Structural studies on AIPL1 and its functional interactions with NUB1 to identify key interacting residues in LCA4

S. Muthukumaran · V. Umashankar ·
Meena Revathi Valliappan

Received: 19 February 2013 / Accepted: 27 February 2013 / Published online: 4 April 2013
© Springer Science+Business Media New York 2013

Abstract Leber congenital amaurosis (LCA) is an autosomal recessive disorder that causes visual impairment in children due to fifteen different gene mutations. Of these, mutations in Aryl-Hydrocarbon Receptor Interacting Protein-like 1 (AIPL1) cause the most severe form of LCA (LCA4) leading to the degeneration of photoreceptor cells. NEDD8 Ultimate Buster 1 (NUB1), a protein that regulates cell cycle progression, interacts with AIPL1 to prevent the over expression of NUB1. In the case of over expression, cell cycle progression is disrupted and may lead to LCA. The studies on interactions between these two proteins will aid in identifying potential modulators for this condition. Since no three-dimensional structure is currently available for these two proteins, in this study we predicted the structures of these two proteins by molecular modelling methods. Moreover, we also modelled the three proven significant mutant forms of AIPL1 spanning the tetratricopeptide domain. Finally, both the modelled wild and mutant structures of AIPL1 (A197P, C239R and G262S) were computationally docked to NUB1, so as to map the potential molecular interactions. This is the first study on modelling the structure–function relationship of AIPL1–NUB1 interactions which shall aid in discovery of novel therapeutic agents.

Keywords Leber congenital amaurosis · AIPL1 · Molecular modelling · Protein–protein docking · Molecular dynamics

Abbreviation

LCA	Leber congenital amaurosis
AIPL1	Aryl-hydrocarbon receptor interacting protein-like 1
NUB1	NEDD8 Ultimate Buster 1
SNP	Single nucleotide polymorphism
NVT	Number of moles volume, temperature
NPT	Number of moles pressure, temperature

Introduction

Leber congenital amaurosis (LCA) is an inherited dystrophy that causes vision loss in infants and children [1]. Of the fifteen genes known, AIPL1 is known to cause the most severe form of LCA. It encodes for essential protein required for the proper functioning of the rod photodiesterase enzyme thereby aiding proper vision [2]. The gene locus of AIPL1 is on chromosome 17p13 [3].

Though the precise functional role of AIPL1 within the photoreceptors remains unknown [1], it has been established that they play a major role in the proper functioning and survival of rod photoreceptor cells [4]. AIPL1 interacts with NUB1 (NEDD8 Ultimate Buster 1), a protein involved in regulation of cell cycle progression. AIPL1 may function in the regulation of cell cycle progression through its interaction with NUB1. Mutations in AIPL1 may lead to photoreceptor cell death during development by disrupting the normal regulation of the cell cycle [5]. PDE6 (rod photodiesterase) is required to convert visual light stimuli into electrical signals in photoreceptor cells. Lack of AIPL1 in the retina shows decreased levels of PDE6 alpha

Sivashanmugam Muthukumaran and Meena Revathi Valliappan contributed equally to this work.

S. Muthukumaran · V. Umashankar (✉) · M. R. Valliappan
Centre for Bioinformatics, Vision Research Foundation,
Sankar Nethralaya, 18, College Road,
Chennai 600 006, India
e-mail: drvus@snmail.org

V. Umashankar
e-mail: vumashankar@gmail.com

M. R. Valliappan
A.C. College of Technology, Anna University,
Chennai 600 025, India

and beta enzymes. It has been found that AIPL1 is essential for the stability of the nascent rod PDE6 polypeptides [6].

The paucity of three-dimensional (3D) structures of AIPL1 and NUB1 has not led to further structure–function studies which shall aid the way to modulation LCA4 [3]. One of the major mutations in AIPL1 is W278X results in the total loss of synthesis of the protein as such [7]. Other mutations: R53W, M79T, V96I, A197P, C239R, G262S, and R302L can significantly alter the functions [5]. In the present study, only the mutations present in the three tetratricopeptide (TPR) domains (A197P, C239R and G262S) were analysed for further in silico approach as these regions span the NUB1 binding region. Of these mutations, only A197P did not reduce the interaction with NUB1 significantly, whilst G262S and C239R reduced the binding of AIPL1 to NUB1, inferred through in vitro studies [5]. However, the possible molecular interactions gained and lost were yet to be deciphered as the structural information for these proteins are not available. Hence, to better understand the macromolecular interactions we attempted this study, whereby, the 3D models of AIPL1 and NUB1 were predicted and docked using structural bioinformatics methods.

Materials and methods

Data sets

The amino acid sequences of AIPL1 (Q9NZN9) and NUB1 (Q9Y5A7) were obtained from UniProtKB (<http://www.uniprot.org>). The UniProt Knowledgebase (UniProtKB) is an expertly curated database, a central access point for integrated protein information with cross-references to multiple sources.

Three-dimensional structure prediction of AIPL1 and NUB1

In general, a good homology model method should be able to reproduce experimental results. However, the model may be accurate if the relationship between the model and the template (PDB database) is close, especially when the sequence identity between the templates is more than 40 %. The sequence retrieved from UniProtKB was subjected to BLASTP [8] against PDB database [9]. The result obtained showed templates only with >40 % homology for both the proteins. Hence, we used I-TASSER to predict the structures based on threading method. Here, the 3D models were predicted based on multiple-threading alignments by LOMETS and iterative I-TASSER assembly simulations [10].

Refinement of modelled structures

The best modelled structures of AIPL1 and NUB1 from I-TASSER were then refined by submitting it to ModRefiner

SERVER [11]. ModRefiner is an algorithm for atomic-level, high-resolution protein structure refinement, which can start from the C-alpha trace, main-chain model or full-atomic model. Both side-chain and backbone atoms were completely flexible during structure refinement simulations, where conformational search is guided by a composite of physics- and knowledge-based force field. The models thus obtained were analysed using procheck server [12] and were found to contain residues in the disallowed region in Ramachandran plot. Therefore, these models were loop refined using MODELLER9v7 [13].

Structure validation

To ensure the quality of structure prediction, the refined models were validated by multiple methods: Procheck, to assess the stereochemical quality of the overall and residue-by-residue geometry, PSIPRED to cross-validate the secondary structure elements [14], QMEAN score, which implements six scoring function terms namely, torsion analysis, pairwise comparison, solvation, all atoms secondary structure specific interaction, secondary structure orientation and solvent accessibility [15]. Moreover, the models were also validated using PROQ server which implements LGscore and MaxSub score [16].

Molecular dynamics and energy minimization

The valid structures were further subjected to energy minimization and molecular dynamics simulation. All simulations were performed with the GROMACS simulation package version 4.5.3 [17] under periodic boundary conditions in a 1.0-nm cubic box. In the first step, the charged proteins were neutralised using genion command. Then energy minimization was performed using GROMOS96 43a1 force field with the steepest descent algorithm with 1000 steps. The position restraining steps were performed using the NVT (number of moles volume, temperature) and NPT (number of moles pressure, temperature) commands for 100 ps. The final simulation step was run for 5 ns.

Mutation studies

The post simulation model of AIPL1 Wild Type was used as template to model the mutant types: A197P, C239R and G262S were modelled using Beta SWISS MODEL server [18]. The mutant models were further geometry-optimised using the GROMACS package.

Mutation analysis

Sequence and structure-based deleterious analysis of ALP1 mutations were performed using SIFT [19], POLYPHEN

[20], I-mutant [21], PopMuSiC [22] and Eris Modelling suite [23], respectively.

Docking studies

Protein–protein docking of modelled AIPL1 (WT and MT) and NUB1 were performed using Hex Protein Docking Server [16]. The complete details of the inter-molecular interactions were analysed using PDBsum [24] and Protein Interactions Calculator (PIC) [25].

Results

Data sets

AIPL1 gene expression is limited to photoreceptors and the pineal gland in humans [26], and it encodes a protein of 384 amino acids in length. The precise functional role of AIPL1 within the photoreceptors still remains unknown. The protein sequence contains three TPR motifs of 34 amino acids each and is thought to serve as an interfaces for inter-molecular interactions. TPR motifs are found in a variety of proteins that mediate of functions, such as protein trafficking or protein folding. TPR motifs are generally present in proteins that form complexes with several other proteins [27]. In addition to the TPR domains, the protein has a 56 amino acids proline-rich region at the carboxyl-terminus in

humans [28]. The proline-rich region is from 328 to 384. AIPL1 may be important in protein trafficking and/or protein folding and stabilisation [5]. Its sub cellular location is in the cytoplasm and nucleus. TPR regions span the amino acid regions: 178 to 211, 230 to 263 and 264 to 296. Sequence conservation is highest across the three AIPL1 TPR motifs and extends across the protein, except for the proline-rich amino acid sequence present only at the C-terminus of primate AIPL1 [28]. NUB1 is a 615-amino-acid-long protein and localised predominantly in nucleus. It is a specific down-regulator of the NEDD8 conjugation system. It recruits NEDD8, UBD and their conjugates to the proteasome for degradation. Hence, understanding AIPL1–NUB1 interactions is crucial for therapeutic interventions.

Three-dimensional structure prediction of AIPL1 and NUB1 and refinement

The sequence retrieved from UniProtKB was subjected to BLASTP [8] against PDB database [9]. The result obtained was less than 40 %, as shown in Fig 1a and b. I-TASSER predicted five models for the proteins studied. Of these, the models with the least C-score [10] were chosen for further studies. The best model for AIPL1 had a least score of −1.15 and NUB1 had −1.75. The Ramachandran plot for AIPL1 showed seven residues in the disallowed region and 85.7 % in the favourable region. NUB1 model showed 14 residues in the disallowed region and 85.4 % in the favourable

Fig. 1 BLASTP Result of **a** AIPL1 (above) and **b** NUB1 (below) against PDB database

1.BLAST RESULT OF a.AIPL1(above) and b.NUB1(below)

The best model (4534th of 5000 models) of AIPL1 after simulation.
Potential energy: (-1.75858e⁺⁰⁵ KJ/mol)

Fig. 2 3D Structure of AIPL1: the secondary structure of AIPL1 is shown as a coloured cartoon: helix in red, sheet in yellow, and coil in green using PYMOL

region. Further, these models were error fixed using Modrefiner and Modeller9v7.

Structure validation

The final refined models of AIPL1 and NUB1 had no residues in the disallowed region with 90.4 and 90.1 % in the favourable regions, respectively. Further, these models were cross-validated for secondary structure elements formation with that of the sequence-based prediction using PSIPRED. QMEAN server provided the significant Qmean score of 0.582 and 0.316 for AIPL1 and NUB1, respectively. Moreover, PROQ server also predicted the structures to be highly plausible, with the LGscore and MaxSub score for AIPL1 model to be 2.192 (good) and 0.315 (correct), respectively. In the case of NUB1 model, LGscore and MaxSub score were 1.195 (good) and 0.074 (correct), respectively. Hence, these best refined models were further used for molecular dynamics simulation.

Molecular dynamics simulation of AIPL1 and NUB1

The predicted structure of AIPL1 WT was subjected to molecular dynamics (MD) Simulation in solvated conditions (40896 SPC216 water molecules). The protein showed a negative charge of $-1.2e$, so to neutralise the solvated system, one molecule of sodium cation (Na^+) was added. Further, the protein was energy-minimised using GROMOS96 43a1 force field with the steepest descent algorithm with 1,000 steps. The position restraining steps were performed using NVT and NPT command for 100 ps under the pressure and temperature conditions of 1 bar and 300 K, respectively. The final energy minimization run was performed 5000 ps. The model with the lowest potential energy ($-1.75858e^{+05}$ KJ/mol) is 4,534th of the 5000 models generated shown in Fig. 2, was taken as the best model for further mutation analysis and docking studies. The root mean square deviation graph (RMSD) graph infers the biologically stable conformation of the structure (Fig. 3a). The root mean square fluctuation graph infers high flexibility at the proline-rich region (328 to 84) than the TPR domains (178–296) (Fig. 3b). In the solvent accessibility surface area (SASA) graph, the overall accessible area is reduced from 151.355 nm² in initial step to 132.5 nm² in final step Fig. 3c.

The predicted structure of NUB1 was subjected to molecular dynamics simulation in solvated conditions (83738 SPC216 water molecules). As the protein had a negative charge of $-12.000e$, 12 molecules of Na^+ were added to neutralise the systems. The MD steps followed for the AIPL1 protein was also adopted for NUB1. The final simulation step was run for 5,000 ps. The model with the lowest potential energy ($-3.58718e^{+06}$ KJ/mol) (4610th of the 5000 models) generated shown in Fig. 4a and was chosen as the best model for further docking studies. The RMSD graph showed stable conformation of the protein in simulated

Fig. 3 Molecular dynamics of AIPL1 for 5 ns. **a** Root mean square deviation analysis, **b** root mean square fluctuation and **c** solvent accessible surface analysis for 5 ns

Fig. 4 **a** 3D structure of NUB1: the secondary structure of NUB1 is shown as coloured cartoon: Helix in red, sheet in yellow, and coil in green using PYMOL. **b** Molecular dynamics of NUB1 for 5 ns. **b** Root mean square deviation analysis and **c** solvent accessible surface analysis for 5 ns

conditions Fig. 4b. In the SASA graph, the area is reduced from 239.968 nm² in initial step to 208.281 nm² in final step shown in Fig. 4c.

Mutation studies

The final refined model of AIPL1 WT obtained after simulation was submitted to the advanced option of Beta SWISS MODEL as a template for modelling the mutants. All the three mutant structures (A197P, C239R and G262S) studied were modelled similarly. Furthermore, all the mutant were models optimised using the GROMACS 4.3.5. The potential energy of A197P was reduced from $-8.29958e^{+03}$ to $-5.660033e^{+04}$ KJ/mol after energy minimization. For C239R, it was reduced from $-6.65945e^{+03}$ to $-5.3687070e^{+04}$ KJ/mol and for G262S it was reduced from $-8.36898e^{+03}$ KJ/mol to $-5.4928875e^{+04}$ KJ/mol.

Sequence- and structure-based mutation analysis

The result for mutation analysis using SIFT, PolyPhen 2, and I-mutant web server shown in Table 1. The energy-minimised mutated models of AIPL1 were submitted in the Eris and PopMuSiC web server. The $\Delta\Delta G$ value of >0 of PopMuSiC server denotes the mutation to be destabilising (Table 2).

Table 1 Sequence-based mutation analysis of AIPL1

Mutation	SIFT (tolerance)	PolyPhen-2.0 (stability)	I-mutant (damaging)
A197P	Intolerable	Probably	Increases
C239R	Intolerable	Probably	Decreases
G262S	Tolerable	Possibly	Decreases

Docking studies

The predicted models of AIPL1 and NUB1 were docked using Hex Protein Docking Server. The best docked complexes were analysed using PDBsum, which showed 52 residues of AIPL1 to be interacting with 55 residues of NUB1 (Fig. 5a, b). In the case of NUB1, the residues ranging from 347 to 601 at the C-terminal were found to be interacting with AIPL1. The three TPR domains of AIPL1, TPR2 and TPR3 were found to confer major interactions with NUB1. Moreover, the complex was also analysed using PIC server which recognises various kinds of interactions such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic–aromatic interactions, aromatic–sulphur interactions and cation–pi interactions between the proteins in the complex as (shown in Table 3). In silico alanine scanning [29] was performed through replacing the significant residues of AIPL1 by alanine with Modeller9v7 and the replaced models were subsequently docked with NUB1 by HEX docking server. This infers the potential of replaced residues in terms of AIPL1–NUB1 complex.

Discussion

Though the interaction between AIPL1 and NUB1 is experimentally proven, the information on the interacting residues

Table 2 Structure-based mutation analysis of AIPL1

Mutations	PopMuSiC ($\Delta\Delta G$ kcal/mol)	Eris ($\Delta\Delta G$ kcal/mol)
A197P	3.04	>10
C239R	1.60	>10
G262S	0.38	-0.09

Fig. 5 **a** Interaction between AIPL1 and NUB1 from PDBsum server. Chain A represents AIPL1 and chain B indicates NUB1. **b** Docking complex of AIPL1 interaction with NUB1. Illustration of docked complex for AIPL1 and NUB1: the AIPL1 region is represented by *red* and *green*, and the NUB1 region represent for *blue* and *yellow* using PYMOL

is yet to be revealed. By this study, we attempted to perform a detailed analysis of AIPL1–NUB1 interactions through computational approaches.

The plausible 3D structures of AIPL1 (WT and MT) and NUB1 were predicted and validated using molecular modeling methods. Furthermore, the impact of the reported pathogenic mutations [5] in AIPL1 on AIPL1–NUB1 interactions was also studied in detail. Previous in vitro studies have shown that A197P mutation does not affect its interaction with NUB1. However, C239R and G262S mutations were shown to affect the NUB1 binding. Interestingly, the present in silico studies also fall in line with documented experimental studies. The residues of AIPL1 contributing to most of the interactions with NUB1 were: Ser 199, Lys 243, Tyr 247, Glu 248, Lys 265, Lys 289, Glu 292 and Glu 294. These residues were also found to be highly significant in our in silico alanine scanning studies.

Therefore, it can be hypothesised that any mutations in the said residues shall affect the interactions with NUB1. The insight from this study shall lead ways to design of novel

Table 3 Analysis of the docked complexes using PIC web server

Interactions	Wild AIPL1–NUB1	A197P–NUB1	C239R–NUB1	G262S–NUB1
Hydrophobic	11	16	5	8
Main chain–main chain hydrogen bond	0	1	2	2
Main chain–side chain hydrogen bond	65	51	19	30
Side chain–side chain hydrogen bond	47	45	27	29
Cation–pi bond	1	3	2	2
Ionic bond	19	8	4	4

therapeutic peptides which mimic the AIPL1 interacting interface of NUB1 which shall potentially inhibit AIPL1–NUB1 interactions, thereby modulating LCA4.

References

- Kaplan J, Bonneau D, Frézal J, Munnich A, Dufier JL. Clinical and genetic heterogeneity in retinitis pigmentosa. *Hum Genet.* 1990;85(6):635–42. PubMed PMID: 2227956.
- Makino CL, Wen XH, Michaud N, Peshenko IV, Pawlyk B, Brush RS, et al. Effects of low AIPL1 expression on phototransduction in rods. *Invest Ophthalmol Vis Sci.* 2006;47(5):2185–94. Erratum in: *Invest Ophthalmol Vis Sci.* 2006 Jun;47(6):2279. PubMed PMID: 16639031.
- Sohocki MM, Perrault I, Leroy BP, Payne AM, Dharmaraj S, Bhattacharya SS, et al. Prevalence of AIPL1 mutations in inherited retinal degenerative disease. *Mol Genet Metab.* 2000;70(2):142–50.
- Kirschman LT, Kolandaivelu S, Frederick JM, Dang L, Goldberg AF, Baehr W, et al. The Leber congenital amaurosis protein, AIPL1, is needed for the viability and functioning of cone photoreceptor cells. *Hum Mol Genet.* 2010;19(6):1076–87. Epub 2009 Dec 30. PubMed PMID: 20042464; PubMed Central PMCID: PMC2830831.
- Akey DT, Zhu X, Dyer M, Li A, Sorensen A, Blackshaw S, et al. The inherited blindness associated protein AIPL1 interacts with the cell cycle regulator protein NUB1. *Hum Mol Genet.* 2002;11(22):2723–33. Erratum in: *Hum Mol Genet.* 2003 Feb 15; 12(4):451. PubMed PMID: 12374762; PubMed Central PMCID: PMC2585502.
- Kolandaivelu S, Huang J, Hurley JB, Ramamurthy V. AIPL1, a protein associated with childhood blindness, interacts with alpha-subunit of rod phosphodiesterase (PDE6) and is essential for its proper assembly. *J Biol Chem.* 2009;284(45):30853–61. Epub 2009 Sep 16. PubMed PMID: 19758987; PubMed Central PMCID: PMC2781484.
- Jacobson SG, Cideciyan AV, Aleman TS, Sumaroka A, Roman AJ, Swider M, et al. Human retinal disease from AIPL1 gene mutations: foveal cone loss with minimal macular photoreceptors and rod function remaining. *Invest Ophthalmol Vis Sci.* 2011;52(1):70–9. Print 2011 Jan. PubMed PMID: 20702822.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
- Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, et al. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D: Biol Crystallogr.* 1998;54(Pt 6 Pt 1):1078–84.
- Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinforma.* 2008;9:40. PubMed PMID: 18215316; PubMed Central PMCID: PMC2245901.
- Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J.* 2011;101(10):2525–34. Epub 2011 Nov 15. PubMed PMID: 22098752; PubMed Central PMCID: PMC3218324.
- Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol.* 1993;231(4):1049–67.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci.* 2007 Nov;Chapter 2:Unit 2.9. PubMed PMID: 18429317.
- McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16(4):404–5.
- Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 2009 Jul;37(Web Server issue):W510–4. Epub 2009 May 8. PubMed PMID: 19429685; PubMed Central PMCID: PMC2703985.
- Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci.* 2003;12(5):1073–86. PubMed PMID: 12717029; PubMed Central PMCID: PMC2323877.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26(16):1701–18.
- <http://beta.swissmodel.expasy.org>
- <http://sift.bii.a-star.edu.sg/>
- <http://genetics.bwh.harvard.edu/pph2/>
- <http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant/>
- <http://babylone.ulb.ac.be/popmusic/>
- <http://troll.med.unc.edu/eris/login.php>
- <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>
- <http://pic.mbu.iisc.ernet.in/>
- Hendrickson A, Bumsted-O'Brien K, Natoli R, Ramamurthy V, Possin D, Provis J. Rod photoreceptor differentiation in fetal and infant human retina. *Exp Eye Res.* 2008;87(5):415–26. Epub 2008 Aug 20. PubMed PMID: 18778702.
- Blatch GL, Lässle M. The tetratricopeptide repeat: a structural motif mediating protein–protein interactions. *Bioessays.* 1999;21(11):932–9. Review. PubMed PMID: 10517866.
- Sohocki MM, Sullivan LS, Tirpak DL, Daiger SP. Comparative analysis of aryl-hydrocarbon receptor interacting protein-like 1 (Aipl1), a gene associated with inherited retinal disease in humans. *Mamm Genome.* 2001;12(7):566–8. PubMed PMID: 11420621; PubMed Central PMCID: PMC2581445.
- Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein protein interfaces. *Sci STKE.* 2004; 2004(219):pl2. PubMed PMID: 14872095.