

## Computational Screening of Rhodopsin Mutations Associated with Retinitis Pigmentosa

Angelo Felline,<sup>†</sup> Michele Seeber,<sup>†</sup> Francesco Rao,<sup>‡</sup> and Francesca Fanelli<sup>\*,†</sup>

Dulbecco Telethon Institute and Department of Chemistry, via Campi 183,  
41100 Modena, Italy, and Laboratoire de Chimie Biophysique/ISIS 8, Université Louis  
Pasteur, allée Gaspard Monge, 67000 Strasbourg, France

Received March 26, 2009

**Abstract:** Retinitis pigmentosa (RP) refers to a group of debilitating, hereditary disorders that cause severe visual impairment in as many as 1.5 million patients worldwide. Rhodopsin mutations account for >25% of the autosomal dominant form of the disease (ADRP). Forty artificial and ADRP-associated mutations located in the second extracellular loop (EL2) that folds into a twisted  $\beta$ -hairpin were screened through replica exchange molecular dynamics (REMD) simulations using the FACTS implicit solvent model. According to in vitro experiments, ADRP-linked mutants fail to express at the plasma membrane and/or to reconstitute with 11-*cis*-retinal, indicative of variable defects in protein folding and/or stability. The computational protocol was first probed on the protein G C-terminal  $\beta$ -hairpin, proving the effectiveness of the implicit solvent model in reproducing the free energy landscape of  $\beta$ -hairpin formation. Eight out of the 40 EL2 mutants resulted in misfolding effects on the native  $\beta$ -hairpin structure, consistent with in vitro evidence that they all share severe impairments in folding/expression. Five mutants displayed moderate misfolding attitudes, whereas the remaining 27 mutants, overall characterized by milder effects on rhodopsin expression, did not perturb significantly the conformational behavior of the native  $\beta$ -hairpin but are expected to exert variably disturbing effects on the native interactions of the loop with the chromophore and/or the surrounding receptor domains. Collectively, the results of this study add structural insight to the poorly resolved biochemical behavior of selected class II ADRP mutations, a fundamental step toward an understanding of the atomistic causes of the disease.

### 1. Introduction

Retinitis pigmentosa (RP) is a group of hereditary human diseases that are characterized by progressive retinal degeneration due to death of the rod photoreceptor cells, the vertebrate photoreceptors dedicated to dim light vision.<sup>1–3</sup> Patients affected by RP display nyctalopia (night blindness), progressive loss of peripheral and, eventually, central vision and the characteristic accumulation of intraretinal pigment deposits, from which the disease gets its name. Despite the high genetic heterogeneity of the RP syndrome, over 120

point mutations have been discovered in the gene of rhodopsin, the visual pigment molecule of rod cells that generates a detectable electrical response following light capture.<sup>4</sup> Although some of the rhodopsin mutations cause autosomal recessive retinitis pigmentosa (ARRP), the vast majority cause the autosomal dominant form (ADRP) of the pathology (collected in part in the rhodopsin mutation database at <http://www.retina-international.com/sci-news/rhmut.htm>).<sup>1–3</sup>

Rhodopsin, the cornerstone of family A of G protein coupled receptors (GPCRs),<sup>5</sup> is a transmembrane receptor protein expressed in the retina and composed of a protein (opsin) and a chromophore. Opsin is an up-and-down bundle of seven transmembrane (TM) helices linked to three intracellular (IL) and three extracellular (EL) loops as well

\* Corresponding author. Tel: +39 059 2055114. Fax: +39 059 37353. E-mail: fanelli@unimo.it.

<sup>†</sup> Dulbecco Telethon Institute and Department of Chemistry.

<sup>‡</sup> Université Louis Pasteur.

**Table 1.** Computational and in Vitro-Determined Indices Concerning Spontaneous and Artificial Mutants of Rhodopsin EL2

mutant <sup>a</sup>	D&K <sup>b</sup> (%)	P&R <sup>c</sup> (%)	Sut <sup>d</sup> (%)	HB1–4 <sub>avg</sub> <sup>e</sup>	rmsd <sup>f</sup>	El2 misfolding <sup>g</sup>	fold <sup>h</sup>	expr <sup>i</sup>	retinal <sup>j</sup>	ref <sup>k</sup>
WT				77.44	88.66					
R177C*	58.80	44.46	<b>44.79<sup>f</sup></b>	44.79	53.76	+++	nd	nd	no	18
R177K	63.19	53.33	<b>59.79</b>	59.79	69.46	+	≈	≈	≈	18
R177Q*	<b>52.40</b>	53.46	52.39	52.40	57.48	++	≈	≈	≈	18
Y178C <sub>ADRP</sub>	74.16	<b>72.14</b>	57.29	72.14	86.49	—	nd	no	no	37, 38
Y178N <sub>ADRP</sub>	70.92	62.49	<b>64.81</b>	64.81	80.49	+	nd	nd	nd	39
P180A <sub>ADRP</sub>				38.70	34.32	+++	no	nd	nd	40
G182S <sub>ADRP</sub>	<b>70.35</b>	71.76	65.73	70.35	81.82	—	nd	—	nd	41, 42
Q184P <sub>ADRP</sub>					67.97	78.49	+	nd	nd	43
C185S	<b>66.44</b>	79.25	66.29	66.44	78.69	+	≈	≈	≈	44, 45
S186P <sup>*</sup> <sub>ADRP</sub>					52.54	61.88	++	nd	no	38, 41
S186W <sup>*</sup> <sub>ADRP</sub>	76.46	53.16	<b>74.50</b>	74.50	90.98	+	nd	nd	nd	46
C187A					67.55	81.22	+	—	nd	—
C187Y <sub>ADRP</sub>	<b>71.59</b>	70.02	73.99	71.59	92.09	—	nd	nd	no	45
G188E <sub>ADRP</sub>	70.05	<b>67.79</b>	65.57	67.79	80.43	+	nd	—	nd	41
G188R <sup>**</sup> <sub>ADRP</sub>	72.53	74.77	<b>72.83</b>	72.83	88.22	+	no	no	no	38, 40, 41, 47
D190A <sup>*</sup> <sub>ADRP</sub>					43.95	51.4	+++	—	nd	—
D190C		51.72	<b>45.60</b>	45.60	53.68	+++	nd	nd	no	18
D190E	<b>77.72</b>	80.14	75.38	77.72	89.55	—	nd	nd	no	18
D190G <sup>*</sup> <sub>ADRP</sub>					40.88	49.03	+++	nd	no	18, 37
D190N <sup>*</sup> <sub>ADRP</sub>	60.09	<b>56.79</b>	45.35	56.79	68.64	++	≈	≈	≈	18, 38
D190Y <sup>*</sup> <sub>ADRP</sub>	—	<b>43.11</b>	53.26	43.11	50.73	+++	nd	no	no	18, 38

<sup>a</sup> Spontaneous (ADRP) and artificial mutants of rhodopsin EL2. A single asterisk means that the energy of the folded state is about 0.5 RT units higher than that of the wild type, whereas double asterisks mean that the lowest energy basin is shifted in between two and three interstrand H-bonds. <sup>b</sup> Average interstrand H-bond probability (HB1–4<sub>avg</sub>) derived from simulation on the mutated side chain rotamer from the D&K library.<sup>29</sup> <sup>c</sup> HB1–4<sub>avg</sub> index derived from simulation on the mutated side chain rotamer from the P&R library.<sup>30</sup> <sup>d</sup> HB1–4<sub>avg</sub> index derived from simulation on the mutated side chain rotamer from the Sut library.<sup>31</sup> <sup>e</sup> Selected HB1–4<sub>avg</sub> index. <sup>f</sup> Fraction of native-like structures, i.e., those characterized by a C<sub>α</sub>-rmsd  $\leq 2$  Å from the native structure. <sup>g</sup> Predicted misfolding effect, based upon REMD simulations. In detail, the symbols “+++”, “++”, “+”, and “—” stand, respectively, for misfolding, moderately misfolding, low misfolding, and non-misfolding. <sup>h</sup> Preservation of the native fold by mutation, according to in vitro experiments. The symbols “no”, “—”, “≈”, and “nd” stand, respectively, for unfolded, misfolded, folded like the wild type, and not determined. <sup>i</sup> Maintenance of the native expression by mutation. The symbols “no”, “—”, “≈”, and “nd” stand, respectively, for not expressed, expressed lower than the wild type, expressed like the wild type, and not determined. <sup>j</sup> Preservation of the native retinal binding ability by mutation. The symbols “no”, “—”, “≈”, and “nd” stand, respectively, for no binding, lower binding than the wild type, wild type-like binding, and not determined. <sup>k</sup> Source of information concerning in vitro data. <sup>l</sup> In bold are the HB1–4<sub>avg</sub> indices selected as the closest to the average value from three independent simulations.

as to an extracellular N-term and an intracellular C-term.<sup>6</sup> In the dark, inactive state of rhodopsin, the chromophore is 11-*cis*-retinal, forming a Schiff base with K296 in H7 (“H” stands for helix). Absorption of a photon provides rhodopsin with the energy to form the active state. Three phases of the activation process can be distinguished: (1) light-induced *cis*–*trans* isomerization of the retinal, (2) thermal relaxation of the retinal–protein complex, and (3) the late equilibria that are affected by the interaction of rhodopsin with the G protein.<sup>7,8</sup> The latter state, metarhodopsin II (Meta II or MII), is in equilibrium with Meta I (MI), which derives from the lumirhodopsin state (LUMI) following a shift of the protonated Schiff base (PSB) from the counterion E113 (in H3) to E181 (in EL2).<sup>7,9</sup>

The spectrum of biochemical and cellular properties of rhodopsin mutations associated with ADRP is quite wide and includes six different classes.<sup>1–3</sup>

ADRP rhodopsin mutations are essentially located in the N-term, EL1, EL2, and the seven-helix bundle. With respect to the extracellular domains, pathogenic mutations concentrate essentially in EL2 (i.e., the S176-T198 sequence), which contains a highly stable twisted β-hairpin that lays alongside the retinal chromophore. Very recent solid-state NMR determinations support the high stability of this rhodopsin segment that seems to change position, rather than conformation, following photoactivation of the pigment.<sup>10</sup> Moreover, five out of the nine amino acid residues predicted to participate in the stability core of rhodopsin by either one

or two different computational methods,<sup>11</sup> i.e., R177<sup>(1)</sup>, P180<sup>(4)</sup>, Q184<sup>(8)</sup>, C185<sup>(9)</sup>, S186<sup>(10)</sup>, and C187<sup>(11)</sup> (each amino acid in the peptide is labeled by two numbers; the first number is the sequential one, whereas the number in parentheses indicates the position of the amino acid residue in the 14-residue β-hairpin), belong to EL2, thus emphasizing the fundamental role of this loop in rhodopsin folding.<sup>11</sup>

This study is part of an ambitious project aimed at structurally characterizing, through molecular simulations, the majority of spontaneous rhodopsin mutations. Within this project, computational protocols are defined ad hoc on the basis of the structural localization and the biochemical classification, if any, of each mutation. In this respect, the 40 mutations considered in this study lay in the β-hairpin portion of EL2 and comprise 15 ADRP mutants, the majority of which falls in the biochemical class II, as they fail to express at the plasma membrane and/or to reconstitute with 11-*cis*-retinal, indicative of variable defects in protein folding and/or stability (Tables 1 and 2 and references therein). The simulated mutants include also all the 19 possible substitutions of E181<sup>(5)</sup>, comprising one ADRP mutation (i.e., E181K) (Table 2).

The effects of the 40 natural and artificial mutations on the structural stability of rhodopsin EL2 were, therefore, studied through parallel replica exchange molecular dynamics (REMD) simulations using the FACTS (fast analytical continuum treatment of solvation) implicit solvent model implemented in the CHARMM biomolecular simulation

**Table 2.** Computational and in Vitro-Determined Indices Concerning Spontaneous and Artificial Mutants of Rhodopsin E181

mutant <sup>a</sup>	D&K <sup>c</sup> (%)	P&R <sup>d</sup> (%)	Sut <sup>e</sup> (%)	HB1–4 <sub>avg</sub> <sup>f</sup>	rmsd <sup>g</sup> (%)	EL2 misfolding <sup>i</sup>	expr. <sup>j</sup> (%)	λ <sub>max</sub> <sup>k</sup> dark (nm)	λ <sub>max</sub> <sup>k</sup> light (nm)	HA <sup>l</sup> react (min)	MII <sup>m</sup> half-life (min)
WT				77.44	88.66		100	501	382	5440 ± 170	12.5 ± 0.5
E181A*				66.37	80.82	+	30–50	499	386	60.1 ± 4.0	17.3 ± 0.7
E181C	72.68	<b>70.70<sup>n</sup></b>	66.97	70.70	80.41	–	30–50	499	383	21.7 ± 2.0	16.7 ± 0.5
E181D	47.50	<b>56.27</b>	60.76	56.27	61.09	+	>50	497	383	1733 ± 357	7.7 ± 0.2
E181F	58.29	78.98	<b>72.45</b>	72.45	81.52	–	30–50	501	375	8.4 ± 0.4	52 ± 18
E181G*				42.74	43.84	+++	30–50	500	384	20.1 ± 4.3	6.5 ± 0.1
E181H*	<b>62.35</b>	67.33	58.65	62.35	76.17	+	30–50	497	388	24.7 ± 5.0	10 ± 2.0
E181I	<b>76.99</b>	75.19	79.09	76.99	86.99	–	30–50	501	384	1.8 ± 0.2	9.0
E181K* <sup>b</sup> <sub>ADRP</sub>	45.97	70.70	<b>70.28</b>	70.28	82.54	+	<20	nd	nd	nd	nd
E181L*	70.85	<b>70.37</b>	63.26	70.37	81.47	+	30–50	502	384	nd	9.6 ± 0.9
E181M	54.19	<b>73.94</b>	75.28	73.94	86.24	–	>50	500	383	5.0 ± 0.6	15.4 ± 0.7
E181N	<b>65.91</b>	54.07	71.38	65.91	76.63	+	30–50	500	384	25.1 ± 3.0	11.0 ± 1.0
E181P** <sup>b</sup>				43.11	52.07	+++	<20	nd	nd	nd	nd
E181Q	65.89	<b>74.41</b>	77.90	74.41	90.13	–	>50	508/5	386	280 ± 9.0	5.2 ± 0.2
E181R* <sup>b</sup>	65.42	<b>65.13</b>	63.18	65.13	77.41	+	<20	nd	nd	nd	nd
E181S*	53.33	<b>54.53</b>	55.26	54.53	64.12	++	30–50	500	383	23.3 ± 1.7	13.0 ± 0.1
E181T*	42.77	<b>51.22</b>	68.86	51.22	62.73	++	30–50	502	383	6.0 ± 0.6	17.3 ± 0.7
E181V	68.88	61.43	<b>61.99</b>	61.99	73.34	+	30–50	489	383	5.7 ± 0.7	12.0 ± 0.5
E181W	60.61	<b>61.93</b>	74.25	61.93	68.97	+	30–50	502	376	111 ± 4.0	27.5 ± 1.0
E181Y*	65.62	<b>73.27</b>	73.50	73.27	91.61	+	30–50	501	392	9.8 ± 0.8	6.9 ± 0.1

<sup>a</sup> Spontaneous (ADRP) and artificial mutants of E181.<sup>36,41,49</sup> A single asterisk means that the energy of the folded state is about 0.5 RT units higher than that of the wild type, whereas double asterisks mean that the lowest energy basin is shifted in between two and three interstrand H-bonds. <sup>b</sup> Mutant did not bind 11-cis-retinal to form a stable pigment.<sup>36</sup> <sup>c</sup> Average interstrand H-bond probability (HB1–4<sub>avg</sub>) derived from simulation on the mutated side chain rotamer from the D&K library.<sup>29</sup> <sup>d</sup> HB1–4<sub>avg</sub> index derived from simulation on the mutated side chain rotamer from the P&R library.<sup>30</sup> <sup>e</sup> HB1–4<sub>avg</sub> index derived from simulation on the mutated side chain rotamer from the Sut library.<sup>31</sup> <sup>f</sup> Selected HB1–4<sub>avg</sub> index. <sup>g</sup> Fraction of native-like structures, i.e. those characterized by a C<sub>α</sub>-rmsd ≤ 2 Å from the native structure. <sup>h</sup> Predicted misfolding effect, based upon REMD simulations. In detail, the symbols “+++”, “++”, “+”, and “–” stand, respectively, for misfolding, moderately misfolding, low misfolding, and non-misfolding. <sup>i</sup> Level of expression compared to the wild type; “nd” stands for not determined.<sup>36</sup> <sup>j</sup> Absorbance wavelength in the dark; “nd” stands for not determined.<sup>36</sup> <sup>k</sup> Absorbance wavelength upon illumination; “nd” stands for not determined.<sup>36</sup> <sup>l</sup> Rates of reaction with hydroxylamine; “nd” stands for not determined.<sup>36</sup> <sup>m</sup> MII decay rates; “nd” stands for not determined.<sup>36</sup> <sup>n</sup> In bold is the representative trajectory of each mutant that was selected to produce HB1–4<sub>avg</sub> values closest to the average value from the three independent simulations.

package.<sup>12</sup> The computational protocol was first probed and optimized on the protein G C-terminal amino acid peptide, which is known to fold into a stable 4:4 type β-hairpin and has been widely used as a model system to test REMD-based folding protocols.<sup>13–17</sup>

This study describes the strategy employed to individuate and differentiate EL2 rhodopsin mutations that would affect the intrinsic stability of the native β-hairpin from mutations expected to variably impair native contacts between the loop and the surrounding receptor domains.

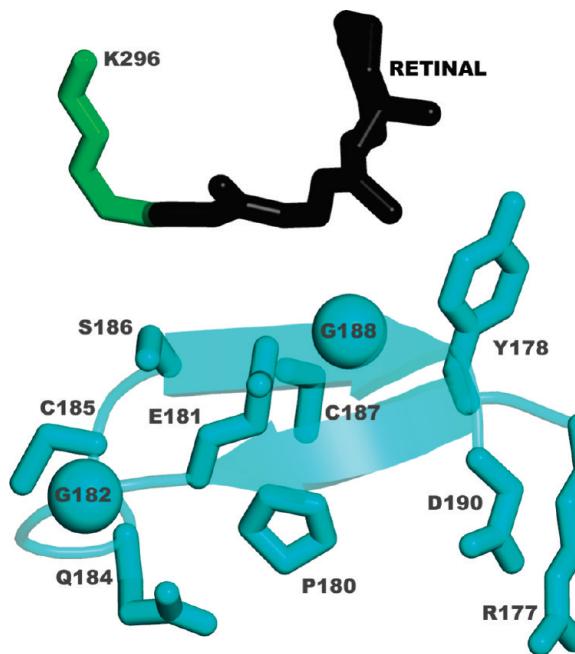
## 2. Computational Details

### 2.1. Structural Analysis of EL2 and Its Mutation Sites.

EL2 of rhodopsin (i.e., the S176-T198 sequence) folds into a highly stable twisted β-hairpin (i.e., the <sup>177</sup>R<sup>YIPEG-MQCSCGID<sup>190</sup></sup> sequence) that makes extensive contacts with the other extracellular domains, on one side, and with the retinal chromophore, on the other one (Figure 1). In particular, it forms a four-stranded β-sheet with the N-terminal tail of the receptor protein, thus making a plug that shields the chromophore from the extracellular (intradiscal) solvent.<sup>6</sup> A cysteine residue from this loop, C187<sup>(11)</sup>, is engaged in a disulfide bridge with C110 on the N-terminal end of H3, thus contributing to the stability of rhodopsin. The turn of the EL2 β-hairpin corresponds to the <sup>182</sup>GMQC<sup>185</sup> amino acid stretch, whereas the N-terminal and C-terminal strands are, respectively, made by the <sup>178</sup>YIPE<sup>181</sup> and <sup>186</sup>SCGI<sup>189</sup> amino acids stretches. The EL2 amino acids directly involved in interactions with retinal belong to the

C-terminal strand and include only S186<sup>(10)</sup>, G188<sup>(12)</sup>, and I189<sup>(13)</sup>. In folded rhodopsin, R177<sup>(1)</sup> and D190<sup>(14)</sup>, the first and last amino acids in the β-hairpin, respectively, are involved in a salt bridge interaction expected to contribute to the stability of the loop.<sup>18</sup>

The 40 EL2 mutations screened in this study comprise 15 ADRP-linked and 25 artificial mutations (Tables 1 and 2). The ADRP-linked mutations concern nonconservative mutations of Y178<sup>(2)</sup> in cysteine and asparagine, P180<sup>(4)</sup> in alanine, E181<sup>(5)</sup> in lysine, G182<sup>(6)</sup> in serine, Q184<sup>(8)</sup> in proline, S186<sup>(10)</sup> in proline and tryptophan, C187<sup>(11)</sup> in tyrosine, G188<sup>(12)</sup> in glutamate and arginine, and D190<sup>(14)</sup> in glycine, asparagine, and tyrosine (Tables 1 and 2), which represent the majority of the amino acids of the EL2 β-hairpin. In contrast, the 25 artificial mutations include cysteine, lysine, and glutamine substitutions for R177<sup>(1)</sup>, serine substitution for C185<sup>(9)</sup>, and 18 different amino acid substitutions for E181<sup>(5)</sup> (i.e., all the possible natural amino acid substitutions except for the ADRP-linked lysine substitution). The E181<sup>(5)</sup> amino acid residue, which points toward the center of the retinal polyene chain, is, indeed, the most investigated residue in the loop. Photochemistry studies indicated the involvement of such amino acid in the counterion switch during the photoactivation of rhodopsin.<sup>9</sup> In fact, according to earlier studies, such a switch was suggested to occur by the proton transfer from E181<sup>(5)</sup>, protonated in the dark state, to E113 (in H3), through an evolving H-bond (HB) network formed primarily with residues of EL2.<sup>9</sup> In fact, in the crystal structures of dark rhodopsin,<sup>19</sup> as well as of the BATHO and LUMI



**Figure 1.** Drawing of the retinal and the EL2  $\beta$ -hairpin extracted from the 1U19 crystal structure of dark rhodopsin. The retinal chromophore in its 11-cis conformation linked to K296 as a Schiff base is drawn as sticks. The retinal is black, the covalently bound K296 is green, and the EL2 cyan. The side chains of the  $\beta$ -hairpin amino acids targeted by in silico mutagenesis are represented as sticks. In this respect, the positions of G182 and G188 are indicated by two spheres positioned on the  $C_{\alpha}$ -atoms. Drawings were done by means of PyMOL 0.99 (<http://pymol.sourceforge.net/>).

photointermediates,<sup>20,21</sup> this anionic amino acid is hydrogen bonded to both Y192 (in EL2) and Y268 (in H6) and is also involved in a water-mediated HB with both S186<sup>(10)</sup> and E113 (in H3) (Figure 1). More recent studies suggest that a change of the PSB counterion from the dark state to MI would not necessarily require a proton transfer.<sup>22</sup>

Collectively, the structural analysis of the EL2  $\beta$ -hairpin in the context of folded rhodopsin in its dark state shows that seven of the spontaneous or artificial mutation sites considered in this study, R177<sup>(1)</sup>, P180<sup>(4)</sup>, G182<sup>(6)</sup>, Q184<sup>(8)</sup>, C185<sup>(9)</sup>, S186<sup>(10)</sup>, and C187<sup>(11)</sup>, belong to the extracellular–TM interface of the retinal binding pocket or to the interface between EL2 and the surrounding domains.

**2.2. REMD Simulation Setup: The GB1 Model System.** The computational protocol was set on the  $\beta$ -hairpin from the C terminus (i.e., G44-E56 segment) of Streptococcal protein G (PDB code 2gb1, i.e., ACE-<sup>41</sup>GEWTYDDATK-TFTVTE<sup>56</sup>-NME, herein named as GB1), patched with acetyl and methyl-amino groups at the N- and C-terminals. The finally selected setup consists of REMD simulations by means of the CHARMM force field (in all-atom mode),<sup>23</sup> using the FACTS implicit solvent model.<sup>12</sup> REMD simulations were carried out using an in-house developed program that calls internally the CHARMM package.<sup>24,25</sup> The algorithm is implemented in C language with the use of the message-passage-interface (MPI) libraries.

A total of 20 replicas were simulated by Langevin dynamics with a friction coefficient of  $5.0 \text{ ps}^{-1}$  and temper-

ature values spanning the interval from 270 to 690 K (270, 283, 298, 313, 328, 345, 363, 381, 400, 421, 442, 465, 488, 513, 539, 566, 594, 625, 656, and 690 K). Each replica was thermalized at its respective temperature for 20 ps with a time step of 2 fs. REMD sampling was carried out for a total of 5 ns using a time step of 2 fs. Transitions between adjacent temperatures were attempted every 0.2 ps and protein configurations were saved every 0.4 ps, giving a total of 0.25 million configurations. This setup was selected following a significant number of trials. The latter included tests of (a) an alternative implicit solvation model, i.e., GBSW;<sup>26</sup> (b) a number of different temperature sets; (c) longer REMD sampling, i.e., 10 and 20 ns; (d) more frequent coordinate saving, i.e., every 0.2, 0.1, and 0.05 ps; and (e) different friction coefficients, i.e., 2.0, 1.0, and  $0.5 \text{ ps}^{-1}$ .

Both the charged and neutral form of the peptide, i.e. protonated at E42, D46, and E56, were considered, leading to 256 and 259 atoms per replica, respectively. The results of simulations on the neutral and charged peptide are comparable, although the ones on the neutral forms are more realistic. Therefore, the results herein presented refer to simulations on the neutralized peptide.

The selected REMD setup is the best compromise between speed of simulation and consistency with NMR determinations as well as explicit water simulations.<sup>14</sup> In this respect, prolonging REMD simulations from 5 to 20 ns did not change significantly the distribution of conformational states. In fact, the average fraction of native contacts at 313 K following 5, 10, and 20 ns REMD simulations are, respectively, 67.2, 68.1, and 68.0. Thus, selection privileged the shortest simulation time. Similarly, denser sampling did not change the outcome. In fact, coordinate saving every 0.4, 0.2, 0.1, and 0.05 ps led, respectively, to 67.2, 68.5, 68.5, and 68.5 fractions of native contacts. On the same line, friction coefficients did not affect the outcome of simulations. Indeed, simulations at friction coefficients equal to 0.1, 0.5, 1.0, 2.0, and  $5.0 \text{ ps}^{-1}$  resulted, respectively, in 67.8, 63.7, 65.1, 68.6, and 67.2 fractions of native contacts. Final selection concerned a friction coefficient of  $5 \text{ ps}^{-1}$ , as it was closer than the others to the viscosity of water ( $10 \text{ ps}^{-1} < \gamma < 100 \text{ ps}^{-1}$ ) although still in the low viscosity regime.<sup>27</sup>

As native contacts stabilizing the GB1 peptide during trajectory analysis we took the following 26 pairs, by considering a distance cutoff of 7.6 Å between the side chain geometric centers, according to previous reports:<sup>28</sup> 4–5, 5–6, 8–9, 15–16, 1–3, 2–4, 3–5, 4–6, 5–7, 6–8, 9–11, 11–13, 12–14, 13–15, 6–9, 7–10, 6–11, 5–11, 4–11, 5–12, 5–13, 3–12, 4–13, 2–13, 3–14 and 2–15.

**2.3. REMD Simulation Setup: The Rhodopsin EL2.** The computational protocol set on the GB1 peptide was then extended to the EL2  $\beta$ -hairpin of rhodopsin, a 14 amino acid peptide from R177<sup>(1)</sup> to D190<sup>(14)</sup> extracted from the crystal structure of dark rhodopsin (PDB code 1U19)<sup>19</sup> and patched with acetyl and methyl-amino groups at the N- and C-terminals, respectively. Comparative REMD simulations were carried out on the wild type and on 40 mutated forms (Tables 1 and 2). The wild type was simulated both in the charged and neutral states, i.e., carrying E181<sup>(5)</sup> in its protonated form. Controversial data support both protonation

and deprotonation of E181<sup>(5)</sup> in the dark state, whereas the deprotonated (charged) form would characterize the active states starting from MI.<sup>9,22</sup> The results shown in this study refer to simulations with protonated E181<sup>(5)</sup>.

For wild type EL2, the total number of atoms per replica was 218. For each replacing amino acid, three different input rotamers were subjected to REMD simulations. These starting conformations were assigned according to the Dunbrack and Karplus (D&K),<sup>29</sup> Ponder and Richards (P&R),<sup>30</sup> and Sutcliffe (Sut)<sup>31</sup> rotamer libraries.

The transition acceptance ratio was around 45%.

**2.4. REMD Simulation Analyses.** For both GB1 and EL2  $\beta$ -hairpins, the energy landscape or potential of mean force (PMF) was calculated from the normalized population densities as previously described<sup>15</sup>

$$\text{PMF} = -\log P(X_1, X_2)$$

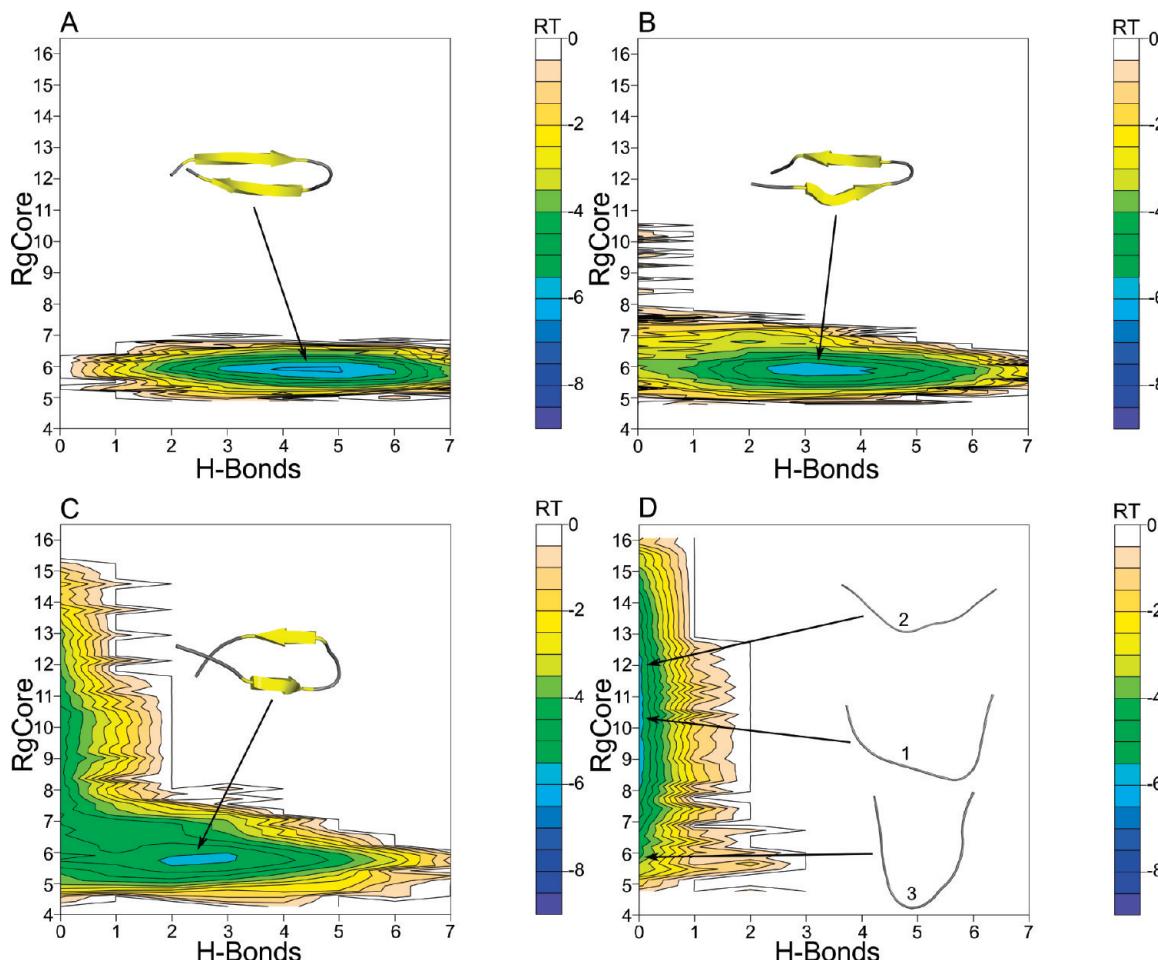
where  $P(X_1, X_2)$  is the normalized probability as a function of  $X_1$  and  $X_2$ , and  $X_1$  and  $X_2$  are parameter sets describing the peptide conformations. In this study, such parameters are the native  $\beta$ -sheet hydrogen bonds (H-bonds) and the geometric radius of gyration of the hydrophobic core (RgCore). In detail, for the 2GB1, the seven native  $\beta$ -sheet hydrogen bonds are E42:N–H/T55:O, T55:N–H/E42:O, T44:N–H/T43:O, T43:N–H/T44:O, D46:N–H/T51:O, T51:N–H/D46:O, and K50:N–H/D47:O. A hydrogen bond was counted if the distance between the O and the N atoms was less than 3.5 Å, and the angle formed by the three atoms (N, H, and O) was larger than 150°. Furthermore, the RgCore was computed on the side chain atoms of the four hydrophobic residues W43, Y45, F52, and V54, to allow for comparisons with the results of previous computational studies.<sup>13,14,16</sup> For rhodopsin EL2, the five native  $\beta$ -sheet hydrogen bonds are D190:N–H/R177:O, I179:N–H/G188:O, G188:N–H/I179:O, E181:N–H/S186:O, and Q184:N–H/E181:O. Rg calculations were limited to the hydrophobic core amino acids rather than to the whole peptide. The RgCore was, thus, computed on the side chain atoms of the following four residues: Y178, P180, C187, and I189.

Cluster analysis of the REMD trajectories was based on the QT clustering algorithm<sup>32</sup> implemented in the Wordom software.<sup>33</sup> In this case study, the algorithm first calculated the  $C_{\alpha}$ -atom root mean square deviation ( $C_{\alpha}$ -rmsd) for each superimposed pair of frames and then it computes the number of neighbors for each frame by using a threshold  $C_{\alpha}$ -rmsd. The frame with the highest number of neighbors is considered as the center of the first cluster. All the neighbors of this configuration are removed from the ensemble of configurations to be counted only once. The center of the second cluster is then determined in the same way as the first cluster, and this procedure is repeated until no more clusters can be found.

### 3. Results

**3.1. The C-Terminal  $\beta$ -Hairpin of Protein G as a Model System.** The computational protocol was set on the C terminal  $\beta$ -hairpin of Streptococcal protein G (i.e., the G44-E56 segment herein named as GB1), extensively used as a

model system to probe *in silico*  $\beta$ -hairpin folding protocols. REMD simulations in implicit solvent, spanning 20 temperatures from 270 to 690 K, produced a free energy contour map at 313 K characterized by a wide unique global energy minimum corresponding to the native-like state, i.e., with five  $\beta$ -sheet H-bonds and a RgCore around 5.92 Å (Figure 2). The latter index was computed on the side chain atoms of W43<sup>(3)</sup>, Y45<sup>(5)</sup>, F52<sup>(12)</sup>, and V54<sup>(14)</sup>. These results overlap significantly with those of explicit water simulations by Zhou and co-workers.<sup>14</sup> Along this line, the probability of finding all five interstrand HBs satisfied reaches the maximum value at temperatures between 270 and 313 K, whereas it drops at temperature values above 520 K (Figure 3A). In this respect, the inner interstrand H-bonds (HB3, HB4, and HB5, Figure 3A) are more persistent than the most external ones (HB1 and HB2, Figure 3A). The HBs that involve the turn (HB6 and HB7), in particular HB7, are the less persistent ones. In summary, at 313 K, the rank order of each H-bond probability (expressed as percentages with respect to the total number of frames) is HB4 > HB5 > HB3 > HB2 > HB6 > HB1 > HB7 (the values being 87.26, 85.62, 82.03, 65.73, 62.15, 47.21, and 14.3, respectively). Interestingly, this rank order is overlapping with that from explicit water simulations.<sup>14</sup> The average HB probability at 313 K from our simulations is 63.47%, slightly higher but characterized by a higher decrease rate compared to that from explicit water simulations.<sup>14</sup> In fact, according to our data, such an index decreases to around 0 at 690 K (i.e., 0.27), whereas it never drops to zero following explicit water simulations.<sup>14</sup> In line with the trend of the HB probability, the  $\beta$ -hairpin population, as accounted for by the fraction of native contacts (see the experimental procedures for its definition), is 67.87% at 270 K, 67.52% at 298 K, and 67.22% at 313 K (Figure 3B). These data are similar to those from explicit water simulations with the OPLSAA all-atom force field, which found populations of native contacts of about 71% at 270 K and 66% at 310 K.<sup>14</sup> Thus, near the biological temperature, our simulations, similar to simulations in explicit water,<sup>14</sup> found a  $\beta$ -hairpin population of the GB1 peptide in reasonable agreement with *in vitro* experiments. Indeed, NMR data found a population of about 80% at 270 K, 50% at 300 K, and 40% at 310 K (as shown above, the fraction of native contacts from computational experiments does not change significantly in the 270–313 K range). The highest discrepancies with NMR data concern temperatures higher than 313 K. Indeed, the temperature-dependent decrease rate of the fraction of native contacts from our simulations is quite slower than that from NMR determinations, though faster than that from explicit water simulations.<sup>14</sup> In fact, according to our computations, the fraction of native contacts decreases significantly after 420 K, reaching the lowest value of 26.89% at 690 K (Figure 3B). In contrast, according to NMR data, the population of native  $\beta$ -hairpin is already 0 around 360 K, whereas the fraction of native contacts from explicit water simulations is still above 35% at 690 K.<sup>14</sup> In line with the trend of the fraction of native contacts, the fraction of native-like structures, i.e., those characterized by a  $C_{\alpha}$ -rmsd  $\leq$  2 Å from the native structure, is 56.4% at 313 K, whereas it



**Figure 2.** Free energy contour maps at different temperatures of the GB1 peptide folding versus the number of native  $\beta$ -sheet hydrogen bonds and the hydrophobic core radius of gyration (RgCore). The seven native  $\beta$ -sheet hydrogen bonds are E42:N–H/T55:O, T55:N–H/E42:O, T44:N–H/T43:O, T43:N–H/T44:O, D46:N–H/T51:O, T51:N–H/D46:O, and K50:N–H/D47:O. A hydrogen bond was counted if the distance between the O and the N atoms was less than 3.5 Å and the angle formed by the three atoms (N, H, and O) was larger than 150°. The geometric radius of gyration (excluding any mass weighing) of the hydrophobic core was computed on the side chain atoms of the four hydrophobic residues W43, Y45, F52, and V54. The contours are spaced at intervals of 0.5 RT. Cartoon representations of the cluster centers from the lowest energy basins (indicated by arrows) are also shown. In detail, (A) the free energy landscape at 313 K is characterized by one broad energy basin whose representative structure holds five H-bonds and a RgCore = 5.58 Å, (B) the free energy landscape at 400 K is characterized by one broad energy basin whose representative structure holds three H-bonds and a RgCore = 5.73 Å, (C) the free energy landscape at 465 K is characterized by one broad energy basin whose representative structure holds three H-bonds and a RgCore = 5.43 Å, and (D) the free energy landscape at 690 K is characterized by a number of energy basins corresponding to zero H-bonds. The representative structures extracted from these basins hold the following RgCore values: (1) 10.37 Å, (2) 12.04 Å, and (3) 5.96 Å.

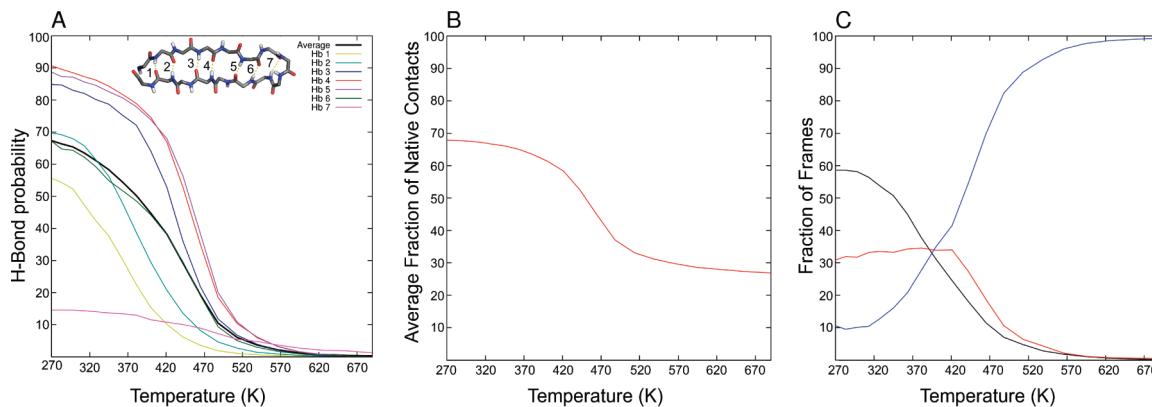
progressively decreases with the increase in temperature (Figure 3C).

Collectively, as already discussed in previous studies,<sup>14</sup> high-temperature discrepancies between in vitro and in silico data are shared by computational experiments using different force fields (i.e., CHARMM, OPLSAA, and AMBER) and different treatments of the solvent and may depend on many factors, including force field parameters, water models (for explicit water simulations), the employment of high pressures at high temperatures in constant volume simulations, or the lack of water-density-dependent parameters in implicit solvent simulations.

**3.2. REMD Simulations on Rhodopsin EL2.** The computational protocol optimized on the GB1 peptide was employed to investigate the effects of 40 point mutations

on the structural stability of rhodopsin EL2 (the S176-T198 sequence) that holds a highly stable twisted  $\beta$ -hairpin (the <sup>177</sup>RYIPEGMQCSCGID<sup>190</sup> sequence). The experimental set consists of 15 spontaneously occurring mutants associated with ADRP and 25 artificial mutants (Tables 1 and 2). Comparative REMD simulations were, hence, carried out on the wild type and the 40 mutant forms.

Similar to the GB1 peptide, the free energy contour map of wild type EL2 at 313 K is characterized by a wide unique global energy minimum corresponding to the native-like state, i.e. with four interstrand H-bonds and a RgCore equal to 5.14 Å (Figure 4). This index was computed on the side chains of Y178<sup>(2)</sup>, P180<sup>(4)</sup>, C187<sup>(11)</sup>, and I189<sup>(13)</sup> from the center (i.e., the most representative frame) of the lowest energy cluster. The lowest energy basin at 313 K progres-



**Figure 3.** Temperature dependence of structural indices. (A) The temperature dependence of the probability of forming the seven native  $\beta$ -sheet H-bonds (E42:N–H/T55:O (HB1), T55:N–H/E42:O (HB2), T44:N–H/T43:O (HB3), T43:N–H/T44:O (HB4), D46:N–H/T51:O (HB5), T51:N–H/D46:O (HB6), and K50:N–H/D47:O (HB7)) is shown. The curves relative to HB1–7 are colored in yellow, cyan, blue, red, violet, green, and magenta, respectively. The thick black line represents the average probability over all native hydrogen bonds. A stick representation of the GB1  $\beta$ -hairpin with the seven native  $\beta$ -sheet H-bonds is shown as well. (B) The average fraction of a set of 26 native contacts as a function of the temperature is shown. The set of contacts includes the following pairs: T44:Y45, Y45:D46, A48:T49, T55:E56, G41:W43, E42:T44, W43:Y45, T44:D46, Y45:D47, D46:A48, T49:T51, T51:T53, F52:V54, T53:T55, D46:T49, D47:K50, D46:T51, Y45:T51, T44:T51, Y45:F52, Y45:T53, W43:F52, T44:T53, E42:T53, W43:V54, and E42:T55. A contact was counted if the distance between the side chain geometrical center of the two residues in each pair was less than 7.5 Å. (C) The temperature dependence of the  $C_{\alpha}$ -rmsd from the starting structure is shown. Curves corresponding to  $C_{\alpha}$ -rmsd thresholds  $\leq 2$  Å,  $> 2$  Å and  $\leq 3$  Å, and  $\geq 3$  Å are colored in black, red and blue, respectively.

sively moves toward zero  $\beta$ -sheet HBs and higher RgCore values with the increases in temperature (Figure 4).

The distorted wild type EL2  $\beta$ -hairpin is characterized essentially by four interstrand HBs (HB1, HB2, HB3, and HB4) and one strand-turn HB (HB5) (Figure 5A). Similar to the GB1  $\beta$ -hairpin, the probability of finding all four interstrand HBs satisfied reaches the maximum value at temperatures between 270 and 320 K, whereas it drops at temperatures above 520 K. Analogously to the GB1  $\beta$ -hairpin, the inner interstrand HBs (HB2, HB3, and HB4, Figure 5A) are more persistent than the most external one (HB1, Figure 5A). The HB5 strand-turn H-bond shows the lowest persistency. In summary, at 313 K, the rank order of each HB probability (expressed as percentages with respect to the total number of frames) is HB2 > HB3 ≈ HB4 > HB1 > HB5 (the values being 82.78, 77.39, 77.15, 72.45, and 3.50, respectively). The average interstrand HB probability (i.e. HB1–4<sub>avg</sub>, that considers only HB1, -2, -3, and -4) at 313 K is 77.44% (Table 1). Consistently, the fraction of native-like structures, i.e., those characterized by a  $C_{\alpha}$ -rmsd  $\leq 2$  Å from the native structure, is 88.66% at 313 K, whereas it progressively decreases with increases in temperature (Figure 5B).

Possible disturbing effects of the 40 EL2 point mutations on the structural features of wild type  $\beta$ -hairpin were, hence, evaluated by comparing the free energy landscapes as well as the probability of interstrand HB formation and the fraction of native-like structures of the mutant trajectories with those of the wild type.

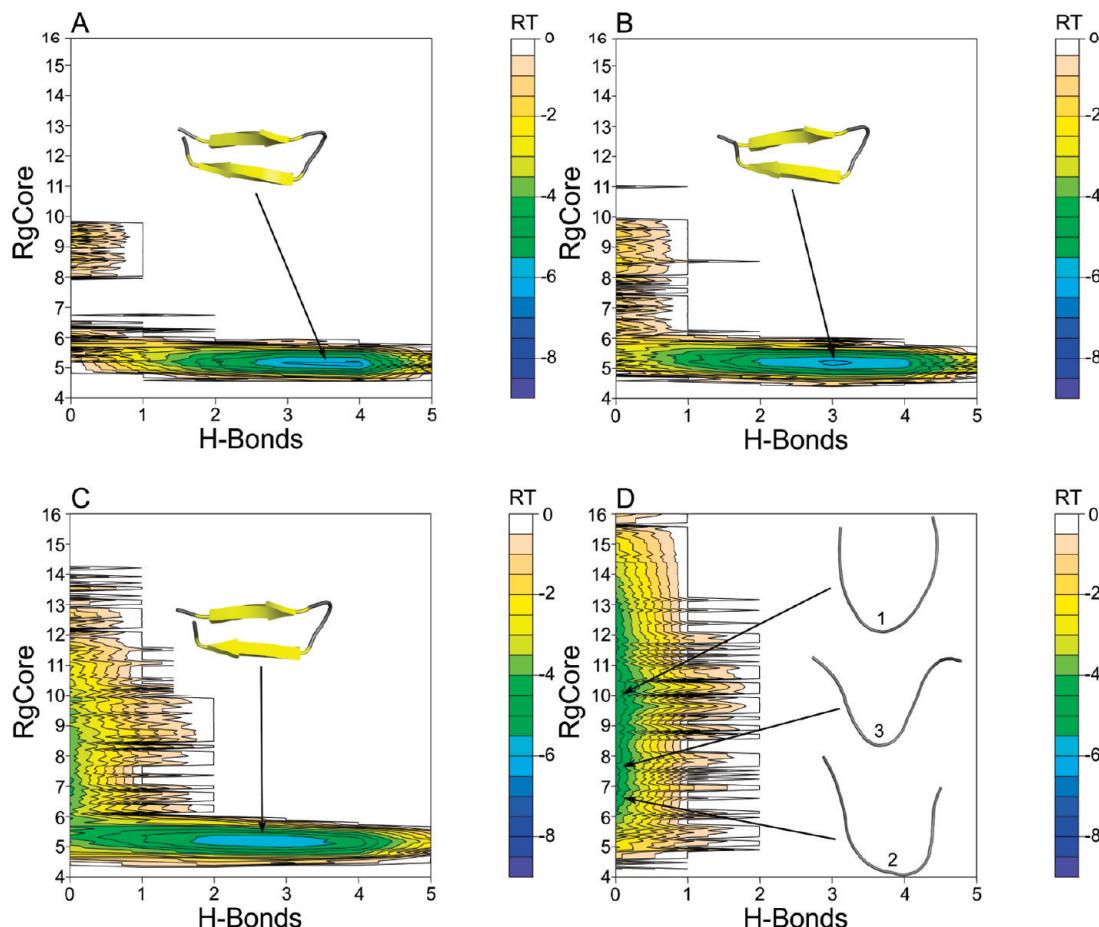
As for the in vitro behavior of the considered mutants, 13 of them are impaired in folding/expression and/or retinal binding (Tables 1 and 2; marked by red color in Figure 6). Furthermore, 17 mutations exert a milder impairing effect on folding/expression (Tables 1 and 2; marked by the yellow

color in Figure 6), whereas seven mutations do not significantly change the expression or chromophore binding compared to the wild type (Tables 1 and 2; marked by the green color in Figure 6). Finally, in three cases, the effect of mutations on the structural stability of rhodopsin is unknown (Tables 1 and 2; marked by the gray color in Figure 6).

For the mutants characterized by rotamers on the replacing amino acid side chains, possible different rotameric states, according to the D&K,<sup>29</sup> P&R,<sup>30</sup> and Sut<sup>31</sup> libraries, were probed as simulation inputs. We then selected as a representative trajectory of each mutant the one that produced HB1–4<sub>avg</sub> values closest to the average value from the three independent simulations (bold numbers in Tables 1 and 2). A strong agreement was generally achieved between the HB1–4<sub>avg</sub> values from at least two of the three independent simulations.

In silico screening showed a spectrum of mutation-induced changes in the HB1–4<sub>avg</sub> index and in the fraction of native-like structures characterizing the wild type at 313 K (i.e., 77.44% and 88.66%, respectively; Figure 6). Given the very high correlation between the fraction of native-like structures and HB1–4<sub>avg</sub> index ( $r = 0.977$ ), we decided to employ the latter as a structural hallmark of mutation effects. Thus, as for HB1–4<sub>avg</sub>, 14 mutants hold wild type-like values (i.e., higher than 70%), 11 mutants hold values between 60% and 70%, seven mutants hold values between 50% and 60%, and eight mutants hold values below 50% (Tables 1 and 2, Figure 6). The latter include the R177C, P180A, E181G, and E181P mutants as well as the nonconservative mutations of D190<sup>(14)</sup> (i.e., A, C, G, and Y substitutions; Tables 1 and 2, Figure 6).

REMD simulations showed also mutation-induced changes in the free energy landscape characterizing the

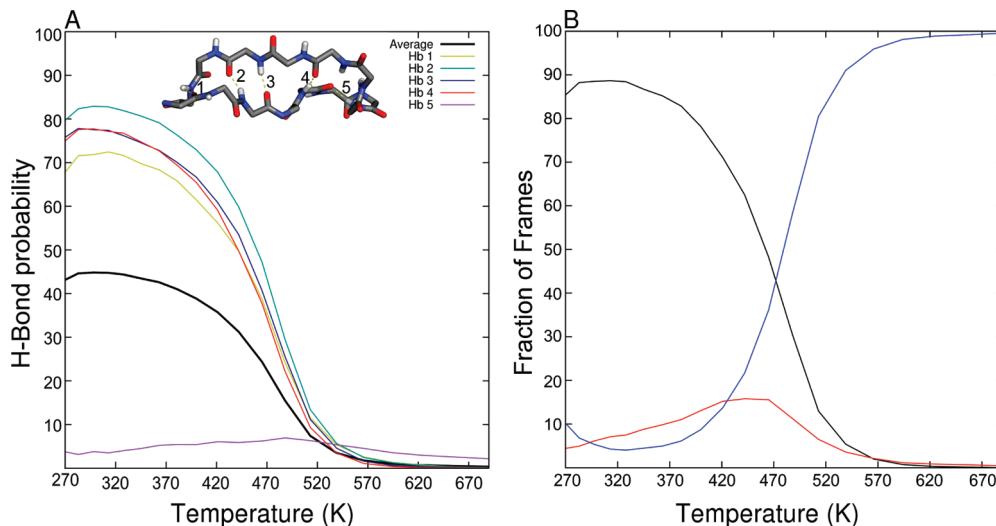


**Figure 4.** Free energy contour maps at various temperatures of rhodopsin EL2 (177–190 sequence) versus the number of native  $\beta$ -sheet hydrogen bonds and the core radius of gyration (RgCore). The five native  $\beta$ -sheet hydrogen bonds are D190:N–H/R177:O, I179:N–H/G188:O, G188:N–H/I179:O, E181:N–H/S186:O, and Q184:N–H/E181:O. The radius of gyration was computed on the side chain atoms of the following four residues: Y178, P180, C187 and I189. Contours are spaced at intervals of 0.5 RT. Cartoon representations of the cluster centers from the lowest energy basins (indicated by arrows) are also shown. In detail, (A) the free energy landscape at 313 K is characterized by one broad energy basin whose representative structure holds four H-bonds and a RgCore = 5.14 Å, (B) the free energy landscape at 400 K is characterized by one broad energy basin whose representative structure holds two H-bonds and a RgCore = 5.33 Å, (C) the free energy landscape at 465 K is characterized by one broad energy basin whose representative structure holds three H-bonds and a RgCore = 5.4 Å, and (D) the free energy landscape at 690 K is characterized by a number of energy basins corresponding to zero H-bonds. The representative structures extracted from these basins hold the following RgCore values: (1) 9.92 Å, (2) 6.57 Å, and (3) 7.88 Å.

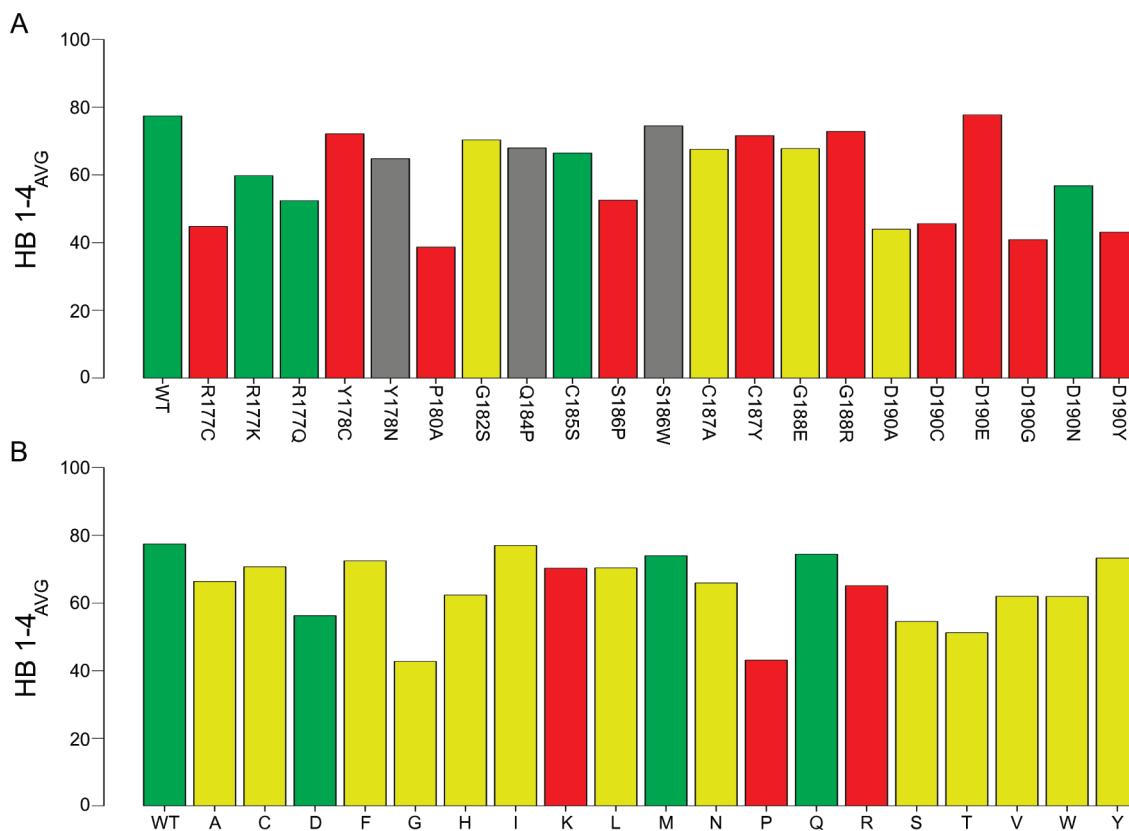
wild type form at 313 K (Figures 6–8, and S1–S7, Supporting Information). Differences between wild type and mutant forms consist of the appearance of alternative higher energy basins at the expense of the native-like state that remains the most populated one. As a consequence, for 17 of the 40 mutants (i.e., marked by an asterisk in Tables 1 and 2) the energy of the folded state is about 0.5 RT units higher than that of the wild type. Furthermore, for the E181P and G188R mutants, the lowest energy basin is shifted in between two and three interstrand HBs, indicative of a misfolding effect of such mutations (marked by two asterisks in Tables 1 and 2). In line with these data, the free energy landscapes of the mutants holding a HB1–4<sub>avg</sub> index lower than 50% are characterized by the appearance of higher energy basins at zero  $\beta$ -sheet HBs and RgCore values above 6 Å (Figures 7, 8, and S1–S7, Supporting Information). These basins generally correspond to

$\alpha$ -helical structures (Figures 7, 8, and S1–S7, Supporting Information).

Collectively, by considering all together the HB1–4<sub>avg</sub> index and the shapes of the free energy landscapes, we could classify EL2 mutants as misfolding (marked as +++ in Tables 1 and 2), moderately misfolding (marked as ++ in Tables 1 and 2), low misfolding (marked as + in Tables 1 and 2), and non-misfolding (marked as – in Tables 1 and 2). In detail, (a) misfolding mutants are characterized by a HB1–4<sub>avg</sub> index below 50% that is generally accompanied by a shift in energy or position of the lowest energy basin (i.e., marked by one or two asterisks, respectively, in Tables 1 and 2); (b) moderately misfolding mutants are characterized by a HB1–4<sub>avg</sub> index between 50% and 60% associated with a shift in energy or position of the lowest energy basin; (c) low misfolding mutants are characterized by a HB1–4<sub>avg</sub> index between 60% and 70%,



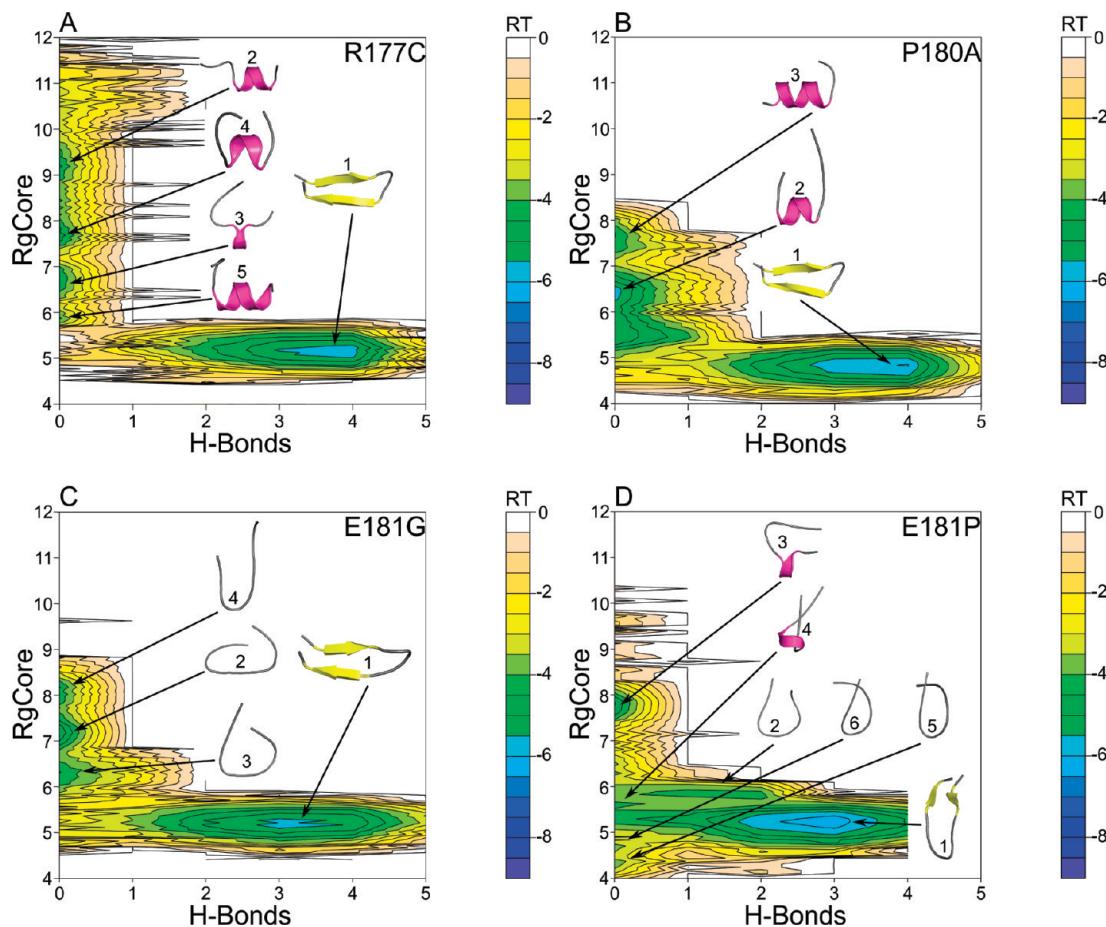
**Figure 5.** Temperature dependence of the probability of forming individual native  $\beta$ -sheet hydrogen bonds and of the  $C_\alpha$ -rmsd for the wild type EL2. (A) The temperature dependence of the probability of forming the five native  $\beta$ -sheet H-bonds (D190:N–H/R177:O (HB1), I179:N–H/G188:O (HB2), G188:N–H/I179:O (HB3), E181:N–H/S186:O (HB4), and Q184:N–H/E181:O (HB5)) is shown. The curves relative to HB1–5 are colored in yellow, cyan, blue, red, and purple, respectively. The thick black line represents the average probability over all native hydrogen bonds. A stick representation of EL2 together with the five native  $\beta$ -sheet hydrogen bonds is shown as well. (B) The temperature dependence of the  $C_\alpha$ -rmsd from the starting structure is shown. Curves corresponding to  $C_\alpha$ -rmsd thresholds  $\leq 2 \text{ \AA}$ ,  $> 2 \text{ \AA}$  and  $< 3 \text{ \AA}$ , and  $\geq 3 \text{ \AA}$  are colored in black, red, and blue, respectively.



**Figure 6.** EL2 wild type and mutant average probabilities of interstrand H-bonds (HB1–4<sub>avg</sub>). The histograms report the average probabilities of the interstrand H-bonds concerning the wild type and all the simulated mutations at EL2 sites other than E181 (A) and all mutations of E181 (B). Color codes refer to in vitro behavior. In detail, (a) red stands for impaired folding/expression and/or retinal binding, (b) yellow indicates a moderate impairing effect on folding/expression and/or chromophore binding, (c) green stands for wild type-like behavior, and (d) gray stands for unknown biochemical effect.

independent of the position and depth of the lowest energy basin, or by HB1–4<sub>avg</sub> above 70% but associated with shifts in energy and/or position of the lowest energy basins; and,

finally, (d) the wild type-like or non-misfolding mutants are characterized by a HB1–4<sub>avg</sub> index above 70% and the lowest energy basin similar to that of the wild type.

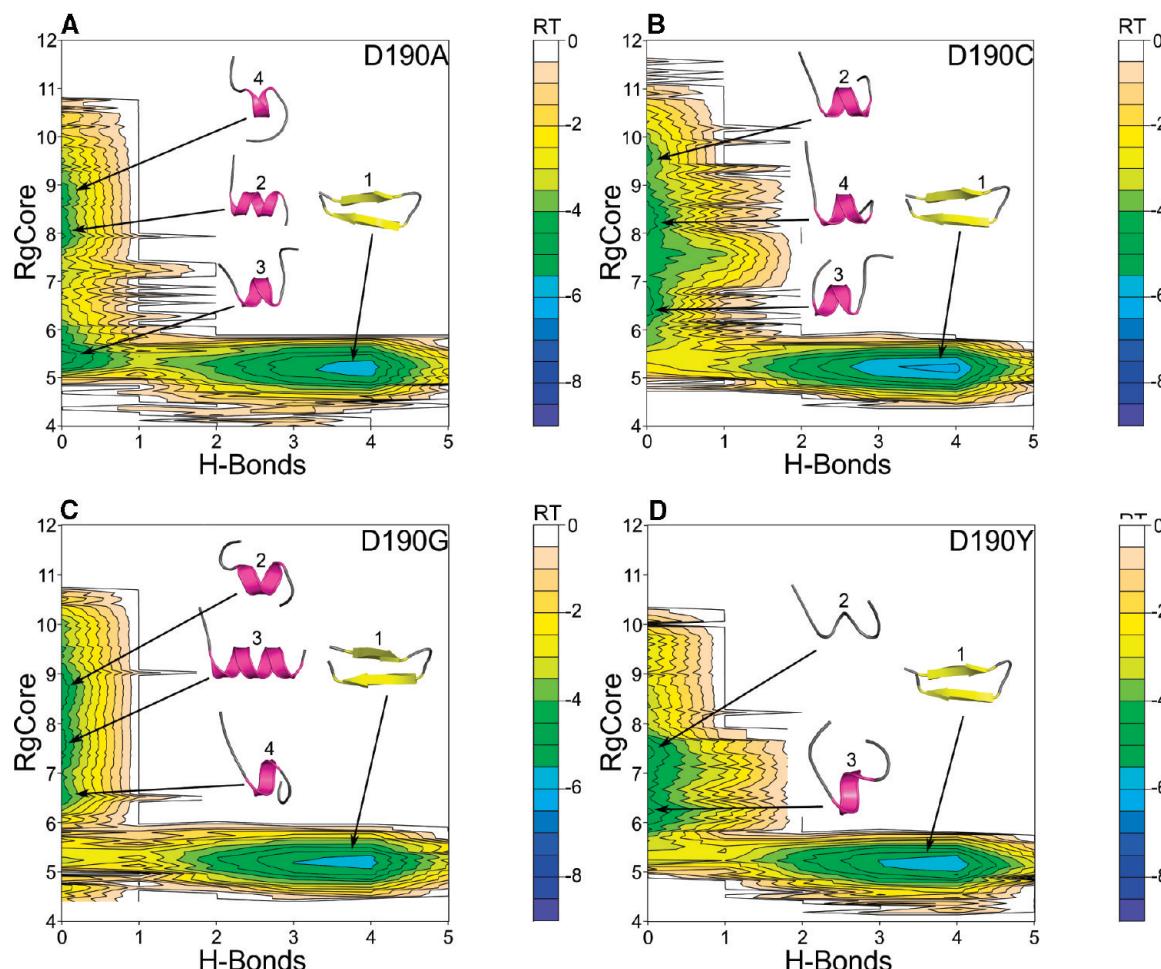


**Figure 7.** Free energy contour maps at 313 K relative to the R177C (A), P180A (B), E181G (C), and E181P (D) mutants, versus the number of native  $\beta$ -sheet hydrogen bonds and the core radius of gyration (RgCore). The general description of this legend is the same as that of Figure 4. The representative structures from each energy basin are shown as cartoons. In detail, (A) for the R177C mutant five structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a native-like  $\beta$ -hairpin, characterized by four  $\beta$ -sheet H-bonds and RgCore = 5.19 Å. The remaining four structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 9.30 Å, (3) 6.68 Å, (4) 7.71 Å, and (5) 5.9 Å. (B) For the P180A mutant, three structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a native-like  $\beta$ -hairpin, characterized by four  $\beta$ -sheet H-bonds and RgCore = 4.93 Å. The remaining two structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 6.46 Å, and (3) 7.59 Å. (C) For the E181G mutant, four structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a misfolded  $\beta$ -hairpin characterized by three  $\beta$ -sheet H-bonds and RgCore = 5.17 Å. The remaining three structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 7.20 Å, (3) 6.36 Å, and (4) 8.13 Å. (D) For the E181P mutant, six structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a misfolded  $\beta$ -hairpin characterized by three  $\beta$ -sheet H-bonds and RgCore = 5.17 Å; structure 2 holds two  $\beta$ -sheet H-bonds and a RgCore = 5.95 Å. The remaining four structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (3) 7.75 Å, (4) 5.40 Å, (5) 4.33 Å, and (6) 4.72 Å.

#### 4. Discussion

RP refers to a group of debilitating, hereditary disorders that cause severe visual impairment in as many as 1.5 million patients worldwide.<sup>1–3</sup> Many genes have been associated with RP, and it exhibits extreme heterogeneity in terms of its severity and mode of inheritance. Although 30 RP genes have been recently identified, there were immensely exciting developments in the study of the ADRP form of the disease.<sup>1–3</sup> Rhodopsin mutations account for >25% of ADRP, and ~100 distinct mutations have been identified throughout the transcript. Mutations have been identified in all of the structural domains of the rhodopsin protein, and although attempts have been made to categorize mutants into six general classes of biochemical defects, many do not fit into predictable groups.

This study is part of a project aimed at structurally characterizing, through molecular simulations, the majority of pathogenic rhodopsin mutations. In this framework, the choice of the approach is dictated by the structural localization and the biochemical effect of a given mutation. The mutations considered in this study concentrate in the structured part of EL2. The crystal structures of dark rhodopsin<sup>6</sup> as well as of the BATHO and LUMI photo-intermediates<sup>20,21</sup> show that this rhodopsin portion folds into a twisted  $\beta$ -hairpin, whose C-terminal strand forms the “floor” of the chromophore binding pocket (if the receptor is seen in a direction parallel to the membrane surface with the intracellular side on top). Although structure determinations of the isolated EL2 fragment are lacking, very recent solid-state NMR determinations support the structural stabil-



**Figure 8.** Free energy contour maps at 313 K relative to the D190A (A), D190C (B), D190G (C), and D190Y (D) mutants, versus the number of native  $\beta$ -sheet hydrogen bonds and the core radius of gyration (RgCore). The general description of this legend is the same as that of Figure 4. The representative structures from each energy basin are shown as cartoons. In detail, (A) for the D190A mutant, four structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a native-like  $\beta$ -hairpin, characterized by four  $\beta$ -sheet H-bonds and RgCore = 5.19 Å. The remaining three structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 8.19 Å, (3) 5.29 Å, and (4) 8.95 Å. (B) For the D190C mutant, four structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a misfolded  $\beta$ -hairpin, characterized by three  $\beta$ -sheet H-bonds and RgCore = 5.44 Å. The remaining three structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 9.68 Å, (3) 6.38 Å, and (4) 8.21 Å. (C) For the D190G mutant four structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a native-like  $\beta$ -hairpin characterized by four  $\beta$ -sheet H-bonds and RgCore = 5.18 Å. The remaining three structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 8.80 Å, (3) 7.89 Å, and (4) 6.72 Å. (D) For the D190Y mutant three structures have been extracted: the structure extracted from the lowest energy basin (1) corresponds to a native-like  $\beta$ -hairpin characterized by four  $\beta$ -sheet H-bonds and RgCore = 5.28 Å. The remaining two structures share zero  $\beta$ -sheet H-bonds and the following RgCore values: (2) 7.32 Å and (3) 6.10 Å.

ity of the loop, by showing that during rhodopsin activation EL2 changes position but not conformation.<sup>10</sup> Furthermore, the web server BETAHAIRPRED (<http://triton.rmn.ifr.csic.es/software/bhairpredv1.0/bhairpred.htm>) predicts the <sup>177</sup>RYIPEGMQCSCGID<sup>190</sup> sequence as prone to form a  $\beta$ -hairpin, suggesting that such a conformational propensity is intrinsic to the primary sequence of the peptide, independent of the environment. These data support the strategy to investigate the potential effects of 40 mutations on the structural stability of the native  $\beta$ -hairpin taken off the protein context. The aim was pursued by comparative parallel REMD simulations using the FACTS implicit solvent model.<sup>12</sup> In line with reduction of the molecular system, the choice of an implicit over an explicit solvent model was dictated by

the need to implement a fast in silico screening approach. Indeed, in this study, computational screening consisted of 101 independent REMD simulations, starting from different input structures, which included two different prototropic forms of E181 for the wild type form and three different rotameric states for all mutated side chains, excluding alanine, glycine, and proline.

The computational protocol was set on the GB1 peptide, a model system for computational experiments on  $\beta$ -hairpin folding. In agreement with previous explicit water simulations,<sup>14</sup> the most populated state of GB1 at 313 K resulted in the native  $\beta$ -hairpin. Moreover, the average fraction of native contacts at 313 K is quite overlapping with the results of explicit water simulations and does not vary in the

270–313 K temperature range (i.e., 67.87% and 67.22%, respectively). Near the biological temperature range, the fraction of native contacts from our experiments is in reasonable agreement with those found by NMR determinations, whereas the highest discrepancies occur at temperatures higher than 313 K. These findings are in line with previous results of computational experiments on the same model system.<sup>14</sup> It is worth noting, however, that temperatures higher than 313 K go beyond the computational screening of rhodopsin mutations, the main goal of this study.

The computational protocol optimized for the GB1 peptide was thus extended to the R177-D190 structured rhodopsin fragment, toward the building of a fast *in silico* screening tool for structure-based reclassification of selected ADRP mutations. The latter, indeed, fall essentially in class II according to poorly defined biochemical behaviors, characterized by more or less pronounced impairment in receptor folding/expression and/or retinal binding (Tables 1 and 2).

Similar to the GB1 peptide, the native state of wild type EL2 prevailed at 313 K. The probability of native interstrands HBs, i.e., the HB1–4<sub>avg</sub> index, and the shape of the free energy landscape were, thus, employed as primary structural hallmarks of the native state in the comparative mutational analysis. These hallmarks were variably perturbed following REMD simulations of the 40 naturally occurring and artificial EL2 mutants considered in this study, resulting in a structural classification of mutational effects in misfolding, moderately misfolding, low misfolding, and non-misfolding (Tables 1 and 2). According to this classification and consistent with *in vitro* evidence of a detrimental effect on folding/expression and/or chromophore binding, misfolding mutations include nonconservative substitutions of the first and last amino acids in the loop, i.e. R177C and D190G, -A, -C, and -Y, as well as the P180A, E181G, and -P mutants. For these mutants, the HB1–4<sub>avg</sub> index is, indeed, below 50% due to increases in the population of non-native states compared to the wild type. Furthermore, the lowest energy basin is shifted toward a lower number of  $\beta$ -sheet HBs or higher energy values. For the majority of these mutants, alternative conformational states essentially include one- or two-turn  $\alpha$ -helices characterized by zero  $\beta$ -sheet HBs and RgCore values above 6 Å (Figures 7 and 8). The formation of  $\alpha$ -helix turns, while compatible with EL2 bridging H4 and H5, as demonstrated by the crystallographic structures of the homologous  $\beta$ 1- and  $\beta$ 2-adrenergic receptors,<sup>34,35,34,35</sup> is expected to perturb the native interactions of the loop with the surrounding domains of rhodopsin. In contrast, the E181G and -P mutants are essentially characterized by misfolded forms of the native  $\beta$ -hairpin rather than  $\alpha$ -helix turns. The misfolding effect of nonconservative mutations of R177<sup>(1)</sup> or D190<sup>(14)</sup> is likely due to a disruption of the native interstrand salt bridge between the two residues, expected to stabilize the native  $\beta$ -hairpin based also upon previous *in vitro* investigations.<sup>18</sup> Differently from the R177<sup>(1)</sup> and D190<sup>(14)</sup> mutants, the misfolding effect of P180A, E181G, and -P may be due to the elimination or introduction of proline or glycine residues that would perturb the native backbone conformational behavior of EL2.

Furthermore, five mutants show moderate misfolding properties, whereas the majority of the mutants, 27 over 40 (68%), show a poor or absent misfolding effect (Tables 1 and 2). With respect to this large subset of mutants, we speculate that their structural effects are to introduce non-natural disulfide bridges or to perturb the EL2–TM and/or EL2–retinal interface rather than the intrinsic folding properties of EL2.

Fifteen of the 27 low or non-misfolding mutants concern E181<sup>(5)</sup>, which was subjected to all possible amino acid substitutions (Table 2). *In vitro* mutational analysis of this highly studied glutamate shows that replacements with lysine, arginine, and proline result in totally impaired receptor expression (Table 2).<sup>36</sup> The remaining 16 mutants expressed and bound 11-*cis*-retinal to form pigments. Such *in vitro* data suggest that E181<sup>(5)</sup> does not contribute significantly to spectral tuning of the ground state of rhodopsin but rather affects the environment of the retinylidene Schiff base in the active MII photoproduct (Table 2).<sup>36</sup>

Consistent with the results of previous analyses,<sup>36</sup> we could not find any significant linear correlation between *in vitro* data on the E181<sup>(5)</sup> mutants (Table 2) and a significant number of descriptors of the physicochemical properties of the amino acids, including hydrophobicity and hydrophilicity parameters, size descriptors, volume and surface area values, solution properties, and chromatographic properties, as well as polarity and polarizability indices (results not shown).

The results of REMD simulations, consistent with *in vitro* data, suggest that the effects of these mutations should be ascribed to perturbations in the network of interactions mediated by such glutamate rather than to a disruption of the native EL2  $\beta$ -hairpin.

## 5. Summary

In this study, 40 rhodopsin mutations, 15 ADRP-linked and 25 artificial, all located in EL2, were screened by REMD simulations. The results of the screening constitute the start of a systematic structure-based reclassification of ADRP mutations.

Eight out of 40 EL2 mutants resulted in strong misfolding effects on the native  $\beta$ -hairpin, consistent with *in vitro* evidence that they all share severe impairments in folding/expression and/or retinal binding. Four of these misfolding mutants, i.e. P180A, and D190A, -G, and -Y, are associated with ADRP. Moreover, five residues displayed moderate misfolding effects and they include two ADRP-linked mutants, i.e. S186P and D190N. The remaining 27 mutants, including nine ADRP-linked mutants and overall characterized by milder effects on rhodopsin expression, did not perturb significantly the conformational behavior of the native  $\beta$ -hairpin. Thus, the computational screening could individuate and differentiate EL2 rhodopsin mutations that would affect the intrinsic stability of the native  $\beta$ -hairpin from mutations expected to variably impair native contacts between the loop and the surrounding receptor domains.

We, therefore, predict that for six out of the 15 ADRP-linked mutants, the structural determinants of the disease are mutation-induced misfolding effects on EL2. A misfolded EL2, being part of the stability core, is expected to undermine

the stability of rhodopsin, consistent with the impaired folding/expression observed for these mutants.

The extensive computational screening carried out in this study relies on strong comparative bases and takes advantage of the use of a fast and effective implicit solvent model. Within a comparative framework, possible overestimations of the native state ensembles can be neglected, as they are expected to be equally shared by wild type and mutant forms and to not affect predictions. The latter, indeed, profit by the internal consistency that characterize any comparative approach aimed at highlighting differences/similarities rather than absolute values/behaviors.

The results of this study add structural insight to the poorly resolved biochemical behavior of selected class II ADRP mutations, a fundamental step toward an understanding of the atomistic causes of the disease.

**Acknowledgment.** This study was supported by a Telethon-Italy grant no. S00068TELU (To F.F.).

**Supporting Information Available:** Additional analysis plots (Figures S1–S7). This information is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- Chuang, J. Z.; Vega, C.; Jun, W.; Sung, C. H. Structural and functional impairment of endocytic pathways by retinitis pigmentosa mutant rhodopsin–arrestin complexes. *J. Clin. Invest.* **2004**, *114*, 131–40.
- Kennan, A.; Aherne, A.; Humphries, P. Light in retinitis pigmentosa. *Trends Genet.* **2005**, *21*, 103–10.
- Mendes, H. F.; van der Spuy, J.; Chapple, J. P.; Cheetham, M. E. Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends Mol. Med.* **2005**, *11*, 177–85.
- Burns, M. E.; Arshavsky, V. Y. Beyond counting photons: Trials and trends in vertebrate visual transduction. *Neuron* **2005**, *48*, 387–401.
- Fanelli, F.; De Benedetti, P. G. Computational modeling approaches to structure–function analysis of G protein-coupled receptors. *Chem. Rev.* **2005**, *105*, 3297–3351.
- Palczewski, K. G protein-coupled receptor rhodopsin. *Annu. Rev. Biochem.* **2006**, *75*, 743–67.
- Okada, T.; Ernst, O. P.; Palczewski, K.; Hofmann, K. P. Activation of rhodopsin: New insights from structural and biochemical studies. *Trends Biochem. Sci.* **2001**, *26*, 318–24.
- McBee, J. K.; Palczewski, K.; Baehr, W.; Pepperberg, D. R. Confronting complexity: The interlink of phototransduction and retinoid metabolism in the vertebrate retina. *Prog. Retin. Eye Res.* **2001**, *20*, 469–529.
- Yan, E. C.; Kazmi, M. A.; Ganim, Z.; Hou, J. M.; Pan, D.; Chang, B. S.; Sakmar, T. P.; Mathies, R. A. Retinal counterion switch in the photoactivation of the G protein-coupled receptor rhodopsin. *Proc Natl Acad Sci U S A* **2003**, *100*, 9262–7.
- Ahuja, S.; Hornak, V.; Yan, E. C.; Syrett, N.; Goncalves, J. A.; Hirshfeld, A.; Ziliox, M.; Sakmar, T. P.; Sheves, M.; Reeves, P. J.; Smith, S. O.; Eilers, M. Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nat. Struct. Mol. Biol.* **2009**, *16*, 168–75.
- Rader, A. J.; Anderson, G.; Isin, B.; Khorana, H. G.; Bahar, I.; Klein-Seetharaman, J. Identification of core amino acids stabilizing rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7246–51.
- Haberthur, U.; Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **2008**, *29*, 701–715.
- Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–73.
- Zhou, R.; Berne, B. J.; Germain, R. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–6.
- Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **2001**, *42*, 345–54.
- Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins* **2004**, *56*, 310–21.
- Evans, D. A.; Wales, D. J. Folding of the GB1 hairpin peptide from discrete path sampling. *J. Chem. Phys.* **2004**, *121*, 1080–90.
- Janz, J. M.; Fay, J. F.; Farrens, D. L. Stability of dark state rhodopsin is mediated by a conserved ion pair in intradiscal loop E-2. *J. Biol. Chem.* **2003**, *278*, 16982–91.
- Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **2004**, *342*, 571–83.
- Nakamichi, H.; Okada, T. Crystallographic analysis of primary visual photochemistry. *Angew. Chem., Int. Ed.* **2006**, *45*, 4270–3.
- Nakamichi, H.; Okada, T. Local peptide movement in the photoreaction intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12729–34.
- Ludeke, S.; Beck, M.; Yan, E. C.; Sakmar, T. P.; Siebert, F.; Vogel, R. The role of Glu181 in the photoactivation of rhodopsin. *J. Mol. Biol.* **2005**, *353*, 345–56.
- MacKerell, A. D. J.; Bashford, D.; Bellott, M.; Dunbrack, R. L. J.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, T. D.; Prodhom, B.; Reiher, W. E. I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- Rao, F.; Caflisch, A. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.* **2003**, *119*, 4035–4042.
- Ceccolini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **2004**, *121*, 10748–10756.
- Im, W.; Lee, M. S.; Brooks, C. L., III. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–702.
- Zagrovic, B.; Pande, V. Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem.* **2003**, *24*, 1432–6.

- (28) Klimov, D. K.; Thirumalai, D. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544–9.
- (29) Dunbrack, R. L., Jr.; Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **1993**, *230*, 543–74.
- (30) Ponder, J. W.; Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **1987**, *193*, 775–91.
- (31) Sutcliffe, M. J.; Hayes, F. R.; Blundell, T. L. Knowledge based modelling of homologous proteins, part II: Rules for the conformations of substituted sidechains. *Protein Eng.* **1987**, *1*, 385–92.
- (32) Heyer, L. J.; Kruglyak, S.; Yooseph, S. Exploring expression data: Identification and analysis of coexpressed genes. *Genome Res.* **1999**, *9*, 1106–15.
- (33) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007**, *23*, 2625–2627.
- (34) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human beta<sub>2</sub>-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
- (35) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhamet-zianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Structure of a beta<sub>1</sub>-adrenergic G-protein-coupled receptor. *Nature* **2008**, *454*, 486–91.
- (36) Yan, E. C.; Kazmi, M. A.; De, S.; Chang, B. S.; Seibert, C.; Marin, E. P.; Mathies, R. A.; Sakmar, T. P. Function of extracellular loop 2 in rhodopsin: glutamic acid 181 modulates stability and absorption wavelength of metarhodopsin II. *Biochemistry* **2002**, *41*, 3620–7.
- (37) Sung, C. H.; Schneider, B. G.; Agarwal, N.; Papermaster, D. S.; Nathans, J. Functional heterogeneity of mutant rhodopsins responsible for autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8840–4.
- (38) Kaushal, S.; Khorana, H. G. Structure and function in rhodopsin. 7. Point mutations associated with autosomal dominant retinitis pigmentosa. *Biochemistry* **1994**, *33*, 6121–8.
- (39) Souied, E.; Gerber, S.; Rozet, J. M.; Bonneau, D.; Dufier, J. L.; Ghazi, I.; Philip, N.; Soubbrane, G.; Coscas, G.; Munnich, A. Five novel missense mutations of the rhodopsin gene in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **1994**, *3*, 1433–4.
- (40) Iannaccone, A.; Man, D.; Waseem, N.; Jennings, B. J.; Ganapathiraju, M.; Gallaher, K.; Reese, E.; Bhattacharya, S. S.; Klein-Seetharaman, J. Retinitis pigmentosa associated with rhodopsin mutations: Correlation between phenotypic variability and molecular effects. *Vision Res.* **2006**, *46*, 4556–67.
- (41) Sung, C. H.; Davenport, C. M.; Nathans, J. Rhodopsin mutations responsible for autosomal dominant retinitis pigmentosa. Clustering of functional classes along the polypeptide chain. *J. Biol. Chem.* **1993**, *268*, 26645–9.
- (42) Colley, N. J.; Cassill, J. A.; Baker, E. K.; Zuker, C. S. Defective intracellular transport is the molecular basis of rhodopsin-dependent dominant retinal degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3070–4.
- (43) Dryja, T. P.; McEvoy, J. A.; McGee, T. L.; Berson, E. L. Novel rhodopsin mutations Gly114Val and Gln184Pro in dominant retinitis pigmentosa. *Invest Ophthalmol. Vision Sci.* **2000**, *41*, 3124–7.
- (44) Kono, M.; Yu, H.; Oprian, D. D. Disulfide bond exchange in rhodopsin. *Biochemistry* **1998**, *37*, 1302–5.
- (45) Hwa, J.; Reeves, P. J.; Klein-Seetharaman, J.; Davidson, F.; Khorana, H. G. Structure and function in rhodopsin: further elucidation of the role of the intradiscal cysteines, Cys-110, -185, and -187, in rhodopsin folding and function. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1932–5.
- (46) Rüther, K.; von Ballestrem, C. L.; Müller, A.; Kremmer, S.; Eckstein, A.; Apfelstedt-Sylla, E.; Gal, A.; Zrenner, E. In *Degenerative Diseases of the Retina*; Anderson, R. E., LaVail, M. M., Hollyfield, J. G., Eds.; Plenum Press: New York, 1995, pp 303–312.
- (47) Liu, X.; Garriga, P.; Khorana, H. G. Structure and function in rhodopsin: correct folding and misfolding in two point mutants in the intradiscal domain of rhodopsin identified in retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4554–9.
- (48) Doi, T.; Molday, R. S.; Khorana, H. G. Role of the intradiscal domain in rhodopsin assembly and function. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4991–5.
- (49) Dryja, T. P.; Hahn, L. B.; Cowley, G. S.; McGee, T. L.; Berson, E. L. Mutation spectrum of the rhodopsin gene among patients with autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9370–4.

CT900145U