

## LACTOFERRIN: ANALYSIS OF THE STRUCTURE PROFILE

Lilia Anghel

*Institute of Chemistry of the Academy of Sciences of Moldova, 3, Academiei str., Chisinau MD-2028, Republic of Moldova  
e-mail: anghel.lilia@gmail.com*

**Abstract.** The multitude of physiological processes in which the binding of iron ions takes part makes its mechanism worth investigating. The multiple sequence alignment method was applied to investigate the structure similarities of five lactoferrin X-ray crystallographic structures and outline the differences and similarities between lactoferrin and serum transferrin. The results of this study provide useful insights into the mechanism of iron-binding of lactoferrin protein molecule.

**Keywords:** iron, protein, lactoferrin, transferrin, multiple sequence alignment.

### Introduction

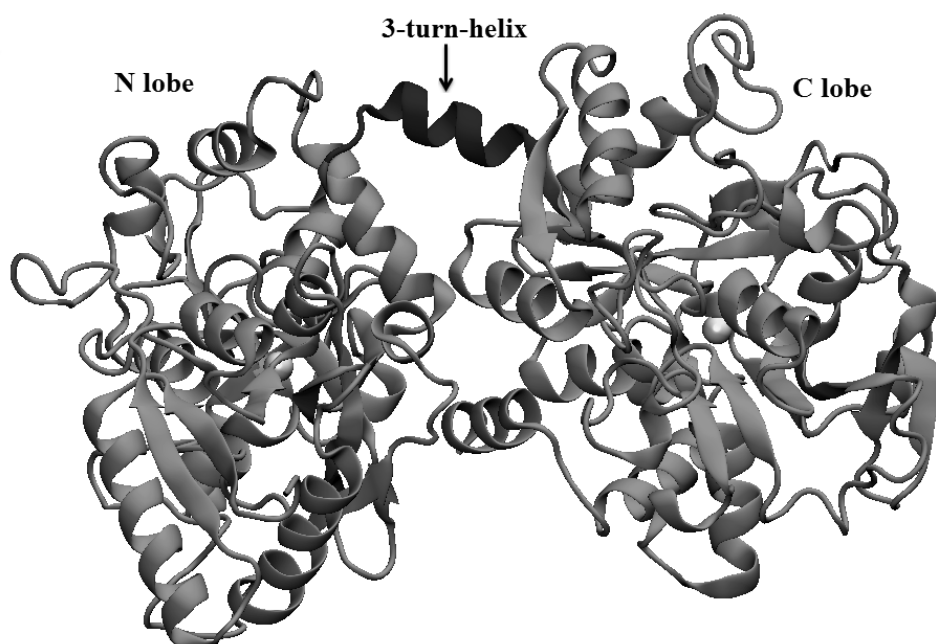
Intracellular iron plays an essential role in different physiological reactions such as metabolic oxidation/reduction reactions or building highly selective biological catalysts – enzymes [1]. Unbounded iron ions could initiate a multitude of deleterious processes in biological systems [2], therefore, cells have developed various systems for iron acquisition and intracellular transport. A common toolkit for iron transport is represented by the family of transferrin. The primary role of this protein family is to control the level of free iron ions in the body, protecting it from possible damages which free iron ions could initiate.

Lactoferrin is a protein that belongs to the family of transferrin. This protein also referred as lactotransferrin, originally was isolated from milk and later it was found in biological fluids such as blood plasma, tears, saliva, pancreatic juice, bile. In blood plasma lactoferrin derives from specific granules of neutrophils but there are evidences that it might be produced by other cells and even microorganisms [3].

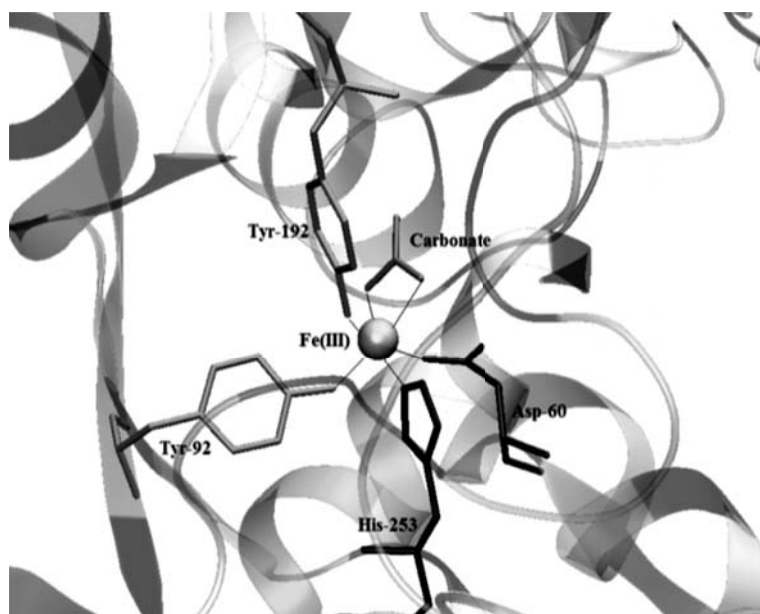
### Structure of lactoferrin

Lactoferrin is a non-haem iron-binding glycosylated protein with a molecular mass of about 80 kDa. Its polypeptide chain consists of about 600-700 amino acid residues. The number of amino acids in the protein structure varies depending on the origin of the molecule. The structure of a lactoferrin molecule is composed of  $\alpha$ -helix and  $\beta$ -sheets which are presented in Figure 1 as ribbons and arrows, respectively.

The protein structure is divided in two lobes, referred as N- and C- lobes which are connected by a 3-turn-helix structure (see Figure1). Both protein lobes share a degree of similarity of about 40% [4]. Each lobe contains an iron-binding site consisting of an aspartic acid, a histidine and two tyrosine residues (Figure 2). Iron coordination in each protein cleft is finished through a carbonate / bicarbonate ion.



**Figure 1.** Structure representation of diferric human lactoferrin (1B0L Protein Data Bank code). The protein is presented in ribbon and ferric ions are presented as van der Waals spheres.



**Figure 2. Graphical representation of the active iron-binding site of the N-lobe of human lactoferrin (1B0L Protein Data Bank code).**

Table 1 presents the bond lengths of ferric ion to coordinated ligands measured on the basis of the X-ray crystallographic structure file of human lactoferrin (1B0L entry).

Even though subsequent studies have shown that lactoferrin molecule is capable of binding other metal ions its affinity for iron ions is greater in comparison to serum transferrin [5-6]. The primary biological role of lactoferrin is to bind iron in physiological fluids, but there is no evidence of transport function that is characteristic to the other members of the transferrin family. This indicates on the existence of a different mechanism of iron-binding. The process of iron binding and release from lactoferrin molecule depends on its molecular properties and on the location where it is expressed. To help gain more insights into the process of iron uptake, it is important to investigate the structure similarities of lactoferrin molecules isolated from different species (see Table 2) and outline the structure differences and similarities between lactoferrin [7] and serum transferrin [12].

*Table 1*

**The bond lengths (Å) of Fe(III) to coordinated ligands in human lactoferrin.**

<i>Bond</i>	<i>N lobe</i>	<i>C lobe</i>
Fe-O:Asp60 (Asp395)	2.146	2.008
Fe-O:Tyr92 (Tyr435)	2.035	2.004
Fe-O:Tyr192 (Tyr 528)	1.817	1.848
Fe-N:His253 (His597)	2.087	2.194
Fe-O1:CO <sub>3</sub> <sup>2-</sup> 695(CO <sub>3</sub> <sup>2-</sup> 696)	2.007	2.286
Fe-O2:CO <sub>3</sub> <sup>2-</sup> 695(CO <sub>3</sub> <sup>2-</sup> 696)	2.138	2.010

### Materials and methods

Multiple sequence alignment (MSA) method is used in sequence and structural assessment of sets of macromolecular biological molecules which have a common evolutionary relationship. The sequence assessment is performed using sequence alignment analysis, which makes use of the sequence information from the protein structure to perform the alignment of the protein molecules and for the evaluation of their degree of similarity. When making a sequence alignment it should also be considered the structural information. The effect of the substitution of an amino acid sometimes within the sequence can cause substantial effects of the function of the protein. Therefore, for a correct alignment, structural information should also be used. The structural assessment is performed using the structure alignment analysis. The output of the implementation of this method offers the possibility to build an accurate profile for structurally related proteins.

The multiple sequence alignment method was applied to investigate the structure similarities and differences of five lactoferrin X-ray crystallographic structures and outline the differences between lactoferrin and serum transferrin. The coordinate files of the investigated proteins were downloaded from Protein Data Bank (PDB) [13]. The PDB codes

are listed in Table 2. Visual Molecular Dynamics (VMD) software [14] was used for all graphical representations. The analysis was performed using MultiSeq 2.0, a unified analysis environment, included in the VMD. MultiSeq 2.0 was developed for the analysis of sequence and structure of proteins and nucleic acids [15]. For the structural data, proteins were aligned using STAMP [16] plug-in and for sequence data, proteins were aligned using ClustalW [17] plug-in. As a result of the implementation of the algorithms include in STAMP and ClustalW plug-ins the following set of parameters was computed for each protein structure:  $Q_{res}$  (structure similarity per residue),  $Q_H$  (structural homology), Percent Identity and RMSD (root mean square deviation). The computed values for these parameters will differ as a result of a different alignment algorithm applied to the same set of proteins. Therefore these parameters can be used to assess the quality of the alignment and common similarities of the analyzed proteins.

$Q_{res}$  is the parameter used to measure the structural similarity of each residue in a set of aligned structures by measuring the backbone carbon distances between a residue and all other residues in the protein, excluding nearest neighbors, to the corresponding distances in a given set of proteins.

In the MultiSeq 2.0 this parameter is defined by the following formula:

$$Q_{res}^{(i,n)} = \aleph \sum_{\substack{proteins \\ (m \neq n)}} \sum_{\substack{residue \\ (j \neq i-1, i, i+1)}} \exp \left[ -\frac{\left( r_{ij}^{(n)} - r_{i'j'}^{(m)} \right)^2}{2\sigma_{ij}^2} \right] \quad (1)$$

where:  $Q_{res}^{(i,n)}$  is the structural similarity of the  $n^{th}$  amino acid residue in the  $n^{th}$  protein;

$r_{ij}^{(n)}$  is the  $C_\alpha - C_\alpha$  distance between residues  $i$  and  $j$  in protein  $n$ ;

$r_{i'j'}^{(m)}$  is the  $C_\alpha - C_\alpha$  distance between residues  $i'$  and  $j'$  in protein  $m$  that correspond to residues  $i$  and  $j$  in protein  $n$ ;

$\sigma_{ij}^2 = |i - j|^{0.15}$  is the variance related to the sequence separation between residues  $i$  and  $j$ ;

$\aleph = \left( (N_{seq} - 1)(N_{res} - k) \right)^{-1}$  is the normalization, with  $N_{seq}$  - the number of proteins in the set and  $N_{res}$  - the number of amino acid residues in protein  $n$ , and  $k = 3$  except when the residue  $i$  is the N or C-terminus of the protein, in this case  $k = 2$  [15].

$Q_H$  measures the homology of two structures by summing the similarity of all residues ( $Q_{res}$ ) and adds a term for each gap (structural deviation) in the alignment, its mathematical expression is presented elsewhere [18]. Percent Identity parameter is derived from the  $Q_H$ .  $Q$  values range from 0 to 1, for  $Q=1$  the proteins are identical. RMSD measures the distances in angstroms between the atoms  $C_\alpha$  of two aligned residues and indicates how well two structures were aligned [15].

Table 2

The PDB code of the X-ray structure files of the proteins used for this study.

PDB code	Description	Reference
1B0L	Diferric human lactoferrin	[7]
1BIY	Diferric buffalo lactoferrin	[8]
1BLF	Diferric bovine lactoferrin	[9]
1I6B	Diferric equine lactoferrin	[10]
1I6Q	Diferric camel lactoferrin	[11]
3QYT	Diferric human transferrin	[12]

## Results and discussion

### Conservation of lactoferrin across the species

The goal of this study was to detect the common similarities of several lactoferrin protein molecules from the alignments of their structures and sequences. A simple structure inspection in the VMD showed that in all five lactoferrin structures iron is bounded to the same four amino acid residues (Figure 2) and one carbonate ion. Each lactoferrin molecule consists of 689 amino acids, except human lactoferrin, which is composed of 691 amino acids.

Representation of the multiple sequence alignment of bovine, buffalo, equine and camel lactoferrin colored by sequence identity is depicted in Figure 3.

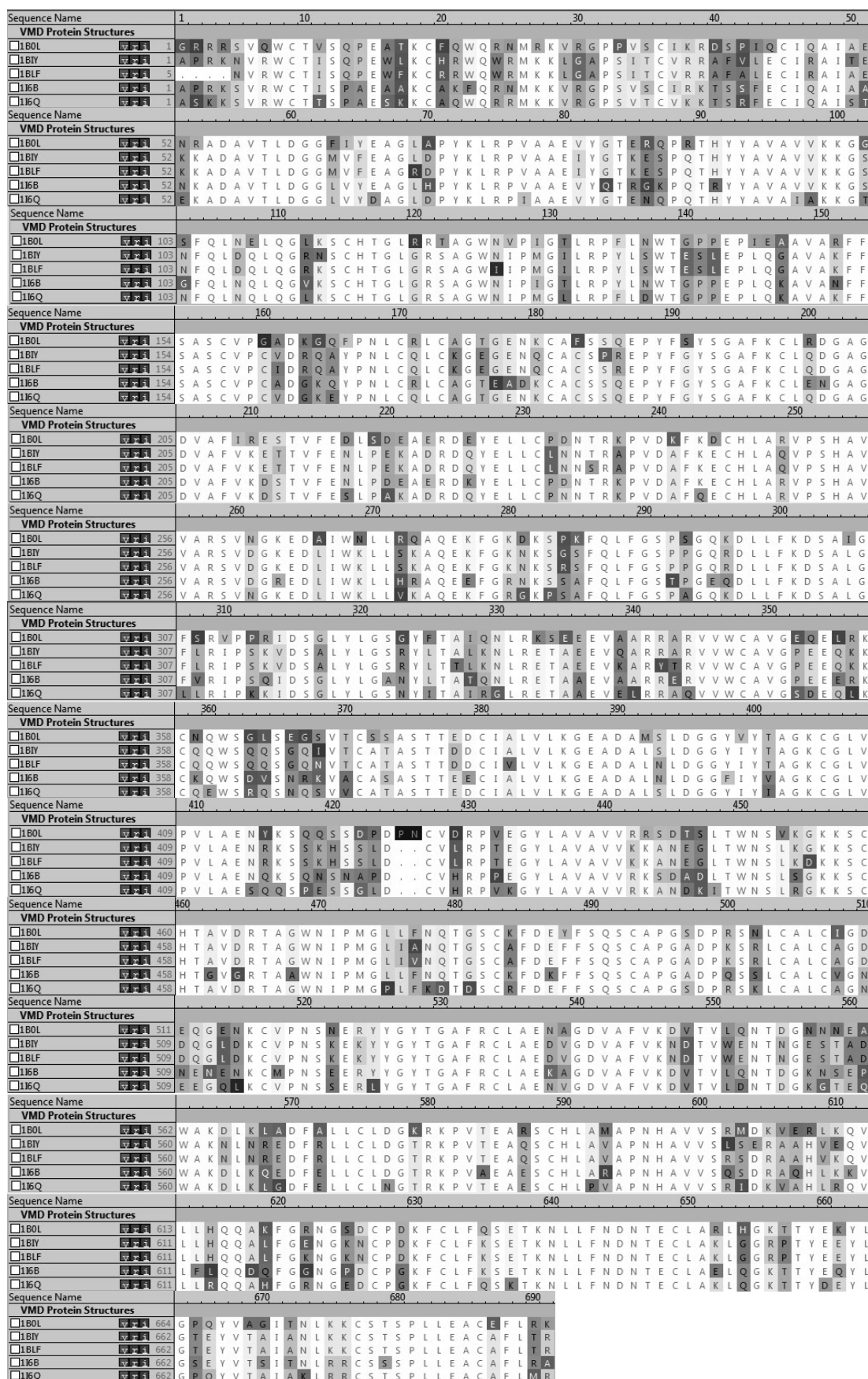


Figure 3. Representation of the multiple sequence alignment of the five studied lactoferrins colored by sequence identity. Variable residues are highlighted in grey.

The structure and sequence and identity parameters for the five studied lactoferrins structures were computed with respect to the structure of human lactoferrin and results of this study are presented in Table 3 and Table 4.

Parameters computed from the structural alignment show that overall the structure of lactoferrin is highly conserved across all species investigated in this study. Only in the case of camel lactoferrin the structural homology parameter is lower in comparison to the rest of proteins,  $Q_H$  is 0.3869 and the Percent Identity is 30.60%.

Table 3

**Structural alignment parameters computed for the studied lactoferrin structures.**

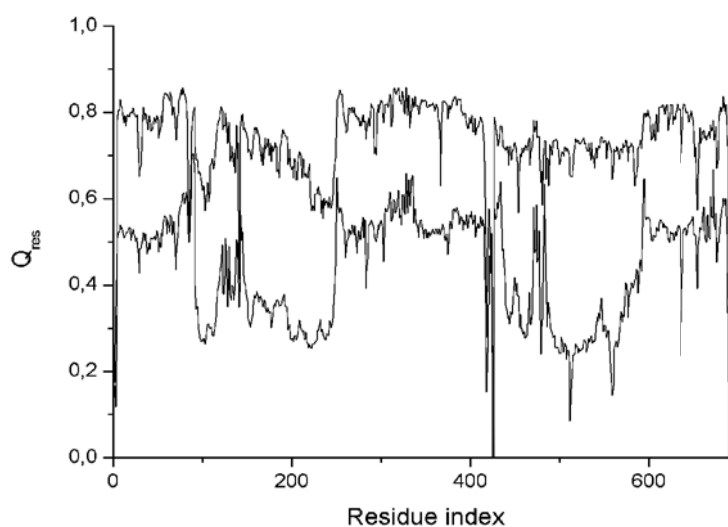
	<i>Buffalo lactoferrin</i>	<i>Bovine lactoferrin</i>	<i>Equine lactoferrin</i>	<i>Camel lactoferrin</i>
$Q_H$	0.8393	0.8160	0.8688	0.3869
RMSD (Å)	1.5354	1.8106	1.2773	1.6093
Percent Identity (%)	69.35	68.88	74.31	30.60

Table 4

**Sequence alignment parameters computed for the studied lactoferrin structures.**

	<i>Buffalo lactoferrin</i>	<i>Bovine lactoferrin</i>	<i>Equine lactoferrin</i>	<i>Camel lactoferrin</i>
$Q_H$	0.8356	0.8154	0.8634	0.4503
RMSD (Å)	1.6093	1.8616	1.1414	12.6343
Percent Identity (%)	70.19	69.61	74.96	73.81

The sequence alignment analysis showing that camel lactoferrin differs from the other structures not only chemically but also the arrangement of the residues in the protein structure differs a lot more (RMSD=12.6343). Figure 4 presents the plot of structure similarity ( $Q_{res}$ ) per residue of the human lactoferrin and camel lactoferrin, computed from the sequence alignment. Even though the  $Q_{res}$  values of both lactoferrins differ, the general pattern presented in Figure 4 is similar in both cases indicating a common structure profile.



**Figure 4. Plot of the structural similarity computed from the sequence alignment of each residue of the human lactoferrin (black) and camel lactoferrin (light grey).**

#### ***Structural comparison of human serum transferrin and human lactoferrin molecules***

Human serum transferrin and human lactoferrin are members of the transferrin family. These proteins are known for their ability of binding iron, although serum transferrin is also responsible for the iron transportation. Both proteins share a similar architecture, with a structure consisting of two lobes (N and C) attached by a linker region. In both proteins an iron ion is bounded to an aspartic acid, two tyrosine, a histidine residues and a carbonate or bicarbonate ion. Although these two proteins share many similarities, there are some structural differences that may influence the mechanism of iron-binding. The most prominent structural difference between these two proteins is the linker region between the two lobes. The lactoferrin linker is an uncompleted  $\alpha$ -helix structure whilst the transferrin linker is represented by an unstructured residue conformation (see Figure 5). Experimental studies [19] have shown that iron-binding process is pH-dependent, meaning that iron will bind to the protein only when the key-residues from the protein structure will have a certain protonation state. For example, in case of transferrin iron-binding is triggered by two lysine residues which are localized in the vicinity of the active site of the N lobe of protein. In acidic medium these two

lysine residues will become charged which will lead to repulsion and forcing the opening of the N lobe of transferrin, this mechanism is known as dilysine trigger. In the lactoferrin structure this mechanism is absent due to the fact that the first lysine is replaced by an arginine residue. The multiple sequence alignment was applied to investigate the degree of similarity between the structures of human lactoferrin and human serum transferrin. Results of this study are presented in Figure 6 and Table 5.



Figure 5. Graphical representation of the linker region in the structure of human lactoferrin (a) and human serum transferrin (b).

Sequence Name	1	10	20	30	40	50
<b>VMD Protein Structures</b>						
1B0L	1	G R R R S V Q W C T I V S Q P E A T K C F Q W Q R N M R K V R . . . G P P V S C I K R D S P I Q C I Q A				
3QYT	1	V P D K T V R W C A V S E H E A T K C Q S F R D H M K S V I P S D G P S V A C V K K A S Y L D C I R A				
Sequence Name	60	70	80	90	100	
<b>VMD Protein Structures</b>						
1B0L	49	I A E N R A D A V T L D G G F I Y E A G L A P Y K L R P V A A E V Y G T E R Q P R T H Y Y A V A V V K				
3QYT	52	I A A N E A D A V T L D A G L V Y D A Y L A P A N N L K P V V A E F Y G S K E D P Q T F Y Y A V A V V K				
Sequence Name	110	120	130	140	150	
<b>VMD Protein Structures</b>						
1B0L	100	K G G S F Q L N E L Q G L K S C H T G L R R T A G W N V P I G T L R P F L N W T G P P E P I E A A V A				
3QYT	103	K D S G F Q M N Q L R G K K S C H T G L G R S A G W N I P I G L L . . . Y C D L P E P R K P L E K A V A				
Sequence Name	160	170	180	190	200	
<b>VMD Protein Structures</b>						
1B0L	151	R F F S A S C V P G A D K G Q F P N L C R L C A G T G E N K C A F S S Q E P Y F S Y S G A F K C L R D				
3QYT	152	N F F S G S C A P C A D G T D F P Q L C Q L C P G . . . . . C G C S T L N Q Y F G Y S G A F K C L K D				
Sequence Name	210	220	230	240	250	
<b>VMD Protein Structures</b>						
1B0L	202	G A G D V A F I R E S T V F E D L S D E A E R D E Y E L L C P D N T R K P V D K F K D C H L A R V P S				
3QYT	198	G A G D V A F V K H S T I F E N L A N K A D R D Q Y E L L C L D N T R K P V D E Y K D C H L A Q V P S				
Sequence Name	260	270	280	290	300	
<b>VMD Protein Structures</b>						
1B0L	253	H A V V A R S V N G K E D A I W N L L R Q A Q E K F G K D K S P K F Q L F G S P S G Q K D L L F K D S				
3QYT	249	H T V V A R S M G G K E D L I W E L L N O A Q E H F G K D K S K E F O L F S S P H G . K D L L F K D S				
Sequence Name	310	320	330	340	350	
<b>VMD Protein Structures</b>						
1B0L	304	A T G F S R V P P R I D S G L Y L G S G Y F T A I Q N L R K . . S E E E V A A R R A R V V W C A V G E				
3QYT	299	A H G F L K V P P R M D A K M Y L G Y E Y V T A I R N L R E G T C P E A P T D E C K P V K W C A L S H				
Sequence Name	360	370	380	390	400	
<b>VMD Protein Structures</b>						
1B0L	353	Q E L R L K C N Q W S G L S E G S V T C S S A S T T E D C I A L V L K G E A D A M S L D G G Y V Y T A G				
3QYT	350	H E R L L K C D E W S V N S V G K I E C V S A E T T E D C I A K I M N G E A D A M S L D G G F V Y I A G				
Sequence Name	410	420	430	440	450	
<b>VMD Protein Structures</b>						
1B0L	404	K C G L V P V L A E N Y K S Q Q S S D P D P N C V D R P V E G Y L A V A V V R R S D T S L T W N S V K				
3QYT	401	K C G L V P V L A E N Y N K . . . S D . . . N C E D T P E A G Y F A V A V V K K S A S D L T W D N L K				
Sequence Name	460	470	480	490	500	510
<b>VMD Protein Structures</b>						
1B0L	455	G K K S C H T A V D R T A G W N I P M G L L F N Q T G S C K F D E Y F S Q S C A P G S D P R S N L C A				
3QYT	446	G K K S C H T A V G R T A G W N I P M G L L Y N K I N H C R F D E F F S E G C A P G S K K D S S L C K				
Sequence Name	520	530	540	550	560	
<b>VMD Protein Structures</b>						
1B0L	506	L C I G D E Q G E N K C V P N S N E R Y Y G Y T G A F R C L A E N A G D V A F V K D V T V L Q N T D G				
3QYT	497	L C M G . . S G L N L C E P N N K E G Y Y G Y T G A F R C L V E K G D V A F V K H Q T V P Q N T G G				
Sequence Name	570	580	590	600	610	
<b>VMD Protein Structures</b>						
1B0L	557	N N E A W A K D L K L A D F A L L C L D G K R K P V T E A R S C H L A M A P N H A V V S R M D K V E				
3QYT	545	K N P D P W A K N L N E K D Y E L L C L D G T R K P V E E Y A N C H L A R A P N H A V V T R K D K E A				
Sequence Name	620	630	640	650	660	
<b>VMD Protein Structures</b>						
1B0L	608	R L K Q V L L H Q Q A K F G R N G S D C P D K F C L F Q S E T K N L L F N D N T E C L A R L H G K T T				
3QYT	596	C V H K I L R Q Q H L F G S N V T D C S G N F C L F R S E T K D L L F R D D T V C L A K L H D R N T				
Sequence Name	670	680	690			
<b>VMD Protein Structures</b>						
1B0L	659	Y E K Y L G P Q Y V A G I T N L K K C S T S P L L E A C E F L R K				
3QYT	647	Y E K Y L G E E Y V K A V G N L R K C S T S S L L E A C T F R R P				

Figure 6. Representation of the multiple sequence alignment of the human lactoferrin (1B0L pdb) and human serum transferrin (3QYT pdb). Variable residues are highlighted.

Structure alignment parameters indicate a visible structural similarity between proteins, and that 41.90% of residues from the structure are identical in both proteins. The sequence alignment parameters present a different picture. The RMSD value of 6.8932 Å, computed from the sequence alignment indicates that the structures of both proteins aren't well aligned and therefore the sequences of these two proteins differ, although the structural homology,  $Q_H$ , is approximately 0.5 and a Percent Identity of 60.49%.

Table 5

**Structure and sequence alignment parameters computed for the human lactoferrin and human serum transferrin structures.**

	$Q_H$	RMSD (Å)	Percent Identity (%)
Structure alignment analysis	0.5251	3.2254	41.90
Sequence alignment analysis	0.5501	6.8932	60.49

## Conclusions

The multiple sequence alignment method was applied to investigate the structure similarities of five lactoferrin X-ray crystallographic structures and outline the differences between lactoferrin and serum transferrin.

Given the obtained results, it can be concluded that overall the structure of lactoferrin is highly conserved across all species investigated in this study. It was identified that only camel lactoferrin structurally differs from other lactoferrins. General pattern of camel lactoferrin is similar to the other four structures, pointing on a common structure profile.

Comparison of the structure human lactoferrin and human serum transferrin has revealed some insights into the architecture of these two proteins. Structure alignment parameters indicate a visible structural similarity between proteins; both proteins share a similar architecture. Structural differences of human lactoferrin and human serum transferrin were assessed using the sequence alignment analysis. Sequence differences are dictated by nucleotide sequence of their genes pointing onto the existence of differences in the three-dimensional structure that determines the proteins activity.

Results of this study increase our understanding of the structural profile of lactoferrin protein and provide useful insights into the iron-binding process.

## Acknowledgements

I would like to acknowledge and express my sincere gratitude to my scientific adviser Acad. Gheorghe Duca for continuous guidance and encouragement throughout every process of my research.

I gratefully acknowledge the Programs of Collaboration JINR-Romania (projects No.81/18.02.2013-37 and No.95/17.02.2014-26 within the JINR Theme No.04-4-1069-2009/2014) and Moldovan State Programs (project No. 14.518.02.05A) for the support.

## References

- Duca, Gh. Homogeneous Catalysis with Metal Complexes: Fundamentals and Applications. Springer Series in Chemical Physics: Berlin Heidelberg, 2012, XII, 478 p.
- Duca, Gh.; Secara, N. The free radicals in biological systems: formation and protection mechanisms. Akademos: Journal of Science, Innovation, Culture and Art, 2010, 4(19), pp. 115-118.
- Levay, P.F.; Viljoen, M. Lactoferrin: a general review. Haematologica, 1995, 80, pp. 252-267.
- Baker, E.N.; Baker, H.M. Molecular structure, binding properties and dynamics of lactoferrin. Cellular and Molecular Life Sciences, 2005, 62, pp. 2531-2539.
- Smith, C.A.; Anderson, B.F.; Baker, H.M.; Baker, E.N. Metal substitution in transferrins: the crystal structure of human copper-lactoferrin at 2.1-Å resolution. Biochemistry, 1992, 31, pp. 4527-4533.
- Loehr, T.M. Ed. Iron Carriers and Iron Proteins, VCH Publishers: New York, 1989, 533 p.
- Sun, X.L.; Baker, H.M.; Shewry, S.C.; Jameson, G.B.; Baker, E.N. Structure of recombinant human lactoferrin expressed in *Aspergillus awamori*. Acta Crystallographica, Section D, 1999, 55, pp. 403-407.
- Karthikeyan, S.; Yadav, S.; Paramasivam, M.; Srinivasan, A.; Singh, T.P. Structure of buffalo lactoferrin at 3.3 Å resolution at 277 K. Acta Crystallographica, Section D, 2000, 56, pp. 684-689.
- Moore, S.A.; Anderson, B.F.; Groom, C.R.; Haridas, M.; Baker, E.N. Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. Journal of Molecular Biology, 1997, 274, pp. 222-236.
- Kumar, P.; Khan, J.A.; Yadav, S.; Singh, T.P. Crystal structure of equine apolactoferrin at 303 K providing further evidence of closed conformations of N and C lobes. Acta Crystallographica, Section D 2002, 58, pp. 225-232.
- Khan, J.A.; Kumar, P.; Srinivasan, A.; Singh, T.P. Protein intermediate trapped by the simultaneous crystallization process. Crystal structure of an iron-saturated intermediate in the  $Fe^{3+}$  binding pathway of camel lactoferrin at 2.7 Å resolution. The Journal of Biological Chemistry, 2001, 276, pp.36817-36823.

12. Yang, N.; Zhang, H.; Wang, M.; Hao, Q.; Sun, H. Iron and bismuth bound human serum transferrin reveals a partially-opened conformation in the N-lobe. *Scientific Reports*, 2012, 2, doi:10.1038/srep00999.
13. Protein Data Bank – an information portal to biological macromolecules <http://www.rcsb.org/pdb/home/home.do>
14. Humphrey, W.; Dalke, A. Schülten, K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 1996, 14, pp. 33-38.
15. Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: unifying sequence and structure data for evolutionary analysis, *BMC Bioinformatics*, 2006, 7(382), pp. 1-11.
16. Russell, R.B.; Barton, G.J. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 1992, 14(2), pp. 309-323.
17. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 1994, 22(22), pp. 4673-4680.
18. O'Donoghue, P.; Luthey-Schulten, Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiology and Molecular Biology Reviews*, 2013, 67(4), pp. 550-573.
19. He, Q.Y.; Mason, A.B.; Tam, B.M. MacGillivray, R.T.A.; Woodworth, R.C. Dual role of Lys206-Lys296 interaction in human transferrin N-lobe: Iron-release trigger and anion-binding site. *Biochemistry*, 1999, 38, pp. 9704–9711.