Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

© Wiley-VCH 2011

69451 Weinheim, Germany

Solution NMR Structure of Proteorhodopsin**

*Sina Reckel, Daniel Gottstein, Jochen Stehle, Frank Löhr, Mirka-Kristin Verhoefen, Mitsuhiro Takeda, Robert Silvers, Masatsune Kainosho, Clemens Glaubitz, Josef Wachtveitl, Frank Bernhard, Harald Schwalbe, Peter Güntert, and Volker Dötsch**

anie_201105648_sm_miscellaneous_information.pdf

Supporting Information

Table of Content

Methods

Figures

Tables

Methods

Cell-free expression and sample preparation. PR retains its native structure in the *E. coli* membrane[1] but is also well expressed in an *E. coli*-based cell-free expression system.[2] For structure determination we chose the continuous-exchange cell-free expression system[3] as it greatly facilitated selective labeling strategies for the assignment process and the sample preparation was less time-consuming than bacterial expression. The cell-free expression system furthermore allowed elimination of the signal peptide sequence of PR[4] without reduction of expression yield and stability. The green-absorbing variant of proteorhodopsin (PR) was therefore cloned without the signal peptide into the pIVEX2.3d vector (Roche Applied Science) optimized for cell-free expression. In addition to the 229 residues from PR (amino acids 21-249) the construct consists of the N-terminal Met, a hexa-His-tag on the C-terminus and a 5 residue long linker region N-terminal to the His-tag resulting in a final size of 241 amino acids, 26 kDa. For assignment of the characteristic H75 a Strep-tag (WSHPQFEK) was cloned at the C-terminal end in place of the His-tag. Continuous-exchange cell-free expression was based on an S30 extract and followed standard protocols.[3] PR was expressed in the presence of 0.6 mM all-*trans* retinal in the detergent mode (D-CF) using 0.4% digitonin (Sigma) mixed with diC7PC (1,2-diheptanoyl-sn-glycero-3-phosphocholine, Avanti Polar Lipids) in a 4:1 molar ratio. Ni-affinity purification was conducted with 0.05% DDM (dodecyl-β-D-maltopyranoside, AppliChem) in the loading and washing buffer and the detergent was exchanged into 0.1% diC7PC prior elution. Strep-tagged PR was purified via a *Strep*-Tactin matrix (IBA GmbH) with 0.1% diC7PC in the washing and elution buffer. Buffer exchange was achieved by either centrifugal concentration devices (10,000 MWCO, Amicon) or gravity flow PD10 Desalting columns (GE Healthcare). NMR samples were buffered with 25 mM NaOAc, pH 5 and 2 mM DTT. The detergent concentration in the NMR sample added up to a final value of 2%. Side chain assignment and ^{13}C-edited NOESY experiments required the use of deuterated acetate (Cambridge Isotope Laboratories, CIL) as well as deuterated detergent of which both forms (D_{26} and D_{39}) were synthesized by Avanti Polar Lipids. The samples were exchanged into deuterated detergent during the detergent exchange step on the Ni-affinity column.

Isotope labeling. Protocols for isotope labeling were adjusted according to the labeling scheme. For uniform ^{15}N or ^{15}N,^{13}C-labeling 20 amino acid mixtures are available (CIL) and were added to a final concentration of 30 mg/ml reaction mixture. In principle this also applies for ^{2}H,^{15}N and ^{2}H,^{15}N,^{13}C labeling schemes, however, to maintain a high deuteration level, cell-free expression was conducted in completely deuterated solutions (except RNAse inhibitor and T7 Polymerase).[5] Selective labeling was achieved by replacing single types of amino acid in the unlabeled amino acid

mixture used for cell-free expression with the labeled equivalent (CIL). In addition to the overlap and relaxation optimized (oro-)SAIL amino acid (SAIL Technologies, Inc.), the remaining 19 amino acids were ^{2}H labeled. The 11,20-^{13}C-labeled retinal was co-translationally bound to K231 by replacing the unlabeled cofactor in the cell-free expression system. Per 3 ml expression volume for one NMR sample 0.5 mg retinal were required.

Sample preparation for PRE measurements. Single-cysteine mutants for PRE measurements were created by QuikChange site-directed mutagenesis (Stratagene). After NiNTA purification, the sample was divided into two parts. One half was processed as described above and used as the reference sample without paramagnetic tag. The second half was tagged with the spin-label MTSL (methanethiosulfonate, Toronto Research Chemicals) by addition of 20-fold excess MTSL to the protein sample in 20 mM Na_2HPO_4, 400 mM NaCl, >0.1% diC7PC at 4°C overnight. Removal of excess spin label together with buffer exchange was achieved using gravity flow PD10 Desalting columns. Samples for PRE measurement were selectively labeled with Ala, Gly, Leu, Phe and Ser, or Met, Phe, Thr, Tyr and Val. In some cases, cysteine mutation or spin label attachment caused structural distortions such as changes in the absorption maximum or severe chemical shift perturbations in the [^{15}N,^{1}H]-TROSY spectra. These mutants were not included in the structure determination.

Sample preparation for RDC measurements. Weak alignment of PR samples was achieved by stretched 4% polyacrylamide gels as alignment medium. Different molar ratios of acrylamide and N,N'-methylene-bisacrylamide were tested (37.5:1, 50:1, 75:1, 150:1), as well as a positively charged gel composition consisting of APTMAC ((3-acrylamidopropyl)-trimethylammonium chloride) and N,N'-methylene-bisacrylamide in a 19:1 molar ratio. Best spectral quality and sufficiently large dipolar couplings were obtained at an acrylamide:N,N'-methylene-bisacrylamide ratio of 150:1. For sample preparation, a commercially available gel system (New Era Enterprises, Inc) was used. Gels were cast with a diameter of 6 mm, dialyzed against an excessive amount of H_2O over night and dried at room temperature. Dried gels were incubated with the PR NMR samples and allowed to swell to its original diameter within 24 h. Afterwards the gels were transferred into NMR tubes with an inner diameter of 4.2 mm. To reduce peak overlap PR was selectively labeled (sample1: Ala, Gly, Leu, Phe and Ser; sample2: Ala, Met, Ile; sample 3: Phe, Thr, Trp, Tyr and Val) and purified as described above.

NMR Spectroscopy. For structure determination NMR samples were buffered in 25 mM NaOAc pH 5 with 2 mM DTT. After concentrating the protein, the amount of the detergent diC7PC was about 2% estimated according to signal intensities in a 1D ^{1}H NMR spectrum. Protein concentrations varied between 0.3 and 0.5 mM protein. Measurements on SAIL-PR were carried out at 313 K to prolong sample lifetime, whereas the sample temperature was adjusted to 323 K for all remaining experiments. As shown by [^{13}C,^{1}H]-HSQC, side chain resonances were not significantly affected by a change in temperature. Measurements were conducted on Bruker Avance spectrometers equipped with cryogenic ^{1}H{^{13}C,^{15}N} probes operating at proton resonance frequencies of 600–950 MHz. A conventional triple resonance xyz-gradient probe was employed for measurements at 500 MHz.

Backbone assignment was achieved using TROSY-type[6] BEST[7] versions of HNCO, HNCA, HN(CO)CA, HNCACB triple resonance experiments and ^{15}N-separated 3D and 4D NOESY experiments (mixing time 180 ms and 200 ms, respectively). Using the known $^{13}C^{\alpha}$ and ^{13}C' chemical shifts, $^{1}H^{\alpha}$ resonance assignments were obtained from carbonyl-detected 3D HCACO,[8] taking advantage of $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ multiple-quantum line-narrowing.[9] The sequence-specific assignment of methionine methyl groups was performed in two steps: A double-sensitivity enhanced 3D (H)CCH-TOCSY[10] was employed for the assignment of β- and γ-CH_2 groups. Methyl groups were then linked to γ-carbons via $^{3}J_{H\varepsilon C\gamma}$ couplings using a modified version of the 3D HMBC experiment.[11] The pulse sequence concatenates long-range and direct HMQC type transfer steps to provide three-bond $^{1}H^{\varepsilon}$-$^{13}C^{\gamma}$ and one-bond $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ correlations in separate dimensions and uses methyl-selective proton pulses to accelerate longitudinal relaxation.[12] The same sequence was employed to correlate Thr methyl groups with α- and β-carbons as well as Ile δ-methyl groups with β- and γ1-carbons and Ile γ2-methyl groups with α-, β- and γ1-carbons. Connectivities between Thr/Val methyl groups and α- and β-carbons were also obtained via $^{1}J_{CC}$ couplings in HMCMCB and HMCMCBCA experiments, again using methyl-selective proton pulses. Applied to an Ala-selectively labeled sample the former experiment provided correlations between β-methyl groups and α-carbons. Assignments from scalar coupling based experiments were confirmed and supplemented with information from a 3D [^{13}C,^{1}H]-HMQC-NOESY-[^{13}C,^{1}H]-HMQC experiment (60 ms mixing time) and a 3D ^{13}C-detected (HCC) version of a ^{13}C-^{13}C NOESY experiment (600 ms mixing time).[13] Side chain assignment of oro-SAIL Leu and oro-SAIL Val was achieved with information from ^{1}H-methyl detected out-and-back experiments,[14] 4D ^{15}N,^{13}C-separated NOESY as well as 3D ^{13}C-separated NOESY. Proton-detected 3D HCACO and (H)CCH-TOCSY experiments were recorded to account for the deuterium isotope shift and to transfer the assignment to uniformly labeled samples. Sequence-specific assignments of aromatic Trp resonances were

obtained with a ALVW-labeled sample by correlating δ1-CH groups with β-carbons using a 3D [^{13}C,^{1}H]-TROSY-HCD(CG)CB experiment.[15] Knowledge of $^{13}C^{\delta 1}$ chemical shifts enabled the assignment of ε1-NH groups by application of a 3D [^{15}N,^{1}H]-TROSY-HNC pulse sequence. Chemical shifts of the remaining indole CH groups resulted from correlations with δ1 and ε1 protons in a 3D NOESY-[^{13}C,^{1}H]-HMQC experiment (50 ms mixing time), which also provided intraresidual NOE contacts to α- and β-protons to support sequential assignments and contributed long-range distance information.

Medium- to long-range NOEs involving methyl groups in ALVW-labeled PR were derived from a 4D SOFAST-methyl-TROSY ^{13}C,^{13}C-separated NOESY experiment (mixing time 50 ms).[12] The same experiment was carried out on a AILMTVW-labeled sample (mixing time 70 ms) using 28%-sparse non-uniform sampling. Data processing of the latter spectrum was achieved using multi-dimensional decomposition (MDD)[16] within the TopSpin 3.1 software (Bruker). Retinal-to-protein NOEs were also contained in these NOESY experiments. Additional NOEs involving protein amide groups came from a 2D ^{15}N-separated, ^{13}C-edited NOESY (mixing time 100 ms).

BEST-[^{15}N,^{1}H]-TROSY experiments were recorded for PRE measurements on selectively labeled samples. An additional spin echo of either 8 μs or 8 ms duration was incorporated into the initial INEPT period for analysis by the two-time-point method.[17] The measurement of $^{1}D_{NH}$ RDCs relied on the F_1 peak displacement of TROSY ("south-east") and semi-TROSY ("north-east") components selected in an interleaved manner with a 2D BEST-[^{15}N,^{1}H]-TROSY type pulse sequence. In order to partially compensate for the differential linewidths the scan repetition time for recording the TROSY component was linearly decreased from 400 ms to 150 ms in concert with t_1[18] while the delay was increased from 400 ms to 1.1 s for recording the semi-TROSY component. The precision of the RDC values is estimated to ± 3 Hz based on duplicate measurements for Ala and Phe residues present in two of the three samples. [^{15}N,^{1}H]-SOFAST-HMQC experiments[19] were used for bicelle samples and the [^{15}N,^{1}H]-TROSY was employed for uniformly labeled samples during titrations with paramagnetic agents. Spectra were processed with Topspin (Bruker) and analyzed in Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

Sample optimization and conformational exchange line broadening. After screening different expression and purification conditions, UV-VIS spectroscopy and size exclusion chromatography suggested that PR was stable and monodisperse when expressed in a 4:1 molar ratio of digitonin/diC7PC in the presence of retinal and subsequently exchanged into the short-chain lipid diC7PC alone (Supporting Information Figure S1). In addition, melting curves obtained by CD spectroscopy showed a sharp transition above 70°C further demonstrating the stable formation of

detergent-solubilized PR (Supporting Information Figure S1). Judged from the size-exclusion elution profile, the proteo-micelle has an approximate molecular weight of 80-90 kDa which is further corroborated by a 2D TRACT (TROSY for rotational correlation times) analysis[20] of selectively ^{15}N-valine labeled PR (Supporting Information Figure S2). The high molecular weight of the PR proteo-micelle necessitated an elevated measuring temperature of 323 K and further improvement of the spectral quality was observed upon lowering the pH of the NMR sample to pH 5 (Supporting Information Figure S2). Nonetheless, conformational exchange of certain amino acids was still present as manifested by a field-dependent line broadening observed in a series of [^{15}N,^{1}H]-TROSY spectra (Figure 2). In this regard, screening of different protein-to-detergent ratios revealed a critical detergent concentration close to the critical micellar concentration of diC7PC which in the experimental setup corresponded to a molar ratio of 1:83. At higher detergent concentrations and molar ratios the spectral quality improved significantly (Supporting Information Figure S3). For structure determination, all measurements were carried out at a detergent concentration of approximately 2% (corresponding to 41.5 mM) lying well above the critical protein-detergent ratio. Nevertheless, the assignment process remained challenging, since conformational exchange was not abolished completely and line broadening occurred predominantly at high field. It is important to note, that the conformational exchange has not only been observed for cell-free expressed PR but also for samples expressed in *E. coli* and purified according to the same protocol.

Flash photolysis experiments. Short-chain phosphocholines have already proven beneficial for a number of membrane proteins also including the seven-transmembrane-helix receptor SRII.[21] Functional evidence of PR solubilized in the diC7PC micelle was obtained from flash photolysis measurements monitoring the PR photocycle (Supporting Information Figure S1). As the photocycle characteristics differ below and above the pKa of the proton acceptor D97, measurements were performed at pH 5 and 9. For pH 9 three transients were measured which are indicative for population changes of the ground state (λ_{Probe} = 500 nm) as well as the blue- and red-shifted photocycle intermediates (λ_{Probe} = 410 nm and 590 nm) (Supporting Information Figure S1). Directly after photoexcitation a positive absorption change is monitored at 590 nm. This contribution can be attributed to the K state. It decays mainly on a 10 µs time scale (τ_1 = 6 µs), in good agreement with former measurements.[22] The subsequently observed positive difference signal is most likely due to the absorption of the late red-shifted product states (N- and O-like photoproducts). Concomitant with the K decay the formation of a positive absorbance change is monitored at 410 nm. Due to its spectral position this contribution must be assigned to the M-like

state with the retinal in the deprotonated Schiff base configuration. The major part of the M state decays within 100 µs (τ_2 = 35 µs). However, a smaller fraction is visible until 1 s. It can therefore be assumed that an equilibrium between the M-like and the late N/O-like states is present at the end of the photocycle. The turn over rate of the PR photocycle is around 1 s^{-1}. The global analysis of the transients reveals a time constant of 660 ms.

Although the same optical geometries and pump pulse energies were used for the measurements at pH 9 and pH 5, the overall signal size is drastically smaller under acidic conditions. In addition, changes of the absorption maximum of the late red-shifted intermediates were noticed. Transients were thus measured for the depopulation of the ground state at 500 nm and the absorption of blue- and red-shifted intermediates at 410 nm, 600 nm and 650 nm (Supporting Information Figure S1). In agreement with former experiments[23] the photocycle is completed slightly faster at pH 5 compared to pH 9 (τ_{TO} (pH 5) = 310 ms, τ_{TO} (pH 9) = 660 ms). As at pH 9, the red-shifted K intermediate (λ_{Probe} = 600 nm) is formed within the excitation and decays with a time constant of τ_1 = 10 µs. The transient absorbance changes at 410 nm provide clear evidence that no blue-shifted M-like intermediate is present after K decay. The small negative signal at 410 nm is most likely due to the bleaching of the $S_0 \rightarrow S_2$ transition. The late N- or O-like intermediates seem to absorb near the ground state and/or most likely do not have a high extinction coefficient since nearly no positive signals are observed at the red wing of the spectrum for times > 50µs. Only at 650 nm a small positive signature is present, which decays on the same time scale as the depopulation of the ground state signature (τ_4 = 310 ms). In addition, the transients between 550 nm and 600 nm exhibit an increasing (τ_2 = 140 µs) and subsequently decreasing dynamics (τ_2 = 5 ms, see λ_{Probe} = 600 nm in Supporting Information Figure S1 G), which points to a superimposed positive contribution.

Measurements at both pH values indicated the existence of functional photocycles as they were also reported in previously published measurements in DDM micelles.[23]

Secondary structure prediction and structure calculation. Prediction of the secondary structure elements of PR employed the program TMHMM (transmembrane prediction using hidden Markov models)[24] and PsiPred protein secondary structure prediction.[25] Topology predictions based on NMR chemical shifts included TALOS+ (Torsion Angle Likeliness Obtained from Shift and Sequence Similarity)[26] and the chemical shift index (CSI) method.[27] The structure of PR was calculated with CYANA based on torsion angle restraints, α-helical hydrogen bond restraints, short-range H^N-H^N NOEs, medium- to long-range NOEs and distance restraints from paramagnetic relaxation enhancement as well as restraints from RDCs. TALOS+ was used to derive torsion angle restraints based on chemical shift data including H^N, N, H^α, C', C^α, C^β where applicable. Hydrogen

bonds were introduced for residues located in α-helices but not for helix E' and the potential π-bulge region in helix G. Medium- to long-range NOEs were given an upper limit of 5 Å. PRE-based distance restraints were derived by a combination of the two-time-point measurement[17] and the method used by Battiste and Wagner.[28] For signals with inherently low signal-to-noise ratio, calculation by the two-time-point method did not provide meaningful data. We thus used the calculated distance restraints of a minor set of resonances with good signal-to-noise ratio to determine a linear relationship of the paramagnetic/diamagnetic peak intensity ratio to the distance. On the basis of this linear approximation, the remaining distance restraints were derived. Resonances that were strongly affected by the spin label were assumed to have distances less than 13 Å. Distances ranging from 13 to 20 Å were converted into upper and lower limits with a ± 4 Å error margin. Signals without a detectable PRE effect were assigned a lower limit of 20 Å. Control experiments on a cysteine-free mutant (C107A, C156A, C175S) were conducted to account for unspecific binding. The experimental RDCs, covering the range form -26.9 to -2.4 Hz, were used to determine the initial alignment tensor magnitude and rhombicity via the histogram method incorporated in CYANA, which were subsequently refined by fitting against the structure. CYANA structure calculations with torsion angle dynamics[29] were performed with 200 conformers. The 20 conformers with the lowest target function values were superimposed with an rmsd to the mean coordinates of 0.87 Å for residues 23-169, 178-208 and 215-247 as identified by CYRANGE[30] and 0.81 Å within the transmembrane region (Supporting Information Table S1). To validate the structure, complete sets of PRE restraints for each cysteine mutant were excluded one by one from the structure calculation. The resulting structures were very similar with rmsd values to the original structures ranging from 0.47 Å to 1.34 Å.

Structural validation. To further validate the structure and to distinguish solvent-accessible extramembrane residues from residues buried inside the micelle, we conducted titration experiments with paramagnetic agents that are either water- or detergent-soluble such as Mn^{2+} and 16-doxy-1-stearic acid (16-DSA), respectively (Supporting Information Figure S7). For Mn^{2+} a stock solution of 100 mM in 25 mM NaOAc pH 5 was prepared and added to final concentrations of 0.05, 0.25, 0.5, 1.5, 2, 3 and 5 mM. Appropriate amounts of 16-DSA were dissolved in methanol, dried under nitrogen-flow and re-dissolved in 25 mM NaOAc pH 5 to give a stock concentration of 40 mM. Titration steps were measured at 1, 2, 4 and 8 mM.

Both agents led to a reduction of signal intensities in [^{15}N,^{1}H]-TROSY spectra due to relaxation enhancement. Titration with the paramagnetic metal ion Mn^{2+} showed broadening of resonances of the N- and C-termini and some loop residues. Altogether, only a minor subset of residues was

affected by the Mn^{2+} ions indicating that the protein is deeply buried within the detergent micelle. One or two helical turns of helices A, B and G are protruding out of the micelle and also the F-G loop is highly solvent exposed. In contrast, titration with 16-DSA had a great impact on the spectra with affected residues forming a belt-like arrangement around the center of the helices (Supporting Information Figure S7).

In another approach to show that the structure presented here is indeed close to the state in lipid bilayers and not an artifact of the detergent micelle, we reconstituted PR in a bicellar system[31] composed of the long-chain lipid DMPC (dimyristoyl-phosphocholine) and the short-chain lipid diC7PC. The presence of a bicelle was confirmed by ^{31}P-NMR. Due to the large size of the proteo-bicelle deuteration of the protein was required to obtain [^{15}N,^{1}H]-correlation spectra. Comparison with the spectra of detergent-solubilized PR revealed broader line shapes for the proteo-bicelle sample but an otherwise very similar peak pattern (Supporting Information Figure S2).

Figures

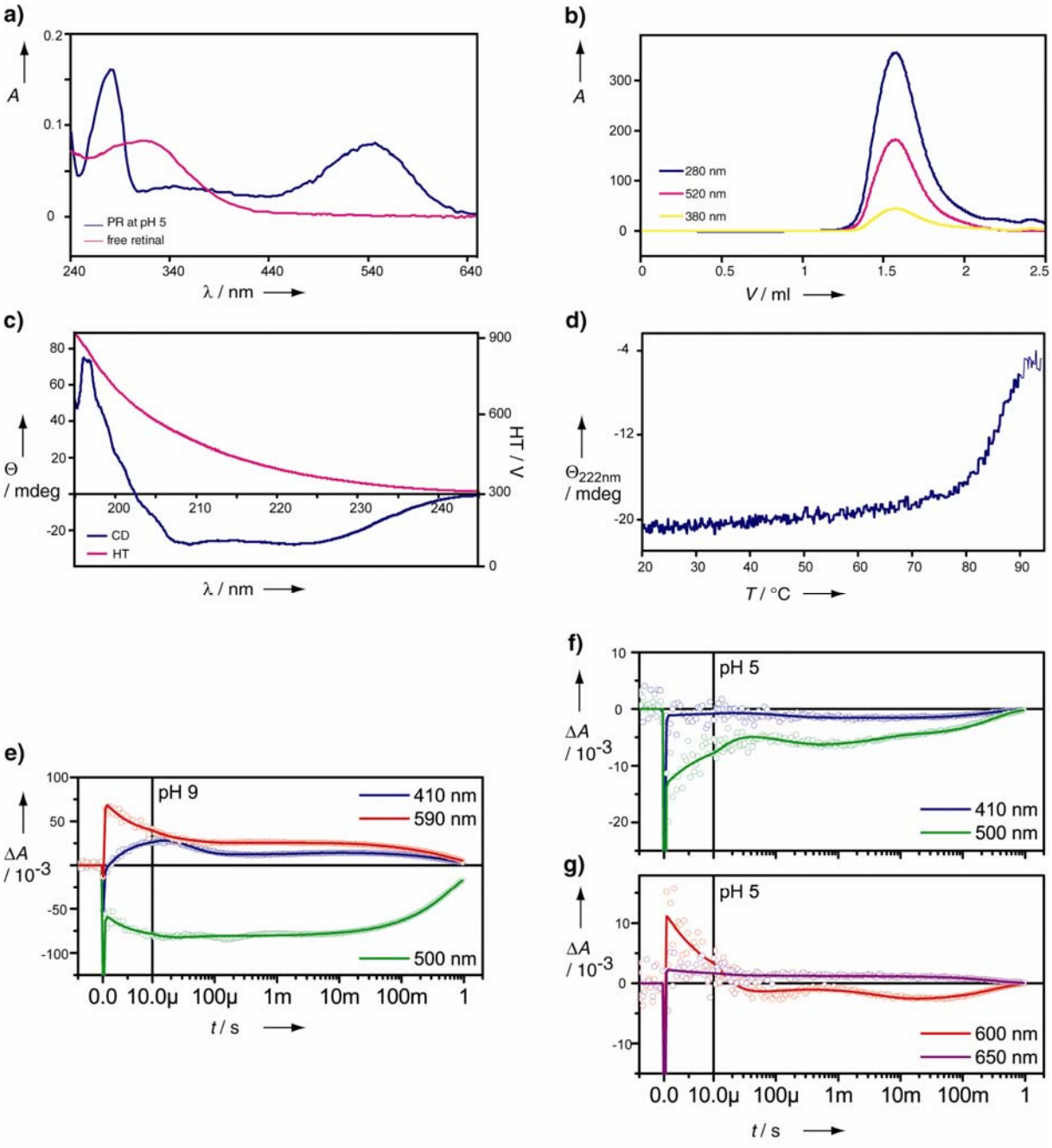
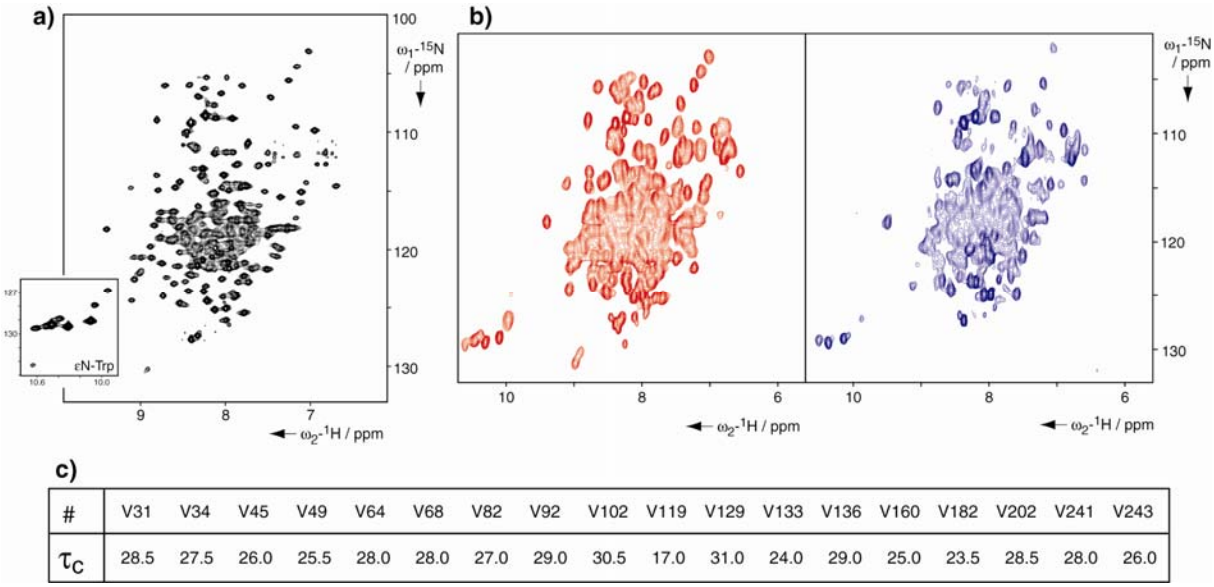


Figure S1. **Characterization of PR in the diC7PC micelle.** a) UV-VIS spectroscopy shows an absorption maximum of 540 nm for the bound retinal cofactor at pH 5. With a ratio of 280 to 540 nm ranging between 1.9 and 2, the sample was highly pure and the majority of the protein retinal bound. b) Size-exclusion chromatography indicates a single Gaussian-shaped peak eluting around 80-90 kDa on a Superdex 200 column calibrated with soluble protein standards. c) CD spectroscopy underlines the α-helical secondary structure and in d) a melting curve measured at 222 nm shows a clear transition above 70°C demonstrating formation of stable, detergent-solubilized PR. Additionally, flash photolysis measurements indicate functional photocycles of PR

at alkaline and acidic conditions. e) Transient absorbance changes (○) of PR at pH 9 after excitation at 520 nm and f) - g) PR at pH 5 after excitation at 540 nm. The transients are indicative for the depopulation of the ground state (500 nm), the absorption of the M-like intermediate (410 nm) as well as the K- and N- and O-like intermediates (590 nm or 600 m and 650 nm, respectively). The solid lines represent the results of the global fits. For a better presentation of the data a part of the laser artifact has been cut off by scaling the y-axis.



c)

#	V31	V34	V45	V49	V64	V68	V82	V92	V102	V119	V129	V133	V136	V160	V182	V202	V241	V243
τ_C	28.5	27.5	26.0	25.5	28.0	28.0	27.0	29.0	30.5	17.0	31.0	24.0	29.0	25.0	23.5	28.5	28.0	26.0

Figure S2. **NMR data of PR in a micelle and bicelle environment.** a) Representative [^{15}N,^{1}H]-TROSY spectrum recorded at 900 MHz of PR in 25 mM NaOAc pH 5, 2 mM DTT and 2% diC7PC at a temperature of 323 K. These conditions allowed the detection of almost all NH backbone resonances. b) [^{15}N,^{1}H]-SOFAST-HMQC spectrum of PR in diC7PC (shown in red) is compared to a spectrum recorded after reconstitution in 12% DMPC/diC7PC bicelles (shown in blue), revealing broader lineshapes in the bicelle spectrum but an otherwise similar peak pattern. Spectra are recorded with identical parameters at 950 MHz. c) Rotational correlation time of PR in diC7PC micelles determined by TRACT (TROSY for rotational correlation time). Experiments were conducted as 2D versions at 313 K using PR selectively labeled with oro-SAIL Val to minimize signal overlap. Peak heights were used for analysis by Sparky yielding individual τ_c values. An overall rotational correlation time of 31 ns at 313 K can thus be assumed explaining the limited sensitivity in crucial three-dimensional experiments.

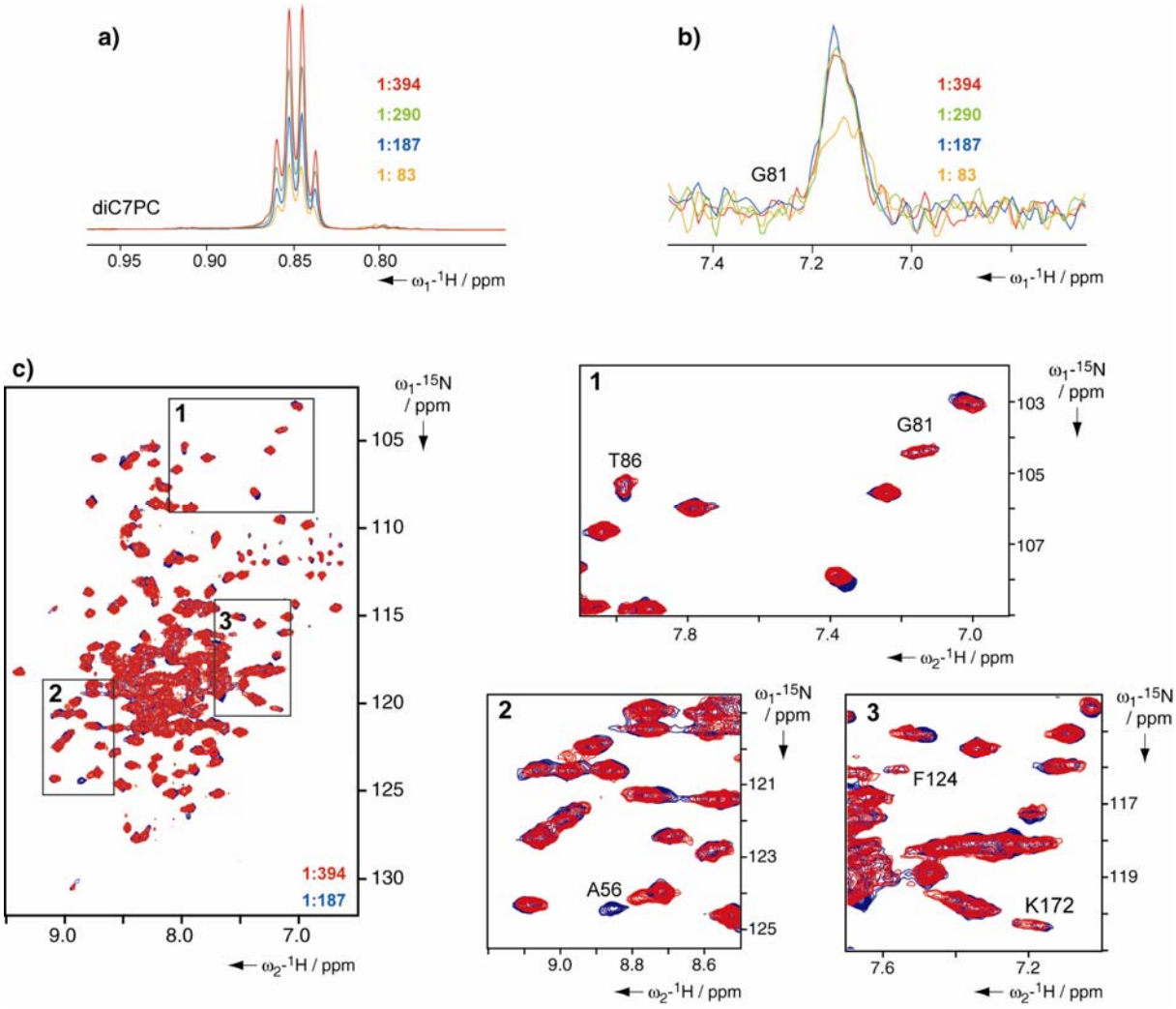


Figure S3. **Optimization of the protein-to-detergent ratio for NMR spectroscopy.** During sample preparation PR was not concentrated after the buffer exchange step resulting in a concentration of 100 µM as determined via the absorbance at 280 nm. To determine the detergent concentration, a series of 1D ^{1}H spectra were recorded at 0.1, 0.5, 1 and 2% diC7PC with identical parameters as the protein ^{1}H spectra. According to the signal intensity of the detergent signals, the initial detergent concentration of the PR sample was estimated to be 0.4%, which thus corresponded to a protein-to-detergent ratio of 1:83. Under these conditions long-term stability of the protein was reduced and precipitation occurred. a) The detergent concentration was increased stepwise and monitored by the signal intensity of the detergent methyl groups. b) The signal-to-noise ratio of G81 did improve significantly after the first detergent step as illustrated by the 1D slices takes from 2D [^{15}N,^{1}H]-TROSY spectra. At higher detergent concentrations no significant changes were observed. c) Overlay of [^{15}N,^{1}H]-TROSY spectra of uniformly ^{15}N-labeled PR at protein-to-detergent ratios of 1:187 ratio (blue) and 1: 394 (red). The lineshapes of resonances affected from line broadening did not improve significantly. Only minor chemical shift changes were observed upon increasing the detergent concentration. An exception is the loop residue A56, which shows a noticeable chemical shift perturbation. All spectra were recorded at 950 MHz.

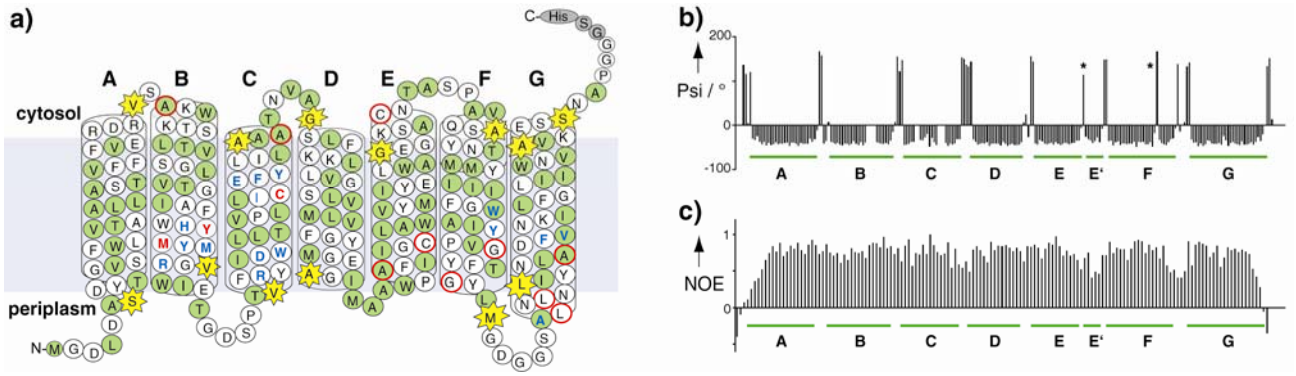


Figure S4. **Secondary structure of PR.** a) Topology model with seven transmembrane helices denoted A–G (A: 25-50, B: 56-86, C: 91-117, D: 123-141, E: 149-167, E': 170-176, F: 180-209 and G: 216-246). Residues that could not be assigned are in bold red letters, those with partial backbone assignments in bold blue. Amino acids for which methyl or aromatic resonances could be assigned are highlighted in green and sites of successful spin label attachment are highlighted in yellow. For all Val, Ala and Met PRE-sites, side chain assignments were obtained as well. Red circles indicate positions where cysteine mutation or spin label attachment caused structural distortions. These mutants were thus not used for structure determination. b) Secondary structure prediction by TALOS+ based on NMR chemical shift information. The predicted average backbone torsion angle ψ was plotted against the protein sequence; green bars below indicate helices A-G. The distortions in helix E at position G169 as well as in helix F at position P201 are indicated by an asterisk. c) Plot of $\{^{1}H\}^{15}N$-heteronuclear NOE against the protein sequence indicating a compact tertiary structure with restricted loop mobility. Slightly increased dynamics can be observed in the extension of helix E and in the loop between helices F and G.

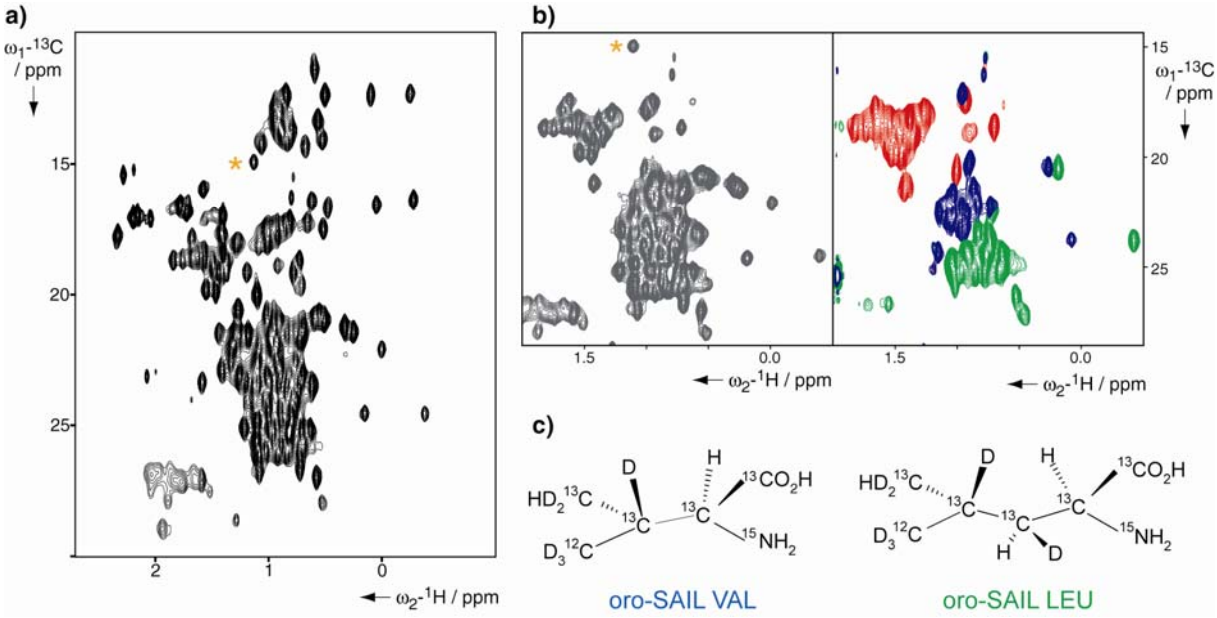


Figure S5. **Side chain assignment of PR.** a) [^{13}C,^{1}H]-SOFAST-HMQC of PR selectively labeled with ^{15}N,^{13}C-AILMTVW. The sample additionally contained 11,20-^{13}C retinal of which the signal of the C_{20}-methyl group is clearly visible in the spectrum, highlighted by the yellow asterisk. b) Comparison of [^{13}C,^{1}H]-HSQC spectra obtained with uniform ALV-labeling (left) and with Leu and Val replaced by the oro-SAIL variant (right). Spectra were obtained from three different samples each with a single amino acid labeled, either ^{15}N,^{13}C-Ala (red), oro-SAIL Val (blue) or oro-SAIL Leu (green). SAIL amino acids drastically reduce the degree of signal overlap as the second prochiral methyl group of Leu and Val is deuterated. The labeling pattern for overlap and relaxation optimized SAIL Leu and Val is shown in c).[32]

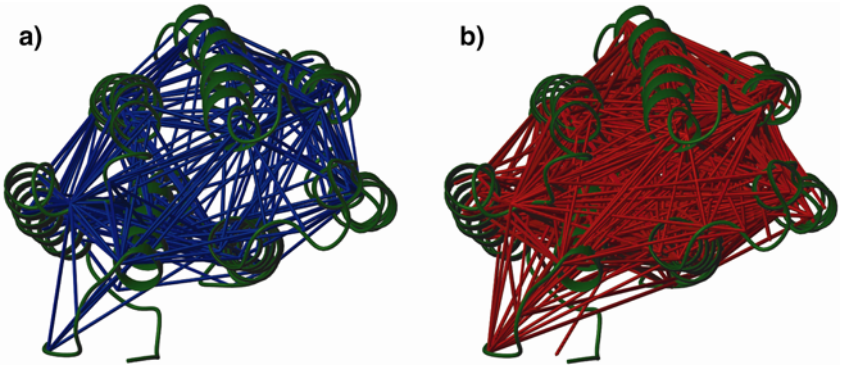


Figure S6. **PRE restraints for the structure determination of PR.** In total, 13 single-cysteine mutants were used for PRE measurements with samples that were selectively labeled to minimize signal overlap. The 290 upper limits form a dense network between the transmembrane helices of PR as illustrated in a). Similarly, the 716 lower limits provide important information for the structure determination as shown in b).

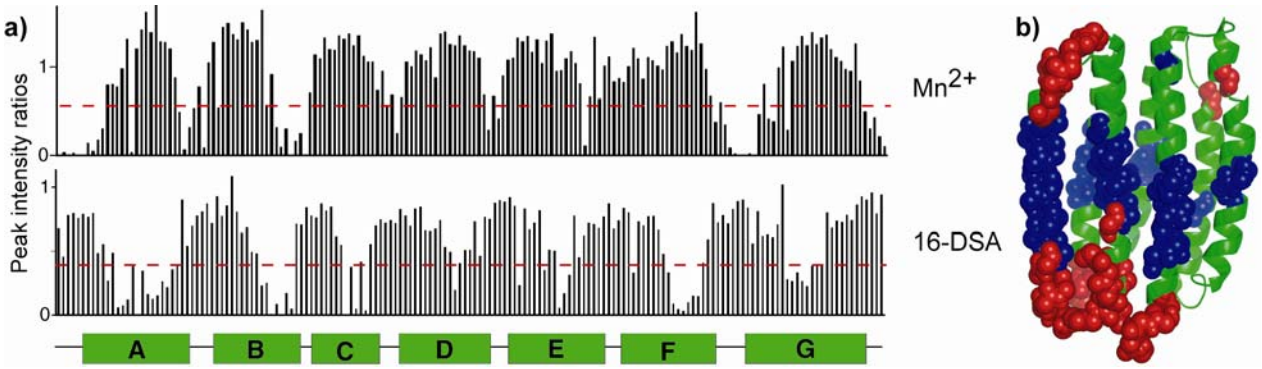


Figure S7. **Topology titrations of PR.** a) Peak height ratios in [^{15}N,^{1}H]-TROSY spectra upon addition of 1.5 mM water-soluble Mn^{2+}, or 1 mM detergent-soluble 16-DSA to detergent-solubilized PR. Green bars below the diagrams indicate the seven transmembrane helices and dashed lines in red indicate the threshold used in b. b) Cartoon representation of PR with red and blue spheres representing residues affected by Mn^{2+} (peak height ratio < 0.5) and 16-DSA (peak height ratio < 0.4), respectively. Regions most affected by Mn^{2+} are clustered around negatively charged amino acids.

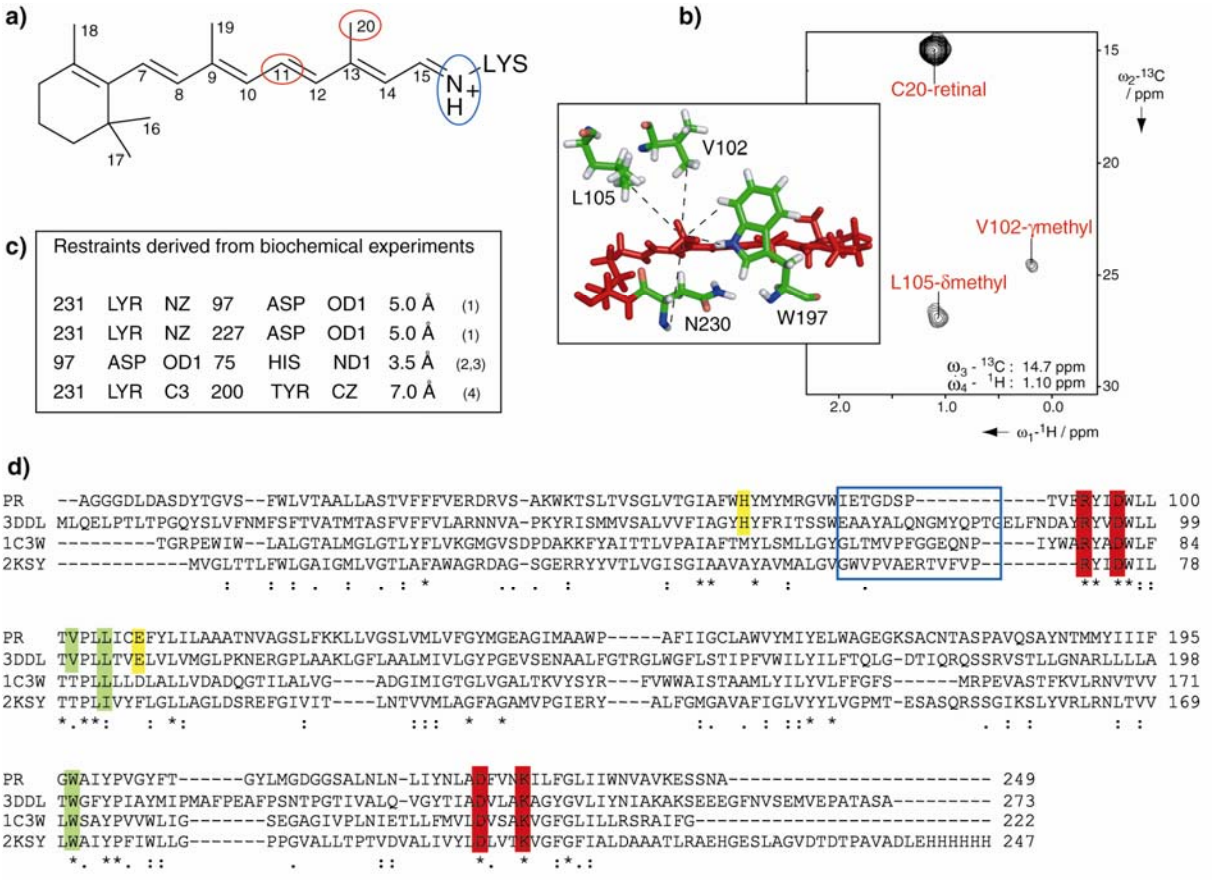


Figure S8. **Retinal assignment and experimental distance restraints.** a) Line broadening and spectral overlap hindered the assignment of the unlabeled cofactor and a small fraction of free retinal, giving rise to sharp and intense signals, made the unambiguous assignment of peaks from the bound cofactor challenging. A ^{13}C-labeled retinal was therefore employed including position C_{11} and C_{20} (red circles) close to the Schiff base (blue circle). In the [^{13}C,^{1}H]-HSQC spectrum the resonance of the methine group was assigned at δ(^{1}H) 5.04 ppm, δ(^{13}C) 140.17 ppm positioned in between the intense signals of H_2O and the glycerol head group of the detergent. Due to a low signal-to-noise ratio and its position close to water, meaningful data related to this cofactor position were not obtained. In contrast, the C_{20} methyl group could be assigned at δ(^{1}H) 1.09 ppm, δ(^{13}C) 14.93 ppm characterized by a reasonable signal-to-noise ratio and unaffected by overlap with detergent signals. A single population of the C_{20} resonance, as observed in our spectra, is expected for a predominantly all-*trans* conformation of the retinal.[33] Both signals were significantly upfield-shifted from the positions of the unbound retinal. b) In a 4D [^{13}C-^{13}C]-separated NOESY with uniformly ^{15}N,^{13}C-ALV labeled PR including the labeled cofactor, two NOEs to the C_{20} methyl group were identified suggesting a distance below 5 Å. In total, 6 distance restraints to the C_{20} retinal position could be derived involving residues V102, L105, W197 and N230. c) Four additional upper limit distance restraints derived from biochemical data were employed to model the retinal binding pocket. References are given below. d) Sequence alignment of PR with its homologues from bacteria (XR, PDB ID: 3DDL) and from archea (BR, PDB ID: 1C3W and SRII, PDB ID: 2KSY).

Highly conserved residues involving the retinal binding site and the primary proton acceptor are highlighted in red. XR and PR share the His residue close to the primary proton acceptor and a Glu as the primary proton donor in place of an Asp in BR (highlighted yellow). Residues that were involved in the experimental distance restraints with the C_{20} methyl group in PR are highly conserved and colored in green. The amino acid sequence forming the anti-parallel β-sheet is surrounded by a blue box. This region is significantly shorter in PR, as compared to the other three sequences, disfavoring formation of an extended β-sheet for a β-turn.

(1) D. Ikeda, Y. Furutani, and H. Kandori, FTIR study of the retinal Schiff base and internal water molecules of proteorhodopsin. *Biochemistry* **2007**, *46*, 18.

(2) F. Hempelmann, *et al*., The His75-Asp97 Cluster in Green Proteorhodopsin. *J. Am. Chem. Soc.* **2011**, *133*, 12.

(3) N. Pfleger, *et al*., Solid-state NMR and functional studies on proteorhodopsin. *Biochim. Biophys. Acta*, **2009**, *1787*, 6.

(4) J.Y. Jung, *et al*., Spectroscopic and photochemical analysis of proteorhodopsin variants from the surface of the Arctic Ocean. *FEBS Letters*, **2008**, *582*.

Tables

Table S1. **Restraints and structure calculation of PR.**

Restraints	
NOE distance restraints	
Sequential (1≤ |i - j| ≤ 2)	239
Medium-range (2 < |i - j| ≤ 4)	50
Long-range (|i - j| > 4)	87
Hydrogen bonds	133
Dihedral angle restraints	
Φ	196
Ψ	196
PRE distance restraints	
Upper limits	290
Lower limits	716
RDC restraints	81
Restraints from biochemical experiments	4
Structure statistics	
No. of conformers calculated/analyzed	200/20
Violations	
Distance restraint violations >0.2 Å	4
Ave. distance viol., Å	0.37 ± 0.03
Dihedral angle restraint violations > 5°	3
Ave. dihedral angle viol., °	7.24 ± 0.91
No. of RDC restraints	4
Ave. RDC viol., Hz	1.81 ± 0.39
Averaged CYANA target function (20/200)	6.06
Averaged rmsd to mean coordinates*	
Backbone, Å	0.81 ± 0.13
Heavy atom, Å	1.42 ± 0.15

* Determined in α-helical regions (25-50, 56-86, 91-117, 123-141, 149-167, 180-209, 216-246)

References

[1] B. R. Kelemen, M. Du, R. B. Jensen, *Biochim. Biophys. Acta* **2003**, *1618*, 25.

[2] P. Gourdon, A. Alfredsson, A. Pedersen, E. Malmerberg, M. Nyblom, M. Widell, R. Berntsson, J. Pinhassi, M. Braiman, O. Hansson, N. Bonander, G. Karlsson, R. Neutze, *Protein Expr. Purif.* **2008**, *58*, 103.

[3] D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, S. Sobhanifar, V. Dötsch, F. Bernhard, *Nat. Protoc.* **2007**, *2*, 2945.

[4] L. Shi, M. A. Ahmed, W. Zhang, G. Whited, L. S. Brown, V. Ladizhansky, *J. Mol. Biol.* **2009**, *386*, 1078.

[5] T. Etezady-Esfarjani, S. Hiller, C. Villalba, K. Wüthrich, *J. Biomol. NMR* **2007**, *39*, 229.

[6] a) K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. U S A* **1997**, *94*, 12366; b) M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 13585.

[7] a) E. Lescop, T. Kern, B. Brutscher, *J. Magn. Reson.* **2010**, *203*, 190; b) P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042.

[8] Z. Serber, C. Richter, D. Moskau, J.-M. Böhlen, T. Gerfin, D. Marek, M. Häberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2000**, *122*, 3554.

[9] K. Pervushin, A. Eletsky, *J. Biomol. NMR* **2003**, *25*, 147.

[10] a) R. Otten, B. Chu, K. D. Krewulak, H. J. Vogel, F. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 2952; b) M. Sattler, M. G. Schwendinger, J. Schleucher, C. Griesinger, *J. Biomol. NMR* **1995**, *6*, 11.

[11] A. Bax, F. Delaglio, S. Grzesiek, G. W. Vuister, *J. Biomol. NMR* **1994**, *4*, 787.

[12] C. Amero, P. Schanda, M. A. Dura, I. Ayala, D. Marion, B. Franzetti, B. Brutscher, J. Boisbouvier, *J. Am. Chem. Soc.* **2009**, *131*, 3448.

[13] M. W. F. Fischer, L. Zeng, E. R. P. Zuiderweg, *J. Am. Chem. Soc.* **1996**, *118*, 12457.

[14] V. Tugarinov, L. E. Kay, *J. Am. Chem. Soc.* **2003**, *125*, 13868.

[15] F. Löhr, V. Katsemi, M. Betz, J. Hartleib, H. Rüterjans, *J. Biomol. NMR* **2002**, *22*, 153.

[16] V. Tugarinov, L. E. Kay, I. Ibraghimov, V. Y. Orekhov, *J. Am. Chem. Soc.* **2005**, *127*, 2767.

[17] J. Iwahara, C. Tang, G. M. Clore, *J. Magn. Reson.* **2007**, *184*, 185.

[18] S. Macura, *J. Am. Chem. Soc.* **2009**, *131*, 9606.

[19] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014.

[20] D. Lee, C. Hilty, G. Wider, K. Wüthrich, *J. Magn. Reson.* **2006**, *178*, 72.

[21] A. Gautier, J. P. Kirkpatrick, D. Nietlispach, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 7297.

[22] T. Friedrich, S. Geibel, R. Kalmbach, I. Chizhov, K. Ataka, J. Heberle, M. Engelhard, E. Bamberg, *J. Mol. Biol.* **2002**, *321*, 821.

[23] F. Hempelmann, S. Hölper, M.-K. Verhoefen, A. C. Woerner, T. Köhler, S. A. Fiedler, N. Pfleger, J. Wachtveitl, C. Glaubitz, *J. Am. Chem. Soc.* **2011**, *133*, 4645.

[24] A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, *J. Mol. Biol.* **2001**, *305*, 567.

[25] L. J. McGuffin, K. Bryson, D. T. Jones, *Bioinformatics* **2000**, *16*, 404.

[26] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213.

[27] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171.

[28] J. L. Battiste, G. Wagner, *Biochemistry* **2000**, *39*, 5355.

[29] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283.

[30] D. K. Kirchner, P. Güntert, *BMC Bioinformatics* **2011**, *12*, 170.

[31] S. F. Poget, M. E. Girvin, *Biochim. Biophys. Acta* **2007**, *1768*, 3098.

[32] T. Ikeya, T. Terauchi, P. Güntert, M. Kainosho, *Magn. Reson. Chem.* **2006**, *44 Spec No*, S152.

[33] N. Pfleger, M. Lorch, A. C. Woerner, S. Shastri, C. Glaubitz, *J. Biomol. NMR* **2008**, *40*, 15.