 °C and pH 5.6 only 29 amides were left in the TOCSY spectrum, namely amides belonging to the following residues: 6–7 and 9 ( $\beta_\text{A}$ ), 38–41 ( $\beta_\text{B}$ ), 46–49 and 52 ( $\beta_\text{C}$ ), 88–90 ( $\beta_\text{G}$ ), 101 ( $\beta_\text{H}$ ), 107–113 ( $\beta_\text{I}$ ), 118–122 and 125 ( $\beta_\text{J}$ ). Amides belonging to  $\beta_\text{D}$ ,  $\beta_\text{E}$ ,  $\beta_\text{F}$  strands and to the two helices exhibited very low protection.

A total of 1705 NOESY cross-peaks obtained with a mixing time of 100 ms were assigned and integrated, yielding 1179 non redundant constraints (345 intraresidue, 319 short, 135 medium and 380 long range contacts). Additional constraints were derived from measurements of 64  $^3J_{\alpha\text{N}}$  coupling constants.

Hydrogen bond formation or solvent exclusion from the amide protons was assumed to account for the slow and medium amide exchange rates. The partners for the detected hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints. The detected 25 hydrogen bonds were enforced by 50 distance restraints.

The statistics of the twenty conformers which resulted from a final DYANA calculation with 100 randomised starting conformations are given in Table 1. Each structure was minimised and the ten structures with the minimum potential energy and lowest number of violations were selected for further analysis (see Table 1). All the structures were analysed with PROCHECK-NMR (Laskowski et al., 1993). For the 10 best Lb-FABP conformers, a root-mean-square deviation (RMSD) value of  $1.24 \pm 0.14$  Å was determined (Figures 3A and 3B). The overall fold of Lb-FABP, common to all known FABPs, consists of ten antiparallel  $\beta$ -strands ( $\beta_\text{A}$ – $\beta_\text{J}$ ) organised in two



**Figure 2.** Amino acid sequence of chicken Lb-FABP and a survey of the observed short- and medium-range NOE connectivities. For the sequential NOEs  $d_{\alpha N}$ ,  $d_{NN}$  and  $d_{\beta N}$ , thick and thin bars indicate strong and weak intensities. Medium range NOEs are indicated by lines connecting the two related residues. Up and down triangle identifies residues with  $J_{\alpha N} < 5.5$  Hz and  $J_{\alpha N} > 7.3$  Hz, respectively.

nearly orthogonal  $\beta$ -sheets that form a  $\beta$ -clam type structure with a gap between  $\beta_D$  and  $\beta_E$  strands. The two helices  $\alpha_1$  and  $\alpha_2$ , inserted between  $\beta_A$  and  $\beta_B$ , close the protein cavity where small hydrophobic ligands, such as long chain saturated and unsaturated fatty acids, can be bound. The atom coordinates have been deposited in the Protein data bank (PDB ID code 1MVG). It is clear from the ensemble of NMR structures that certain regions display a greater backbone RMSD deviation, either because of the lack of structurally relevant NOEs or of a greater flexibility of these regions or both. The less defined regions comprise residues 21–24 ( $\alpha_1$ – $\alpha_2$  loop), 54–56 ( $\beta_C$ – $\beta_D$  loop), 71–77 ( $\beta_E$ – $\beta_F$  segment) and 91–95 ( $\beta_G$ – $\beta_H$  segment). All these residues are located at the level of the so-called ‘portal area’ at the entrance of the internal cavity. H/D exchange data are consistent with a higher flexibility in the mentioned regions.

Two distinct hydrophobic clusters were identified, on the basis of NOE data. Cluster I, at the bottom

**Table 1.** Structure statistics of chicken Lb-FABP NMR final structures<sup>a</sup>

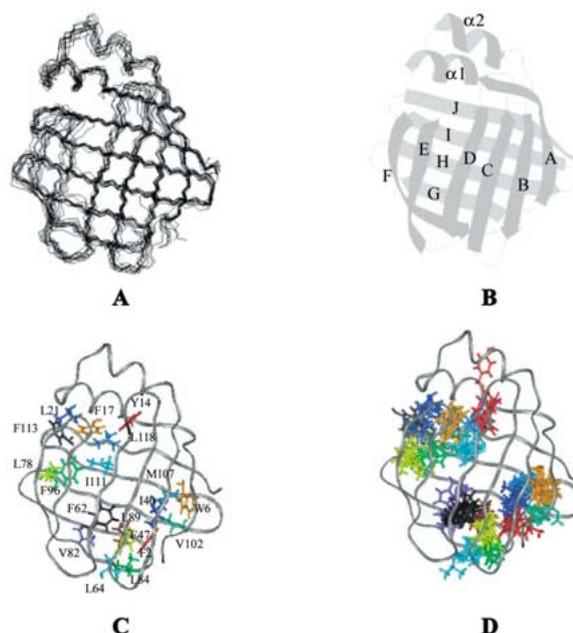
DYANA target function ( $\text{\AA}^2$ ) <sup>b</sup>	$1.10 \pm 0.31$ (0.57...1.86)
NOE violations $> 0.2$ $\text{\AA}$	$1.3 \pm 1.3$ (0...4)
AMBER Energy (Kcal mol <sup>-1</sup> )	$-533 \pm 14$ (-567...-508)
RMSD, N, C $^\alpha$ , C' ( $\text{\AA}$ ) <sup>c</sup>	$1.24 \pm 0.14$ (0.95...1.49)
RMSD, N, C $^\alpha$ , C' ( $\text{\AA}$ ) <sup>d</sup>	$1.00 \pm 0.12$ (0.66...1.29)
RMSD all heavy atoms ( $\text{\AA}$ ) <sup>c</sup>	$2.06 \pm 0.16$ (1.60...2.36)

<sup>a</sup>Average values  $\pm$  standard deviations for the final structures, minimum and maximum values for individual conformers in parentheses.

<sup>b</sup>Analysis of the best 20 DYANA structures before energy minimization.

<sup>c</sup>RMSD values for residues 3–125.

<sup>d</sup>RMSD values for residues 3–20, 25–53, 58–70, 78–90, 96–125.



**Figure 3.** Superposition of the final ten best chicken Lb-FABP structures (A); a ribbon drawing of the structure with the lowest energy (B); in colour are the hydrophobic side-chains of residues belonging to cluster I and II (C); superposition of apo-FABPs residues corresponding to residues belonging to chicken Lb-FABP cluster I (D). The superimposed structures are: 1eal, 3ifb, 1ael, 1bwy, 1g5w.

of the hydrophobic cavity, included residues F2 (N terminus), W6 ( $\beta_A$ ), I40 ( $\beta_B$ ), F47 ( $\beta_C$ ), F62 ( $\beta_D$ ), L64 (CD loop), V82 and L84 ( $\beta_F$ ), L89 ( $\beta_G$ ), V102 ( $\beta_H$ ) and M107 ( $\beta_I$ ). Cluster II, at the opening of the calyx between the two helices and the barrel, included residues Y14 and F17 ( $\alpha_1$  helix), L21 ( $\alpha_1$ – $\alpha_2$  loop), L78 ( $\beta_F$ ), F96 ( $\beta_H$ ), I111 and F113 ( $\beta_I$ ) and L118 ( $\beta_J$ ) (see Figure 3C).

## Discussion and conclusions

A multiple sequence alignment obtained with ClustalW (<http://www.ebi.ac.uk/clustalw/>) using all the known sequences of basic-type and non basic FABPs indicated that the non basic protein with the highest similarity is pig ileal lipid binding protein (ILBP) (43% identity, 58% similarity). The structural comparison of Lb-FABP with all apo-FABPs of known structure (ileal binding protein, human and rat intestinal, bovine and human heart FABP) (Zhang et al., 1997; Hodson and Cistola, 1997; Lassen et al., 1995; Lücke et al., 1996, 2001), performed through the VAST search service (<http://www.ncbi.nlm.nih.gov/Structure/VAST>), indicated that ILBP has the highest structural similarity, exhibiting an RMSD for alignable regions of 1.61 Å. The RMSD with other apo-FABPs was higher and comprised between 2.02 and 3.16 Å.

The detected flexible segments of Lb-FABP are located at the portal area and, interestingly, correspond closely to the regions of ILBP involved in bile salts binding (Lücke et al., 1996). This observation suggests that Lb-FABP flexible regions have a functional role related to binding and that bile acids might be Lb-FABP putative ligands. In this line it is worth to stress that the residues important for bile acid binding in ILBP (W49, K77, F79, Y97, H99 and E110) (Lücke et al., 2000) are mostly conserved in Lb-FABP (V49, K76, L78, F96, H98 and E109) but not in chicken L-FABP (I49, E76, A78, L97, S99 and H110). In addition Lb-FABP, together with ILBP, displays a relatively low fatty acid binding affinity ( $K_d$  in the micromolar range) differently from other liver and liver basic proteins displaying much higher affinities for fatty acids ( $K_d$  in the nanomolar range) (Beringhelli et al., 2001; Lücke et al., 1996). These observations further support that fatty acids are not the primary ligands for this protein and that chicken liver and chicken liver basic proteins have different functional roles.

Cluster I, located at the bottom of the internal cavity, appears well conserved in the FABP family, independently from sequence similarity (Figure 3D). When hydrophobic residues of cluster I were superimposed to residues forming the internal hydrophobic cluster in bovine  $\beta$ -Lactoglobulin (BLG) (Ragona et al., 1997), a lipocalin belonging, together with FABPs, to the calycin structural superfamily, it was observed that seven out of eleven residues exhibited a similar orientation in the two proteins, thus indicating that this well conserved cluster is relevant to the calycins stability.

Cluster II involving residues belonging to C-terminal strands and to  $\alpha 1$  and  $\alpha 2$  helices, located at the entrance of the internal cavity, is much less conserved among different FABPs and may play some role in the binding specificity observed for different FABPs. Indeed a different mode of binding and a different stoichiometry was observed for FABPs from different tissues, with Lb-FABP exhibiting a different stoichiometry and a different binding mode with respect to L-FABP (Beringhelli et al., 2001). Further NMR investigations on the recombinant apo and holo Lb-FABP are now in progress to clarify mode, dynamics and specificity of binding.

## Acknowledgements

This work was supported by MIUR (ex 40% 2000), INFM and by the Italian National Council of Research 'Progetto Finalizzato Biotecnologie'. Anna Tramontano is acknowledged for helpful discussions.

## References

- Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **5**, 1–10.
- Beringhelli, T., Goldoni, L., Capaldi, S., Bossi, A., Perduca, M. and Monaco, H.L. (2001) *Biochemistry*, **40**, 12604–12611.
- Börchers, T. and Spener, F. (1993) *Mol. Cell. Biochem.*, **123**, 23–27.
- Bratt, T. (2000) *Biochim. Biophys. Acta*, **1482**, 318–326.
- Di Pietro, S.M., Veerkamp, J.H. and Santomè, J.A. (1999) *Eur. J. Biochem.*, **259**, 127–134.
- Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 263–296.
- Hodson, M.E. and Cistola, D.P. (1997) *Biochemistry*, **36**, 2278–2290.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.
- Lassen, D., Lücke, C., Kveder, M., Mesgarzadeh, A. Schmidt, J.M., Specht, B., Lezius, A., Spener, F. and Rüterjans, H. (1995) *Eur. J. Biochem.*, **230**, 266–280.
- Lücke, C., Rademacher, M., Zimmerman, A., van Moerkerk, H.T.B., Veerkamp, J.H. and Rüterjans, H. (2001) *Biochem. J.*, **354**, 259–266.
- Lücke, C., Zhang, F., Hamilton, J.A., Sacchettini, J.C. and Rüterjans, H. (2000) *Eur. J. Biochem.*, **267**, 2929–2938.
- Lücke, C., Zhang, F., Rüterjans, H., Hamilton, J.A. and Sacchettini, J.C. (1996) *Structure*, **4**, 785–800.
- Ragona, L., Pusterla, F., Zetta, L., Monaco H.L. and Molinari, H. (1997) *Fold. Des.*, **2**, 281–190.
- Scapin, G., Spadon, P., Pengo, L., Mammi, M., Zanotti, G. and Monaco, H.L. (1988) *FEBS Lett.*, **240**, 196–200.
- Schievano, E., Mammi, S. and Peggion, E. (1999) *Biopolymers*, **5**, 1–11.
- Zhang, F., Lücke, C., Baier, L.J., Sacchettini J.C. and Hamilton J.A. (1997) *J. Biomol. NMR*, **9**, 213–228.