Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0304.04**

AN IN-SILICO BASED IDENTIFICATION OF DISTANT HOMOLOGS OF COCKROACH MILK PROTEIN

Tushar Mandloi, Ajay Bhaskar, Pranav Thakkar, Jenny Tamboli, Selvaa Kumar C*

School of Biotechnology and Bioinformatics, D.Y. Patil University, CBD Belapur, Navi Mumbai, India.

ABSTRACT: Objective: The cockroach milk protein (CMP) is found in the mid-gut of a Pacific Beetle Viviparous cockroach in liquid state. This becomes crystals due to the ingestion by embryos in the midgut. The main aim of this study is to identify distant homologs of the CMP from sequence and structure perspective. **Methods:** In this study we carried Fold based approaches like pDomTHREADER and pGenTHREADER to identify distant homologs. Multiple sequence alignment (MSA) was performed to identify sequence identity between homologous proteins. PRINTS, PRODOM and PROSITE were used for protein domain analysis. Structural superimposition of CMP with Bacterial Outer Membrane Lipoprotein (Blc) of *E. coli* and Apolipoprotein D (ApoD) of Homo sapiens were carried out using Chimera. Finally, the domain mobility was studied using iMODs online server. **Results:** Based on the fold based method, CMP is closely related to Blc-protein (*E.Coli*). From alignment perspective, CMP and ApoD exhibits higher sequence similarity compared to Blc and Lazarillo (*Schistocerca americana*). Furthermore, protein domain of ApoD was identified to be of lipocalin family as per the PROSITE database search. From structural perspective, both CMP and ApoD superimposes well with each other. Furthermore, CMP shows lesser domain mobility compared to Blc and ApoD protein due to the presence of proline residues. This data reveals significant evolution of proteins during speciation. **Conclusion:** Significant structural similarities were observed between CMP and ApoD. The sequential and structural aspects of cockroach milk protein holds further scope for comparative analysis between lipocalin families based on different adaptation or protein modifications.

KEYWORDS: Cockroach milk protein, Diploptera punctate, Lipocalin, Apolipoprotein D

***Corresponding Author: Dr. Selvaa Kumar C** Ph.D.

School of Biotechnology and Bioinformatics, D.Y. Patil University, CBD Belapur, Navi Mumbai, India.

* Email Address: selvaakumarc@gmail.com

1.INTRODUCTION

Cockroaches are classified into oviparous, ovoviviparous and viviparous types. Firstly, in oviparous cockroaches the eggs are well protected by oothecal deposited on a substrate. Secondly, in ovoviviparous the eggs are protected by ootheca are deposited in female the brood sac (Uterus) of the female. Here the embryos gets protection and water inside the brood sac, [1]. Thirdly, in viviparous cockroach lay their eggs with little yolk, which is in to the nourishment of the developing offspring done from the brood sac. *Diploptera punctate*, is the only known species reported till date with yolk. During gestation, milk replaces the yolk present in embryo gut [2] in crystal form with a lipoclain fold [3]. Regarding lipoclain, they are the small extracellular proteins critically involved in retinol transport, prostaglandin synthesis, invertebrate cryptic coloration, olfaction and pheromone transport regulation of cell homoeostasis and the modulation of the immune response [3, 4]. This protein accumulates mutations within the sequence and try to conserve 3D structure with an eight stranded continuously hydrogen bonded antiparallel β-barrel that has an internal ligand-binding site [5,6]. Apart from cockroaches, these folds were also observed in Bacterial Outer Membrane Lipoprotein (Blc) of *E. coli* [7] critically involved in lipid storage. Basically, they are expressed during stress, when the cell needs maintenance or when it suffers physical damage [8]. The homologs of Blc were reported in Laziarillo in insects and Apolipoprotein D (ApoD) in mammals [9]. Lazarillo protein is a cell surface glycoprotein, which is expressed, in the embryonic stage of grasshopper in their nervous system, and it belongs to the lipocalin family [10]. ApoD is associated with ligand transport between the cells within an organ. Furthermore, they scavenge ligand within the organ for transport to the blood [7]. The present study focuses on sequence and structural similarities of ApoD and Blc proteins with cockroach milk protein from *in-silico* perspective.

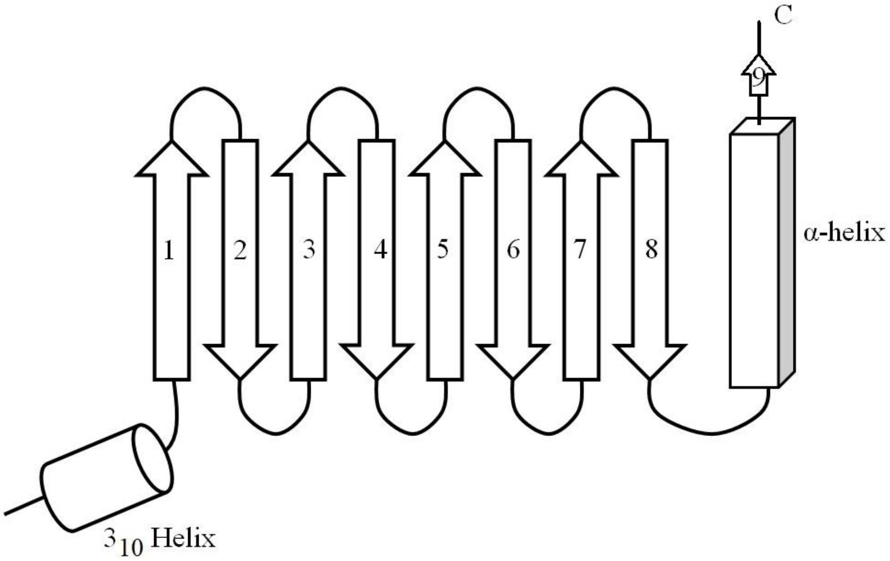


Figure 1 Lipocalin Fold structure. They have nine β sheets represented using arrows, which are antiparallel in nature, one α helix at c terminal end represented as a cuboid and one 3_{10} helix represented as cylinder.

2. MATERIALS AND METHODS

A) Identification of distant homologous through fold based methods

The protein sequence of Cockroach Milk Protein (CMP) (Uniprot ID: Q6SVB6) was downloaded from Uniprot [11]. The crystal structure with higher resolution of 1.2 Å (4NYQ [3]) was downloaded from Protein Data Bank (www.rcsb.org) [12]. We performed fold based search to identify potential distal structural homologs of CMP using pGENTHREADER [13] and pDOMTHREADER [13]. pGenTHREADER, is a new protein fold recognition method that is both fast and reliable. This method uses a traditional sequence alignment algorithm with the help of which it generates alignments, which are then evaluated by a method derived from threading techniques. The pDomTHREADER includes PSIPRED [14] based structure identification which incorporates both the sequence and structural data directly and thus improves sensitivity and selectivity. Finally, pGenTHREADER is for remote homology detection.

B) Multiple Sequence Alignment

Based on the generated hits from fold based method the protein sequences of ApoD of *Homo sapiens* (Uniprot ID: P05090), Lazarillo protein of *Schistocerca americana* (Uniprot ID: P49291) and Blc of *E. coli* (Uniprot ID: P0A901) were downloaded. All three sequences along with the CMP were considered for Multiple Sequence Alignment using ClustalW [15].

C) Domain Analysis

To identify the potential domains of ApoD and CMP here we opted for databases like PRINTS [16], ProDom [17] and PROSITE [18]. PRINTS database is a collection of various protein family fingerprints. The conserved domain of a protein is also known as fingerprints, which provides distinctive signatures for particular protein families and structural/functional domains. ProDom is the database that searches the set of protein domain families according to the query. PROSITE is the database which is used to find the protein domains, Motifs and Protein Families.

D) Structural Superimposition

The crystal structure of ApoD (PDB id: 2HZQ [19], CMP and Blc (PDB id: 1QWD [8]) were downloaded from Protein Data Bank. Furthermore, using MatchMaker tool of Chimera [20], the structural superimposition of CMP with Blc and CMP with ApoD were performed consecutively. Basically, this generated pairwise sequence alignments and then fits the aligned residue pairs. During structural superimposition, Needleman-Wunsch algorithm and BLOSUM 30 matrix was used while rest all parameters were kept as default.

E) Mobility Analysis

To understand the protein flexibility of CMP, Blc and ApoD, here we opted for iMODS online server [21]. This tool helps to understand the protein flexibility using the Normal Mode Analysis with internal coordinates. Given an input structure, the server helps to model, visualize and analyse functional collective motions. Default parameters were used for mobility analysis. Arrow field representation shows the movement of individual residues and affine arrows shows the complete domain mobility.

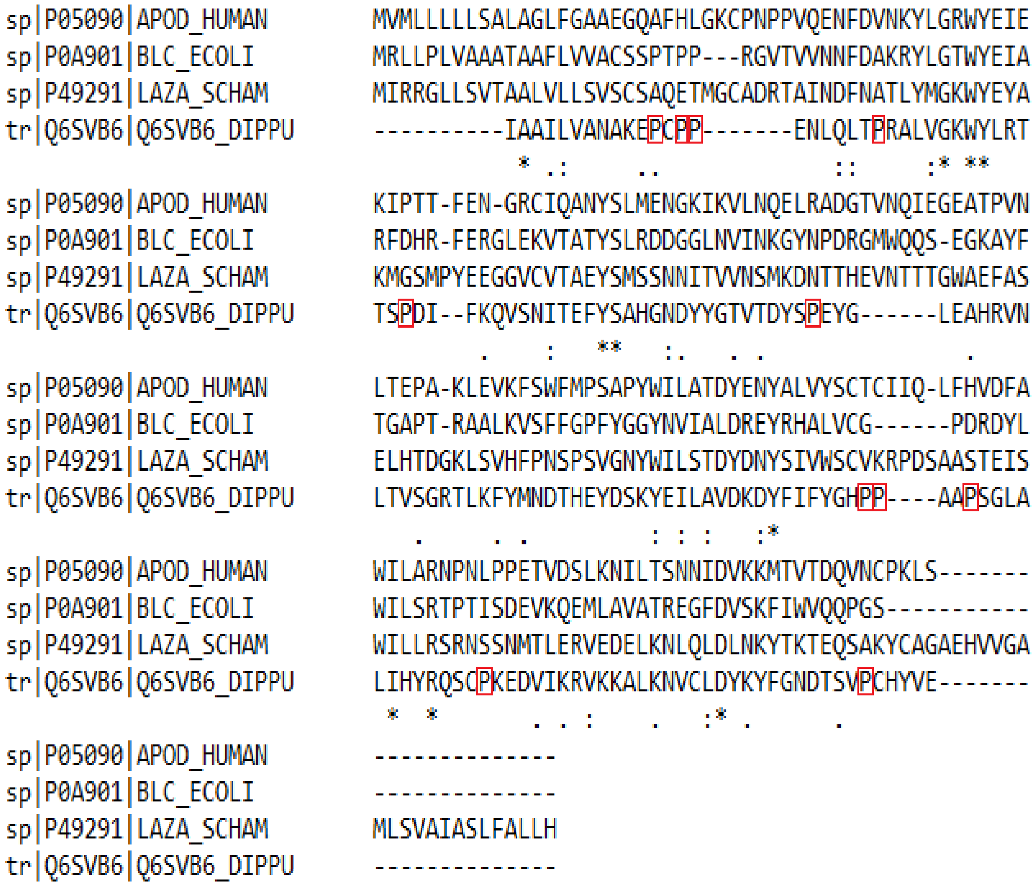
4.RESULTS AND DISCUSSION

A) Distant homolog identification through fold based methods

The CMP protein was submitted to pGENTHREADER and pDOMTHREADER. The generated results from pGENTHREADER show that 1QWD (Blc - *E.Coli*) with a medium Confidence has a net score of 42.98 with an overall identity of 14.6%. Conversely, the pDOMTHREDER reports a high confidence score. Their net score was 4.388.

B) Multiple Sequence Alignment

All four proteins were considered for Multiple Sequence Alignment. Identical residues were ten in number which were sparingly distributed during the alignment. However, the conservative substitutions were twelve in number. Regarding the pairwise scoring of the aligned proteins, CMP and ApoD shows higher score compared to CMP and Blc protein (Table 1). These alignment scores tell us that ApoD has higher sequence similarity with CMP than Blc and Lazarillo protein. It was seen that the CMP has 11 proline residues present as compared to other sequences, (Blc has 9, ApoD 12, Lazarillo 4), making it rigid structure [22].



```
sp|P05090|APOD_HUMAN        MVMLLLLLSALAGLFGAAEGQAFHLGKCPNPPVQENFDVNKYLGRWYEIE
sp|P0A901|BLC_ECOLI         MRLLPLVAAATAAFLVVACSSPTPP---RGVTVVNNFDAKRYLGTWYEIA
sp|P49291|LAZA_SCHAM        MIRRGLLSVTAALVLLSVSCSAQETMGCADRTAINDFNATLYMGKWYEYA
tr|Q6SVB6|Q6SVB6_DIPPU      ----------IAAILVANAKEPCPP-------ENLQLTPRALVGKWYLRT
                                       * .:     ..             ::     :* **

sp|P05090|APOD_HUMAN        KIPTT-FEN-GRCIQANYSLMENGKIKVLNQELRADGTVNQIEGEATPVN
sp|P0A901|BLC_ECOLI         RFDHR-FERGLEKVTATYSLRDDGGLNVINKGYNPDRGMWQQS-EGKAYF
sp|P49291|LAZA_SCHAM        KMGSMPYEEGGVCVTAEYSMSSNNITVVNSMKDNTTHEVNTTTGWAEFAS
tr|Q6SVB6|Q6SVB6_DIPPU      TSPDI--FKQVSNITEFYSAHGNDYYGTVTDYSPEYG------LEAHRVN
                                    .    :   **   :.   . .               .

sp|P05090|APOD_HUMAN        LTEPA-KLEVKFSWFMPSAPYWILATDYENYALVYSCTCIIQ-LFHVDFA
sp|P0A901|BLC_ECOLI         TGAPT-RAALKVSFFGPFYGGYNVIALDREYRHALVCG------PDRDYL
sp|P49291|LAZA_SCHAM        ELHTDGKLSVHFPNSPSVGNYWILSTDYDNYSIVWSCVKRPDSAASTEIS
tr|Q6SVB6|Q6SVB6_DIPPU      LTVSGRTLKFYMNDTHEYDSKYEILAVDKDYFIFYGHPP----AAPSGLA
                               .     . .         : : :   :*

sp|P05090|APOD_HUMAN        WILARNPNLPPETVDSLKNILTSNNIDVKKMTVTDQVNCPKLS-------
sp|P0A901|BLC_ECOLI         WILSRTPTISDEVKQEMLAVATREGFDVSKFIWVQQPGS-----------
sp|P49291|LAZA_SCHAM        WILLRSRNSSNMTLERVEDELKNLQLDLNKYTKTEQSAKYCAGAEHVVGA
tr|Q6SVB6|Q6SVB6_DIPPU      LIHYRQSCPKEDVIKRVKKALKNVCLDYKYFGNDTSVPCHYVE-------
                             *  *       . . :    .   :* .      .

sp|P05090|APOD_HUMAN        --------------
sp|P0A901|BLC_ECOLI         --------------
sp|P49291|LAZA_SCHAM        MLSVAIASLFALLH
tr|Q6SVB6|Q6SVB6_DIPPU      --------------
```

Figure 2 Multiple sequence alignment of CMP, ApoD, Blc and Lazarillo protein using ClustalW. Proline residues are shown in red boxes. Total number of Proline residues observed in CMP was 11.

Table 1 Multiple Sequence Alignment Scores of the four sequences: CMP, Blc, ApoD and Lazarillo protein.

Seq. No.	Uniprot ID	Sequence Name	Length	Pairwise alignment	Alignment Score
1.	Q6SVB6	DIPPU_CMP	164 aa	-	-
2.	P0A901	BLC_ECOLI	177 aa	(Seq. 1: Seq. 2)	12
3.	P05090	APOD_HUMAN	189 aa	(Seq. 1:Seq. 3)	17
4.	P49291	LAZA_SCHAM	214 aa	(Seq. 1:Seq. 4)	10

C) Domain Analysis

An in-silico based domain search was performed using PRINTS, PRODOM and prosite for ApoD, and CMP. On performing PRINTS by submitting human Apolipoprotein D, we found that seven motif regions with their lengths ranging from 12 to 23 (Table 2.a) and three motifs with Lipocalin fingerprint (Table 2.b), but there were no hits for CMP in PRINTS database.

Sr No	**Motif**	**Width**	**St**	**Int**
1	GQAFHLGKCPNPPVQ	15	20	20
2	FDVNKYLGRWYEIEKIP	17	37	2
3	FENGRCIQANYSLMENGKIKVLN	23	56	2
4	NLTEPAKLEVKF	12	98	19
5	APYWILATDYENYALVYSCT	20	116	6
6	LFHVDFAWILARNPNLP	17	140	4
7	ILTSNNIDVKKMTVTDQVNC	20	166	9

Table 2 (a) Different motifs found in Apolipoprotein D in PRINTS database. The location of the motif within the sequence (ST), and the interval between the adjacent motifs (INT) are shown in the above table along with the length of the motifs.

Motif Length	**Idscore**	**Pfscore**	**Pvalue**	**Sequence**	**Score**
1 of 3	34.82	272	9.03e-04	NKYLGRWYEIEKI	13
2 of 3	47.71	373	1.18e-06	ILATDYENYALVY	13
3 of 3	23.73	161	2.69e-04	ILARNPNLPPETVDSL	16

Table 2 (b) Lipocalin fingerprints in ApoD present in PRINTS database, along with the motif score and p-value.

To identify domain arrangements within known or unknown families, automated sequence comparisons were performed by SWISS-PROT through PRODOM database. Each entry provides a multiple sequence alignment of homologous domains and a family consensus sequence. ProDom domain results are shown in Table 3, with its positions, scores and E values of ApoD. The ProDom domain with the score of 289 from position 118 to 182 was found out to be related to a lipocalin protein Blc with 89 % positives and 84% identities. Domain with score of 291 belongs the ApoD itself.

Position	ProDom Domain	Score	E Value
23-134	#PDB8Y573	193	6e-16
23-80	#PDB0N7D2	101	0.001
25-66	#PDC6I5H8	110	3e-05
26-71	#PD416277	102	0.0009
27-154	#PD907964	107	0.0003
33-53	#PD381306	113	9e-06
35-189	#PDA068U6	160	5e-11
36-109	#PDA1G229	163	2e-12
36-110	#PDA1G9Z3	115	9e-06
36-71	#PDA2B6O8	105	0.0001
40-76	#PDA1R5J1	100	0.0007
55-110	#PDB0G029	291	3e-31
118-182	#PDA1F827	289	7e-31
118-165	#PD093265	112	1e-05
121-185	#PDB7E382	137	6e-09
121-151	#PDC6C416	97	0.002
121-185	#PDB3G7L6	130	6e-08
160-187	#PDC6O916	137	3e-09

Table 3 ProDom domains producing High-scoring Segment Pairs, arranged according to their positions. Expect value (E) tells us the number of hits that can be expected to see while searching a database. Score tells about the identities and positives between the query sequence and the ProDom domains. Higher the score higher is the identities and positives percentage and vice versa.

Finally, PROSITE database was used to extract domain sequence of Apo D. On submitting the sequence, we obtained optimum result by confirming lipocalin as domain name with ~ 15 length and its sequence NFDVNKYLGRWYEI.

D) Structural Superimposition

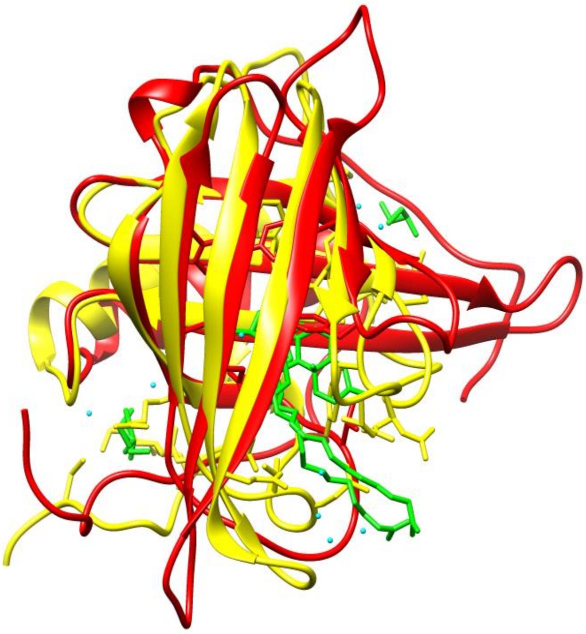


Figure 3 a Superimposed structure of ApoD and CMP.

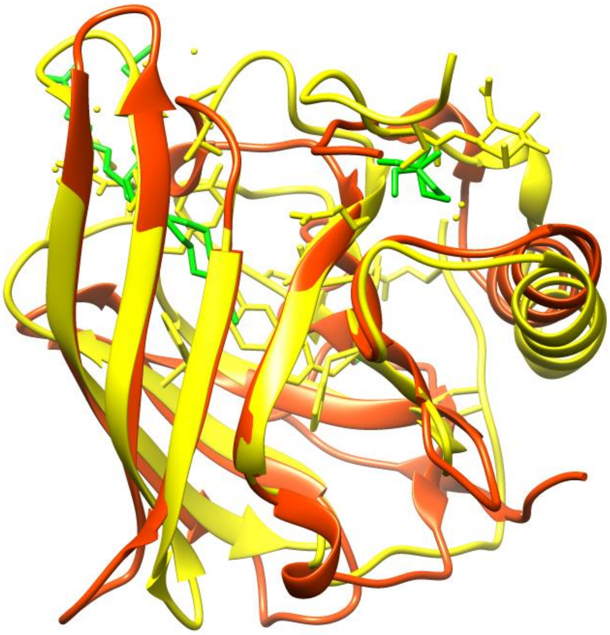
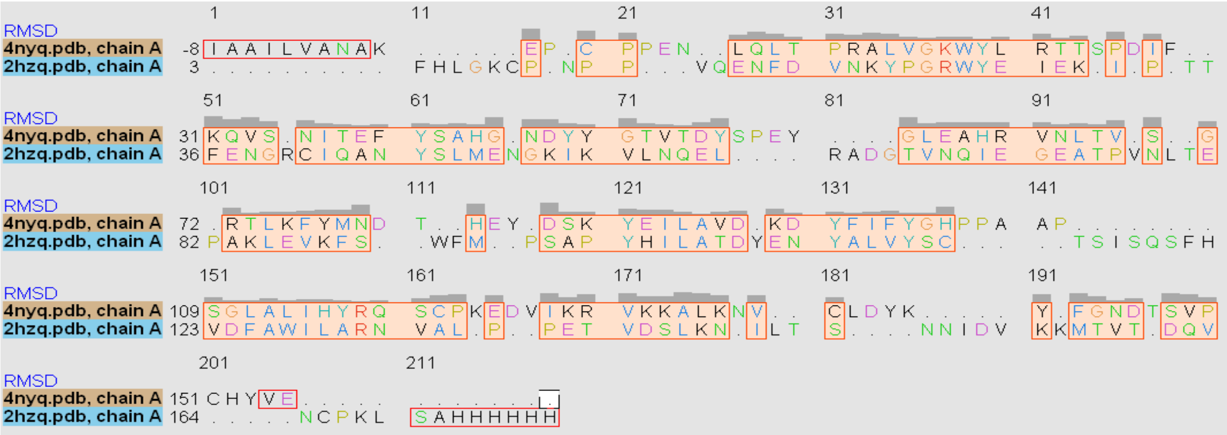


Figure 3 b Superimposed structure of Blc and CMP.

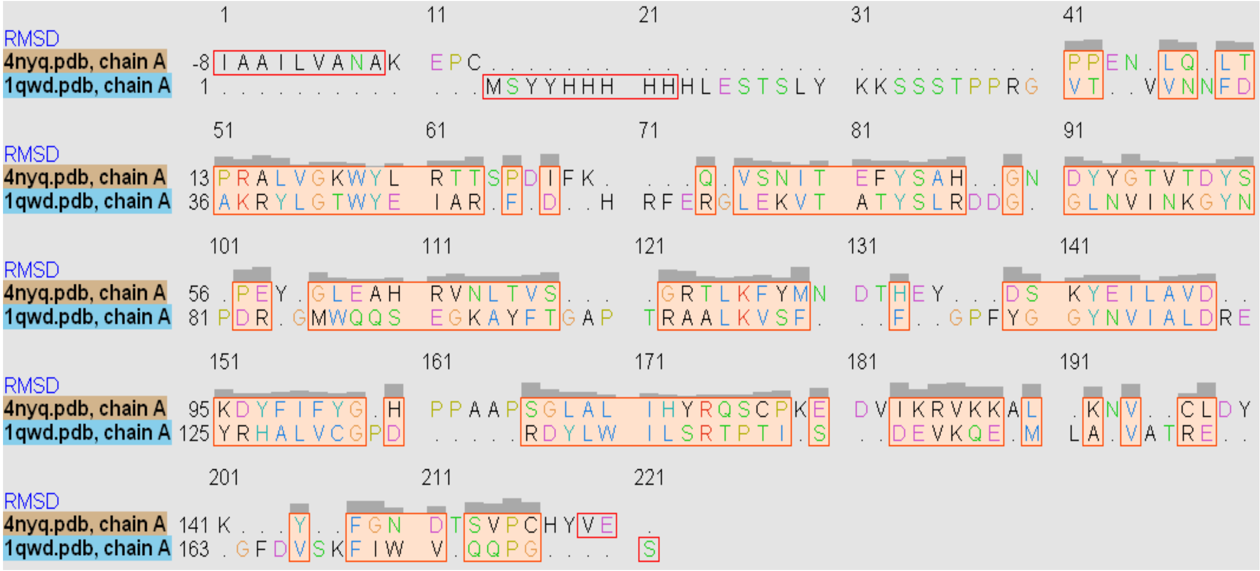
Structural comparison of CMP with ApoD was performed using Matchmaker tool of Chimera (Figure 3.a) which gives a RSMD score of 0.98 Å between the two crystal structures. Certain regions of ApoD which did not superimpose with CMP include Ile32-Ile 42, Leu 61-Thr66, Val77-Ala83; Ser90-Ser95 and Cys114-His122 (Figure 4.a). Interestingly, these residues were the part of loops in the structure of ApoD. Structural comparison of CMP with Blc was also performed (Figure 3.b). RSMD value observed was 1.12 Å. Regions in Blc that did not superimpose with CMP are Lys18-Gly27, Asn32-Leu40, Phe49-Val60, Thr96-Ala102, and Gly132-Tyr137 (Figure 4.b). These residues are part of loops in Blc structure.

(a)



```
                        1          11         21         31         41
RMSD
4nyq.pdb, chain A   -8  IAAILVANAK ......EP.C PPEN..LQLT PRALVGKWYL RTTSPDIF..
2hzq.pdb, chain A    3  .......... FHLGKCP.NP P...VQENFD VNKYPGRWYE IEK.I.P.TT

                        51         61         71         81         91
RMSD
4nyq.pdb, chain A   31  KQVS.NITEF YSAHG.NDYY GTVTDYSPEY ....GLEAHR VNLTV.S..G
2hzq.pdb, chain A   36  FENGRCIQAN YSLMENGKIK VLNQEL.... RADGTVNQIE GEATPVNLTE

                        101        111        121        131        141
RMSD
4nyq.pdb, chain A   72  .RTLKFYMND T..HEY.DSK YEILAVD.KD YFIFYGHPPA AP........
2hzq.pdb, chain A   82  PAKLEVKFS. .WFM..PSAP YHILATDYEN YALVYSC... ..TSISQSFH

                        151        161        171        181        191
RMSD
4nyq.pdb, chain A  109  SGLALIHYRQ SCPKEDVIKR VKKALKNV.. CLDYK..... Y.FGNDTSVP
2hzq.pdb, chain A  123  VDFAWILARN VAL.P..PET VDSLKN.ILT S....NNIDV KKMTVT.DQV

                        201        211
RMSD
4nyq.pdb, chain A  151  CHYVE..... ........
2hzq.pdb, chain A  164  .....NCPKL SAHHHHHH
```

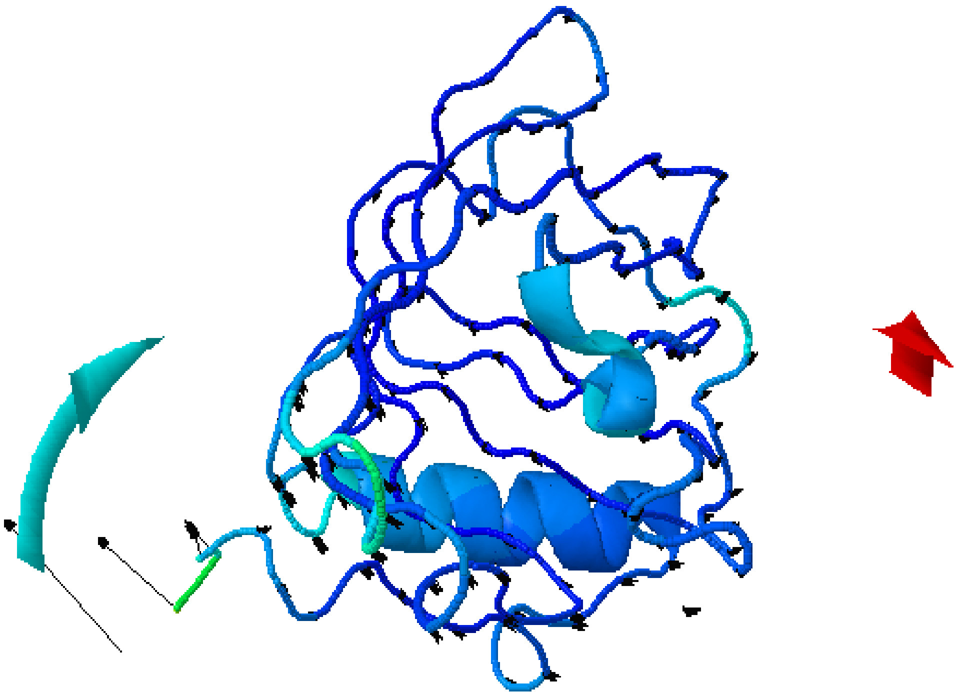
(b)



```
                       1          11         21         31         41
RMSD
4nyq.pdb, chain A   -8 IAAILVANAK EPC....... .......... .......... PPEN.LQ.LT
1qwd.pdb, chain A    1 .......... ...MSYYHHH HHHLESTSLY KKSSSTPPRG VT..VVNNFD

                       51         61         71         81         91
RMSD
4nyq.pdb, chain A   13 PRALVGKWYL RTTSPDIFK. ...Q.VSNIT EFYSAH..GN DYYGTVTDYS
1qwd.pdb, chain A   36 AKRYLGTWYE IAR.F.D..H RFERGLEKVT ATYSLRDDG. GLNVINKGYN

                       101        111        121        131        141
RMSD
4nyq.pdb, chain A   56 .PEY.GLEAH RVNLTVS... .GRTLKFYMN DTHEY...DS KYEILAVD..
1qwd.pdb, chain A   81 PDR.GMWQQS EGKAYFTGAP TRAALKVSF. ..F..GPFYG GYNVIALDRE

                       151        161        171        181        191
RMSD
4nyq.pdb, chain A   95 KDYFIFYG.H PPAAPSGLAL IHYRQSCPKE DVIKRVKKAL .KNV..CLDY
1qwd.pdb, chain A  125 YRHALVCGPD .....RDYLW ILSRTPTI.S ..DEVKQE.M LA.VATRE..

                       201        211        221
RMSD
4nyq.pdb, chain A  141 K...Y..FGN DTSVPCHYVE .
1qwd.pdb, chain A  163 .GFDVSKFIW V.QQPG.... S
```

Figure 4: Structure based sequence comparison of (a) CMP with ApoD (b) CMP with Blc. The residues superimposing with the residues of CMP are shown in red box. Residues that did not overlap were investigated separately.

E) Mobility Analysis



(a)

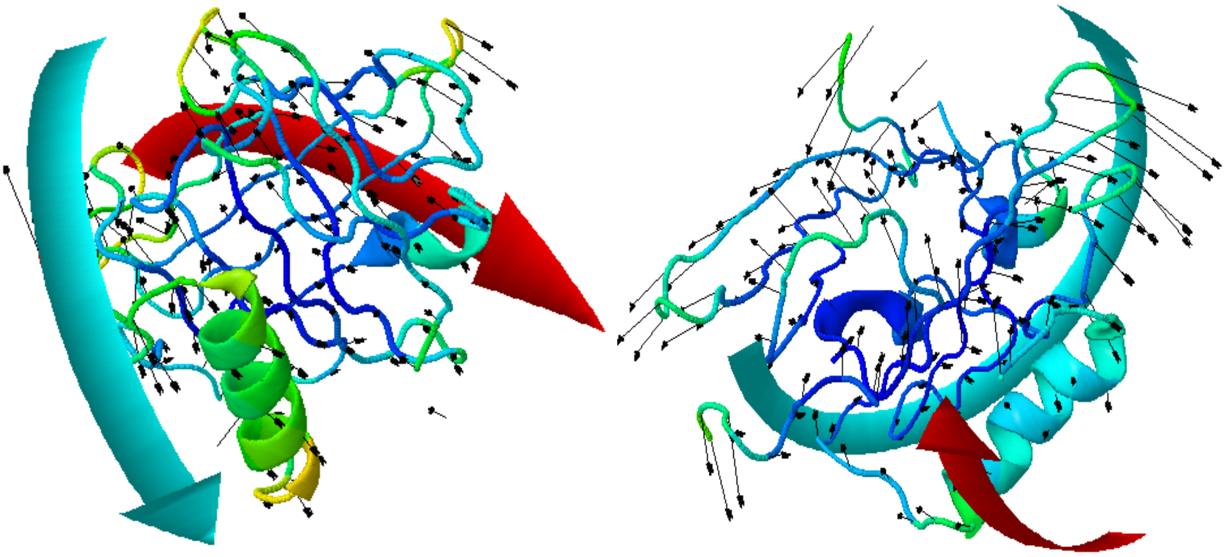
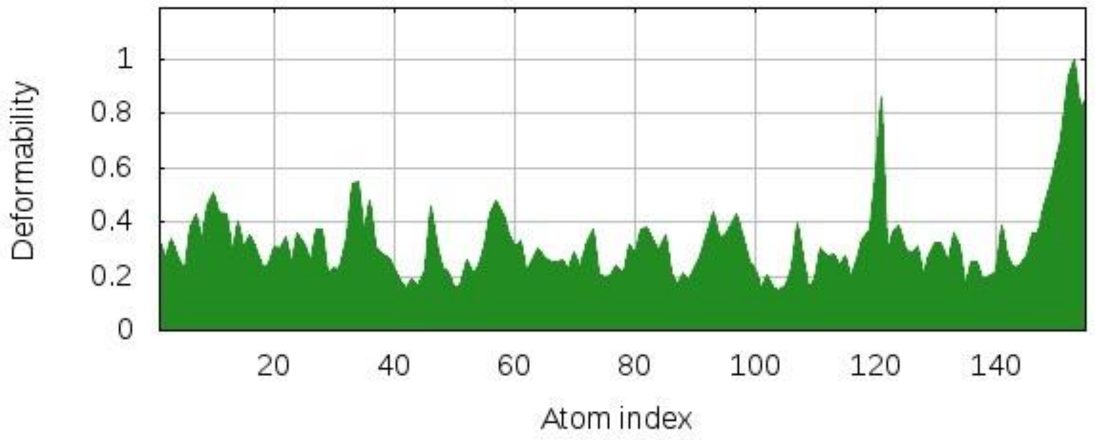
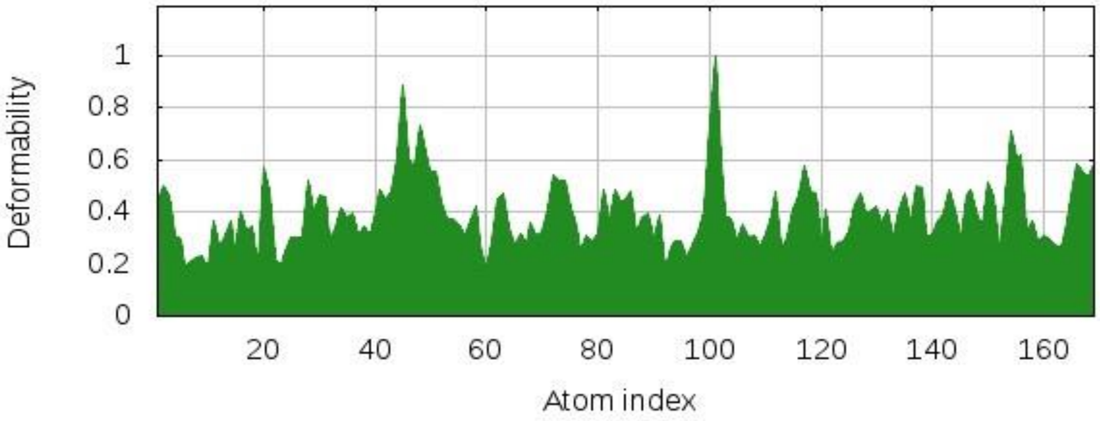


Figure 5: Mobility analysis using iMODS. (a) CMP, (b) Blc and (c) ApoD. Domains with maximum mobility are shown. Complete domain mobility is represented by blue and red color arrows. Black colored arrows show the movement of individual residues.

(a)



(b)



(c)

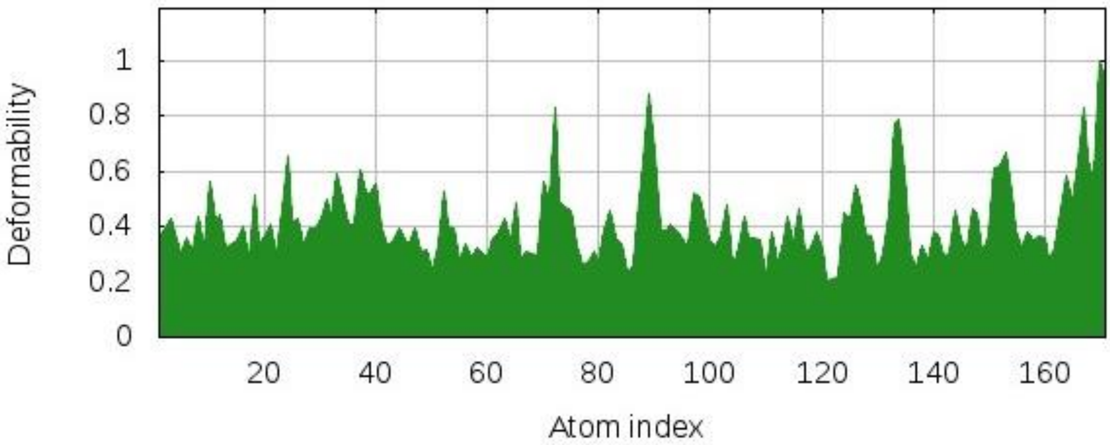


Figure 6: Deformity graph showing deformability of individual atoms. Residues with higher value of deformability may be a part of hinges. (a) Deformability graph for CMP. (b) Deformability graph for Blc. (c) Deformability graph for ApoD.

The overall mobility analysis of ApoD with CMP and Blc with CMP was carried out using iMODS tools. Here we observed that ApoD (Figure 5.c) and Blc (Figure 5.b) have higher mobility compared to CMP (Figure 5.a). As per the generated report, CMP residues have lesser deformability than ApoD and Blc. A single hinge region was observed in Blc with value of deformability more than one (Figure 6.b) and ApoD has a hinge value of deformability 0.8 (Figure 6.c). The location of the chain 'hinges' can be derived from high deformability regions. Blc and ApoD have similar hinge pattern i.e. around 100 residue but in CMP it has been shifted towards residue index 120 (Figure 6.a).

4. CONCLUSION

The *Diploptera punctuate*, a viviparous cockroach species, have the lipocalin like milk protein. Our studies showed that cockroach milk protein (CMP) and Bacterial lipocalin (Blc) of E.coli were closely related from structural perspective with lesser sequence similarity. After performing multiple sequence alignment, we found that CMP has more sequence similarity to Apolipoprotein D (ApoD) of humans with alignment score 17 than that of Blc. And Prosite pattern revealed that the ApoD belong to lipocalin domain. When these three proteins were compared with respect to their structures the RSMD value of CMP against ApoD using matchmaker was found out to be 0.98 Å while with Blc 1.12 Å, which shows that there is lesser structural deviation observed between CMP and ApoD compared to Blc. Finally, after observing significant structural similarity between CMP and ApoD, we performed various mobility test of CMP, ApoD and Blc.Domain mobility among the three proteins concluded that there is more of positive correlation in CMP and ApoD than with Blc. Also, CMP has less mobility due to presence of excess proline residues making them more reigid. Whereas, ApoD and Blc-E.coli shows more domain flexibility.Further studies on lipocalin family and significant study in amino acids involved will pave a way for future researchers to understand their evolution in a better way.

ACKNOWLEDGEMENT

The authors acknowledge the management of D.Y. Patil University, CBD Belapur for providing the facilities to carry out this work. This project has not been funded by any external agencies.

CONFLICT OF INTEREST

The authors have no conflict of interest.

REFERENCES

1. Evans, L. D., and Stay, B. 1989. Humoral induction of milk synthesis in theviviparous cockroach Diploptera punctata.Invertebr. Reprod. Dev. 15:171–176.
2. Evolution of a novel function: nutritive milk in the viviparous cockroach, Diploptera punctate Anna Williford, Barbara Stay, and Debashish Bhattacharya Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA
3. Banerjee S, Coussens NP, Gallat FX, Sathyanarayanan N, Srikanth J, Yagi KJ, Gray JS, Tobe SS, Stay B, Chavas LM, Ramaswamy S. Structure of a heterogeneous, glycosylated, lipid-bound, in vivo-grown protein crystal at atomic resolution from the viviparous cockroach Diploptera punctata. IUCrJ. 2016 Jul 1;3(4):282-93.
4. D R Flower The lipocalin protein family: structure and function. Biochem J. 1996 Aug 15; 318(Pt 1): 1–14.
5. Flower DR. Multiple molecular recognition properties of the lipocalin protein family. J Mol Recognit. 1995 May-Jun;8(3):185–195.
6. Ganfornina MD, Sánchez D, Bastiani MJ. Lazarillo, a new GPI-linked surface lipocalin, is restricted to a subset of neurons in the grasshopper embryo. Development. 1995 Jan;121(1):123–134.
7. Russell E Bishop, Christian Cambillau, Gilbert G. Privé, Derek Hsi, Desiree Tillo, and Elisabeth R. M. Tillier. Bacterial Lipocalins: Origin, Structure, and Function. Madame Curie Bioscience Database; 2000-2013.
8. Valerie Campanaccia; Didier Nurizzob; Silvia Spinellia, Christel Valenciaa; Mariella Tegonia, Christian Cambillaua. The crystal structure of the Escherichia coli lipocalin Blc suggests a possible role in phospholipid binding. 4 March 2004.
9. Campanacci Valérie,Bishop Russell E.,Blangy Stéphanie,Tegoni Mariella and Cambillau Christian(2006), The membrane bound bacterial lipocalin Blc is a functional dimer with binding preference for lysophospholipids, FEBS Letters, 580, doi: 10.1016/j.febslet.2006.07.086
10. Development. 1995 Jan;121(1):135-47. Developmental expression of the lipocalin Lazarillo and its role in axonal pathfinding in the grasshopper embryo. Sánchez D1, Ganfornina MD, Bastiani MJ.

11. The UniProt Consortium UniProt: the universal protein knowledgebase Nucleic Acids Res. 45: D158-D169 (2017).
12. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
13. Lobley, A., Sadowski, M.I. & Jones, D.T. (2009) pGenTHREADER and pDomTHREADER: New Methods For Improved Protein Fold Recognition and Superfamily Discrimination. Bioinformatics. 25, 1761-1767.
14. Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.
15. Marsden, R.L., McGuffin, L.J. & Jones, D.T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Science, (2002) 11, 2814-2824.
16. Teresa K. Attwood Alain Coletta Gareth Muirhead Athanasia Pavlopoulou Peter B. Philippou Ivan Popov Carlos Romá-Mateo Athina Theodosiou Alex L. Mitchell. The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012 Database (Oxford) (2012) 2012: bas019.DOI:https://doi.org/10.1093/database/bas019 Published:14 April 2012
17. [BRU] Catherine Bru, Emmanuel Courcelle, Sébastien Carrère, Yoann Beausse, Sandrine Dalmar, and Daniel Kahn (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res. 33: D212-D215.
18. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. “New and continuing developments at PROSITE” Nucleic Acids Res. (2012): D344-7.
19. Eichinger, A., Nasreen, A., Kim, H.J., Skerra, A. Structural insight into the dual ligand specificity and mode of high density lipoprotein association of apolipoprotein d. (2007) J.Biol.Chem. 282: 31068-31075
20. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." J. Comput. Chem. (2004) 25:1605-1612.
21. López-Blanco JR, Aliaga JI, Quintana-Ortí ES and Chacón P. iMODS: Internal coordinates normal mode analysis server. Nucleic acids research. (2014) 42:W271-6.
22. George RA, Heringa J. An analysis of protein domain linkers: their classification and role in protein folding. Protein Engineering, Design and Selection. 2002 Nov 1;15 (11):871-9.