A Computational Approach From Gene to Structure Analysis of the Human ABCA4 Transporter Involved in Genetic Retinal Diseases

Alfonso Trezza,[1] Andrea Bernini,[1] Andrea Langella,[1] David B. Ascher,[2] Douglas E. V. Pires,[3] Andrea Sodi,[4] Ilaria Passerini,[5] Elisabetta Pelo,[5] Stanislao Rizzo,[4] Neri Niccolai,[1] and Ottavia Spiga[1]

[1]Department of Biotechnology Chemistry and Pharmacy, University of Siena, Siena, Italy
[2]Department of Biochemistry and Molecular Biology, University of Melbourne, Bio21 Institute, Parkville, Australia
[3]Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil
[4]Department of Surgery and Translational Medicine, Eye Clinic, Careggi Teaching Hospital, Florence, Italy
[5]Department of Laboratory Diagnosis, Genetic Diagnosis Service, Careggi Teaching Hospital, Florence, Italy

Correspondence: Ottavia Spiga, Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Via A. Moro 2, 53100, Siena, Italy;
ottavia.spiga@unisi.it.

Submitted: May 2, 2017
Accepted: September 14, 2017

Citation: Trezza A, Bernini A, Langella A, et al. A computational approach from gene to structure analysis of the human ABCA4 transporter involved in genetic retinal diseases. *Invest Ophthalmol Vis Sci*. 2017;58:5320–5328. DOI:10.1167/iovs.17-22158

Purpose. The aim of this article is to report the investigation of the structural features of ABCA4, a protein associated with a genetic retinal disease. A new database collecting knowledge of ABCA4 structure may facilitate predictions about the possible functional consequences of gene mutations observed in clinical practice.

Methods. In order to correlate structural and functional effects of the observed mutations, the structure of mouse P-glycoprotein was used as a template for homology modeling. The obtained structural information and genetic data are the basis of our relational database (ABCA4Database).

Results. Sequence variability among all ABCA4-deposited entries was calculated and reported as Shannon entropy score at the residue level. The three-dimensional model of ABCA4 structure was used to locate the spatial distribution of the observed variable regions. Our predictions from structural in silico tools were able to accurately link the functional effects of mutations to phenotype. The development of the ABCA4Database gathers all the available genetic and structural information, yielding a global view of the molecular basis of some retinal diseases.

Conclusions. ABCA4 modeled structure provides a molecular basis on which to analyze protein sequence mutations related to genetic retinal disease in order to predict the risk of retinal disease across all possible ABCA4 mutations. Additionally, our ABCA4 predicted structure is a good starting point for the creation of a new data analysis model, appropriate for precision medicine, in order to develop a deeper knowledge network of the disease and to improve the management of patients.

Keywords: ABCA4 protein, ABC transporter, mutations database

ABCA4, also known as ABCR or the Rim protein, is a retinal-specific member of the ABCA subfamily of adenosine triphosphate (ATP)-binding cassette (ABC) transporters. ABC transporters constitute one of the largest protein superfamilies.[1] These evolutionary, highly conserved transmembrane proteins use the energy of ATP hydrolysis to translocate a broad cluster of molecules across cell membrane, functioning as either importers or exporters. ABCA4 is predominantly expressed in vertebrate rod and cone photoreceptor cells, and it is mainly localized in outer segment disk membranes[2]; the protein is also marginally present in the brain.[3–5]

In recent years, significant progress has been made in our understanding of roles and mechanisms for ABCA4 activity. Several biochemical studies have been carried out to determine its natural substrate, now accepted to be all-*trans*-retinal and *N*-retinylidene-phosphatidylethanolamine (*N*-retinylidene-PE), a conjugate of all-*trans*-retinal and phosphatidylethanolamine).[6–12]

Photoreceptor-specific expression of ABCA4 and its localization[6] confirms the suggestion that ABCA4 may transport a substrate that is critical for photoreceptor function.[5,12,13] Analysis of ABCA4 knockout mice,[9,14–17] together with biochemical studies, has led to a conceptual model for the role of ABCA4 as a retinoid transporter in the visual cycle. This suggests that ABCA4 may act as an ATP-dependent flippase that translocates *N*-retinylidene-PE, promoting the recycling of all-*trans*-retinal released from photobleached rhodopsin via the retinoid cycle. Compromised ABCA4 transporter activity may lead to the accumulation of toxic all-*trans*-retinal derivatives in rods and cones, leading to apoptosis of the supporting retinal pigment epithelium cells and, ultimately, degeneration of photoreceptors.[7,14,15]

ABCA4 was the first ABCA-transporter that has been linked to various genetic retinal diseases, such as (Stargardt disease, cone dystrophy, and retinitis pigmentosa).[18–23] Some ABCA4 sequence variants also have been considered as a risk factor for the development of AMD,[24,25] a frequent progressive degener-

ation of central retina, with a high prevalence in elderly people.[26,27]

Stargardt disease (STGD1) is the most common inherited juvenile macular degeneration and is associated with a reduction of central visual acuity.[28] On funduscopy, it is characterized by macular atrophy, often surrounded by typical fishtail yellowish deposits at the posterior pole and in midperiphery (flecks).[29] Cone dystrophy is characterized by the degeneration of cone photoreceptors with photophobia, reduced visual acuity, and abnormal color vision.[30,31] Fundus appearance may sometimes be normal, but it may often show a macular dystrophy. The main feature of retinitis pigmentosa is progressive photoreceptor atrophy (first rods and then cones), with night blindness and visual acuity and visual field loss. In most cases, patients show typical pigmented deposits on fundus examination.[32,33] The association of ABCA4 with AMD is more controversial, as some studies report higher frequency of ABCA4 mutations for AMD in comparison with a control population,[24,25] even though these results have not been confirmed by other investigations.[34]

Human ABCA4 is a large single polypeptide of 2273 amino acids organized as two structurally related tandem-arranged halves.[1] Each half consists of three structural domains, where the first is a transmembrane domain (TMD) comprising six transmembrane α-helices (TMs), two single TMs are at positions 22–42 (TM1) and 1377–1397 (TM7), delineating the beginning of the two halves, respectively. The second hydrophobic region is formed by the other five TMs, that is, TM2-6 at positions 653–899 and TM8-12 at positions 1670–1964. The second domain is a lumen domain, containing a large exocytoplasmic domain (ECD), that is, ECD1 50–645 and ECD2 1442–1644 positioned between TM1 and TM2–6 and TM7 and TM8–12.[35] The third domain, located on the cyroplasmic side of disk membrane at positions 958–1176 and 1934–2147, respectively,[36] is a large soluble-protein region containing the nucleotide-binding domain (NBD), responsible for ATP hydrolysis. TMDs are homologous only with a small subfamily of ABC transporters, containing the drug-binding sites and forming the structural system for translocation.[12,35,36] ECDs do not show a significant sequence similarity to known proteins, except for other ABCA4 sequences from different species. They are glycosylated[37] and are probably involved in the interactions with other proteins.[36] NBDs are the only regions of ABC transporters that are structurally highly conserved.[12] Based on functional and mutagenesis studies, NBD1 is responsible for basal ATPase activity, whereas NBD2 produces the retinal-stimulated increase in activity.[38,39] Suggested mechanism of transport proposed is based on the alternating access model, established for smaller ABC transporters.[40] Furthermore, the observation that ABCA4 translocates the substrate from the luminal to the cytoplasmic side of the rod outer segment (ROS) disk allowed us to consider ABCA4 an importer, which makes this protein unique among known eukaryotic ABC transporters.[41] In importers, the alternating access model suggests a high-affinity binding site located across the membrane from NBDs. NBD binding probably increases ATP affinity, favoring their interaction and a conformational transition within the TMDs leading to closure of the high-affinity substrate binding site and translocation of the substrate to the low-affinity site located on the cytoplasmic side of the membrane.[42]

Although biochemical aspects of ABCA4 have been well characterized, many questions remain unanswered, such as (1) NBDs' communication with TMDs, (2) TMDs' drug specificity, or (3) ECDs' functional and structural role. To address these questions and correlate the amino-acidic sequence mutations to functionality and phenotype of patients, we present here two important resources. The first of these is the development of a new relational database (ABCA4Database) including all available genetic information and identification of well-defined data about the protein, including variability, homology models, and amino acid residue properties.[40,41] In this work, by performing sequence comparison of human ABCA4 to other ABC family members, a 3D model of ABCA4 protein has been obtained. The x-ray 3D structure of mouse P-gp was considered a very suitable template for human ABCA4 modeling. Thus, the resulting structure is proposed as a rational basis on which to correlate missense mutations related to genetic retinal disease with the functional phenotypes observed in patients. At present, a clear correlation between ABCA4 genotype and retinal phenotypes has not been well established,[42–46] and only one model based on a paper from Maugeri et al.[47] has been proposed. Some authors report higher frequency of severe retinal phenotypes in patients carrying null ABCA4 variants,[48,49] and some specific genotypes have been associated with peculiar clinical features.[50] Despite substantial progress in determining the causal genetic variations, some ambiguities are still present because in clinically confirmed ABCA4 disease no mutations are found, suggesting that a significant revision of diagnostic screening and assessment of ABCA4 variations is needed in retinal diseases etiology.[51]

In the present work, by focusing on the structural distribution of ABCA4 mutations and, therefore, their effects on the protein environment, we aimed to contribute to a deeper understanding of genotype/phenotype correlation in genetic retinal diseases.

Materials and Methods

Reconstruction of the Phylogenetic Tree and Residue Conservation

ABCA4 gene and amino acid sequence analyses were performed by using the following software packages and Web-accessible programs: NCBI Blast,[52] MUSCLE,[53] COBALT,[54] and Scorecons.[55,56] In order to reconstruct a phylogenetic tree of ABCA4, we searched for homologs using PSI-BLAST.[52] The resulting 100 sequences were aligned with MUSCLE,[53] and this alignment was fed to COBALT[54] to build a complete phylogenetic tree. The evaluation of residue conservation among ABCA4 sequences was carried out on the basis of a subset of nine whole protein sequences aligned with MUSCLE.[53] Residue conservation was scored by using the Scorecons[55] server. Sequence variability, s_v, where $s_v = 1 - Ss$ and Ss is the normalized Shannon's sequence entropy score given by Scorecons, was calculated for the entire ABCA4 human sequence. The scoring method is based on a classification of residues into one of seven types, a convention that follows that of Mirny and Shakhnovich.[57]

Model Building and Validation

Chain A from the crystal structure of mouse *P-gp63* (PDB ID: 3G60) was used as the main template for the homology model. The UniProt database (an open-source database by the UniPort Consortium, http://www.uniprot.org, in the public domain) sequences of mouse *P-gp* (P06795) and human ABCA4 (P78363) were used for alignment. The alignment was obtained with the Clustal X tool[58] using the default parameters with the blosum62 substitution matrix.[59] The percentage of similarity (45%) between the two sequences makes this template a potential candidate to elaborate a 3D model of ABCA4 by comparative modeling. Each ABCA4 half-transporter was modeled separately using DeepView/Swiss-PdbViewer v. 3.7 (SPDV).[60] Additional loop regions of ABCA4 were

reconstructed from the Loops in Proteins database. The subunit linker was not modeled. Instead, ECDs were modeled using the threading bioinformatic approach with Phyre v.2.[61] To each ABCA4, half the corresponding ECD subunit was linked and then the half-transporter models were reconstructed as a dimer to reproduce the ABCA4 subunit interface, refined to remove steric clashes upon energy minimization. All the images were prepared with PyMOL.[62]

The backbone root-mean-square deviation (RMSD) between P-gp (3G61)[63] and the ABCA4 model was 2.5 nm. Due to the lack of a template structure, the RMSD could not be calculated for residues 50–645, 900–957, and 1442–1644. The quality of the modeled domains was assessed using local factors such as packing quality, backbone conformation, bond length, and side chain planarity. The Protein Model Portal[64] showed that the ABCA4 model was of comparable quality to P-gp, while the analysis of the Ramachandran plot in Swiss-PdbViewer showed that over 90% of the residues fall within favorable regions, increasing our confidence in the backbone geometry. The obtained ABCA4 model structure was fitted by using Chimera software v.1.10.160 into the EMD ID 5497 with a reported resolution of 18 Å.[65]

Predicting Protein Structure Perturbation

Three computational methods were used to predict and analyze the effects of over 370 missense mutations on a structural basis: mCSM,[66] Site-Directed Mutator,[67] and DUET.[68] The effect of the mutations was assessed in the context of the molecular interactions of the wild-type residues, and those programs were used to predict the effects of the mutations on protein thermal stability.

ABCA4Database

We manually curated a comprehensive table of germline and somatic ABCA4 mutations obtained from numerous sources, including original articles, the retinal disease related to ABCA4 from different public sources as Retina International Database,[69] the Human Gene Mutation Database,[70] and LOVD ABCA4 Database.[71] Details of mutations not included in these sources were obtained from the original reference by Department of Laboratory Diagnosis, Genetic Diagnosis Service, Careggi Teaching Hospital. For each missense mutation, the new database shows different features: phenotype of retinal disease, sequence entropy/variability, structural localization, amino-acidic accessibility, structural stability, ligand binding, genetic correlations, references, etc. MySQL falls into the RDBMS family, a synonymous term for Relational DataBase Management System. In particular, it is entirely open source, and it has been the world's most widely used RDBMS, open-source and client-server model RDBMS. SQL stands for Structured Query Language. MySQL is frequently chosen for use in web applications and is a central component of the widely used LAMP open-source web application software stack (LAMP, a synonym for Linux, Perl/PHP/Python).

Results

We first constructed, with the available experimental and phenotypic data, an inclusive database of ABCA4 missense mutations (ABCA4Database, http://www.sbl.unisi.it/abca4/abcr_mainlist.php, in the public domain). In order to identify the most important residues for ABCA4 structure and function, we have investigated protein residue conservation among a subset of homolog ABC proteins in several vertebrates. Understanding the most conserved regions of the protein would suggest the role of each mutation for the various retinal diseases. The first step was to search for ABCA4 homologs using PSI-BLAST.[52] The 100 obtained sequences were selected to perform multiple alignments and to reconstruct a phylogenetic tree (Fig. 1) using the COBALT tool.[54]

It is apparent that the sequences that are closest to human ABCA4 are the nine ABCA4 proteins belonging to several primates. We decided to focus on these sequences and to analyze the residue conservation using Shannon's information theoretical entropy.[55,56] Accordingly, we have assessed the level of sequence conservation related to each position in the alignment, the similarity, identity, and entropy per position. Shannon entropy value (H) reflects the variability of every residue between compared sequences, and consequently, their conservation. The returned scores are normalized in order to have conserved columns with a score of 1 and diverse columns with a score of 0. For our analysis, these values, ranging from 0 to 1, were inverted, with 0 for invariant and 1 for hypervariable protein sequence positions, respectively. Thus, we have analyzed the entire ABCA4 sequence to gain information on the presence of critical positions along the polypeptide chain. The obtained entropy plot allowed us to have an idea of the amount of variability at each residue position within the alignment, having a measure of the lack of predictability for an alignment position. After that, sequence variability, s_v, where $s_v = 1 - S_s$ and S_s is the normalized Shannon's sequence entropy score, was calculated, and results for each ABCA4 domain are shown in Figure 2.

ECD1 and ECD2 contain, respectively, 145 (24%) and 105 (32%) residues having $s_v = 0$. These values increase to 51 (40%) and 63 (50%) for TMD1 and TMD2, respectively, and to 124 (53%) and 120 (51%) for NBD1 and NBD2, respectively. The values are up to 304 (49%) and 178 (54%) for ECD1 and ECD2, to 75 (62%) and 84 (70%) for TMD1 and TMD2, and to 158 (69%) and 172 (75%) for NBD1 and NBD2 for s_v smaller than 0.2. This finding shows that, although conserved residues exist in all six domains, it is evident that the TMDs and NBDs are the most highly conserved. We decided to analyze the distribution of sequence entropy by using the quartiles method, and in particular, we considered the value of the third quartile (Q_3). Q_3 values were obtained for the entire sequence distribution and for each individual domain, as reported in Figure 3.

The Q_3 overall sequence value was 0.29, consistent with the singular regions of Figure 3, while for some parts it is much smaller, as in the case of TM2, TM5, and TM10 helices with 0.05, 0.02, and 0.01 Q3 values, respectively, and the two NBDs that have Q_3 values of 0.17 and 0.13.

The analysis of sequence entropy within each domain allows us to understand which are the regions of ABCA4 most susceptible to the action of a single mutation, and considering the incidence of missense variants, we can obtain the maximum for the position with low-entropy values, that is, for highly conserved residues. Consequently, the incidence of mutations appears to have a downward trend in correspondence with increasing entropy values, demonstrating that most disease mutations were located at conserved residue positions.

To understand the spatial agreement between the sequence and the structure of a protein and to explore the influence of mutations associated with retinal disease on the structure and functionality of ABCA4 protein, a homology model was generated (see Materials and Methods). To obtain an accurate model of ABCA4 we separately considered the three domains of each half of the protein. The TMDs and NBDs were modeled using as a template the structure of mouse P-gp (PDB ID code 3G60 chain A), a member of the ABC superfamily.[65] The structure of P-gp represents a nucleotide-free inward-facing conformation arranged as two "halves." Since an important aspect of homology modeling is the sequence alignment

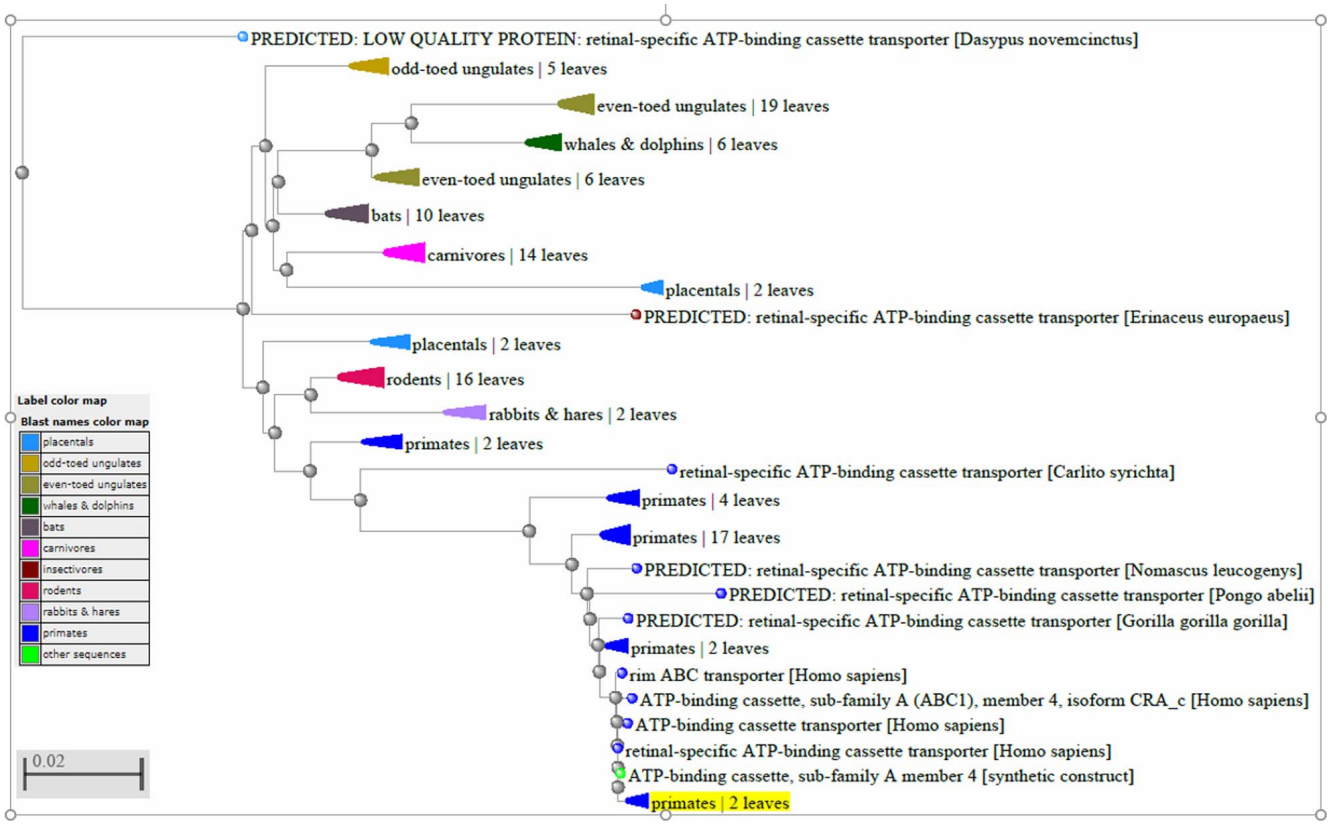

Figure 1. The phylogenetic tree was obtained from multiple alignment of 100 homology sequences of human ABCA4 by using as parameters: tree method COBALT Tree; Max Seq Difference 0.85; Distance Grishin (protein); Sequence Label Taxonomic Name (if available); Collapse Mode Blast Name. *Arrows* represent a branch of collapsed multisequences, and *dots* represent single sequences.

between the template and target proteins,[72] we highlight the reasonable degree of sequence identity between the two proteins (35% for NBDs and 30% for TMDs). The mouse P-gp could not be used as a template for ECDs because suitable sequence alignment with ABCA4 could not be found. Thus, a different ABCA4 homology modeling approach for this region was adopted.[73,74] A full atom model for NBDs and TMDs was generated by manual alignment between the secondary structure elements predicted by PsiPred runs[75] and those observed in the template (3G60), enhancing the correspondence of identical and positively conserved residues of about 40% and 70% for NBDs and 35% and 60% for TMDs. On the other hand, the ECDs were modeled via the threading bioinformatic approach[73,74] with Phyre v. 2.0.[61] The geometrical and stereochemical features were calculated for validation by the Protein Model Portal validation system.[64] The results show that more than 90% of the torsion angles were within the expected Ramachandran regions, all bond distances and angles

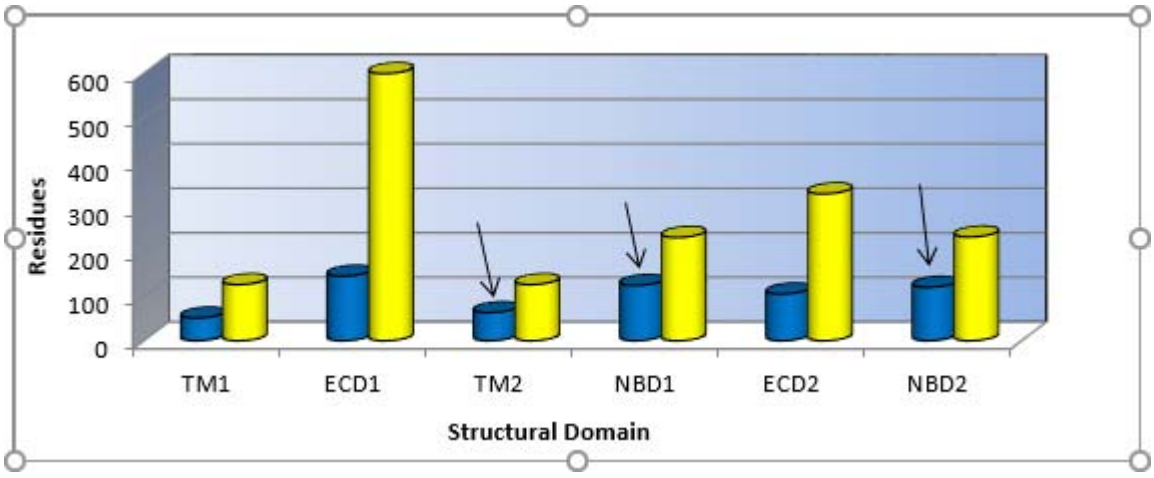

Figure 2. Histogram of residues for each structural domain. *Blue* bins represent the number of amino acids with S_{v0} variability equal to 0, *yellow* bins refer to the amino acid dimension of each domain. *Arrows* highlight the structural domains with a total number of conserved positions higher than 50%.

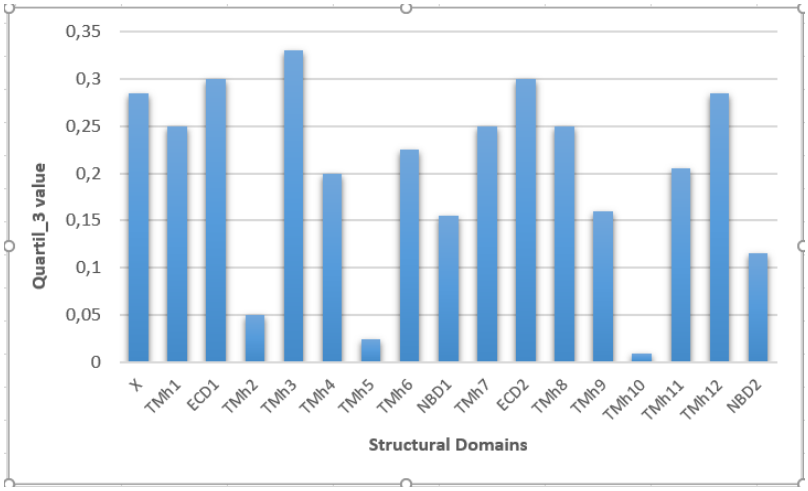

FIGURE 3. Histogram of Quartil_3 values for each structural domain. Every bin represents a structural region; TMDs are divided into different TMhelices (TMhs). X domain represents the residual, not structured portion of the ABCA4 sequence.

lie within the allowable range of the standard dictionary values, and the atom chirality of the models is correct, indicating that our ABCA4 model is reasonable in terms of geometry and stereochemistry.

The primary 3D model of ABCA4 (ECDs+TMDs+NBDs) from the structural modeling is shown in Figure 4.

Recently, the Electron Microscopy Map of ABCA4 in complex with adenyl-imidodiphosphate (AMPPNP) was published in the EMDB (http://www.emdatabank.org, in the public domain) with ID EMD-5498.[72] The 3D structural analysis of ABCA4 by electron microscopy shows, quite clearly, the three portions of our protein: ECDs, TMDs, and NBDs.[65] By using Chimera v.1.10.1[76] "Fit in Map" command, our 3D-modeled coordinates were transformed by fitting with the EMD-5498 density map into a plot whose 0.71 cross-correlation coefficient suggests a good reliability of our results (Fig. 5). The obtained superposition can be used to get an overall idea of the correct architecture of our model.

We then tried to predict possible pathogenic damages, using mCSM[66] and DUET,[68] with the structural analysis of missense mutations as collected in our ABCA4Database, allowing their classification based on special deleterious effects due to backbone rigidity or, conversely, loss of backbone

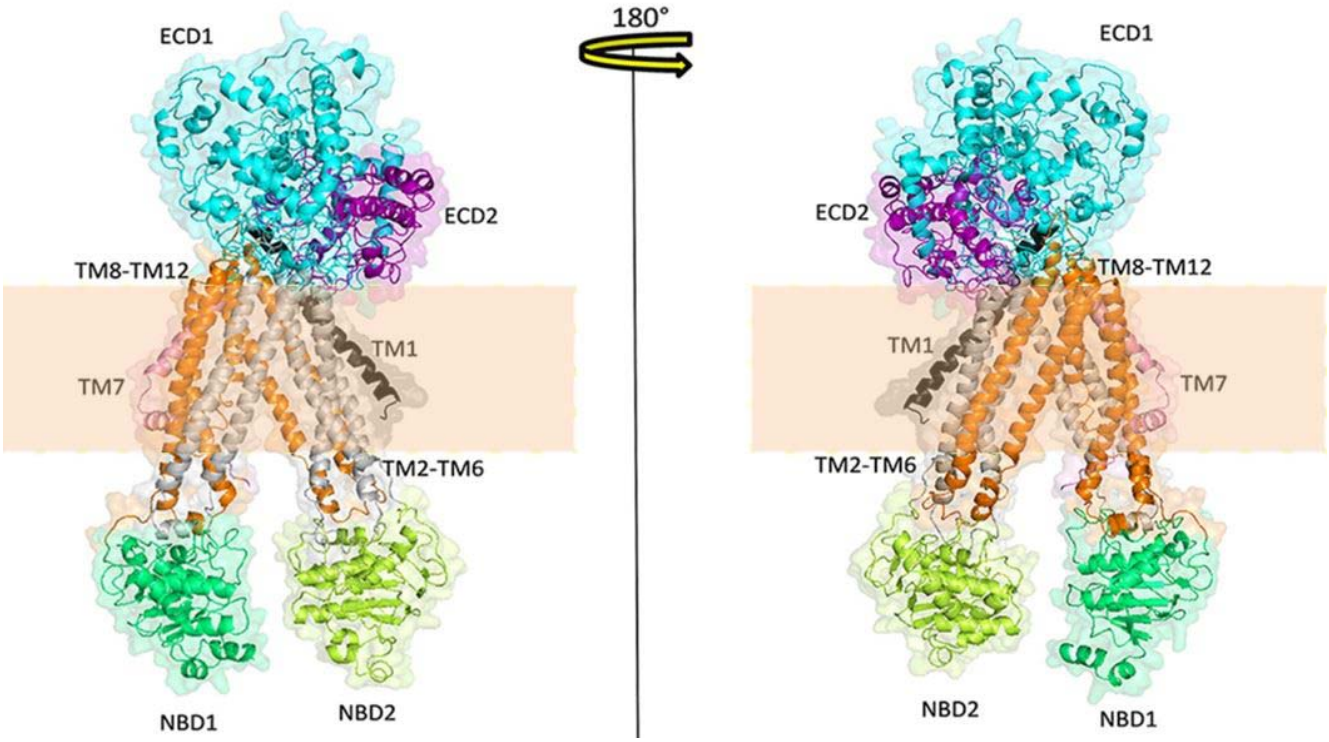

FIGURE 4. Surface and cartoon representation of human ABCA4 3D structural model overview. Human ABCA4 structure is reported in surface and cartoon representation. The TM1, ECD1, TM2–TM6, NBD1, ECD2, TM7, TM8–TM12, and NBD2 domains are labeled accordingly. The TM regions are depicted with *dotted lines*.

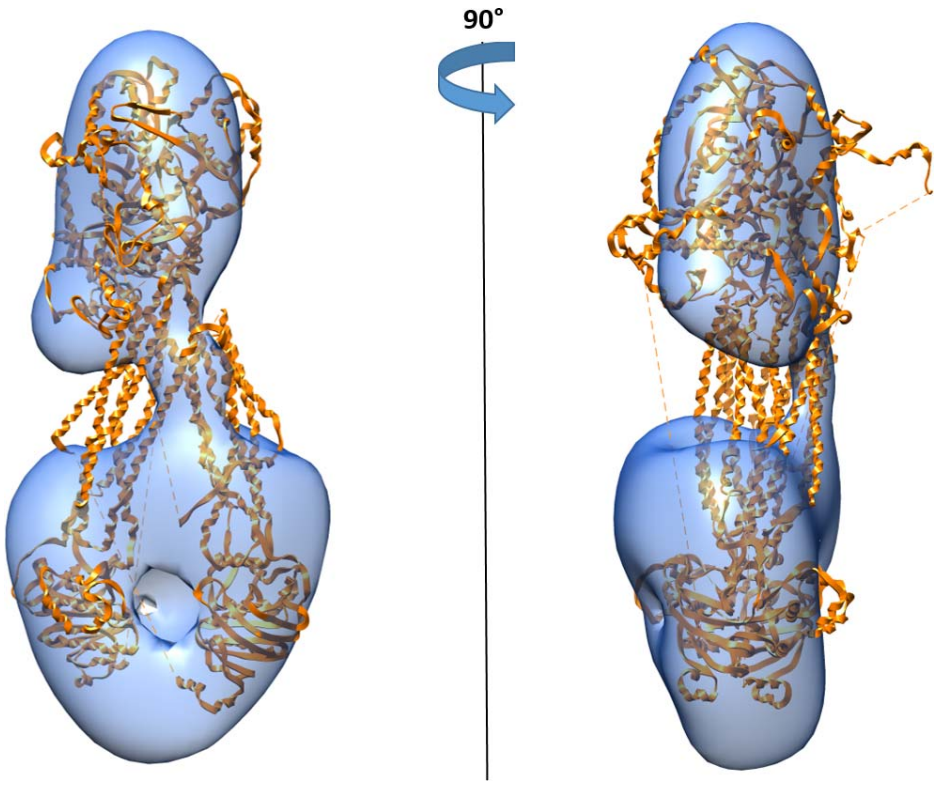

FIGURE 5. 3D structural analysis of ABCA4 by electron microscopy and homology modeling. The electron microscopy map is shown with the modeled structure of ABCA4 (*yellow*) fitted into ECD/TMD/NBD region (*blue*) using Chimera. The TM region is depicted with *dotted lines*.

rigidity, which can alter structural stability or residues' binding capability (Fig. 6).

DISCUSSION

We have initially focused our attention on testing residue conservation along the whole ABCA4 protein sequence. This is a crucial step in order to identify the most important regions for protein functionality and thus to understand the effects that mutations associated with retinal disease might have on it. Mutations occurring in a conserved site are assumed to have a greater impact on the correct protein folding[77,78] and, consequently, to alter its functionality. Indeed, phylogenetically conserved sequences in divergent species probably identify a protein region of extreme importance for structural stability. On the contrary, in sequence regions exhibiting low-residue conservation, the effect of a mutation could not severely damage protein structure, possibly preserving its normal activity. By scoring sequence entropy for ABCA4, we observed

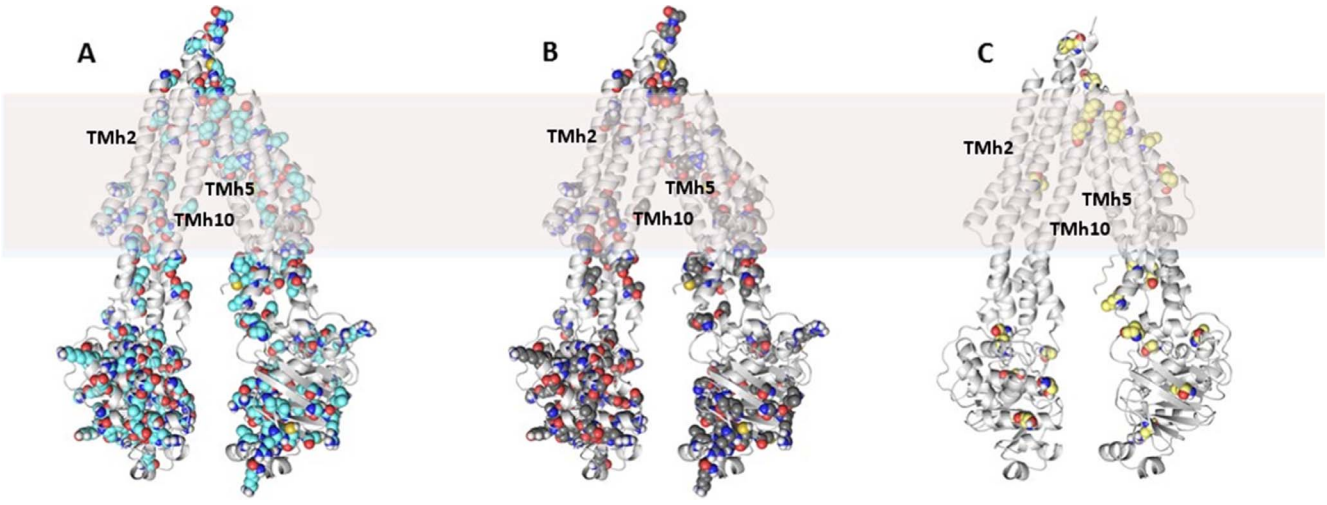

FIGURE 6. Cartoon representation of NBD and TM domains with identified missense mutations represented by balls colored as conventional CPK molecular models: *red* (oxygen), *white* (hydrogen), and *blue* (nitrogen atoms). The carbon skeleton is colored differently for each image: *cyan* (**A**), *black* (**B**), and *yellow* (**C**). In **A** and **B** are highlighted all missense mutations that can have structural stability effects predicted with mCSM and DUET tools, respectively. Those mutations that can alter backbone structure, as predicted by DUET, are reported in **C**.

that in general the TMDs and the NBDs show about 20% more residue conservation than do the ECDs (Fig. 3). This is not unexpected, as TMDs are domains that keep the ABCA4 protein embedded in the membrane and NBDs are responsible for ATP binding, a key process for the retinal translocation. On the other hand, the ECDs do not exhibit the same degree of residue conservation and show only some consecutive residues with a very low sequence variability. This leads us to suggest that mutations in those domains are more likely to be tolerated. ECDs participating in substrate translocation process with their highly mobile hinge domains[79] can more easily accommodate mutations without disrupting functionality. The 3D structure analysis of the ABCA4 protein is necessary to understand location and spatial distribution of the observed variable regions. In the absence of high-resolution experimentally derived structures, homology models provide a suitable 3D protein map; of course, as structural predictions, homology models require progressive refinement throughout whenever new structural/functional results become available. The crucial factors determining the quality of homology models are the sequence alignment accuracy and the quality and resolution of the template structure.

Currently, there are nine x-ray structures for full-length ABC transporters, but only two are of mammalian origin: the human ABCB10[80] and the mouse ABCB1,[81] also known as P-gp. We then decided to use as a template the mouse transporter because it is a single-chain protein organized in two homologous parts, but not identical ones, as in the human protein.[63] The P-gp (ABCB1) and ABCA4 are members of the same protein family, having the same architecture, except for the exocytoplasmic regions that are absent in P-gp. For this reason, we have used the sequence alignment as a strong basis for accurate homology modeling of TMDs and NBDs; ECDs, instead, were modeled with the threading bioinformatic approach.[61,71]

Thus, the obtained 3D structure is a first approximation to the actual protein structure. Nevertheless, the predicted structure may constitute a useful starting point to understand the complete structural picture of ABCA4 at the different stages of the catalytic cycle, to characterize its functional states, to identify possible conformational changes, and to detect binding regions and cavity pockets. The latter features can guide further investigations on the role of residues that are located at the NBD:NBD, TMD:TMD, and NBD:TMD interfaces, explaining experimental data coming from mutation, cross-linking, and photo-labeling studies. Based on our 3D protein structure, we can differentiate between residues directly involved in particular functions and possibly influenced residues due to their close proximity. For every single mutation, the ABCA4Database contains information about the 3D local structure and the corresponding Shannon's entropy,[55] providing a fast check on mutation effects at atomic resolution. It must be underlined that ABCA4 mutations responsible for pathologic states mainly occur in highly conserved regions. Some of these conserved regions are directly exposed to the protein surface, therefore influencing interactions with other molecules or with the solvent. Other conserved regions where mutations are found lie buried inside ABCA4 domains, yielding protein structural variant with altered conformations and, hence, reduced stability. We have also considered the relationship between the most conserved regions and the secondary structural elements, such as the transmembrane helices of the two TMDs. Most of the residues constituting these helices are highly conserved, confirming their importance for the structural stability of the transporter inside the membrane. We have used a bioinformatics approach to correlate the observed ABCA4 mutations with protein structural and functional features. Protein structure predictive tools yielded results for the various ABCA4 domains with different reliabilities, which are low for ECDs. Nevertheless, the ABCA4 model offers a powerful template to predict the effects of variants responsible for ABCA4-related diseases. Thus, a better understanding of the correlation between a single mutation and a specific pathogenic phenotype, with significant prognostic implications in clinical practice, will be possible, including the identification of new drug-binding sites, supporting the efforts toward medical treatment of genetic retinal diseases.

Acknowledgments

The authors thank Francesco Poggialini and Simone Cirri for their valuable assistance in protein modeling procedures. The authors also thank Elena and Dario Sennati for their insightful contribution to this work.

Supported by the Jack Brockhoff Foundation (JBF 4186) (DBA), a C.J. Martin Research Fellowship from the National Health and Medical Research Council of Australia (APP1072476) (DBA), the René Rachou Institute (IRR/FIOCRUZ Minas) (DEVP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (DEVP), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (DEVP).

Disclosure: **A. Trezza**, None; **A. Bernini**, None; **A. Langella**, None; **D.B. Ascher**, None; **D.E.V. Pires**, None; **A. Sodi**, None; **I. Passerini**, None; **E. Pelo**, None; **S. Rizzo**, None; **N. Niccolai**, None; **O. Spiga**, None

References

1. Illing M, Molday LL, Molday RS. The 220-kDa Rim protein of retinal rod outer segments is a member of the ABC transporter superfamily. *J Biol Chem*. 1997;272:10303–10310.
2. Dean M, Annilo T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu Rev Genomics Hum Genet*. 2005;6:123–142.
3. Bhongsatiern J, Ohtsuki S, Tachikawa M, Hori S, Terasaki T. Retinal-specific ATP-binding cassette transporter (ABCR/ABCA4) is expressed at the choroid plexus in rat brain. *J Neurochem*. 2005;92:1277–1280.
4. Tachikawa M, Watanabe M, Hori S, et al. Distinct spatiotemporal expression of ABCA and ABCG transporters in the developing and adult mouse brain. *J Neurochem*. 2005;95:294–304.
5. Warren MS, Zerangue N, Woodford K, et al. Comparative gene expression profiles of ABC transporters in brain microvessel endothelial cells and brain in five species including human. *Pharmacol Res*. 2009;59:404–413.
6. Ahn J, Molday RS. Purification and characterization of ABCR from bovine rod outer segments. *Methods Enzymol*. 2000;315:864–879.
7. Sun H, Molday RS, Nathans J. Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific ATP-binding cassette transporter responsible for Stargardt disease. *J Biol Chem*. 1999;274:8269–8281.
8. Beharry S, Zhong M, Molday RS. N-retinylidene-phosphatidylethanolamine is the preferred retinoid substrate for the photoreceptor-specific ABC transporter ABCA4 (ABCR). *J Biol Chem*. 2004;279:53972–53979.
9. Molday RS. ATP-binding cassette transporter ABCA4: molecular properties and role in vision and macular degeneration. *J Bioenerg Biomembr*. 2007;39:507–517.
10. Molday RS, Zhong M, Quazi F. The role of the photoreceptor ABC transporter ABCA4 in lipid transport and Stargardt macular degeneration. *Biochim Biophys Acta*. 2009;1791:573–583.

11. Sullivan JM. Focus on molecules: ABCA4 (ABCR)–an import-directed photoreceptor retinoid flipase. *Exp Eye Res*. 2009; 89:602–603.
12. Vasiliou V, Vasiliou K, Nebert DW. Human ATP-binding cassette (ABC) transporter family. *Hum Genomics*. 2009;3: 281–290.
13. Allikmets R, Shroyer NF, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science*. 1997;277:1805–1807.
14. Weng J, Mata NL, Azarian SM, Tzekov RT, Birch DG, Travis GH. Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in abcr knockout mice. *Cell*. 1999;98:13–23.
15. Mata NL, Weng J, Travis GH. Biosynthesis of a major lipofuscin fluorophore in mice and humans with ABCR-mediated retinal and macular degeneration. *Proc Natl Acad Sci U S A*. 2000;97: 7154–7159.
16. Mata NL, Tzekov RT, Liu X, Weng J, Birch DG, Travis GH. Delayed dark-adaptation and lipofuscin accumulation in abcr+/– mice: implications for involvement of ABCR in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2001;42:1685–1690.
17. Radu RA, Mata NL, Nusinowitz S, Liu X, Travis GH. Isotretinoin treatment inhibits lipofuscin accumulation in a mouse model of recessive Stargardt's macular degeneration. *Novartis Found Symp*. 2004;255:51–63; discussion 63–67, 177–178.
18. Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*. 1997;15: 236–246.
19. Cremers FP, van de Pol DJ, van Driel M, et al. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet*. 1998;7:355–362.
20. Maugeri A, Klevering BJ, Rohrschneider K, et al. Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet*. 2000;67: 960–966.
21. Michaelides M, Hardcastle AJ, Hunt DM, Moore AT. Progressive cone and cone-rod dystrophies: phenotypes and underlying molecular genetic basis. *Surv Ophthalmol*. 2006;51: 232–258.
22. Hamel CP. Cone rod dystrophies. *Orphanet J Rare Dis*. 2007; 2:7.
23. Passerini I, Sodi A, Giambene B, Mariottini A, Menchini U, Torricelli F. Novel mutations in of the ABCR gene in Italian patients with Stargardt disease. *Eye (Lond)*. 2010;24:158–164.
24. Allikmets R, Shroyer NF, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science*. 1997;19;277:1805–1807.
25. Allikmets R. Further evidence for an association of ABCR alleles with age-related macular degeneration. The International ABCR Screening Consortium. *Am J Hum Genet*. 2000; 67:487–491
26. Stargardt K. Ueber familiare, progressive degeneration in der Makulagegend des Auges. Albrecht von Graefes Arch Klin Ophthalmol. 1909;71:534–550
27. Klein R, Peto T, Bird A, Vannewkirk MR. The epidemiology of age-related macular degeneration. *Am J Ophthalmol*. 2004; 137:486–495.
28. Allikmets R, Singh N, Sun H, et al: A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*. 1997;15: 236–246.
29. Walia S, Fishman GA. Natural history of phenotypic changes in Stargardt macular dystrophy. *Ophthalmic Genet*. 2009;30: 63–68.
30. Simunovic MP, Moore AT. The cone dystrophies. *Eye (Lond)*. 1998;12:553–565.
31. Miraldi Utz V, Coussa RG, Marino MJ, et al. Predictors of visual acuity and genotype-phenotype correlates in a cohort of patients with Stargardt disease. *Br J Ophthalmol*. 2014;98: 513–518.
32. Hamel C. Retinitis pigmentosa. *Orphanet J Rare Dis*. 2006;11; 1:40.
33. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. *Lancet*. 2006;18;368:1795–1809.
34. Guymer RH, Héon E, Lotery AJ, et al. Variation of codons 1961 and 2177 of the Stargardt disease gene is not associated with age-related macular degeneration. *Arch Ophthalmol*. 2001;119:745–751.
35. Tsybovsky Y, Molday RS, Palczewski K. The ATP-binding cassette transporter ABCA4: structural and functional properties and role in retinal disease. *Adv Exp Med Biol*. 2010;703: 105–125.
36. Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol*. 2009;10:218–227.
37. Bungert S, Molday LL, Molday RS. Membrane topology of the ATP binding cassette transporter ABCR and its relationship to ABC1 and related ABCA transporters: identification of N-linked glycosylation sites. *J Biol Chem*. 2001;276:23539–23546.
38. Biswas EE, Biswas SB. The C-terminal nucleotide binding domain of the human retinal ABCR protein is an adenosine triphosphatase. *Biochemistry*. 2000;39:15879–15886.
39. Biswas EE. Nucleotide binding domain 1 of the human retinal ABC transporter functions as a general ribonucleotidase. *Biochemistry*. 2001;40:8181–8187.
40. Kos V, Ford RC. The ATP-binding cassette family: a structural perspective. *Cell Mol Life Sci*. 2009;66:3111–3126.
41. Linton KJ, Higgins CF. Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch*. 2007;453:555–567.
42. Kjellström U. Association between genotype and phenotype in families with mutations in the ABCA4 gene. *Mol Vis*. 2014; 20:89–104.
43. Gemenetzi M, Lotery AJ. Phenotype/genotype correlation in a case series of Stargardt's patients identifies novel mutations in the ABCA4 gene. *Eye (Lond)*. 2013;27:1316–1319.
44. Hargitai J, Zernant J, Somfai GM, et al. Correlation of clinical and genetic findings in Hungarian patients with Stargardt disease. *Invest Ophthalmol Vis Sci*. 2005;46:4402–4408.
45. Simonelli F, Testa F, Zernant J, et al. Genotype-phenotype correlation in Italian families with Stargardt disease. *Ophthalmic Res*. 2005;37:159–167
46. Oldani M, Marchi S, Giani A, et al. Clinical and molecular genetic study of 12 Italian families with autosomal recessive Stargardt disease. *Genet Mol Res*. 2012;11:4342–4350.
47. Maugeri A, van Driel MA, van de Pol DJ, et al. The 2588G–>C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with stargardt disease. *Am J Hum Genet*. 1999;64:1024–1235
48. van Driel MA, Maugeri A, Klevering BJ, Hoyng CB, Cremers FP. ABCR unites what ophthalmologists divide(s). *Ophthalmic Genet*. 1998;19:117–122.
49. Riveiro-Alvarez R, Lopez-Martinez MA, Zernant J, et al. Outcome of ABCA4 disease-associated alleles in autosomal recessive retinal dystrophies: retrospective analysis in 420 Spanish families. *Ophthalmology*. 2013;120:2332–2337.
50. Fujinami K, Lois N, Mukherjee R, et al. A longitudinal study of Stargardt disease: quantitative assessment of fundus autofluorescence, progression, and genotype correlations. *Invest Ophthalmol Vis Sci*. 2013;54:8181–8190.

Investigative Ophthalmology & Visual Science

51. Zernant J, Lee W, Collison FT, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet*. 2017;54:404–412.
52. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
53. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–1797.
54. Papadopoulos JS, Agarwala R. COBALT: constraint-bead alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23:1073–1079.
55. Shannon CE. The mathematical theory of communication. 1963. *MD Comput*. 1997;14:306–317.
56. Valdar WS. Scoring residue conservation. *Proteins*. 2002;48:227–241.
57. Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol*. 1999;291:177–196.
58. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. *Trends Biochem Sci*. 1998;23:403–405.
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410.
60. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18:2714–2723.
61. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc*. 2009;4:363–371.
62. The PyMOL Molecular Graphics System, Version 1.8. Schrödinger, LLC.
63. Stenham DR, Campbell JD, Sansom MS, Higgins CF, Kerr ID, Linton KJ. An atomic detail model for the human ATP binding cassette transporter P-glycoprotein derived from disulfide cross-linking and homology modeling. *FASEB J*. 2003;17:2287–2289.
64. Arnold K, Kiefer F, Kopp J, et al. The Protein Model Portal. *J Struct Funct Genomics*. 2009;10:1–8.
65. Tsybovsky Y, Orban T, Molday RS, Taylor D, Palczewski K. Molecular organization and ATP-induced conformational changes of ABCA4, the photoreceptor-specific ABC transporter. *Structure*. 2013;21:854–860.
66. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*. 2014;30:335–342.
67. Topham CM, Srinivasan N, Blundell TL. Prediction of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng*. 1997;10:7–21.
68. Pires DEV, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res*. 2014;42:W314–W319.
69. Retina International. Mutations of the ATP-binding Cassette Transporter Retina (ABCR, ABCA4) [database online]. Zurich, Switzerland. Available at: http://www.retina-international.org/files/scinews/abcrmut.htm. Accessed September 30, 2017.
70. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: 2008 update. *Genome Med*. 2009;1:13.
71. Cornelis SS, Bax NM, Zernant J, et al. In silico functional meta-analysis of 5,962 ABCA4 variants in 3,928 retinal dystrophy cases. *Hum Mutat*. 2017;38:400–408.
72. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993;234:779–815.
73. Bernini A, Spiga O, Ciutti A, et al. Prediction of quaternary assembly of SARS coronavirus peplomer. *Biochem Biophys Res Commun*. 2004;325:1210–1214.
74. Bernini A, Spiga O, Venditti V, et al. Tertiary structure prediction of SARS coronavirus helicase. *Biochem Biophys Res Commun*. 2006;343:1101–1104.
75. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292:195–202.
76. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.
77. Pandurangan AP, Ascher DB, Thomas SE, Blundell TL. Genomes, structural biology, and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. *Biochem Soc Trans*. 2017;45:303–311.
78. Pandurangan AP, Ochoa-Montano B, Ascher DB, Blundell TL. SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res*. 2017;45:229–235.
79. Pollock NL, McDevitt CA, Collins R, et al. Improving the stability and function of purified ABCB1 and ABCA4: the influence of membrane lipids. *Biochim Biophys Acta*. 2014;1838(1 Pt B):134–147.
80. Shintre CA, Pike AC, Li Q, et al. Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc Natl Acad Sci U S A*. 2013;110:9710–9715.
81. Aller SG, Yu J, Ward A, et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*. 2009;323:1718–1722.

Investigative Ophthalmology & Visual Science