Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 119—127

http://www.ripublication.com/jcib.htm

Comparative Modeling Studies of C-type Lectin Domain Family 2 Member D Spliced Isoforms through 3D Structures

Poornima Chauhan[1]*, Ashish Kumar Jaiswal[2], Shipra Gupta[3]** and Abhishek Kr. Malakar[4]****

[1]Department of Bioinformatics,
College of Applied Education & Health Sciences, Meerut, India
[2]Bioimedical Informatics Centre, NIRRH-Mumbai, India
[3]Bioinformatics Lab, Biotech Park- Lucknow, India
[4]National Bureau of Fish Genetic Resources, Lucknow, India
**Corresponding Author E-mail: bi.poornima@gmail.com*
**Reviewer

Abstract

C-type lectin domain family 2 member D acts as a receptor for KLRB1 that protects target cells against natural killer cell-mediated lysis. C-type lectin domain family 2 member D exists in six isoforms produced by alternative splicing**.** In this study information from the transcript and other studies is combined with tertiary structure information derived from homology models, as structure can reveal the fine details of how molecules perform their biological functions. Through the comparative modeling method we have predicted splice variants models. A three-dimensional structural model of C-type lectin domain family 2 member D from the *Homo sapiens* and *Mus musculus* were generated using the crystal structure of Human Cd69 from homo sapiens as template.

Various models of splice variants for target proteins were produced. On the basis of energy, G-factors, residues in the most favored region & disallowed region in Ramachandran plot. The best models were selected of *Homo sapiens* and *Mus musculus*. In present study we consider predicted models were compared with the reference protein isoform and found that little variation in protein sequence convert secondary structural elements to another while in some isoform structure remain the same. The visualization of these possible structures provides new insights about their functionality and can aid the design of further experiments. The predicted model was further analyzed to locate the residues in the active sites those provide interactions with the ligand.

Keywords: C-type lectin; CLC2D; PDB, *Homo sapiens* & *Mus musculus.*

Introduction
Calcium-dependent (C-type) lectins are a family of lectins which share structural homology in their high-affinity carbohydrate-recognition domains [1]. Carbohydrate-recognition domains of C-type (Ca^{2+}-dependent) serve as prototypes for an important family of protein modules. Only some domains in this family bind Ca^{2+} or sugars. A comparison of recent structures of C-type lectin-like domains reveals diversity in the modular fold, particularly in the region associated with Ca^{2+} and sugar binding. C-type lectin-like domains associate with each other through several different surfaces to form dimers and trimers, from which ligand-binding sites project in a variety of different orientations [2]. Carbohydrate-recognition domains of C-type lectins form a subgroup of a larger family of protein domains that are called C-type lectin–like domains (CTLDs) [3]. The CTLD-containing proteins are divided into several families according to primary protein structure and architecture [4]. The classification of CTLD-containing proteins into families is further supported by analysis of the genomic organization of CTLDs [5]. Within the super family of C-type lectin–like proteins, a heterogeneous group of proteins has been described, consisting of a single CTDL found in the absence of other protein domains[6][7][8]. Several alternatively spliced transcript variants have been identified in human, but the full-length nature of every transcript has not been defined [9]. These activities could influence the resultant immune response and can, in certain circumstances, lead to autoimmunity and disease [10]. In the present study we analyzed a C-type lectin domain family 2, member D (gene clc2d) of human and mouse. This gene encodes a member of the natural killer cell receptor C-type lectin family. The encoded protein inhibits osteoclast formation and contains a transmembrane domain near the N-terminus as well as the C-type lectin-like extracellular domain.

The aim of the present study is to derive a putative three-dimensional structure for C-type lectin domain family 2, member D (gene name clc2d) of human and mouse based on the crystal structure of Human Cd69 (PDB code: 1fm5) and active site analysis. Comparative analysis of various isoforms of the given protein in human has also been done. The 3D structures of a protein are of great assistance when planning experiments aimed at the understanding of protein function and during the drug design process.

Material and Methods
The primary protein sequences of human and mouse C-type lectin domain family 2 member D were downloaded from GenBank (NCBI) under the gene name clc2d_human and CLC2D_mouse respectively. The retrieved target sequences were of 191(human) and 207(mouse) amino acids long, stored in FASTA format. The modeling began with a template structure search.

Sequence comparison

Protein BLAST [11] was used to compare target sequence against the structures available in PDB. Alignment confirmed that the Human Cd69 shares the highest percentage residue identity with the query sequences. Template sequence was made up of 199 amino acids but crystal structure was available only of a small fragment (85-199) in PDB. In alignment the human sequence from 74 to 191amino acids and the mouse sequence from 79-207 amino acids were shown similarity against template structure. The build both the human and the mouse C-type lectin domain family 2 member D model using Human Cd69 as templates. The human CLC2D exists in six isoforms produced by alternative splicing.

Model building

Model building was performed using the program MODELLER9v6 program [12] with multiple cycles of refinement with conjugate gradient minimization and molecular dynamics with simulated annealing. Several models at various refinements level and library schedules were generated. The program models protein structure by satisfaction of spatial restraints starting from the extended polypeptide chain conformation. For homology modeling of human and mouse, these restraints were derived from aligned sequences and template structures of Human Cd69 from Protein Data Bank (1fm5a and 3hupa).

Model evaluation

The models were then analyzed further and validated using Ramachandran [13]. The amino acid environment was evaluated using ERRAT plots [14], which assess the distribution of different types of atoms with respect to one another in the protein model and energy minimization performed by Verify3D (15). The model structure was visualized and superimposed with template using Pymol and VMD. Active site analysis was carried out using WHAT IF (http://swift.cmbi.ru.nl/servers) server.

Results & Discussion

BLAST search was performed for C-type lectin domain family 2 member D of *H. sapiens* and *Mus musculus* across from Protein Data Bank and significant similarities were found with several proteins belonging to the C-type lectin domain family (Table 1). Among them two proteins 1fm5a and 3hupa were selected for the further procedure. Homology percent was calculated with the help of GENEBEE tool to select the best homolog among the selected templates (Table 2).

Table 1: Proteins found with BLAST search producing significant alignments with CLC2D human against PDB.

DB ID Source	Description	Identity (%)	E-value
1FM5_A	Chain A, Crystal Structure Of Human Cd69	62	6e-26

3HUP_A	Chain A, High-Resolution Structure Of The Extracellular Domain Of Human Cd69	62	1e-25
1E87_A	Chain A, Human Cd69-Trigonal Form	59	3e-25
1DV8_A	Chain A, Crystal Structure Of The Carbohydrate Recognition Domain Of The H1 Subunit Of The Asialoglycoprotein Receptor	50	2e-13
1T8C_A	Chain A, Structure Of The C-Type Lectin Domain Of Cd23	32	5e-11
Significant alignments with CLC2D_mouse against PDB			
1FM5_A	Chain A, Crystal Structure Of Human Cd69	60	2e-21
3HUP_A	Chain A, High-Resolution Structure Of The Extracellular Domain Of Human Cd69	60	5e-21
1E87_A	Chain A, Human Cd69-Trigonal Form	55	3e-20
3KQG_A	Chain A, Trimeric Structure Of Langerin	75	2e-10
3P7G_A	Chain A, Structure Of The Human Langerin Carbohydrate Recognition Domain In Complex With Maltose	60	2e-10

Table 2: Homology percentage of mouse and isoforms of human using GENEBEE tool.

Target proteins	1FM5_A	3HUP_A
CLC2D_HUMAN (Q9UHP7.1)	44.7	36.7
CLC2D_ HUMAN (Q9UHP7.2)	41.6	33.9
CLC2D_ HUMAN (Q9UHP7.3)	39.0	31.5
CLC2D_ HUMAN (Q9UHP7.4)	27.7	21.2
CLC2D_ HUMAN (Q9UHP7.5)	39.2	43.9
CLC2D_ HUMAN (Q9UHP7.6)	22.6	25.0
CLC2D_MOUSE	41.8	31.4

Among the selected templates the extracellular domain of Human Cd69 from *H. sapiens* (PDB code: 3hup) has high E value and less homologues in BLAST search than Crystal Structure of Human Cd69 (PDB code: 1fm5a) even sequence coverage was same. Multiple sequence alignment of human and mouse along with Human Cd69 sequence was performed (Fig. 1). On the basis of these results, we can conclude that LEPR protein may be folded similarly to these proteins.

T-COFFEE, Version_8.93 (Thu Aug 5 18:09:23 CEST 2010)
Cedric Notredame
CPU TIME:0 sec.
SCORE=98

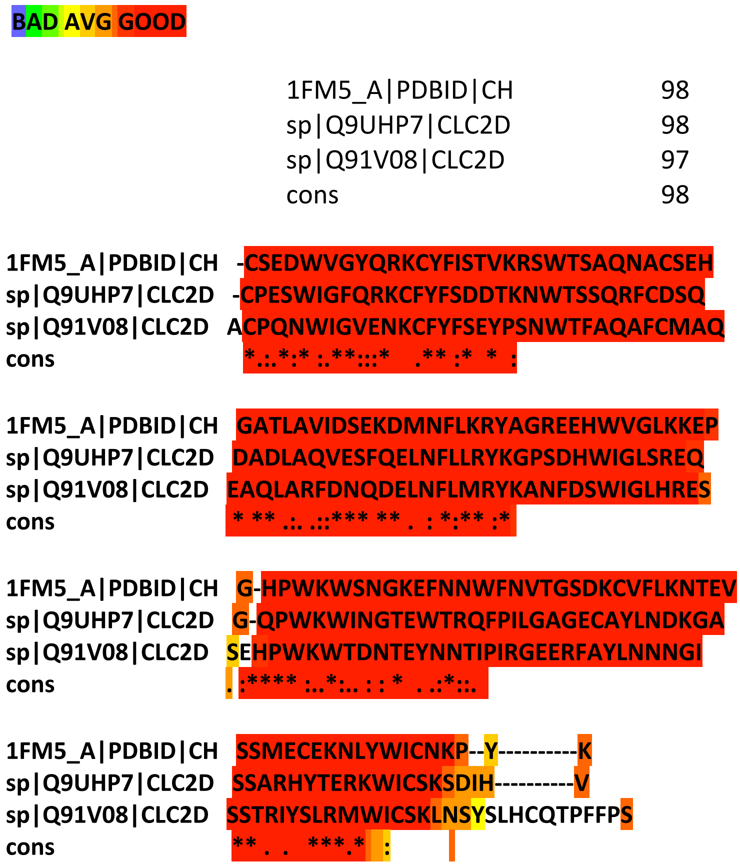

Figure 1: Sequence alignment between CLC2D_human E1, CLC2D_mouse and human Cd69 (PDB: 1FM5_A).

MODELLER 9V6 was used for generation of models and a total of 14 models were generated. The comparable Ramachandran plot characteristic and the goodness factors confirm the quality of the modeled structure. An analysis of templates and the predicted structures using PROCHECK [16] indicates that the quality of the presented model is satisfying (Table 3). It does not differ significantly from those of the templates. A criterion to select the best model was energy that tends to minimum to provide stability to the molecule and core region in the Ramachandran plot should be equal to template or more than 90%.

Table 3: Parameters for structural verification of the predicted models.

Model No.	Core region in Ramachandran plot	Disallowed region in Ramachandran plot
CLC2D_HUMAN (Q9UHP7.1)	91.3%	1.0%

CLC2D_ HUMAN (Q9UHP7.2)	81.6%	1.1%
CLC2D_ HUMAN (Q9UHP7.3)	92.9%	0.0%
CLC2D_ HUMAN (Q9UHP7.4)	85.0%	0.8%
CLC2D_ HUMAN (Q9UHP7.5)	91.6%	0.0%
CLC2D_ HUMAN (Q9UHP7.6)	81.6%	1.1%
CLC2D_MOUSE	82.9%	1.7%
1FM5_A	89.4%	1.0%

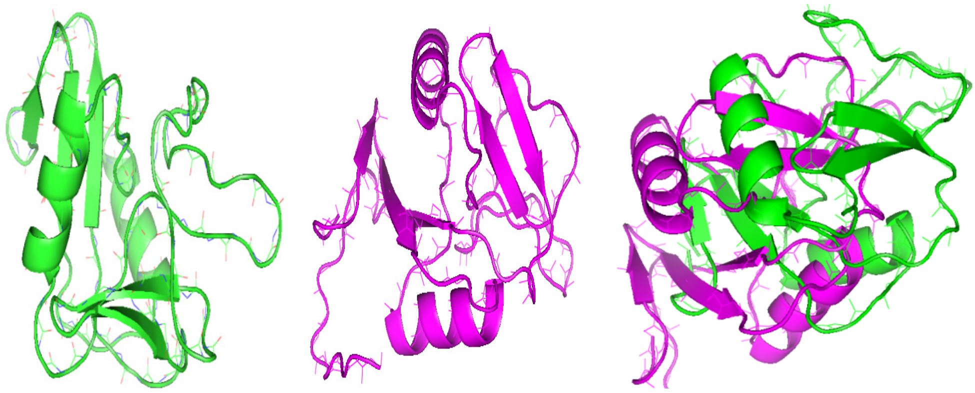

Figure 2: (A) - 3D structure of predicted CLC2D_Human (sequence region 74 to 191) (B) - 3D structure of predicted CLC2D_Mouse (sequence region 79-207) (C) - Superposition between the CLC2D_human and CLC2D Mouse models.

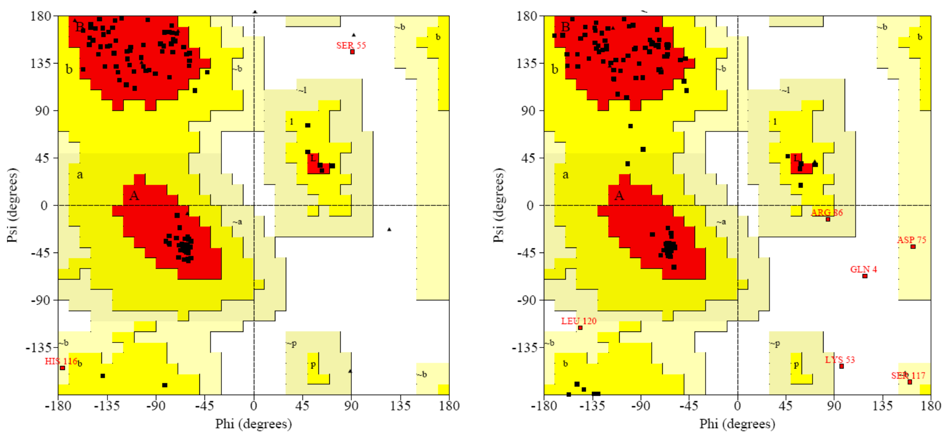

Figure3: Structural verification of the predicted models through Ramachandran Plot (a) CLC2D_human (b) CLC2D_Mouse.

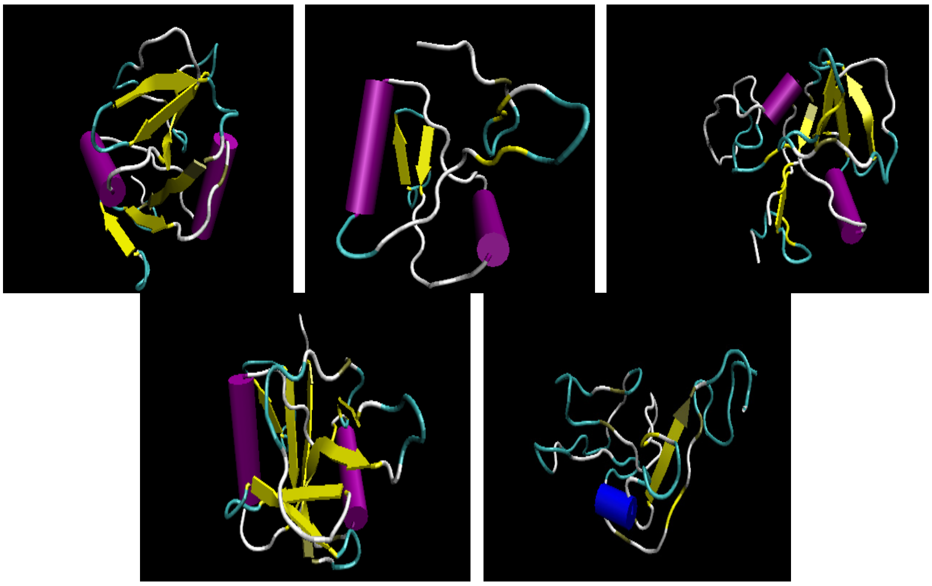

Figure 4: Predicted 3D structure of five isoforms of CLC2D_Human.

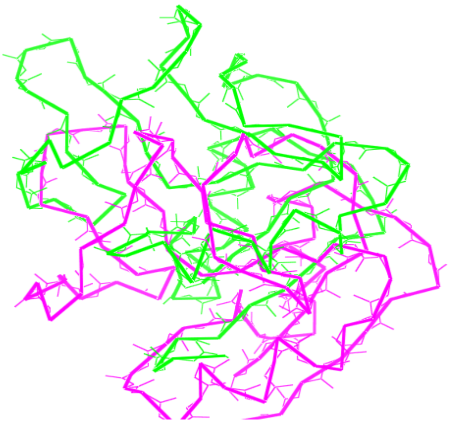

Figure 5: Superposition of the c-type lectin domains from the model of human and mouse E1 (CLC2D_human in green; CLC2D_mouse E1 in magenta).

The predicted model was further analyzed to locate the residues in the active sites those provide interactions with the ligand (Fig.4). In whole protein sequence the active site residues are present as:

Model	Active sites
Isoform 1	V113,F116,L119,N120, 126,P127,S128,D129,W132,I133,L135,G159-A160,G161,A164,I65Y,166L,167N,168D,169K,170G,A171,S172,S173,A174,R175,T178,R180
Isoform 2	C75,P76, F81,Q82,R83,K84,C85,F86, C103, D104, Q105, D106, A107, D108,L109,A110,Q111,V112,E113,S114,F115,Q116,E117,L118,I33,L1

	35,W144,I145,N146,G147,T148,E149,W150,T151,R152,Q153,L154,V155, L164, Q171, Y184, P185, G186, S187, R188, R189, V190, C191, L192, F193,E194
Isoform 3	C75,P76, F81,Q82,R83,K84,C85,F86, C103, D104, Q105, D106, A107, D108,L109,A110,Q111,V112,E113,S114,F115,Q116,E117,L118,I33,L135,W144,I145,N146,G147,T148,E149,W150,T151,R152,Q153,L154
Isoform 4	H2,D3,N5,N6,V7,P13,S14,E15,L16,P17,A18,N19,P20,G21,C22,H24,S25,K26,H28,S29,I30,K31,A32,T33,L34,I35,W36,R37,L38,F39,F40,L41,I42,M43,F44,L45,T46,I47,I48,V49,C50,G51,M52,V53,A54,A55,Q72,G124,R127,V128,C129,L130,F131
Isoform 5	T47,Q48,R49,K50,C51,F52, Q71,D72,A73,D74,L75,A76,Q77,V78,E79, I103,N104,G105,T106,C151,S152,K153,S154,D155,I156,H157,V158
Isoform 6	H2,N6,V7,E8,K9,D10,I11,T12,P13,S14,E15,P30,S31,V32,C33,L34,Q35,A36,A37,C38,P39,E40,S41,W42,I43,S55,D56,D57,T58,K59,V78,E79, S 90,L84,V85,S86,Y87,P88,G89,S90,R91,R92,V93,E97

Discussion

3D structure of predicted CLC2D_Human and CLC2D_Mouse is shown in Fig 2. Both the structures are superimposed also that denotes few variations between them. The predicted structures are verified by Ramachandran plot (fig. 3).The plot is divided into four parts, among them core region and disallowed region are considered (Table 3). After the analysis of various isoform of CLC2D_Human (fig. 4), we found that variation in isoform 2 changed sequences from 155 to 191 but structure was remaining unchanged. While in isoform 3 such region was deleted and isoform 4 had substitutions from 120 to 191 (fig. 6). In isoform 5 sequence region from 21 to 57 was missing, F82 substituted to FFT and Q141 to QFT. Six isoform was shown large variations. In fig 4 purple colour shows ß sheets, α sheets represented by yellow colour and rest region represented loops

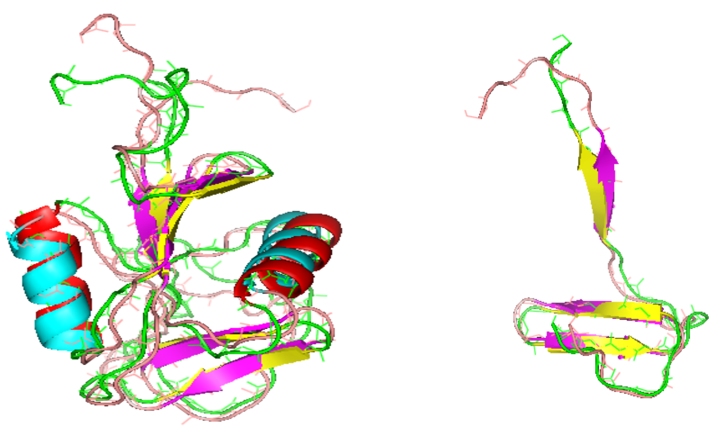

Figure 6: (a) Superposition of the CLC2D_human isoform1 and isoform2 (b) Same structural element in both isoforms (155 to 191 regions).

Acknowledgments
I owe a great thanks to researchers who are continuously working in this area of science, I am thankful to bioinformatics servers and tools developed in different organizations as NCBI, BLAST, PDB, ExPASY server, Modeller, SPDV etc. and also one major search engine Google, because without their help this study could not be completed, Also I want to express my feelings to some of my true friends and well-wisher helped and supported me during the writing of this manuscript.

References

[1] http://www.rndsystems.com/molecule_group.aspx?g=640&r=1
[2] Kurt Drickamer. C-type lectin-like domains
[3] *Dodd R.B., Drickamer, K. (2001) Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. Glycobiology, 11(5), 71R–79R.*
[4] *Drickamer, K. (1993a) Evolution of* Ca^{2+}*-dependent animal lectins. Prog. Nucleic Acid Res. Mol. Biol., 45, 207–232.*
[5] *Gabius, H.J. (1997) Animal lectins. Eur. J. Biochem., 243, 543–576.*
[6] *Lasserre, C., Christa, L., Simon, M.T., Vernier, P., and Brechot, C. (1992). A novel gene (HIP) activated in human primary liver cancer. Cancer Res., 52, 5089–5095.*
[7] *Mann, K. and Siedler, F. (1999). The amino acid sequence of ovocleidin 17, a major protein of the avian eggshell calcified layer. Biochem. Mol. Biol. Int., 47, 997–1007.*
[8] *Weiss, I.M., Kaufmann, S., Mann, K., and Fritz, M. (2000) Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc Haliotis laevigata. Biochem. Biophys. Res. Commun., 267, 17–21.*
[9] provided by RefSeq
[10] *Browjn D. Gordon.*(January 2006). Dectin-1: a signalling non-TLR pattern-recognition receptor., *Nature Reviews Immunology* 6, 33-43
[11] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., (1997).Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25: 3389-3402
[12] Sali A and Blundell TL. J Mol Biol 1993; 234:779.
[13] RAMPAGE Server. http://ravenbioccam.ac.uk/rampage.php [accessed June 23, 2009].
[14] Colovos, C., Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, 1993, *2*, 1511-1519.
[15] Ramachandran G.N., Ramakrishnan C., Sasisekharan V., (1963). Stereochemistry of polypeptide.