J Mol Model (2007) 13:891–896
DOI 10.1007/s00894-007-0201-z

ORIGINAL PAPER

Rational proteomics of PKD1. I. Modeling the three dimensional structure and ligand specificity of the C_lectin binding domain of Polycystin-1.

Vladimir Pletnev · Robert Huether · Lukas Habegger · Wayne Schultz · William Duax

Received: 13 November 2006 / Accepted: 12 March 2007 / Published online: 17 April 2007

Abstract Polycystin-1 (Pc-1) is the 4303 amino acid multi-domain glycoprotein product of the polycystic kidney disease-1 (PKD1) gene. Mutations in this gene are implicated in 85% of cases of human autosomal dominant polycystic disease. Although the biochemistry of Pc-1 has been extensively studied its three dimensional structure has yet to be determined. We are combining bioinformatics, computational and biochemical data to model the 3D structure and function of individual domains of Pc-1. A three dimensional model of the C-type lectin domain (CLD) of Pc-1 (sequence region 405–534) complexed with galactose (Gal) and a calcium ion (Ca^{+2}) has been developed (the coordinates are available on request, e-mail: pletnev@hwi.buffalo.edu). The model has α/β structural organization. It is composed of eight β strands and three α helices, and includes three disulfide bridges. It is consistent with the observed Ca^{+2} dependence of sugar binding to CLD and identifies the amino acid side chains (E499, H501, E506, N518, T519 and D520) that are likely to bind the ligand. The model provides a reliable basis upon which to map functionally important residues using mutagenic experiments and to refine our knowledge about a preferred sugar ligand and the functional role of the CLD in polycystin-1.

V. Pletnev
Institute of Bioorganic Chemistry RAS,
Ul. Miklukho-Maklaya, 16/10,
117997 Moscow, Russia

R. Huether · L. Habegger · W. Schultz · W. Duax (✉)
Department of Structural Biology, Hauptman-Woodward Medical Research Institute & SUNY at Buffalo,
700 Ellicott St., Buffalo, NY 14203, USA
e-mail: duax@hwi.buffalo.edu

Keywords PKD1 disease · Polycystin 1 · C_lectin binding domain · 3D modeling · Bioinformatics

Introduction

Autosomal Dominant Polycystic Kidney Disease is one of the most common human genetic diseases with an incidence rate of about 1:1000 (reviewed in [1]). The disease is characterized by the accumulation of fluid filled cysts in the kidney as a result of perturbed cellular transport caused by undifferentiated cell growth of the renal tubules. It is mainly caused by mutations of either the PKD1 gene on chromosome 16 or the PKD2 gene on chromosome 4. The disease usually results in the loss of renal function requiring dialysis or a kidney transplantation.

The PKD1 gene protein product, polycystin-1 (Pc-1), includes 4303 amino acid residues in a single chain [Swiss-Prot-TrEMBL: P98161] and has been found to regulate the cell cycle [2, 3]. The PKD2 gene product, polycystin-2, has a length of 968 residues [Swiss-Prot-TrEMBL:Q13563] and has been identified as an intracellular calcium release channel. Both products are under extensive study by a variety of experimental methods. Pc-1 contains a large N-terminal extracellular region, 11 predicted transmembrane spanning segments and a relatively short intracellular C-termini, which has been suggested to interact with polycystin-2 (Pc-2). The extracellular region spans more than 3000 residues, contains a number of adhesive domains that implicate Pc-1 in cell-cell and cell-matrix interactions, as well as cell proliferation, migration, differentiation and maturation of the tubules of the kidney [4]. The functions of each of the domains of Pc-1 are being characterized biochemically but their three dimensional organization remains unknown. To date, only the NMR structure of

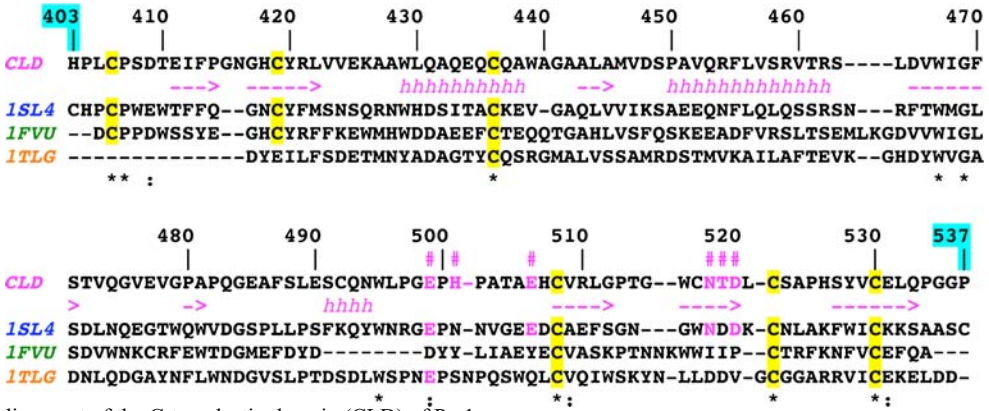

Fig. 1 Sequence alignment of the C-type lectin domain (CLD) of Pc-1 (fragment 405–534) and binding proteins in the 1SL4, 1FVU and 1TLG crystal structures. The proposed CLD carbohydrate binding residues are labeled by #. The Cys residues forming the S-S bridges (406–419, 436–530, 508–522) are highlighted in yellow. *cis* bonds (E499–P500), α helices (429–438, 450–462, 491–494) and β strands (411–414, 417–422, 443–445, 465–471, 480–481, 508–512, 516–520, 527–533) are also identified

one human PKD repeat module (residues 275–354) is in the protein database (PDB accession code: 1B4R) [5].

The complexity of the protein will require an interdisciplinary strategy to gain a complete understanding of its structure and function. Our approach is to model the 3D structures of separate domains of Pc-1 based on proteins of homologous sequences for which crystal structures have been reported. The individual domains will be assembled to model the entire multidomain structure using a combination of bioinformatics, computational biochemistry and molecular graphics methods.

The C-type lectin domain (CLD) of polycystin-1 is situated at residues 405–534 and is encoded by exons 6 and 7 [3] of the PKD-1 gene. This carbohydrate recognition domain was shown to interact with extracellular matrix glycoproteins and the corresponding recombinant product exhibits binding specificity for galactosyl (Gal) and glucosyl (Glc) groups in the presence of Ca^{+2} [6].

Methods

The coordinates and the corresponding amino acid sequences of the homologous templates were taken from the Protein Data Bank [7] and the combined SWISS-PROT/TrEMBL database [8] respectively. The program CLUSTALW (version 1.81) [9] was used for pair-wise sequence alignment. The ExPASy tools [10] were used for bioinformatic characterization and translation of the coding sequence. The graphics program CHAIN (version 7.2) [11] was used for 3D modeling/docking of CLD complex. The 'molrep' program of the crystallographic package CCP4 (version 5.0.2) [12] was used for 3D superposition of protein structures. Molecular mechanics energy optimization of the model structure was performed by the XPLOR_3.1 [13] program with CHARMM [14] parameterization. In addition to the graphics program CHAIN, the programs LIGPLOT (version 4.0) [15] and HBPLUS (version 3.06) [16] were used to study protein-substrate

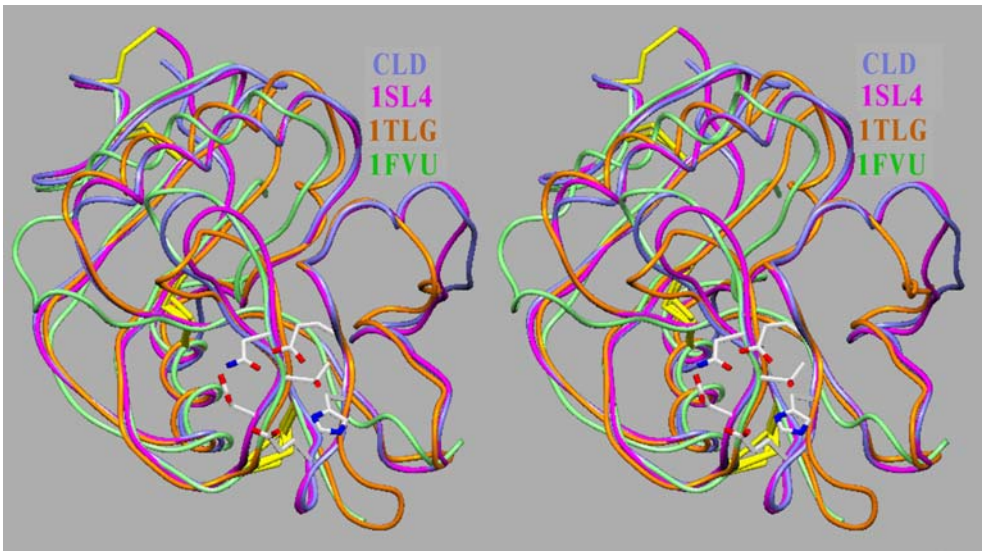

Fig. 2 A SETOR [17] stereo view of the superimposed 3-D molecular structures of 1SL4 (magenta), 1FVU (green), 1TLG (orange) and the CLD model (purple)

Table 1 The position of spatially overlapped disulfide bridges in the C type lectin domains

S-S bridges				
	CLD	1SL4	1TLG	1FVU
1.	406–419	256–267	–	402–413
2.	436–530	284–377	21–119	430–521
3.	508–522	356–369	96–111	498–513
4.	–	253–384	–	

interactions and generate schematic diagrams. The graphics program SETOR (version 4.14.21k) [17] was used to prepare three-dimensional stereo illustrations of the enzyme ternary complex with cofactor and substrate. The Homology-Derived Secondary Structure of Proteins (HSSP) program (Version 1.1) [18] was used to generate a set of aligned proteins sharing ≥30% identity with the sequence of proteins whose 3D crystal structure is known [18].

Results and discussion

A search of the PDB database using our software found three structures with significant sequence homology with CLD. The three structures are the C-type carbohydrate recognition domain of the dendritic cell receptor, DC-SIGN, complexed with tetramannose and Ca^{+2} [PDB:1SL4] [19], a C-type lectin from *Polyandrocarpa misakiensis* complexed with D-galactose and Ca^{+2} [PDB:1TLG] [20] and an uncomplexed C-type lectin botrocetin from the venom of the snake *Bothrops jararaca* [PDB:1FVU] [21]. The pair-wise sequence alignments of 1SL4, 1TLG and 1FVU against CLD have 25, 18 and 27% sequence identity respectively.

A multiple alignment of the sequences of the three PDB structures with that of CLD is illustrated in Fig. 1. The four sequences share only six amino acid residues. When the α-carbon skeletons of the PDB structures are superimposed the observed similarity of their folds (Fig. 2) supports the conclusion that CLD has the same fold and any of the three crystal structures can be used to model the CLD structure. All three PDB structures share two disulfide bridges and two of them (1SL4 and 1FVU) share a third (Table 1).

In vitro biochemical experiments indicate that sugars bind to CLD, 1SL4 and 1TLG in a calcium dependent manner [6, 19, 20]. The similar locations of the sugar moieties and the Ca^{+2} ions in 1SL4 and 1TLG unequivocally identifies the most probable site of carbohydrate binding in CLD. The output from HSSP [18] based on 1SL4 and 1FVU structures retrieved respectively 510 and 478 aligned sequences with ≥ 30% sequence identity. The residues with conservation > 80% from both subsets have been used to derive the fold fingerprint, C-X(10,12)-C-Y-X(8)-W-X(6)-C-X(6,8)-L-X(22,28)-W-X-G-X(14)-G-X(28,37)-C-X(7)-C. The following PROSITE search based on this fingerprint revealed polycystin-1 among selected 189 nonredundant structures.

The 1SL4 structure has been chosen as the most reliable structure to model the CLD 3-dimensional fold in detail. Of the PDB sequences examined, the CLD is most similar to the fold of the dendritic cell receptor (Table 1). After threading the CLD sequence through the 1SL4 skeleton, 1000 cycles of the total potential energy optimization were applied to the model. CLD exhibits binding specificity for galactosyl and glucosyl groups in the presence of Ca^{+2} [6]. The positions of Gal and Ca^{+2} in the CLD modeled complex were determined from the superimposed 1TLG crystal structure and locally adjusted to the stereochemical environment in the CLD carbohydrate binding site on a graphics terminal.

In our structural analysis we have used the Pc-1 sequence numbering indicated in Fig. 1. The CLD model has an α/β fold composed of three α-helices (429–438, 450–462, 491–494) and eight β-strands (411–414, 417–422, 443–445, 465–471, 480–481, 508–512, 516–520, 527–533) (Figs. 1 and 3). The structure is stabilized by three conserved S-S bridges 406–419, 436–530, and 508–522 and has a *cis* peptide bond before Pro500. The carbohydrate binding site in the CLD model is located on the surface of the loop region. It consists of a primary binding site with high sugar specificity and binding affinity

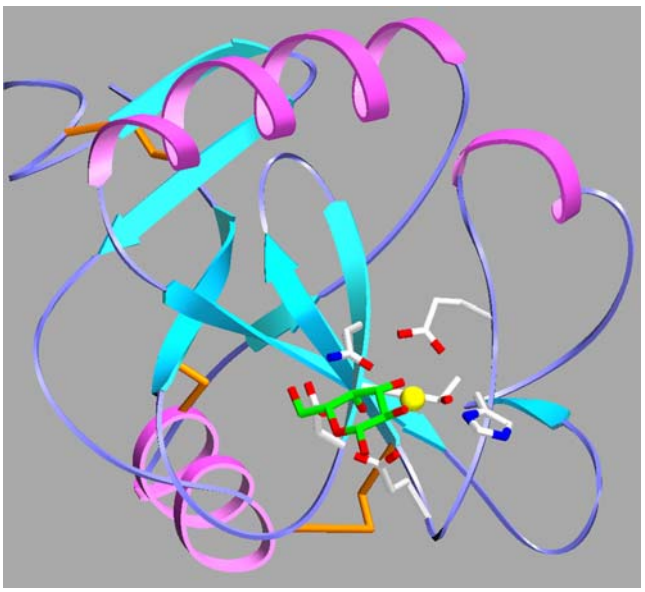

Fig. 3 A SETOR [17] ribbon representation of the 3-D fold of the C-type lectin domain from Pc-1. The α-helices, β-strands and S-S bonds are shown in magenta, blue and orange respectively. The carbohydrate binding subsite with Gal (carbons in green) and Ca^{+2} (in yellow) is provided by six binding shell residues (carbons in white)

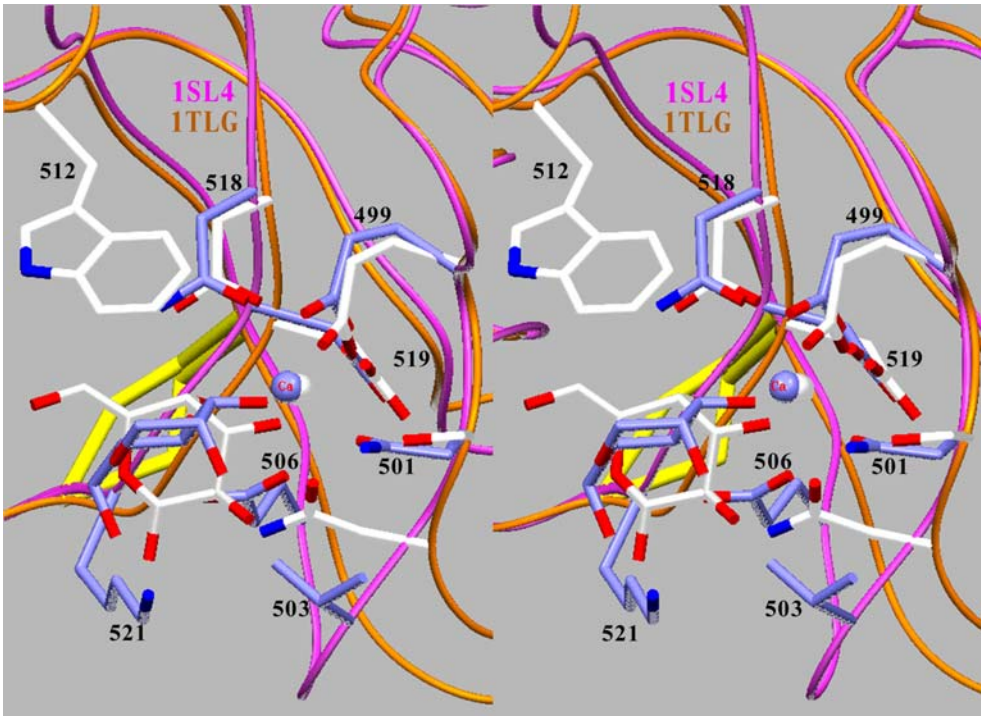

Fig. 4 A SETOR [17] stereo view of the carbohydrate binding sites in the superimposed 1SL4 and 1TLG complexes. Backbones are shown in magenta and orange. Carbon atoms of the ligands and side chains-in violet and white respectively. The positions of the binding residues are labeled with CLD numbering (shown in black) related to Fig. 1

and a few less specific secondary subsites for accommodation of additional sugar moieties from linear or branched oligosaccharides. The proposed CLD sugar binding site is lined by the residues D465, L496, E499, H501, A503, E506, R510, G512, P513, T514, W516, N518, T519, D520, L521, A524, and H526. The shape of the binding site in 1TLG is somewhat different due to conformational differences at sections 503–506 and 512–515.

When the 1SL4 and 1TLG structures are superimposed the Ca^{+2} ions positions are 0.4Å apart and the sugar rings are related to one another by∼90° rotation around the axis joining the C3 and C5 atoms (Fig. 4). The corresponding pairs of hydroxyl groups 3 and 4 occupy similar positions (∼0.7Å and 0.5Å apart, respectively) in the two structures

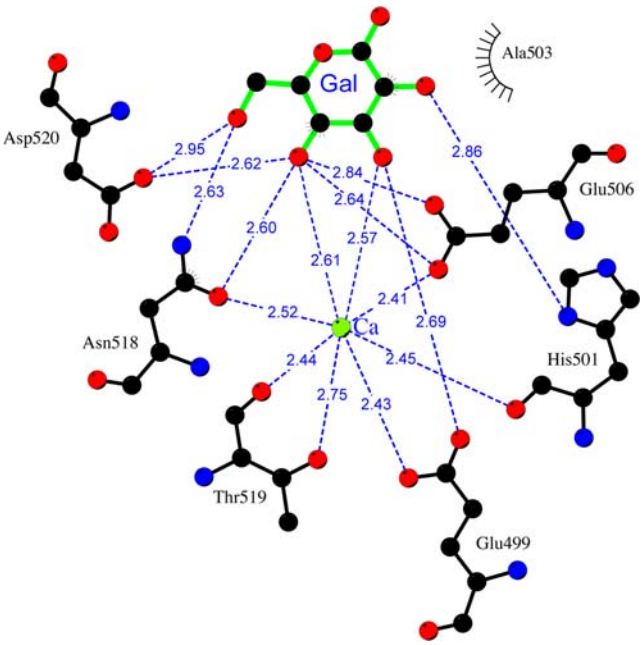

Fig. 5 Schematics (prepared by programs LIGPLOT/HBPLUS [15, 16]) of a 3.9Å binding shell around the Gal sugar coordinated with Ca^{+2} in the primary carbohydrate-binding subsite of the CLD. The bonds are shown by dashed blue lines. The black eyelash shape marks the residue making hydrophobic contacts with Gal

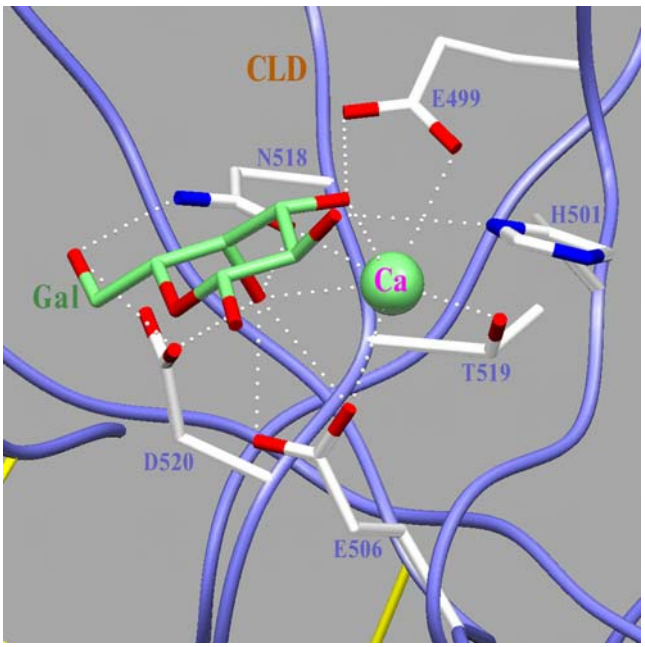

Fig. 6 A view (prepared by program SETOR [17]) of the CLD primary carbohydrate binding subsite with bound Gal and Ca^{+2}

and are involved in specific H-bonding with identical or stereochemically similar side chains at positions 499, 506, and 518 (CLD numbering) (Figs. 1, 4). In 1SL4 and CLD the Asp520 side chain (Val in 1TLG) provides an additional contribution to binding. The residue composition of the primary carbohydrate binding site of CLD more closely resembles that of 1SL4 than 1TLG. In the CLD complex we cannot exclude the possibility of mannosyl oligosaccharides binding with an alternative orientation of the sugar ring and possible participation of adjacent amino acids in binding additional sugars.

A three-dimensional model for the binding of a Gal molecule and a Ca^{+2} ion to CLD is presented in Fig. 6 and a schematic illustration of the distances between the interacting residues of the protein, the calcium ion and the galactose is shown in Fig. 5. In the model six residues (Glu499, His501, Glu506, Asn518, Thr519, and Asp520) form H-bonds with Gal and coordination bonds with the Ca^{+2} ion. The Glu499, His501, Glu506, and Asn518 residues of the 495–507 loop and 516–520 β strand simultaneously coordinate the Ca^{+2} ion and form H-bonds with Gal. The fifth residue, Thr519, coordinates with the calcium and the sixth residue, Asp520, forms a hydrogen bond to Gal.

The range of Ca^{+2} to oxygen distances in our model is consistent with the average geometry observed for Ca^{+2} coordination in protein crystal structures [22, 23].

The identities of four of the six amino acids (Glu499, Glu506, Asn518 and Asp520) that bind the sugar are conserved between CLD and the 1SL4 (Fig. 1). Experimental data suggests that the CLD has specificity for glucosyl and galactosyl groups [6] and does not bind the monosaccharide mannose (Man). The opposite configuration of the OH group at C4 atom in Glc, compared to that in the Gal epimer, has little affect on binding and retains most of the binding contacts at slightly different orientation of the sugar ring. A primary difference between Glc/Gal and Man is a difference in configuration of the C2 hydroxyl of the sugar. A change in the C2 configuration of Gal in the CLD complex model (Fig. 6) would result in disruption of the H-bond with His501 and introduce stereochemical conflict with Ala503. On the other hand, the second possible orientation, observed for the Man moiety in 1SL4, cannot be completely excluded as part of an oligomeric carbohydrate ligand.

It may be worth noting that a hypothetical 129 amino acid protein product could arise from an alternate reading of the sense strand of the CLD coding sequence. A hypothetical protein corresponding to a two residue frame shift reading of the coding strand of the DNA for CLD has 29% sequence identity to a fragment of DNA polymerase from *Thermus aquaticus* of known crystal structures [PDB: 1BGX]. Although the 3D folds of the CLD domain and 1BGX fragment have different connectivity of secondary structural elements, they present similar 3D architecture, a central core built from a four strand β-sheet flanked by three α-helices with a reverse direction chain trace.

Conclusions

The three dimensional structure of the C-lectin binding domain (CLD) of polycystin-1 (Pc-1; sequence region 405–534) complexed with galactose and a Ca^{+2} ion has been modeled (the coordinates are available on request, e-mail: pletnev@hwi.buffalo.edu). The CLD fold is most similar to the fold of the C-type carbohydrate recognition domain in the dendritic cell receptor, DC-SIGN [PDB: 1SL4] and is composed of eight β-strands and three α-helices. Our 3D model reveals the probable three-dimensional fold of the protein, identifies the carbohydrate binding site and illustrates the nature of the Ca^{+2} dependent ligand binding. It supports ligand specificity for mono and oligosaccharides having glucosyl and/or galactosyl groups observed in vitro but does not exclude the possibility of the binding of oligosaccharides with mannosyl moieties. The consistency between empirical evidence that the binding of sugars to the lectin binding domain of Pc-1 is Ca^{+2} dependent and the fact that two PDB structures used to model CLD have sugar molecules bound in a calcium dependent manner in an identical location is strong support for our model and suggests a calcium dependence of sugar binding at this site in vivo. This is particularly so in light of the fact that calcium transport properties of Pc-2 appear to be activated via interaction between Pc-1 and the N terminal helix of Pc-2. It is conceivable that there is a connection between the Ca^{+2} transport of Pc-2 and the Ca^{+2} binding dependence of the CLD domain of Pc-1. The model gives a reliable structural basis for rational design of mutagenic experiments to map functionally important residues of CLD and refine our knowledge of its functional role in polycystin-1.

Acknowledgments This research was supported by the Arrison Foundation. The graphics assistance of Melda Tugas is gratefully acknowledged.

References

1. Igarashi P, Somlo S (2002) J Am Soc Nephrol 13:2384–2398
2. Wilson PD (2001) J am Soc Nephrol 12:834–845
3. Delmas P, Padilla F, Osorio N, Coste B, Raoux M, Crest M (2004) Biochem Biophys Res Comm 322:1374–1383
4. Al-Bhalal L, Akhtor M (2005) Adv Anal Pathol 12:126–133
5. Bycroft M, Bateman A, Clarke J, Hamill SJ, Sandford R, Thomas RL, Chothia C (1999) EMBO J 18:297–305
6. Weston BS, Bagneris C, Price RG, Stirling JL (2001) Biochem Biophys Res Comm 1536:161–176

7. Bernstein FC, Koetzle TF, Williams GJB, Meyer ER, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1997) J Mol Biol 112:535–542
8. Bairoch A, Apweiler R (2000) Nucleic Acids Res 28:45–48
9. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680
10. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) Nucleic Acids Res 31:3784–3788
11. Sack JS (1988) J Mol Graphics 6:224–225
12. Collaborative Computational Project Number 4 (1994) Acta Cryst D Biol Crystallogr 50(Pt5):760–763
13. Brunger AT (1992) *X-PLOR* (version 3.1) Manual, Yale University, New Haven, CT
14. MacKerell AD, Brooks Jr B, Brooks CL, Nilsson L, Roux B, Won Y, Karplus M (1998) In: Schleyer PR et al (ed) The Encyclopedia of Computational Chemistry Vol 1, John Wiley & Sons, Chichester, pp 271–277
15. Wallace AC, Laskowski RA, Thornton JM (1995) Protein Eng 8:127–134
16. McDonald IK, Thornton JM (1994) J Mol Biol 238:777–793
17. Evans SV (1993) J Mol Graphics 11:134–138
18. Sander C, Schneider R (1991) Protein Struct Funct Genet 9:56–68
19. Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K (2004) Nature Struct Mol Biol 11:591–598
20. Poget SF, Legge GB, Proctor MR, Butler PJG, Bycroft M, Williams RL (1999) J Mol Biol 290:867–879
21. Sen U, Vasudevan S, Subbarao G, McClintock RA, Celikel R, Ruggeri ZM, Varughese KI (2001) Biochemistry 40:345–352
22. Yang W, Lee HW, Hellinga H, Yank J (2002) Protein Struct Funct Genet 47:344–356
23. Nayal N, Di Cera E (1994) Proc Nat Acad Sci USA 91:817–821