Effects of pathological mutations on the stability of a conserved amino acid triad in retinoschisin

Franca Fraternali[a], Luigi Cavallo[b], Giovanna Musco[c,*]

[a] National Institute for Medical Research, Department of Mathematical Biology, The Ridgeway, Mill Hill, London NW7 1AA, UK
[b] University of Salerno, Dipartimento di Chimica, via Salvador Allende, I-84081 Baronissi (SA), Italy
[c] Dulbecco Telethon Institute, Neurobiologia cellulare e molecolare, DIBIT, Via Olgettina 58, 20132 Milano, Italy

Received 17 March 2003; revised 10 April 2003; accepted 12 April 2003

First published online 14 May 2003

Edited by Giovanni Cesareni

Abstract **A three-dimensional model has been calculated for the discoidin domain of retinoschisin (RS1), the protein involved in the X-linked juvenile retinoschisis. The model allows for a mapping of the pathological retinoschisis missense mutations and a rationale for the structural effects of an evolutionary conserved surface exposed triad (W122–R200–W163). Molecular dynamics simulations of the triad mutants models, together with ab initio energy calculations of the complexes corresponding to the triad show that the observed pathological mutations sensibly destabilize local interactions and the entire fold. Moreover the presented model reveals evidence of a putative site for membrane association.**

Key words: Retinoschisis; Discoidin domain; Pathological mutation; Membrane binding site; Molecular dynamics; Ab initio energy calculation

1. Introduction

X-linked recessive juvenile retinoschisis (XLRS; OMIM 312700) is the most common cause of hereditary progressive macular degeneration in males [1,2]. XLRS is characterized by splitting within the inner retinal layers, resulting in cystic degeneration of the central retina and consequent vision deterioration early in life. The disease is caused by defects in the gene [3,4] encoding retinoschisin (RS1), a retinal secreted photoreceptor disulfide-linked oligomeric protein [5–7] of unknown function. The mature protein (23 kDa) consists almost entirely of a highly conserved discoidin domain [4]. The domain is present in single or multiple copies in extracellular or transmembrane proteins implicated in cell adhesion, cell–cell interaction [8,9] such as the coagulation factors V (FaV) [10] and VIII (FaVIII) [11], milk fat globule, neuropilin [12] and neurexins.

A wide spectrum of genetic mutations have been found in retinoschisis patients, ranging from point mutations to frame shifts or truncations [4], the majority ($>80\%$) are clustered in the conserved discoidin domain, thus suggesting its crucial importance for the normal function of RS1.

Although it is relatively easy to rationalize the detrimental effects of a truncated protein on its function, it is less straightforward to explain the effect of missense mutations as they may be variously due to incorrect three-dimensional (3D) folding or to functional impairment. Structural analysis of RS1 can therefore contribute to a better comprehension of its function and provide a molecular explanation for the structural/functional effects of the mutations involved in RS1 pathology.

Unfortunately the 3D structure of RS1 is still unknown, thus in this paper we present a homology model of the discoidin domain of RS1 based on the X-ray structure of the C2 domain of FaV (37.7% sequence identity with RS1) which allowed us to rationalize the possible structural basis of the protein defect underlying the RS1 pathology.

The model allowed us to map on the structure the clinically relevant point mutations. In particular we have focused our attention on pathological mutations involving a conserved surface exposed amino acid triplet (W122–R200–W163) which appears to be crucial for the correct function of the discoidin domain [13]. The guanidinium group of the R200 is sandwiched between the two aromatic systems of W122 and W163, thus generating a particularly stable cation–Π interaction [14,15]. With the help of molecular dynamics (MD) simulations on the RS1 model and its mutants we can investigate the local and overall effects of the pathological triad mutations on the stability of the protein. Moreover ab initio calculations in vacuo and in implicit solvent can contribute to a more detailed investigation on the nature of the intramolecular interactions.

2. Materials and methods

2.1. Homology modeling

Modeling of both human RS1 (residues 62–222 of the SwissProt sequence O15537) and the mutants (W122C, R200C, W163C) was accomplished using the X-ray structure of the C2 domain of FaV as template (pdb code 1CTZ, open conformation). Sequence alignment of 38 discoidin domain family members [including human X-linked juvenile retinoschisis precursor protein (AF014459), mouse X-linked juvenile retinoschisis precursor protein (AF073780), fRS1fugu X-linked juvenile retinoschisis precursor protein (AF146687), tyrosine receptor kinase (Q16832), epithelial discoidin domain receptor 1 precursor (Q08345), adipocyte transcription factor (JC5256), contactin associated protein (U87223), milk fat globule membrane protein (Q08431), silkworm hemocytin (S52093), FaV precursor (M14335), FaVIII precursor (P00451) neuropilin, AF018956, slime mold discoi-

*Corresponding author.
E-mail addresses: ffranca@nimr.mrc.ac.uk (F. Fraternali), lcavallo@unisa.it (L. Cavallo), musco.giovanna@hsr.it (G. Musco).

Abbreviations: CAO, contact accepted mutation; MD, molecular dynamics; RS1, retinoschisin; FaV, coagulation factor V

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00433-2

din I chain A (J01282)] was performed with PRALINE [16] by giving a special weight to the highly conserved cysteines both in the N-terminal and the C-terminal regions. This allowed us to build a model with a disulfide bridge between residues C63 and C219, as suggested by the C2-FaV crystal structure. In order to evaluate the areas of reliability of the alignment we used contact accepted mutation (CAO) [17]. The CAO score measures the probability of a mutation from a contact residue pair in a reference structure (here FaV) to the aligned residue pair in the model structure (here RS1) (Fig. 1). Positive CAO scores indicate residue contact conservation, negative scores indicate an improbable substitution event. Alignment scores from a contact-based mutation model reflect both the sequence and structural quality of residue (pair) matches. 3D models (of the wild-type RS1 and of the three mutants RS1W122C, RS1W163C and RS1R200C) were computed with the HOMOLOGY module of INSIGHTII (Accelerys, San Diego, CA, USA). The quality of the structure was assessed with WHATIF OLDQUA and NEWQUA utilities [18]. Solvent accessibilities were computed with MOLMOL [19] and POPS [20].

2.2. MD

Simulations were performed with GROMACS 3.1.3 package [21]. Trajectories were calculated in a periodic rectangular box (8.0, 8.0, 8.0 Å) of explicit SPC water molecules. All bonds were constrained by the linear constraint solver (LINCS) algorithm. Temperature was kept constant to 300 K. Isotropic pressure coupling of the system was applied. Fast particle-mesh Ewald (PME) electrostatics was applied. Solvent was equilibrated by 500 ps of solute positions restrained MD, followed by 1 ns unrestrained MD with a 2 fs time step for each system. Essential degrees of freedom were obtained according to the essential dynamics (ED) method [22]. The diagonalization of the covariance matrix of the backbone coordinate fluctuations produced a set of eigenvalues (amount of motion) associated to eigenvectors (direction of motion). The two extreme projections along the first eigenvector for 1 ns MDs were calculated for each system.

2.3. Ab initio calculations

The coordinates of the sidechains of the W122–R200–W163 sandwich (and respective mutations) were extracted from the generated models and considered as starting point for calculations of the ab initio interaction energies [23]. The four complexes have been identified as WRW, CRW, WRC and WCW, where the sequence numbering is always in the order 122–200–163. For each complex hydrogen atoms were added and their coordinates were optimized in the gas-phase with the hybrid B3LYP DFT approach, using the 6-31G(d,p) basis set. For all the complexes, we calculated the rigid body BSSE-free [24] MP2 energy of interaction of the residue in the middle with both lateral residues of the sandwiches W**R**W, C**R**W, W**R**C and W**C**W. We refer to the generic A**X**B complex in the energy decomposition formula. Under these assumptions the interaction energy E_b becomes, $E_b = E_{AXB} - E_{A-B} - E_X - E_{BSSE}$, where E_{AXB} is the energy of the A**X**B complex, E_{A-B} and E_X are the energies of the isolated A-B and **X** parts in which the A**X**B complex was rigidly fragmented. E_{BSSE} is the basis sets superposition [24]. Within this definition, stabilizing interactions correspond to negative E_b. For these calculations the 6-311+G(d,p) basis set was used. Solvent effects were evaluated using the polarizable continuum PCM model [25,26], which has been tested in several studies on biological systems [27,28]. The Gaussian98 package was used for all the calculations [29].

3. Results and discussion

The core of the RS1 discoidin domain consists of an eight-stranded distorted β-barrel motif formed by a five-stranded antiparallel β-sheet which packs onto a three-stranded antiparallel β-sheet. The final model of RS1 is similar to the C2 domain of FaV template with a root mean square deviation of 1.09 Å for the superposed Cα atoms of the β-sandwich (1288 atoms) (Fig. 1). The quality values according to the OLDQUA and NEWQUA WHATIF routines [18] are −1.81 and −3.19 respectively (−1.3 and −2.45 for FaV) which are acceptable considering a 37.7% of sequence identity. Similar values have been observed for the NEWQUA values in previous studies on immunoglobulin and fibronectin modules from Titin [30,31], while the OLDQUA obtained in that case were around −1.0 on average.

The model provides a basis for a 3D mapping of the reported retinoschisis pathological mutations (X-linked retinoschisis sequence variation database, http://www.dmd.nl/rs/), scattered all over the structure. Several substitutions affect positions comprising buried conserved bulky residues, such as W96R, W112C, L127P, I199T, P203L, L216P which are expected to cause cavities or introduce positive charges in the hydrophobic core of the structure, thus destabilizing the fold.

Among the pathological surface mutations, which might have a functional relevance, we identified a spatial cluster involving an evolutionary conserved triad (W122–R200–W163) in which the guanidinium group of the arginine is sandwiched between the indole moieties of the two tryptophans. The crucial role of this triad common to all discoidin domain family members is confirmed by the observation that mutations inside the triplet motif of FaVIII are known to cause hemophilia A [13] and to destabilize the fold of the C2 domain of FaV [32]. In the case of retinoschisis single point mutations to cysteine of each amino acid of the triad cause the pathology.

Moreover the mutated triad falls into high CAO-scored residue contact conservation region (Fig. 1), with residue 163 displaying the highest probability for mutations (lowest CAO score) compared to the other two members of the triad. MD simulations show that the most flexible segments comprise residues 115–121 and 157–165 which correspond respectively to the lower and upper loops generating the pseudo-cavity in which the sandwich is located. ED analysis shows that a significantly higher amount of motion along the first eigenvector is experienced by RS1 with respect to the other mutated structures, for which the principal motions are distributed along more eigenvectors. All these evidences are graphically rendered in Fig. 2, where 'sausage' representations give a pictorial view of the distribution of essential motions for each system. The highest mobility is found for the upper loop of W163C, which is the region of high probability of mutation according to CAO analysis. Analysis of the energy contributions from MD simulations showed that the most significative changes amongst models are found in the electro-

Table 1
Energy differences of the average electrostatic components of the energy (kcal/mol) calculated on the last 500 ps

Molecule complex	ΔE_{El-sr} (Prot–Prot)	$\Delta E_{El-1,4}$ (Prot–Prot)	ΔE_{El-sr} (Prot–Solv)
RS1	0 (6377.1)	0 (3816.9)	0 (3599.5)
W122C	56.9	−65.9	22.2
W163C	74.2	37.0	27.3
R200C	109.7	−46.2	−160.4

Absolute values in parentheses. Prot–Prot stands for protein–protein, Prot–Solv stands for protein–solvent contribution, sr for short range.

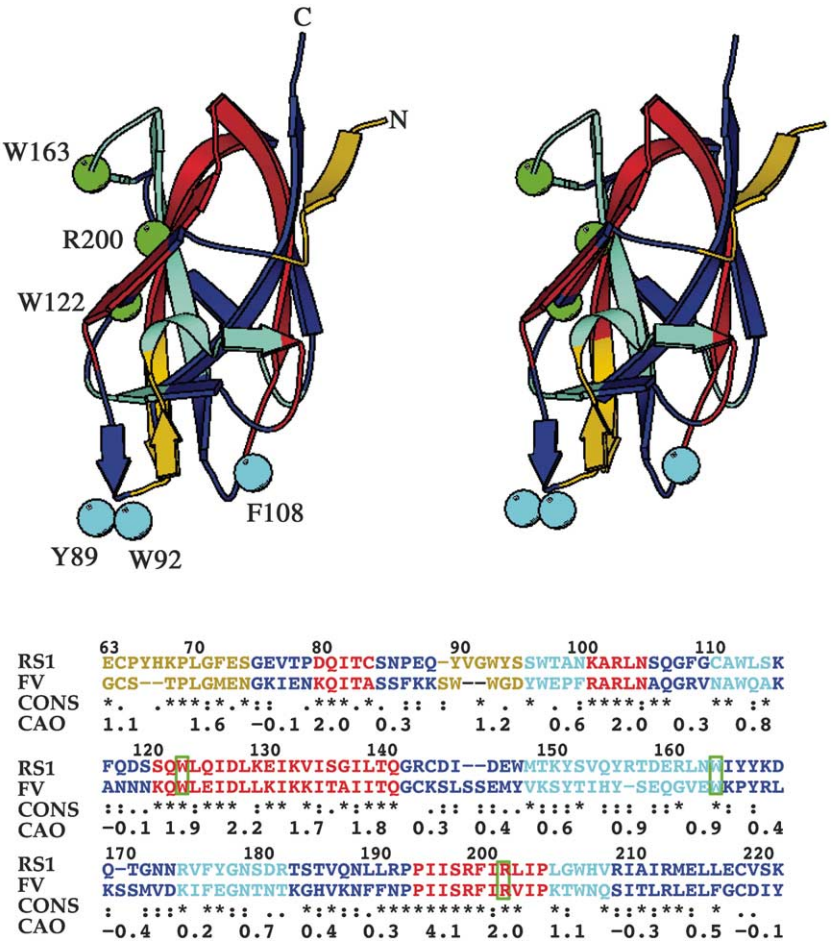

Fig. 1. Sequence alignment used to generate the RS1 model. The CAO relative scores are reported in the alignment. Residues with scores within 0.5 are colored in blue, between 0.6 and 1.0 are colored in cyan, between 1.1 and 1.6 in yellow and higher than 1.7 in red. Green boxes indicate residues of the triad. Molscript [37] stereo-representation of the RS1 model color-coded according to the CAO scores. In green and cyan are highlighted CPK balls of the Cα of the triad (green) and of other solvent exposed mutation sites (cyan).

static contributions. In Table 1 average values over the last 500 ps for the short-ranged electrostatic terms are reported. We only report these terms in order to directly compare their trends with the ab initio interaction energies of the triad presented thereafter. Energetic contributions relative to the RS1 model have been arbitrarily set as reference values. The most stable intramolecular protein–protein contribution is that of RS1, while that of R200C is about 100 kcal/mol less stable. The R200C mutations sensibly destabilize the entire molecule, due to the break up of the WRW cation–Π interaction. On the other hand, R200C is strongly stabilized by electrostatic protein–solvent interactions (with an energy gain of about 160 kcal/mol). During the simulation, in fact, water can easily penetrate into the pseudo-cavity of the WRW sandwich, once the arginine has been substituted by a cysteine. This causes a local disruption of electrostatically favorable protein–protein contacts, largely compensated by protein–solvent interactions.

Calculated gas-phase ab initio interaction energies E_b reported in Table 2 clearly indicate that the interaction of the R sidechain with the external residues in the W**R**W, C**R**W and W**R**C complexes is roughly 15–20 kcal/mol stronger than the interaction of the cysteine sidechain with the external tryptophan residues in the W**C**W complex, which is unstable already in the gas-phase. The remarkably higher stability of the complexes with an R in the middle position, relative to the complex with a C at the same position, can be ascribed to the charged nature of the R sidechain that gives rise to stabilizing cation/Π interactions [14,15].

The same trend is observed in solution-like phase. Although implicit solvent effects destabilize all the considered complexes, those with an R in the middle remain stable also in solution (negative E_b), whereas the W**C**W complex is clearly unstable (positive E_b). Solvent effects reduce the stability of the W**R**W complex by roughly 15 kcal/mol, while the stability of the C**R**W and W**R**C complexes is reduced by about 10 kcal/mol. The higher solvent effects on the W**R**W complex can be related to the higher amount of solvent accessible surface area buried upon formation of the complex for the arginine, roughly 40% of the area of the isolated R sidechain, relative to the amount of area buried upon formation of the W**R**C and C**R**W complexes, roughly 30–35%.

Analysis of the accessible surface area indicated a prominent exposed hydrophobic region rich in aromatic residues encompassing amino acids Y89, V90, W92, Y93, F108, I144.

This patch is highly reminiscent of the hydrophobic surface around the spikes observed for both FaV (W26, W27, M83, L79) [10,33] and FaVIII (V2223, M2199, F2200, L2251, L2252) [11,34], which are likely candidates for phospholipid bilayer binding. Membrane involvement seems to be a com-

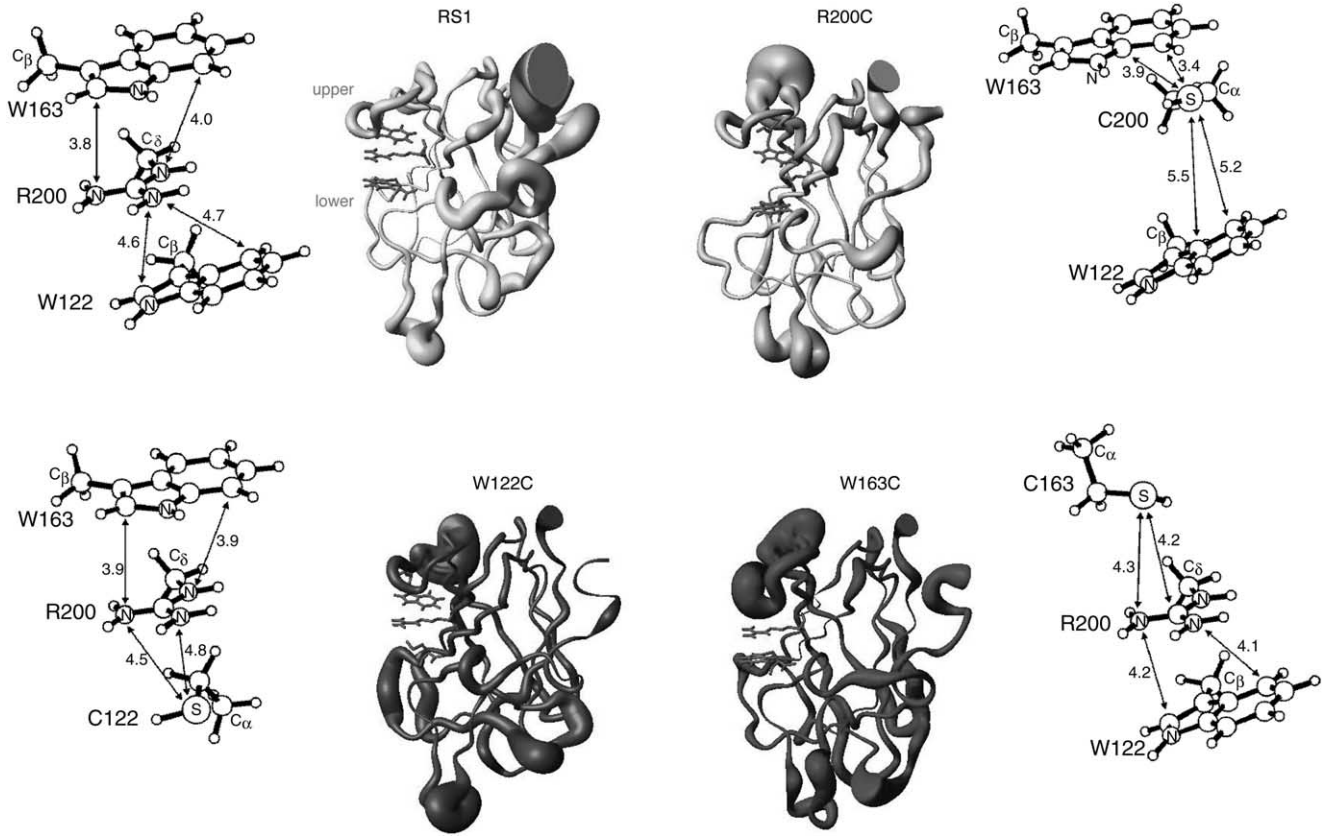

Fig. 2. MOLMOL sausage rendering of the two extreme projections along the first eigenvector of the ED analysis for the studied molecules. The sausage representation draws spline with variable radius for a given bundle of structures. In this representation regions with higher deviation from the average are thick, while well superimposed regions are thin. For each system the final configuration from the ab initio calculation of the corresponding triad is reported with some geometrical parameters.

mon feature for several members of the discoidin domain family. Biochemical and immunocytochemical experiments showed that RS1 is present in the membrane fraction [35,36] thus suggesting a role for RS1 in cell–cell interaction or adhesion. We propose therefore a molecular model for this interaction (Fig. 3) which could be very similar to the one suggested for FaV in the open conformation, where the well conserved R213 (corresponding to R150 in FaV) could act as anchoring point allowing Y89, W92, I144 (corresponding to residues W26, W27, L79 in FaV) to immerse their sidechain in the apolar membrane core. Remarkably three aromatic residues (Y89, W92, F108) of these putative membrane interaction spikes are sites of pathological mutations (Fig. 1) and are consistently changed into cysteines (http://www.dmd.nl/rs/). It is therefore tempting to speculate that these mutations might alter the anchoring properties of RS1 to the membrane thus disrupting its putative role in maintaining retinal cell architecture. Future work will concern MD simulations of this system and the analysis of the stability of the protein–membrane interactions in the wild-type and mutated forms.

4. Conclusions

RS1 pathologically relevant mutations of a phylogenetically conserved WRW triad have been analyzed in detail by means of MD simulations and ab initio interaction energy calculations. The triad forms a sandwich hold in place by a cation–Π interaction, whose energy trend obtained from ab initio calculations remarkably fits short-ranged intramolecular electrostatic contributions. The MD calculations show that single point mutations of the triad are all less stable than the native RS1. In particular the R200C mutation is the least energetically favored one and the relative complex is unstable. Breaking the cation–Π interaction causes penetration of water that partially disrupts the local environment of the sandwich and infers enlarged flexibility to the neighboring loops. At molec-

Table 2
Absolute, E_b, and relative, ΔE_b, interaction energies in kcal/mol of the central residue **X** with the A-B part in the A**X**B complexes (see Section 2)

Complex	E_b (gas)	ΔE_b (gas)	E_b (water)	ΔE_b (water)
W**R**W (RS1)	−20.0	0	−4.4	0
W**R**C (W122C)	−14.8	5.2	−2.7	1.7
C**R**W (W163C)	−13.1	6.9	−2.9	1.5
W**C**W (R200C)	−0.5	19.5	1.5	5.9

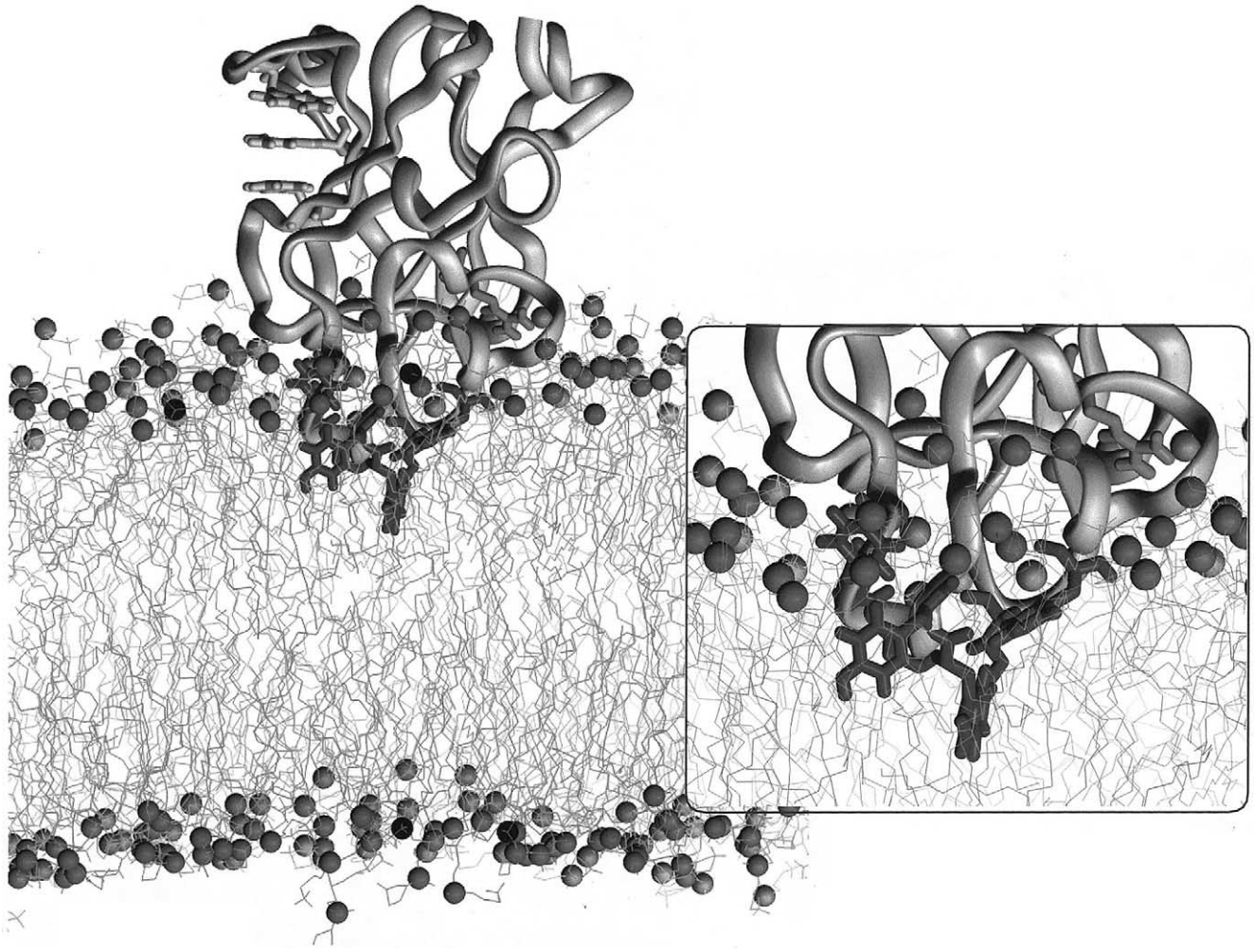

Fig. 3. Proposed binding site for the membrane anchoring of RS1. Phosporous atoms from the lipid membrane are represented in CPK. The sidechains of the hydrophobic patch (Y89, V90, W92, Y93, F108, I144) are represented in grey.

ular level this analysis offers a rationale for the observed misfunctioning of mutated retinoschisis, which seems to be correlated with the compromised stability of the WRW triad.

Moreover the analysis of the surface of the RS1 model allowed us to identify an exposed hydrophobic patch, thus suggesting a possible role of RS1 in phospholipid membrane interaction, in a manner similar to what previously was proposed for FaV [10] and FaVIII [11,14].

This in silico study represents a valuable starting point to encourage experimentalists to study the structure of RS1 and the perturbations caused by the pathological mutations. Moreover this structure prediction might be useful for an audience of molecular genetists hopefully guiding and inspiring new biochemical experiments aiming to disentangle retinoschisis function.

Acknowledgements: The authors like to thank Jaap Heringa and Jens Kleinjung for suggestions with the use of PRALINE and CAO packages. L.C. thanks Drs. M. Cossi and G. Scalmani of the University of Napoli for useful suggestions on the PCM calculations, and the CIMCF of the Università Federico II of Napoli for technical support. G.M. thanks Dr. S. Banfi for stimulating discussions.

References

[1] Forsius, H., Krause, U., Helve, J., Vuopala, V., Mustonen, E., Vainio-Mattila, B., Fellman, J. and Eriksson, A.W. (1973) J. Ophthalmol. 8, 385–393.
[2] George, N.D., Yates, J.R. and Moore, A.T. (1996) Arch. Ophthalmol. 114, 274–280.
[3] Sauer, C.G., Gehrig, A., Warneke-Wittstock, R., Marquardt, A., Ewing, C.C., Gibson, A., Lorenz, B., Jurklies, B. and Weber, B.H. (1997) Nat. Genet. 17, 164–170.
[4] The Retinoschisis Consortium (1998) Hum. Mol. Genet. 7, 1185–1192.
[5] Weber, B.H., Schrewe, H., Molday, L.L., Gehrig, A., White, K.L., Seeliger, M.W., Jaissle, G.B., Friedburg, C., Tamm, E. and Molday, R.S. (2002) Proc. Natl. Acad. Sci. USA 99, 6222–6227.
[6] Wang, T., Waters, C.T., Rothman, A.M., Jakins, T.J., Romisch, K. and Trump, D. (2002) Hum. Mol. Genet. 11, 3097–3105.
[7] Grayson, C., Reid, S.N., Ellis, J.A., Rutherford, A., Sowden, J.C., Yates, J.R., Farber, D.B. and Trump, D. (2002) Hum. Mol. Genet. 9, 1873–1879.
[8] Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R. and Bucher, P. (1998) Protein Sci. 7, 1626–1631.
[9] Vogel, W. (1999) FASEB J. 13, S77–S82.
[10] Macedo-Ribeiro, S., Bode, W., Huber, R., Quinn-Allen, M.A., Kim, S.W., Ortel, T.L., Bourenkov, G.P., Bartunik, H.D.,

Stubbs, M.T., Kane, W.H. and Fuentes-Prior, P. (1999) Nature 402, 434–439.
[11] Pratt, K.P., Shen, B.W., Takeshima, K., Davie, E.W., Fujikawa, K. and Stoddard, B.L. (1999) Nature 402, 439–442.
[12] Lee, C.C., Kreusch, A., McMullan, D., Ng, K. and Spraggon, G. (2003) Structure 11, 99–108.
[13] Liu, M.L., Shen, B.W., Nakaya, S., Pratt, K.P., Fujikawa, K., Davie, E.W., Stoddard, B.L. and Thompson, A.R. (2000) Blood 96, 979–987.
[14] Dougherty, D.A. and Stauffer, D.A. (1990) Science 250, 1558–1560.
[15] Ma, J.C. and Dougherty, D.A. (1997) Chem. Rev. 97, 1303–1324.
[16] Heringa, J. (2002) Comput. Chem. 26, 459–477.
[17] Lin, K., Kleinjung, J., Taylor, W.R. and Heringa, J. (2003) Comp. Biol. Chem, in press.
[18] Vriend, G. (1990) J. Mol. Graph. 8, 52–56.
[19] Koradi, R., Billeter, M. and Wuthrich, K. (1996) Mol. Graph. 14, 51–5, 29–32.
[20] Fraternali, F. and Cavallo, L. (2002) Nucleic Acids Res. 30, 2950–2960.
[21] Lindahl, E., Hess, B. and van der Spoel, D. (1995) Comp. Phys. Comm. 91, 43–56.
[22] Amadei, A., Linssen, A.B. and Berendsen, H.J. (1993) Proteins 17, 412–425.
[23] Foresman, J.B. and Frisch, A.E. (1996) Exploring Chemistry with Electronic Structure Methods, 2nd edn., Gaussian Inc., Pittsburgh, PA.
[24] Boys, S.F. and Bernardi, F. (1970) Mol. Phys. 19, 553–556.
[25] Tomasi, J. and Persico, M. (1994) Chem. Rev. 94, 2027–2094.
[26] Cossi, M., Barone, V., Cammi, R. and Tomasi, J. (1996) Chem. Phys. Lett. 255, 327–335.
[27] Rega, N., Cossi, M. and Barone, V. (1998) J. Am. Chem. Soc. 120, 5723–5732.
[28] Arnaud, R., Adamo, C., Cossi, M., Milet, A., Vallée, Y. and Barone, V. (2000) J. Am. Chem. Soc. 122, 324–330.
[29] Gaussian98 Users Manual, Gaussian Inc., Pittsburgh, PA, 2001.
[30] Fraternali, F. and Pastore, A. (1999) J. Mol. Biol. 290, 581–593.
[31] Amodeo, P., Fraternali, F., Lesk, A.M. and Pastore, A. (2001) J. Mol. Biol. 311, 283–296.
[32] Kim, S.W., Quinn-Allen, M.A., Camp, J.T., Macedo-Ribeiro, S., Fuentes-Prior, P., Bode, W. and Kane, W.H. (2000) Biochemistry 39, 1951–1958.
[33] Fuentes-Prior, P., Fujikawa, K. and Pratt, K.P. (2002) Curr. Protein Pept. Sci. 3, 313–339.
[34] Gilbert, G.E., Kaufman, R.J., Arena, A.A., Miao, H. and Pipe, S.W. (2002) J. Biol. Chem. 277, 6374–6381.
[35] Molday, L.L., Hicks, D., Sauer, C.G., Weber, B.H. and Molday, R.S. (2001) Invest. Ophthalmol. Vis. Sci. 42, 816–825.
[36] Wang, T., Waters, C.T., Rothman, A.M., Jakins, T.J., Romisch, K. and Trump, D. (2002) Hum. Mol. Genet. 11, 3097–3105.
[37] Per, J.K. (1991) J. Appl. Crystallogr. 24, 946–950.