Molecular modeling of retinoschisin with functional analysis of pathogenic mutations from human X-linked retinoschisis

Y.V. Sergeev1,*, R.C. Caruso2, M.R. Meltzer1, N. Smaoui3, I.M. MacDonald4 and P.A. Sieving1

1Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA, 2Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA, USA, 3GeneDx, Gaithersburg, MD, USA and 4University of Alberta, Royal Alexandra Hospital, Edmonton, Canada

Received October 28, 2009; Revised December 21, 2009; Accepted January 7, 2010

Gene mutations that encode retinoschisin (RS1) cause X-linked retinoschisis (XLRS), a form of juvenile macular and retinal degeneration that affects males. RS1 is an adhesive protein which is proposed to preserve the structural and functional integrity of the retina, but there is very little evidence of the mechanism by which protein changes are related to XLRS disease. Here, we report molecular modeling of the RS1 protein and consider perturbations caused by mutations found in human XLRS subjects. In 60 XLRS patients who share 27 missense mutations, we then evaluated possible correlations of the molecular modeling with retinal function as determined by the electroretinogram (ERG) a- and b-waves. The b/a-wave ratio reflects visual-signal transfer in retina. We sorted the ERG b/a-ratios by patient age and by the mutation impact on protein structure. The majority of RS1 mutations caused minimal structure perturbation and targeted the protein surface. These patients’ b/a-ratios were similar across younger and older subjects. Maximum structural perturbations from either the removal or insertion of cysteine residues or changes in the hydrophobic core were associated with greater difference in the b/a-ratio with age, with a significantly smaller ratio at younger ages, analogous to the ERG changes with age observed in mice with no RS1-protein expression due to a recombinant RS1-knockout gene. The molecular modeling suggests an association between the predicted structural alteration and/or damage to retinoschisin and the severity of XLRS as measured by the ERG analogous to the RS1-knockout mouse.

INTRODUCTION

Retinoschisin (RS1), a retinal secreted photoreceptor disulfide-linked oligomeric protein, is expressed exclusively in the retina and pineal gland and functions as an adhesion molecule, preserving the structural and functional integrity of the retina. The pathogenic inherited mutations in this gene cause X-linked recessive retinoschisis (XLRS; OMIM312700). XLRS is a form of juvenile macular and retinal degeneration that affects males and causes schisis, or splitting, within the retinal layers leading to early and progressive vision loss. Affected males have reduced scotopic b-wave amplitudes but relatively preserved a-waves of the electroretinogram (ERG) indicating predominantly an inner retinal abnormality (1). The splitting of retinal layers is associated with changes in retinoschisin structure (2). The RS1 gene has six exons that encode a protein sequence of 224 amino acids.

A number of mutations in RS1 co-segregate with XLRS providing strong evidence that the disease is caused by mutations in the RS1 gene (3). To date, about 170 unique sequence variations in the RS1 protein have been reported in XLRS patients and presented in the Leiden Open Variation Database (http://www.dmd.nl/rs/index.html). More than 60 mutations resulting in premature stop codons and/or frameshifts are known to produce a truncated protein. About 90 missense variants have been reported in RS1 to date. Although inherited mutations in the RS1 gene have been associated with an increased lifetime risk of XLRS, the functional impact of most missense variants that result in a single amino acid change is less well defined.

*To whom correspondence should be addressed at: OGVFB/NEI, Bldg. 10, Rm. 10B10, 9000 Rockville Pike, Bethesda, MD, USA. Tel: +1 3015947053; Fax: +1 3014020204; Email: sergeevy@nei.nih.gov

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
A promising approach to study a functional role of missense variants is to complement epidemiological, clinical and biochemical analysis with indirect approaches such as in silico protein structure—function studies and computational medicine (4,5). This analytical design creates the potential for disease risk assessment at the atomic level (6). Structural changes associated with missense mutations might impact protein folding, protein–protein interaction sites, solubility or stability of protein molecules. The structural effect of mutational changes can be analyzed in silico on the basis of three-dimensional structure, multiple alignments of homologous sequences, and molecular dynamics (MD) (5,7).

At present, most mutations, predominantly missense, have been found in exons 4–6 encoding the discoidin domain, suggesting a biological importance for this part of the protein (1,2). Mature retinoschisin structure (23 kDa) consists of Rs1 (residues 24–63) and the highly conserved discoidin domain (residues 64–224) (8). The discoidin domain is present in single or multiple copies in extracellular or transmembrane proteins implicated in cell–cell adhesion, and cell–cell interaction such as the coagulation factors V (FaV) and VIII (FaVIII), milk fat globule, neuropilin and neurexins, and is implicated in phospholipid binding (9). Discoidin domains comprise a similar structural fold (10). In different species, they have 47 highly conserved residues (~30% sequence identity), including 2 cysteines forming a disulfide bridge and several tryptophans maintaining a stable hydrophobic core of a discoidin domain β-sandwich (11).

Here, we report molecular modeling of RS1 and consider perturbations caused by mutations found in human XLRS subjects. In 60 XLRS patients who share 27 missense mutations, we evaluated possible correlations of the modeling with retinal function as determined by retinal clinical electrophysiology, i.e. the ERG a- and b-waves. The b/a-wave ratio reflects retinal visual-signal transfer through the synapse between the photoreceptors (a-wave) and the bipolar cells (b-wave) of the rod circuit. We sorted the ERG b/a-ratios by patient age and by the mutation impact on protein atomic structure, modeled by homology, using a score combining protein stabilization energy and a Grantham (4,5). This analytical design creates the potential for disease risk assessment at the atomic level (6). Structural changes associated with missense mutations might impact protein folding, protein–protein interaction sites, solubility or stability of protein molecules. The structural effect of mutational changes can be analyzed in silico on the basis of three-dimensional structure, multiple alignments of homologous sequences, and molecular dynamics (MD) (5,7).

RESULTS

Structural validation of genetic mutations

Individual RS1 gene mutations in 36 patients were ascertained through the National Eye Institute (NEI) Clinic, National Institutes of Health (see the Materials and Methods section). These mutations are included in the Data set I and listed in Table S1 (Supplementary Material). In only 25 XLRS cases within this data set, the same missense mutation is shared by two or more patients of different age. All 17 listed mutations were located in exons 4–6 corresponding to the discoidin domain. Among these mutations, only four gene mutations (c.288G>C, c.424 C>T, c.657C>G and c.668G>A) demonstrated a new change of an amino acid to or from a cysteine residue. These corresponding protein missense changes (W96C, C142R, C219W and C223Y) affect thiol exchange in RS1. Transypthon 96 and cysteines 142, 219 and 223 are conserved in 18 sequences of RS1. Mutations in these positions have negative Blosum70 scores and show very significant Grantham scores (12) correlating protein residue substitution frequencies with composition, polarity and molecular volume (see Supplementary Material, Fig. S1). This finding suggests that four new missense mutations found in this study could potentially affect the stability of retinoschisin structure.


The retinoschisin sequence consists of a leader sequence (residues 1–24) performing the function of a signal peptide, the Rs1 domain located in residues 25–64 and the discoidin domain occupying residues 65–224 (8). The location of cysteine residues in the structure is important to understand retinoschisin function (2,8,13–15). Mature RS1 structure contains 10 cysteines shared in an even proportion between Rs1 and discoidin domains. The structure of the discoidin domain and the positions of five cysteines are well established and reported in several cases (8,14,16). The other five cysteines are located in the Rs1 domain of mature retinoschisin. In order to predict the location of these cysteines, we generated a structure of mature retinoschisin. Briefly, the amino acid sequence was implied in the PHYRE search to find an appropriate structural template. Five best protein structures with best scores were found in this search: a bovine lactadherin C2 domain (PDB file: 2pqz, domain B), the discoidin domain of the coagulation factor Va, FA-5/8 C-terminal domain (PDB files: 1sdd, domain B; 1czz, domain A, and 1kex) and the discoidin domain of ddr2 (PDB file: 2z4f, domain A). Structures 1kex and 2z4f have demonstrated lower sequence identity to the RS1 sequence (<30%) and higher E-values (>1E–19). Structure 1czz, implied as a
structural template (16), demonstrated a lower sequence identity (36%) and higher E-value (3.9E−20). Finally, the C2 domain of 1sdd with the sequence identity 39% and E-value 1.4E−20 has been chosen as a structural template.

Mature RS1 structure, built using this structural template, is demonstrated in Figure 1. In order to achieve an appropriate molecular geometry, the structure of monomeric RS1 was equilibrated using MD in water. The MD-equilibrated structure of mature RS1 protein and the surrounding water molecules are in dark green and light grey, respectively (Fig. 1A). The structure of a single discoidin domain (yellow) superimposed on the MD-equilibrated RS1 structure is shown in dark green (Fig. 1B). The majority of dihedral angles characterizing the polypeptide chain fold are located in allowed areas of the Ramachandran plot (Fig. 1C), suggesting a properly refined stereochemistry for the molecule. The corresponding multiple sequence alignment of the mature RS1 and two other sequences used for homology modeling demonstrates residue conservation within the discoidin domains (Fig. 1D). The structure of the predicted discoidin domain of RS1 is consistent with an eight-stranded distorted L-barrel motif described previously (8,14,16). The structure of the barrel is stabilized by hydrogen bonds similar to that of other known β-barrel proteins (17). Although in the human sequence, only 7 out of 10 cysteines are conserved in the RS1 family of proteins, multiple sequence alignment data show that only 2 cysteine residues (C63 and C219) are conserved within a broader family of proteins sharing a common structure to the discoidin domain (Fig. 1D). These cysteines form a disulfide bridge C63–C219 between N- and C-termini in discoidin domains (11).

The expected functional roles for 15 missense mutations from Data set I are described in Table 1. Possible functions for the rest of Data set I were described previously (8,13–15). Here, we consider six missense mutations associated with a change from or to cysteine (Fig. 2A). Three missense mutations were related to the removal of reactive Cys and either insertion of bulky Trp in position C219, or positively charged Arg in position C142 in the area of spike 3, or bulky Tyr in position C223 at the C-terminus. A possible structural effect caused by the removal Cys and insertion of Trp in position 219 is illustrated in Figure 2C. The missense mutation C219W could affect the formation of a disulfide bridge C63–C219 conserved within discoidin domains. In a

Figure 1. Structure of mature RS1 obtained by homology modeling. (A) The MD-equilibrated structure of mature RS1 protein in water is shown in dark green. Water molecules surrounding RS1 are shown in light grey. (B) Superposition of the MD-equilibrated mature RS1 protein (dark green) and the structure of the discoidin domain (yellow) predicted using the protein homology/analogy recognition engine (PHYRE) (http://www.sbg.bio.ic.ac.uk/~phyre/secretindex.cgi). (C) Ramachandran’s plot of the MD-equilibrated RS1 protein showing the quality of the structure refinement. (D) Multiple sequence alignment of the mature RS1 sequence, light chain B of the bovine coagulation factor V (PDB file: 1sdd_B), and the heavy L-chain of the blood coagulation factor Va (PDB file: 1fv4_L).
Table 1. Structure-based analysis predicts structural changes for 15 missense variants from the NEI data set (Data set I) affecting the XLRS phenotype

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Location and predicted effect of mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>S73P</td>
<td>Change to proline could change stability of amino-terminal loop. A mild change with a Blosum70 score $-1$. Mutant variant predicted to be secreted</td>
</tr>
<tr>
<td>W96C</td>
<td>Buried residue located in the loop connecting β2 and β3 strands and physically buried in the cavity between spike 1 and spike 2. This change affects the thiol reactivity of protein. Residue is buried inside of protein molecule (ASA = 9 Å²). Severe change (Blosum70 score $-3$) suggests protein misfolding and no homo-oligomer secretion</td>
</tr>
<tr>
<td>W96R</td>
<td>Buried residue located in the loop connecting β2 and β3 strands and physically buried in the cavity between spike 1 and spike 2. Introduces a positive charge into the cavity. Less severe compared with that of W96C</td>
</tr>
<tr>
<td>R102Q</td>
<td>The replacement breaks five H-bonds stabilizing R102 at the protein surface. These H-bonds are connecting NH1 atom of R102 with main chain O atoms of F71 (3.2 Å) and V76 (2.8 Å), NH2 atom of R102 with main chain O atoms of V76 (2.9 Å) and N99 (3.4 Å), and NE atom of R102 and main chain O atom of N99 (3.1 Å). Severe change which could affect the protein fold</td>
</tr>
<tr>
<td>N179D</td>
<td>Buried residue located in the β7 strand. Accessibility of D179 is 1 Å² suggesting that this could affect protein conformation due to a high propensity of the Asp residue to be exposed at the protein surface</td>
</tr>
<tr>
<td>P192S</td>
<td>Exposed residue located in the loop connecting β7 and β8 strands. H-bond connecting N atom of P192 and OH atoms of Y166 is broken in the mutant protein and a new H-bond between OH group of Y166 and N atom of R171 is created. The unique conformation formed by P192 and P193 is destroyed. This could affect the protein fold</td>
</tr>
<tr>
<td>P192T</td>
<td>Together with P193, surface residue P192 defines an almost 90° turn of the loop connecting β7 and β8 strands. This change will make the conformation of the 90° turn less rigid and could affect protein fold</td>
</tr>
<tr>
<td>P193S</td>
<td>Surface residue P193 located next to P192 in the loop connecting β7 and β8 strands. Mutational change to Ser shows the ASA change $\Delta$ASA = −22 Å³. S193 becomes buried (ASA = 0 Å³). This change will make conformation of the 90° turn less rigid and could affect protein conformation</td>
</tr>
<tr>
<td>R209C</td>
<td>Mutation dramatically changes the residue surface area from the exposed, ASA = 137 Å² (Arg), to the buried, ASA = 29 Å² (Cys). In addition, H-bond connecting atom OE1 of residue Q117 and N atom of residue L210, and H-bond connecting 1NE2 atom of residue Q117 and O atom of residue V208 are broken. Severe change which could affect a native protein fold</td>
</tr>
<tr>
<td>R213W</td>
<td>Surface residue, ASA = 119 Å³ (Arg), located in the last β strand. The mutation to Trp buries this residue, ASA = 3 Å³ (Trp). This changes protein stability and surface charge</td>
</tr>
<tr>
<td>C219W</td>
<td>Buried residue C219 is located at the C-terminus. Intra-molecular disulfide bond C219-C63 is broken. W219 is stabilized by a new H-bond between N atom of W219 and O atom of V132. This could affect protein conformation</td>
</tr>
<tr>
<td>C223Y</td>
<td>Residue C223 is located at the C-terminus and together with C59 forms disulfide bond defining the stability of discoidin domain. Replacement of C223 with tyrosine breaks the S–S bond and decreases stability of the domain. Very severe change which could affect the protein conformation</td>
</tr>
</tbody>
</table>

Mutations shown by bold italic letters were considered as new, if they are not listed in the X-linked retinoschisis sequence variation database, http://www.dmd.nl/rs/. Amino acid residue was considered located at the protein surface if the accessible surface area (ASA) was $>30$ Å²; otherwise, it was considered as buried. Changes due to mutation in ASA, dihedral angles $\Phi$ and $\Psi$ are shown as $\Delta$ASA, $\Delta$Φ and $\Delta$Ψ, respectively. Functional and structural effects for missense variants of Data set I: E72K, Y89C, W92C, R102W, L127P, N136T, R141H, R200C and C219G (Supplementary Material, Table S1), were described earlier (13,14,16) and not included in the table.

Similar way, the missense mutation C223Y disrupts the inter-molecular disulfide bond between residues C59 and C223 (13). Using the FoldX potential, we estimate that this protein stability alteration could increase within a value between 15.3 and 33.5 kcal mol$^{-1}$ (18,19) and could significantly destabilize a native protein structure.

Three missense mutations (W96C, R141C and R209C) are predicted to have a significant structural impact related to the replacement of bulky residues Trp or Arg by a new reactive cysteine. The replacement of a Trp residue located in position 96 of the hydrophobic core by cysteine might introduce a new disulfide bond between $\beta$7 atoms of Cys96 and Cys142, separated by a 5.2 Å distance, respectively (Fig. 2B). Mutation R141C creates an additional cysteine at position 141 located next to Cys142, potentially affecting the formation of a disulfide bond between Cys110 and Cys142, previously described as an intra-molecular S–S bond (13). Hence, the creation of a new disulfide bond might stabilize the protein conformation in the non-native state. In general, a positive change of free energy of protein stabilization, which is necessary to mutate a residue to cysteine, was estimated within the limits of 13.6–26.6 kcal mol$^{-1}$ for mutants from Data set I. Therefore, a new disulfide bond could significantly interfere with an essential pattern of disulfide bonds in native retinoschisin by forming improper disulfide bonds within or between protein molecules.

Phenotypic data

The functional role of missense mutations and their effect on the disease phenotype was evaluated using clinical visual electrophysiology. Indeed, in retina electrophysiology, signal transmission between rod photoreceptors and the post-synaptic bipolar cell neurons is measured by the amplitudes of the a-wave (20,21) and b-wave (22–25), respectively. Conversely, in ~30% of the cases, the ERG shows a pattern of an ‘electro-negative waveform’, where the positive-sloping b-wave fails to rise back or to exceed the baseline from the initial negative a-wave configuration (26). In the dark-adapted ERG for bright stimuli, the b-wave amplitude is normally greater than the a-wave, but in XLRS, the b-wave naturally is reduced (27). Although substantial ERG variability is observed (26,28), the amplitude of the a-wave is reduced in up to 30% of
XLRS-affected males (26). In the RS1-KO mouse model, the disease process definitely involves the rod photoreceptors (29–31). Therefore, normalizing the bipolar cell b-wave signaling to the a-wave helps to account for abnormalities originating from the rod photoreceptors.

In this work, disease severity was estimated in 36 XLRS patients using the b/a-wave ratio (32). Ages of patients, a- and b-wave amplitudes and b/a-wave ratios are represented in Table S1 (Supplementary Material) and are included in Data set I. Ages of patients ranged from 6 to 62 years and phenotypic b/a-ratios for patients ranged from 0.5 to 1.82 (normal range: 1.42–2.64). In a total of 25 cases, repetitive data were obtained from two or more patients per mutation.

Additional amplitudes of a-, b-waves and b/a-wave ratios (Data set II) for 26 patients, aged 2–70 years old, were obtained from the literature (26,28,33–37) and are shown in Table S2 (Supplementary Material). ERG amplitudes and the b/a-wave ratios were measured by a consistent technique, as in Data set I, with the correction-derived values of the b/a-wave ratios ranging from 0.24 to 0.84. The corrected data set shows an association with published phenotypic data with a correlation coefficient of 0.89 and \( P = 0.0001 \) (Supplementary Material, Fig. S1), demonstrating that the corrected b/a-wave ratios included in Data set II also describe a disease phenotype and are consistent with the b/a-wave ratios from Data set I.

Changes with age of b/a-wave ratios normalized with respect to the mean normal b/a-wave ratio are shown in Figure 3. The b/a-wave ratios from Data sets I and II are represented by open squares and open circles, respectively. These normalized b/a-wave ratios for XLRS-affected individuals are in the range between 0.2 and 0.95, in contrast to that of unaffected individuals which range between 0.7 and 1.5. This clearly demonstrates a pathogenic character to missense mutations for the majority of patients. Individual b/a-wave ratios do not, however, show a recognizable tendency to change with age (Fig. 3B). To understand the role of pathogenic mutations in XLRS, we used an MD-equilibrated model of retinoschisin tertiary structure in order to evaluate the energetics and severity of structural changes caused by missense mutations.

**Impact of missense mutations evaluated at the atomic level**

We sorted the ERG b/a-wave ratios by patient age and by the impact of mutation on protein atomic structure defined by homology modeling. In order to reduce the effect of statistical errors and to increase a signal-to-noise ratio in experimental ERG data, we divided all the data sets into four groups by age (younger and older) and by mutation severity (weak and severe), as shown in Table 2. Average ages for younger and older individuals were 18–19 and 38–48 years for 25 patients with repetitive data from the NEI study. In the larger group of 61 patients, younger and older were 12–13 and 37–40 years old, respectively. The mutation effect was evaluated for each mutation separately using a computed impact (CI) score (see
All mutations were ranked according to their CI. We analyzed mutant residue accessibilities in the retinoschisin described by b/a-wave ratios is related to missense changes, using a CI score, are associated with the XLRS severity. Predicted perturbations in protein structure, which were described for younger individuals. These ERG data suggest that pre-slight decrease with CI seems to be similar to that described less significant for older patients, a common trend showing a considerably for younger individuals accompanied by a corresponding rise of CI score, a measure of the structural severity of the mutation (Fig. 4). In this model, the decline in a-wave amplitude with age coincided with the loss of outer nuclear layer cells, whereas the decline in the b-wave and the b/a-wave ratio was associated with increasing severity of schisis cavities in younger ages.

Amplitudes of the a-waves decreased with age for both weak and severe mutations and amplitudes of b-waves decreased with age for weak mutations and showed no change for severe (0.52/0.53) mutations. The corresponding b/a-wave ratios, averaged in each of the four groups, and b- and a-waves amplitudes are shown in Figure 4 and Table 2, respectively. Presented data demonstrate a significant consistency between results for each group from the data sets of 26 and 61 patients, respectively.

The outcome is that this model identifies an association between predicted changes in retinoschisin protein structure caused by missense variants and the clinical phenotype of XLRS as captured by b/a-ratios.

**DISCUSSION**

Results of our modeling indicated that less than half of RS1 mutations targeted the protein surface and caused minimal structural perturbation with quite similar b/a-ratios across younger and older subjects. Maximum structural perturbations, from removal or insertion of cysteine residues or change in the hydrophobic core, caused a considerable difference in the b/a-ratio with age and a significantly smaller ratio at younger ages. The modeling suggested an association between the predicted alteration and/or damage to the retinoschisin structure due to mutations and the severity of XLRS disease measured by the ERG. This was analogous to ERG changes with age observed in mice carrying an Rs1 knockout gene. This indicates that Rs1 mutations differentially affect the XLRS phenotype.

**Genotype–phenotype relationships**

Previous studies have looked for genotype–phenotype relationships in XLRS but none have been identified (38–40). Missense, splice site, frameshift, insertion and deletion mutations all result in a similar phenotype. No correlation has been found between disease severity characterized by visual acuity (36,39,41–44) and mutation type, either for small patient numbers or in a larger study of 86 XLRS patients in whom causative Rs1 mutations were identified (45). We used the Ganzfeld ERG as an estimator of global retinal function to characterize the XLRS subjects. The ERG of XLRS typically shows a relatively preserved a-wave, whose...
leading edge is generated by photoreceptor activation, but a subnormal b-wave that is generated by activation of depolarizing bipolar cells. The relative abnormality in synaptic transfer from photoreceptors to bipolar cells was estimated by the b/a-wave ratio. Therefore, the severity of the retinal function abnormality due to XLRS may be quantified by assessing the degree of abnormality of the b/a-wave ratio. We found an apparent association between the effect of mutations at a molecular level and the XLRS phenotype characterized by the ERG b/a-wave ratio.

Severe structure perturbations caused by cysteines and changes in hydrophobic core

An interesting implication when using computed severity scores is that mutations in hydrophobic core and/or affecting cysteines are more severe than mutations on the protein surface. Indeed, in the endoplasmic reticulum (ER), proteins fold into their native conformations and undergo different post-translational modifications and the formation of disulfide bonds. The primary sequence of protein is the major determinant for proper protein fold (46). The protein folds along several competing pathways into intermediate non-native structures with decreasing free energies until it achieves a conformation with the lowest energy to form a protein with native interactions (47). Misfolded or intermediate non-native proteins never achieve this lowest energy minimum. Hydrophobic core pathogenic mutations produce a severe effect and cause the protein to stay in an intermediate conformation with a higher free energy. Genetic modification of cysteine residues could cause formation of non-native proteins. Indeed, in the ER, the formation of a disulfide bond is catalyzed by protein disulfide isomerase (PDI), a cellular chaperone with foldase function to maintain a native protein fold (48,49). PDI catalyzes oxidation and reshuffling (isomerization) of disulfides in substrate proteins by using the catalytic CxxC motif (50). Current data suggest that PDI docks close to two cysteines in the substrate protein, forms a native disulfide bridge between those cysteines and possibly stabilizes native contacts in that area by utilizing a chaperone function. Hence, we speculate that genetic mutations associated with cysteine removal might interfere with PDI binding in that area. This might cause the appearance of non-native contacts which previously were corrected by the foldase action. The insertion of a new cysteine could also perturb a protein fold due to formation of an additional non-native disulfide bond. Hence, non-native contacts could misfold the protein into a non-native conformation with an intermediate free energy.

Several major mechanisms leading to degradation of misfolded protein are currently known. Approximately 30% of cytosolic proteins carry KFERQ-targeting motifs in their sequences allowing for selective targeting of cytosolic proteins to the lysosome for degradation by chaperone-mediated autophagy (51). The absence of the specific targeting motif in the RS1 sequence suggests another option. As a result, misfolded or unassembled proteins are retained in the ER bound to chaperones or lectins until they are delivered to the cytosol for degradation in the ubiquitin–proteasome pathway (52). Mutant proteins with significant changes in protein folding caused by missense mutations or deletions could generate proteins susceptible to hydrolysis similar to that of human hemoglobin in which about one-fifth of hundreds of missense proteins undergo rapid degradation (52). Therefore, the severe mutant variants are expected to have a markedly perturbed non-native protein fold causing misfolding in the ER and later digestion in the proteasome. This hypothesis is supported by numerous observations on retinoschisins modified by severe mutations and expressed in cell culture (Cos cells) which show the modified RS1 bands to be absent on SDS–PAGE or native gels and associated with a loss of protein due to misfolding and rapid degradation (8,13–15).

Retinoschisin secretion

The CI score derived from the protein structure could be useful to estimate the effect of pathogenic mutation in experiments on retinoschisin secretion in cell culture. Although numerous experiments have demonstrated that retinoschisin is secreted as a monomer, dimer or octamer in model cell systems (13,14), in vivo in photoreceptors and bipolar cells, this protein is secreted as a homo-oligomeric complex. Oligomerization is critical for the functioning of retinoschisin as an extracellular adhesion
was approved by the CNS Institutional Review Board of the

MATERIALS AND METHODS

XLSR phenotypic data

Clinical data from 36 male XLRS patients, aged 7–62 years old, are shown in Table S1 (Supplementary Material). This study was approved by the CNS Institutional Review Board of the NIH (03-E1-#0033). Informed consent was obtained from each participating subject. Pedigrees of the XLRS families showed an X-linked pattern of transmission of the trait in affected males. All the XLRS males recorded in Figure 3 showed macular changes typical for XLRS (schisis cavities in younger patients and atrophic changes in older patients).

Ophthalmologic examination of all patients included best-corrected visual acuity, kinetic perimetry with a Goldmann perimeter using standardized light targets V4e, 14e and 12e, slit-lamp biomicroscopy, fundus examination and photography. The central absolute luminance threshold was measured with a Goldmann–Weekers adaptometer (Haag-Streit, Berne, Switzerland). The maculae of both eyes were examined using an Optical Coherence Tomography 2010 scanner (Zeiss Humphrey Instruments, San Leandro, CA, USA). Three overlapping 7 mm horizontal scans were obtained, from the optic disc, thorough the macular area and temporal to the macula.

Full-field ERGs were recorded at NEI using a UTAS 2000 system (LKC Technologies, Gaithersburg, MD, USA) with Burian–Allen bipolar contact lens electrodes on each eye and a ground electrode on the forehead. Pupils were dilated with topical tropicamide 1% and phenylephrine 2.5%, and the patient dark-adapted for 30 min before testing. ERG recordings followed the 2004 ISCEV standard technique (53). Dark adapter ERG responses were amplified, at 0.3–500 Hz, digitized and stored. Stimulus intensity for the combined response (‘standard flash’) was 1.636 cd m⁻², within the ‘standard flash’ range of ISCEV 2004 (53). Amplitude was measured from a-wave thorough to b-wave peak, and ERG a-wave and b-wave amplitudes (microvolts) and implicit times (milliseconds) were compared with 96 unaffected normal individuals.

For the data obtained from other investigators, recording conditions were described with a varying degree of detail in their publications. Normalization was used to minimize the effect of varying recording conditions among researchers.

XLRS severity was evaluated using full-field ERGs. Full-field ERG in XLRS usually has an almost normal a-wave amplitude and a recognizable reduction in the amplitude of the scotopic b-wave elicited by stimulation with a white single flash (27,42,54,55). The b-wave amplitude was measured at the first conspicuous positive peak following the a-wave not corresponding to an oscillatory potential, and this was used to calculate the b/a-wave ratio. We used the b/a-wave ratio to quantify the severity of XLRS for 36 patients carrying 17 different pathogenic mutations (Supplementary Material, Table S1).

Figure 4. ERG b/a-wave ratio, CI severity caused by missense mutations and patient age of XLRS patients. Average ERG b/a-wave ratios obtained for patients divided into four groups by the value of CI (weak and severe) and by patient age (young and old) as described in Table 2. ERG b/a-wave ratios for each group are shown by black spheres. Points obtained for the selected group of 26 patients from Data set I (NEI Clinic only) and for 61 patients (Data sets I and II together) are connected by dashes (Supplementary Material, Table S1).
conditions are available upon request. Information about missense mutations found here is summarized in the first two columns of Table S1 (Supplementary Material).

**Structure and MD of retinoschisin**

The structure of mature human retinoschisin structure represented by amino acid residues from 21 to 222 of the SwissProt sequence O15537 was modeled by homology. The protein homology/analogy recognition server (PHYRE, ver. 2) available on the Web (http://www.sbg.bio.ic.ac.uk/~phyre/secretindex.cgi) was used to search for the structural template. The X-ray structure of the C2 domain of FaV was chosen as the structural template (Brookhaven protein database, PDB code: 1sdd, chain B4) (56) to obtain the model of RS1. Briefly, the primary sequences of RS1 and 1sdd aligned by the method of Needleman and Wunsch (57) and incorporated in the program Look, version 3.5.2 (58), were used for the

---

**Figure 5.** Missense mutations are divided in two groups by their severity. (A) Mutated residue accessibility was calculated by the surface accessibility algorithm (63,64). Severity of structural change due to missense mutation was estimated by using the CI score (top scale) as described in the Materials and Methods section. The amino acid residue was considered as buried in the hydrophobic core if the residue accessibility was < 40 Å² (dashed line), otherwise the residue was considered as exposed at the surface (> 40 Å²). Residues were divided in two groups by the computed severity value and considered as weak (0.22–0.43, light brown) or as severe (0.43–0.71, light cyan). (B) The majority of weak mutations (light brown) are localized at the protein surface. (C) The majority of severe mutations are buried and/or related to the change from/to cysteine residue (light cyan). (b and c) Molecular surfaces for projections of the molecule shown in (B) and (C), respectively.
Molecular modeling of missense mutations

The structure of the MD-equilibrated RS1 was used to generate in silico the structural models of proteins modified by missense mutations. The program Look, version 3.5.2, was applied to generate the conformation of missense mutations and refine them by self-consistent ensemble optimization (61), which applies the statistical mechanical mean-force approximation iteratively to achieve the global energy minimum structure. Finally, predicted structures of each mutant protein were regularized by an energy minimization procedure as described above for the native RS1 in the presence of water for the final step similar to that of described in previous paragraph.


Evaluation of mutation impact using an atomic structure of protein

The rapid evaluation of free energy change due to missense mutations could be accessed with the FoldX force field, based on an empirical energy function derived from experimental work on proteins (19). The FoldX free energy (in kcal mol\(^{-1}\)) is the combination of solvation energy contributions from hydrophobic and polar groups of protein; a Van der Waals energy term taking into account experimental transfer energies from water to vapor; the energy of hydrogen bonds with regards to simple geometric considerations; electrostatic coulomb free energy terms; and the crude entropy of protein chain to obtain a measure of free energy and free energy of the steric overlaps between atoms in protein structure. The predictive power of this approach has been successfully tested with experimental thermodynamics data on 667 mutant proteins (18) from the ProTherm database (62) and suggested to be useful in protein design algorithms. Recently, a combination of MD folding simulations and FoldX potential has been suggested as an approach of in silico analysis of missense substitutions (5).

The amino acids difference formula was derived by Grantham (12) to identify a combination of chemical factors like composition, polarity and molecular volume which correlate best with protein residue substitution frequencies. The Grantham difference for each amino acid pair is a useful parameter to characterize a physiochemical differences between wild-type and missense variant amino acid residues (4,5).

Therefore, a CI of missense mutation could combine the change in energy of protein stabilization \(\Delta \Delta G_{\text{stab}}\) and a Grantham distance. In a Euclidean space having these two categories as axes, the CI will show the distance between two amino acids involved in the mutation change. The axes are made orthogonal to facilitate the calculation of the difference in distance between amino acids. The overall difference, which is characterized by the CI score, \(c_{\text{mut, wt}}\), will estimate the overall difference between two amino acids for the same position in wild-type (wt) and mutant (mut) protein as

\[
c_{\text{mut, wt}} = \alpha \left[ \left( \frac{\Delta \Delta G_{\text{stab}}}{\text{sd}} \right)^2 + \left( \frac{GD}{\text{sd}} \right)^2 \right]^{1/2}.
\]  

Figure 6. The CI score could be used to classify missense mutations as affecting or not affecting the secretion of an octameric complex. The CI scores are presented by filled black circles for each mutation (lower panel). In the upper panel, results for the octamer (8-mer) secretion in cell culture obtained from the literature (8,14,15) and predicted from CI are shown by open black squares and by grey stars, respectively. Missense mutations were only considered with consistent secretion reported at least by two different laboratories. A threshold of 0.34 was chosen empirically in order to obtain a best match in data on octamer secretion.
Here $\Delta G_{\text{stab}} = \Delta G_{\text{wt}} - \Delta G_{\text{mut}}$ is the change in free energy of protein stabilization obtained for each mutation; GD is a Grantham distance. Both categories are normalized by their corresponding standard deviations (sd) calculated independently for each category over all 55 mutants. Factor $\alpha$ is a scale factor ($\alpha = 0.15$) in order to fit normalized values of b/a-ratios. The CI score, ci(mut, wt), was calculated for each mutation and compared with the severity of XLRS estimated from ERG data.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We thank Dr Kei Shinoda, Department of Ophthalmology, Oita University, Japan, for providing published ERG data; Delphine Blain, Leanne Reuter and Patrick Lopez, National Eye Institute, NIH, for their help with ERG data; Sean Finnegan from the National Eye Institute for his valuable editorial help; and Dr Peter J. Munson, Mathematical and Statistical Computing Laboratory, CIT/NIH, for valuable discussions. We are also thankful for the use of commercial molecular modeling software on the network through the Center of Molecular Modeling, CIT/NIH.

Conflict of Interest statement. None declared.

**FUNDING**

This work was supported by National Institutes of Health (Z01-EY000476-01 to Y.V.S., and Z01-DC000065-08 to P.A.S.).

**REFERENCES**


