J Mol Model (2005) 12: 49–64
DOI 10.1007/s00894-005-0004-z

ORIGINAL PAPER

Birgit Schlegel · Wolfgang Sippl · Hans-Dieter Höltje

Molecular dynamics simulations of bovine rhodopsin: influence of protonation states and different membrane-mimicking environments

Received: 2 January 2005 / Accepted: 1 July 2005 / Published online: 25 October 2005

Abstract G-protein coupled receptors (GPCRs) are a protein family of outstanding pharmaceutical interest. GPCR homology models, based on the crystal structure of bovine rhodopsin, have been shown to be valuable tools in the drug-design process. The initial model is often refined by molecular dynamics (MD) simulations, a procedure that has been recently discussed controversially. We therefore analyzed MD simulations of bovine rhodopsin in order to identify contacts that could serve as constraints in the simulation of homology models. Additionally, the effect of an N-terminal truncation, the nature of the membrane mimic, the influence of varying protonation states of buried residues and the importance of internal water molecules was analyzed. All simulations were carried out using the program-package GROMACS. While N-terminal truncation negatively influenced the overall protein stability, a stable simulation was possible in both solvent environments. As regards the protonation state of titratable sites, the experimental data could be reproduced by the program UHBD (University of Houston Brownian Dynamics), suggesting its application for studying homology models of GPCRs. A high flexibility was observed for internal water molecules at some sites. Finally, interhelical hydrogen-bonding interactions could be derived, which can now serve as constraints in the simulations of GPCR homology models.

B. Schlegel · H.-D. Höltje
Department of Pharmaceutical Chemistry,
Heinrich-Heine University of Düsseldorf,
Universitätsstrasse 1, Geb. 26.23/O2,
40225 Düsseldorf, Germany
Tel.: +49-211-8113854
Fax: +49-211-8113847

W. Sippl (✉)
Department of Pharmaceutical Chemistry,
Martin-Luther-University Halle-Wittenberg,
Wolfgang-Langenbeck-Str. 4, 06120 Halle, Saale, Germany
E-mail: wolfgang.sippl@pharmazie.uni-halle.de

Keywords Molecular dynamics · Protonation · G-protein coupled receptors · Membrane mimics · GROMACS

Abbreviations GPCR: G-protein coupled receptor · DPPC: Dipalmitoylphosphatidylcholine · POPC: Palmitoyloleoylphosphatidylcholine · DMPC: Dimyristoylphosphatidylcholin · RMSD: Root-mean-square deviation (nm)

Introduction

G-protein coupled receptors (GPCRs) are a protein family of outstanding pharmaceutical interest and comprise important members such as dopamine, serotonin, muscarinic, histamine or opiate receptors, just to name some examples. More than half of the therapeutically used drugs target these transmembrane proteins [1], all of which share a common mechanism of action transducing an extracellular signal into an intracellular response. Sequence analysis led to the assumption that the human genome could code for approximately 1000 GPCRs [2], a large number of which are still orphan receptors. A major breakthrough in the understanding of this receptor family was achieved in 2000, when the crystal structure of bovine rhodopsin was resolved [3], and for the first time detailed structural insights were gained.

The main obstacles in the crystallization of other members of this protein family are the usually low-expression levels, and consequently the problems in generating sufficient protein material for a crystallization procedure as well as the difficulties in the crystallization process itself [4]. Thus, it cannot be expected that crystal structures of many other GPCRs will become available in the near future.

Sequence analysis of GPCRs has shown that all members of this family share a 7-helix bundle as well as certain motifs (e.g., D(E)RY in helix 3, NPXXY in helix 7) and numerous highly conserved amino acids at the

transmembrane level [5]. The conserved fold of GPCRs represents the basis for applying the approach of comparative (homology) modeling, and the generation of a number of meaningful models of other members of this protein family stresses the success of this strategy [1, 6–8]. Initially, these rhodopsin-derived models were mainly limited to the study of antagonist-receptor interactions as bovine rhodopsin was crystallized in its inactive state, while the success of such models for studying agonist-receptor interactions has been discussed controversially [6]. However, increasing knowledge of the activation process of bovine rhodopsin has also led to new strategies in the generation of activated-models, taking into account structural changes that supposedly occur near the binding pocket of rhodopsin-like receptors upon activation, e.g., the anticlockwise rotation of helix 6 [6].

The initial model generation is often followed by molecular dynamics (MD) simulations of the receptor–ligand complexes in order to study the dynamic properties and to allow for structural adaptations resulting in energetically favorable models of the receptor studied. The gain in computational power has resulted in increasingly sophisticated simulation systems with ''natural'' lipid bilayers [e.g., dipalmitoylphosphatidylcholine (DPPC), palmitoyloleoylphosphatidylcholine (POPC), or dimyristoylphosphatidylcholin (DMPC)] being currently the state of the art. A membrane model is characterized by a set of force-field parameters that has to be carefully adapted in order to observe realistic physical properties during an MD-simulation. Parameter sets for a hydrated DPPC membrane models as used in this study were published recently [9] for the ffgmx force field or for derivatives of the ffG43a1 force field [10, 11], both implemented in the GROMACS package [12, 13].

The scope of the work presented here was to compare the influence of different simulation setups on the final outcome using the crystal structure of bovine rhodopsin as a prototype of GPCRs. The variations applied to the starting structures were thereby carried out in order to later rationalize the simulation setup of GPCR homology models. Thus, the influence of factors such as N-terminal truncation, simulation environment, choice of protonation states for titratable sites and the position of internal water molecules was analyzed. Although for the structure of bovine rhodopsin, the conformation of the N-terminal ending, the protonation states and the position of a number of internal water molecules are known, it is interesting to evaluate the effect of these parameters on the structural stability during an MD-simulation. The variations considered in this study were chosen because of the difficulty usually related with the proper incorporation of these factors into a homology model.

Many homology models of GPCRs, which are often generated automatically, account only for the transmembrane helices [14]. Inspection of the rhodopsin structure and mutation studies have, however, shown that residues, especially from the second extracellular loop, can be involved in ligand binding [7, 15] or are responsible for receptor-subtype specificity [16, 17]. Thus, neglecting the second extracellular loop is generally not advisable. Recent publications [18] have shown that loop generation is a time-consuming error-prone process. For long loops, such as the third intracellular loop of many rhodopsin-like receptors that can comprise more than 100 amino acids, no reliable procedures exist currently, so that a truncation of this loop to the corresponding length of the bovine rhodopsin structure is the only practicable strategy. Since the intracellular loop region is the site of G-protein interaction, which is not essential in the design of agonists/antagonists that bind in the extracellular part of the helix bundle, this simplification of the model seems to be justified as long as care is taken to avoid distortions of the position of the helices upon loop construction. Until recently, various residues located in the intracellular loop region could not be resolved in the crystal structure of bovine rhodopsin due to their high flexibility. In the structure 1U19, which was published recently by Okada et al. [19], co-ordinates could also be determined for this region. Still, it remains unlikely that this part of the protein has a defined structure. The construction of the extracellular loop region in proximity to the binding pocket, however, is more challenging, as this part is potentially involved in ligand binding and has a defined conformation that remains highly stable under simulation conditions. The N-terminal ending comprising 33 residues in the bovine rhodopsin structure presents a challenging task during the generation of homology models because of its length and low-sequence homology. In the case of, e.g., biogen aminergic receptors, the N-terminal ending has not been reported to be involved in ligand binding so that the importance of an explicit consideration is not known. Thus, the effect of an N-terminal truncation upon the overall protein integrity was tested in order to assess the possibility of potential neglecting this region.

Another question that has to be addressed is the nature of the simulation system chosen. In the most simplified case, a simulation in vacuum will be carried out. In order to account for the neglected solvent environment, the dielectric constant can be adjusted in most simulation packages. However, it is usually not possible to apply different dielectric constants as would be required for a transmembrane protein that interacts with both an aqueous and an apolar core-membrane environment. As described by Mehler et al. [18], there is a strong tendency in vacuum that polar or charged residues on the protein surface do not exhibit extended conformations as can be expected for these residue types in contact with water and ions, but fold back on the protein surface, producing artifacts in the hydrogen-bonding pattern. In order to consider a more realistic environment of transmembrane proteins explicitly, several strategies have been published using a CCl_4/water [20], an octane/water [21], or a phospholipid/water [21, 22] environment. Given that the proper function of a protein often requires a specific lipid composition, the use of CCl_4, octane, or saturated lipid molecules as

membrane mimics will apparently represent a simplification of the real circumstances. Still, a CCl_4/water and a saturated phospholipid/water system were chosen to simulate a model of bovine rhodopsin in order to analyze the influence of membrane mimics of different complexity and dissimilar physical properties.

A further point to be considered is the state of protonation of the buried residues in a protein, as the pKa of titratable residues can shift remarkably when exposed to the protein interior of considerably lower dielectric constant. The calculation of pKa-shifts for a protein is inseparable from an investigation of the presence of internal water molecules due to the possible interdependence. In the simulations described here, the choice of an inappropriate state of protonation had a pronounced effect on the structural integrity of the protein and resulted in an overall distortion of the hydrogen-bonding pattern, including distortions in the backbone regions. Regarding the water molecules located in the protein interior, an explicit consideration did not necessarily result in an improvement of the structural preservation, as will be described in the results section.

Several long simulations of the structure of bovine rhodopsin have been published recently by several groups [23, 24]. In the work of Woolf et al., the main focus was set on a postulated correlation between conformational changes of retinal and surrounding amino acid residues and a larger scale conformational change of the entire protein structure. It was speculated that fluctuations observed during a simulation of the dark-adapted rhodopsin (with retinal in its 11-*cis* conformation) could be part of the pathway to the light-adapted state. Additionally, interactions of the protein structure with the surrounding lipid and water environment were analyzed during the course of the simulation. Huber et al. investigated the complex interactions between the rhodopsin structure and the lipid environment using a POPC membrane mimic. Additionally, the dynamic behavior of internal water molecules and structural fluctuations of the entire rhodopsin structure were analyzed. Interestingly, Huber et al. did not report a change in the hydrogen-bond pattern after 10 ns of simulation period that was observed in the simulation of Woolf et al. A possible explanation could be the different system setups, which varied in the number of water molecules taken into consideration or protonation states for titratable groups in the protein interior. Because of the chaotic behavior of MD-simulations, small variations in the setup can result in different pathways through the energy landscape, thus, resulting in different conformations sampled during the course of an MD-simulation.

Although simulations of bovine rhodopsin already exist, the main focus is usually set on rhodopsin-specific questions such as activation or interaction with the lipid environment. In the present work, the simulations carried out are seen from the viewpoint of a homology modeler using the structure of bovine rhodopsin as a template structure. Thus, the main interest was to evaluate as to which circumstances lead to the most stable simulation and the consequence of simulation setups differing from this "optimal setup." The most stable simulation was also analyzed with regard to the interhelical contacts that seem to be necessary for retaining the rhodopsin protein in its inactive state. Identification of amino acids participating in this off-state hydrogen-bonding pattern could facilitate the generation of GPCR homology models. Namely, if an amino acid is likely to be involved in an interhelical contact, this will narrow down the conformational space accessible for this residue and favor side-chain conformations fulfilling these requirements. This strategy is apparently based on the assumption that the off-state hydrogen-bonding pattern is conserved within the rhodopsin-like GPCR family. This seems, however, likely due to the large number of highly conserved residues within the transmembrane region. Flohil et al. have suggested that unconstrained molecular-dynamics techniques might not be an appropriate method for homology model refinement and they propose constraining those parts of the model that have a high likelihood to be modeled correctly [25]. However, this strategy might be problematic when applied to GPCR homology models as there is evidence that some helices are devoid of kinks present in the template structure. Although individual helices are modeled correctly, the helix bundle might still change its topology, which would be impeded by position restraints. Thus, an alternative might be to restrain only interhelical contacts that occur at high frequency during a simulation of the template protein and are either highly conserved throughout the protein family or can be mapped easily in the target protein structure. For the identification of these interhelical contacts, the simulation was only analyzed up to a point where significant deviations from the crystal structure occurred. Apparently, the crystal structure does not represent a native structure due to packing effects and crystallization adjuvants. However, packing effects due to protein–protein interactions will be restricted mainly to loop regions and residues on the protein surface, which is supported by the low-structural rearrangements that have been observed at sites where heavy-metal ions had been localized previously in the simulation presented here and in Ref. [24]. It can therefore be expected that the interhelical hydrogen-bond pattern is not drastically influenced during a crystallization procedure, which is further supported by the low-structural rearrangement within the transmembrane regions reported in all simulations of bovine rhodopsin published and the persistence of hydrogen bonds observed during simulations with low-structural deviations from the crystal structure. Besides, structural rearrangement due to the removal of crystallization constraints has been reported to occur rather fast during an MD-simulation [25]. For the sake of completeness, it must be mentioned that in some cases where unconstrained MD-simulations were used for homology model refinement, a significant improvement could be achieved [26]. The required time-scales were, however, quite large (10–100 ns) and the focus was set to ab initio protein

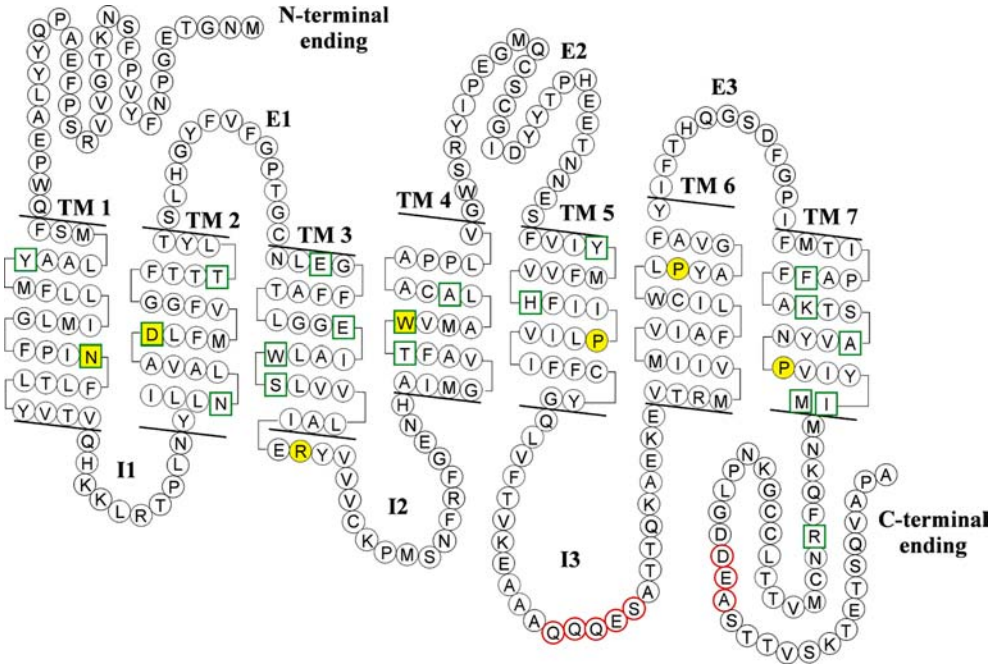

Fig. 1 Snake plot of the bovine rhodopsin protein. The most conserved residue within each transmembrane region (*TM*) is *highlighted yellow* (corresponding to position 50 in the Baldwin numbering scheme [5]). Residues in the third intracellular loop (I3) and C-terminal ending, which could not be resolved in the crystal structure 1HZX, are indicated by *red circles*. Residues that were found to be consistently involved in interhelical hydrogen bonds during MD-simulations are written in *green squares*. The length of transmembrane segments (as described in Swissprot Entry PO2699) is indicated by *parallel lines*

models that tended to rearrange structurally during the simulation.

Materials and methods

Generation of model structures of Opsin–Retinal

The co-ordinates of the crystal structure 1HZX [27] of bovine rhodopsin with a resolution of 2.8 Å were used as a starting point for all simulations described. No N-terminal acetylation, glycosylations, or a C-terminal palmitoyl moiety were considered. Missing residues (Fig. 1) in the A-chain (236–240, 331–333) localized in the third intracellular loop and the C-terminal ending, respectively, were added using the Loop Search and Splice Repair utilities of the Insight 2000 Homology module [28], where residues were added following a distance-matrix approach. Compared to the recently published crystal structure 1U19, where co-ordinates for the entire rhodopsin structure could be resolved [19], the modeled residues showed only small deviations from the experimental structure. All further calculations were carried out with the GROMACS simulation package using the ffG43a1 force field. In order to generate the protonated Schiff base that links the retinal with lysine 296/7.43,[1] the predefined retinal parameters (RTOL) in the ffG43a1.rtp file were adjusted and a new residue type (LYX) was defined in analogy to the predefined lysine (LYS) residue so that a covalent linkage could be set. Several models were generated that varied in their sequence length, the protonation state of the residues D83/2.50, E122/3.37, and E181/4.70 (E2-loop) and a potential consideration of internal water molecules as resolved in the crystal structure 1L9H [29].

Molecular dynamics simulation of Opsin–Retinal in a CCl_4/water box

Two CCl_4/solvent boxes with dimensions $8.56 \times 6.45 \times 8.09$ and $8.56 \times 6.45 \times 9.01$ nm were generated to simulate the truncated and entire rhodopsin models, respectively. Therefore, first, a CCl_4-box of the dimensions $8.56 \times 6.45 \times 3$ nm was generated and filled with CCl_4 molecules up to the correct density of 1.596 g cm^{-3} [20]. Then water molecules were added to fill the box. After 5000 steps of steepest-descent minimization, an MD-simulation was carried out for 1 ns using the parameters listed in Table 1. Parameters for the CCl_4 molecule were predefined in the ffG43a1.rtp file. However, in order to preserve the tetrahedral structure of the CCl_4 molecules, an additional "virtual" bond was set between chlorine atoms 4 and 5 (see ffG43a1.rtp file for atom numbering scheme). As an application of the Particle Mesh Ewald (PME) method [30] for calculation of Coulomb-interactions requires the system to be neutral, sodium ions were added to yield an uncharged system. Additionally, sodium and chlorine ions were added to yield a final concentration of 156 mVal/l, as an explicit consideration of ions has been shown to have beneficial effects on secondary-structure stability [31]. Co-ordinates of the Opsin–Retinal complex and the equilibrated box were merged with the command genbox that automatically removes solvent molecules whose van-der-Waals radius would overlap with the protein structure. In all calculations,

[1]Numbering scheme corresponding to Baldwin et al. [5]: the most conserved residue in each transmembrane segment is assigned position 50 (Fig. 1). The first number refers to the helical segment

Table 1 GROMACS parameters for the simulation setup of a model of bovine rhodopsin in a CCl_4/water environment

$dt = 2/4$ fs	
Neighbor-searching parameters	
nstlist	5
nstype	Grid
pbc	*xyz*
rlist	0.9 nm
Parameters for calculation of electrostatics and vdW interactions	
Coulomb-type	PME
vdW-type	Cut-off (*r*: 0.9 nm)
T-coupling	Berendsen (310 K, τ: 0.1)
p-coupling	No
Constraints	Hydrogen bonds (LINCS)

the Opsin–Retinal complex was initially tethered with a force of 1000 kJ mol^{-1} nm^{-2} in *xyz* directions. This tether was reduced stepwise (each 100 ps) from 1000 to 500 to 200 to 100 kJ mol^{-1} nm^{-2} [resulting in characteristic plateaus of the root-mean- square-deviation value (RMSD) at the beginning of each simulation]. This proceeding leads to superior results in terms of lower RMSD-values compared to a procedure where an unconstrained MD-simulation is carried out immediately following the initial minimization. Subsequently, an unconstrained dynamics simulation was carried out. No pressure coupling was applied in simulations using the CCl_4/solvent boxes; instead the right density was adjusted and the volume kept fixed. This strategy, which introduced artificially high values for pressure, seemed, however, justified from preliminary investigations where simulation of a CCl_4-box with pressure coupling resulted in a 10.2% deviation from the correct density. As deviations of the density are expected to influence the simulation result in a more pronounced manner than deviations of pressure, this strategy was given preference. Table 1 gives an overview of the parameters used for the simulations.

Molecular dynamics simulation of Opsin–Retinal in a DPPC/water box

As mentioned before, membrane models are characterized by the set of force-field parameters adjusted in order to reproduce realistic physical properties. In order to preserve the comparability of MD-simulations, we adapted the ffgmx-DPPC parameters used by Tieleman et al. [32] following the procedure described in Refs. [10, 11], resulting in lipid parameters optimized for application of the ffG43a1 force field that we had previously used for the simulations in the CCl_4/water environment. Starting from a pre-equilibrated box of 128 DPPC and 3655 water molecules [33], the box was enlarged in the *z* direction to yield the dimensions 6.56×6.59×9.21 nm, resulting in a system of 128 DPPC and 7637 water molecules. For the insertion of the Opsin–Retinal complex, a modified version of the mdrun program [34] was used to introduce a cylindrical hole of radius 2 nm (force constant for repulsive force: 50 kJ mol^{-1} nm^{-1} ; hx/hy/hz = 3.3/3.3/4.6). Subsequently, the co-ordinates of this box and of the Opsin–Retinal complex were merged and the resulting system equilibrated for 1 ns in order to allow for an adjustment of the box size to yield correct values for density and pressure. Sodium ions were again added to obtain a neutral system and, additionally, sodium and chlorine ions were added to yield a final concentration of 156 mVal/l.

In all calculations, the Opsin–Retinal complex was initially tethered with a force of 1000 kJ mol^{-1} nm^{-2} in *xyz* directions. This tether was reduced stepwise (each 100 ps) from 1000 to 500 to 200 to finally 100 kJ mol^{-1} nm^{-2} until an unconstrained dynamics simulation was carried out (see Table 2 for parameters).

Calculation of pKa-shifts of titratable amino acid residues

pKa-shift calculations were carried out for the complete Opsin–Retinal models in the absence and presence of the internal solvent molecules resolved by Okada et al. [29] using the program UHBD [35]. Additional residue types for lysine 296/7.43 and retinal were defined and added to the pkaS.dat database. All histidines were assigned with the type HisA. Calculations were carried out at 310 K and an ionic strength of 150 mM, setting the dielectric constants to 80 for solvent and 20 for the protein interior [36]. All cysteins (except for those involved in the disulfide linkage) were included in the calculation. Four grids (2.5, 65/65/65; 1.2, 40/40/40; 0.75, 40/40/40; 0.25, 40/40/40) were used and the maximum number of iterations was set to 300.

Results

Simulation setup for the analysis of interhelical contacts

The positional RMSD from the experimental structure was used as a measure for structural preservation. As stated in Ref. [26], the exclusive use of the RMSD value

Table 2 GROMACS parameters for the simulation setup of a model of bovine rhodopsin in a DPPC/water environment

$dt = 2/4$ fs	
Neighbor-searching parameters	
nstlist	5
nstype	Grid
pbc	*xyz*
rlist	0.9 nm
Parameters for calculation of electrostatics and vdW interactions	
Coulomb-type	PME
vdW-type	Cut-off (*r*: 0.9 nm)
T-coupling	Berendsen (310 K, τ: 0.1)
p-coupling	Berendsen (isotropic, 1 bar, τ: 0.5)
Constraints	Hydrogen bonds (LINCS)

as an indicator of structural changes might be too vague. Thus, similar to [26], the hydrogen-bond pattern was analyzed during the course of the simulation. In order to ease the survey of the hydrogen bonds, the contacts analyzed were restricted to interhelical interactions, which proved to be a good measure for structural preservation. The simulation setup defined through the following parameters resulted in a minimal deviation from the original crystal structure during the first 5 ns.

- Consideration of the entire protein sequence.
- Residues D83/2.50 and E122/3.37 in the protonated (neutral) state.
- No consideration of internal water molecules.
- DPPC/water environment.
- 2 fs time step.

One of the goals of this work was to retrieve a set of interhelical contacts that could be used as constraints for the simulations of GPCR homology models. The criteria for an observed contact being appropriate for defining such a constraint were that the contact must be stable during a simulation that resulted in only minor deviations from the experimental structure.

With the above setup, the RMSD within the helical backbone region remained stable for approximately 4200 ps after the initial equilibration phase (Fig. 2, red curve; Fig. 3a, b). The observed RMSD value of 0.15 nm, together with the high similarity of the side-chain conformations compared to the crystal structure can be considered sufficient to consider the model suitable for analyzing the dynamic behavior of interhelical hydrogen bonds.

The RMSD within the entire backbone region (Fig. 2, blue curve) of the bovine rhodopsin model was considerably higher due to deviations that occurred especially in the intracellular loop region. This flexibility is in agreement with the experimental finding that this part of the protein is difficult to crystallize due to its high

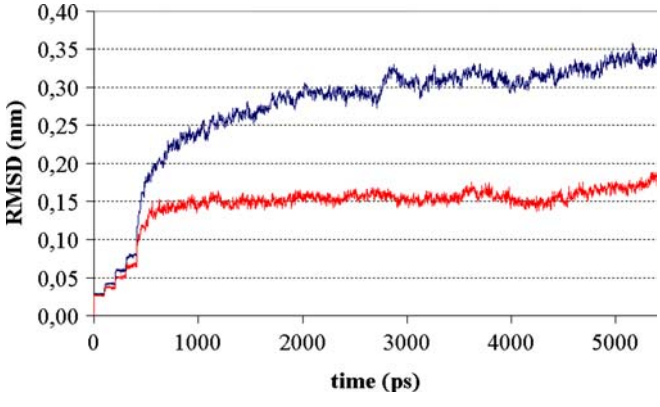

Fig. 2 RMSD (root-mean-square deviation)-course within the helical region (*red curve*) and the entire backbone region (*blue curve*) during the simulation of a model of bovine rhodopsin comprising the entire sequence in a DPPC (dipalmitoylphosphatidylcholine)/water environment. Residues D83/2.50 and E122/3.37 were considered in their neutral state. Internal water molecules were neglected

flexibility and other MD-simulations of bovine rhodopsin already mentioned. In contrast, the N-terminal end and the extracellular loop region including the second extracellular loop remained highly stable during the simulation, which is again in agreement with Huber et al. [24]. The largest deviation in this region consisted in an inward shift of the loop connecting helices 6 and 7 with a maximal displacement of 5.4 Å. Minor changes also occurred within the initial stretch of 10 amino acids of the N-terminal ending. After 4700 ps, a rise in RMSD can be observed which is related to the formation of a hydrogen bond between residues Y223/5.58 and R252/6.35. Formation of a hydrogen bond between the highly conserved residue Y223/5.58 [5] and different residues of helix 6 has also been observed in other simulation setups. In the crystal structure, Y223/5.58 is not involved in any hydrogen bond and points towards residue E232 located in the third intracellular loop. When the MD-simulation is prolonged up to 25 ns, Y223/5.58 becomes permanently involved in hydrogen-bond interactions with various residues from helix 6. Although changes in the hydrogen-bond pattern were observed for residues of helix 6, larger conformational changes of helix 6 were not observed during the simulation, which could not be interpreted as a partial activation, as reported by Woolf et al. In any case, it is doubtful if the complex switch from the off-state hydrogen-bond pattern to the activated state and larger conformational changes can be studied at all reasonably by means of MD-simulation, where neither the *cis–trans* isomerization of retinal nor adaptations of protonation states can be accounted for. Tyrosine 223 has been explicitly mentioned in the work of Huber et al. Here, however, Y223 exhibited a movement towards the lipid–water interface. This supports the statement of Flohil et al. that sampling of the conformational space of a protein is too slow to allow sufficient sampling at the time scale of 10s of nanoseconds. If, however, the amount of configuration space sampled cannot be estimated, it will be difficult to judge the significance of the output of individual MD-simulations. Additionally, it cannot be proven that configurations observed during a simulation are at all of biological relevance. Given that the bovine rhodopsin structure might be present as a dimer and that there are several indications for a G-protein precoupling for other GPCR proteins, the possibility exists that simulations of a simplified isolated rhodopsin structure will lead to artificial protein conformations. Given these uncertainties, it seems justified to use the similarity to the experimental crystal structure as a criterion for defining the time period appropriate for an analysis of interhelical contacts. The temporary equilibration phase can be seen also when the potential energy is plotted as a function of time (Fig. 4). Apparently, the formation of the interaction between helices 5 and 6 is accompanied by a decrease in potential energy of the protein structure. Although the lapse of potential energy has been used as criteria to estimate equilibration, it has been discussed whether total or group energies are a useful measure for

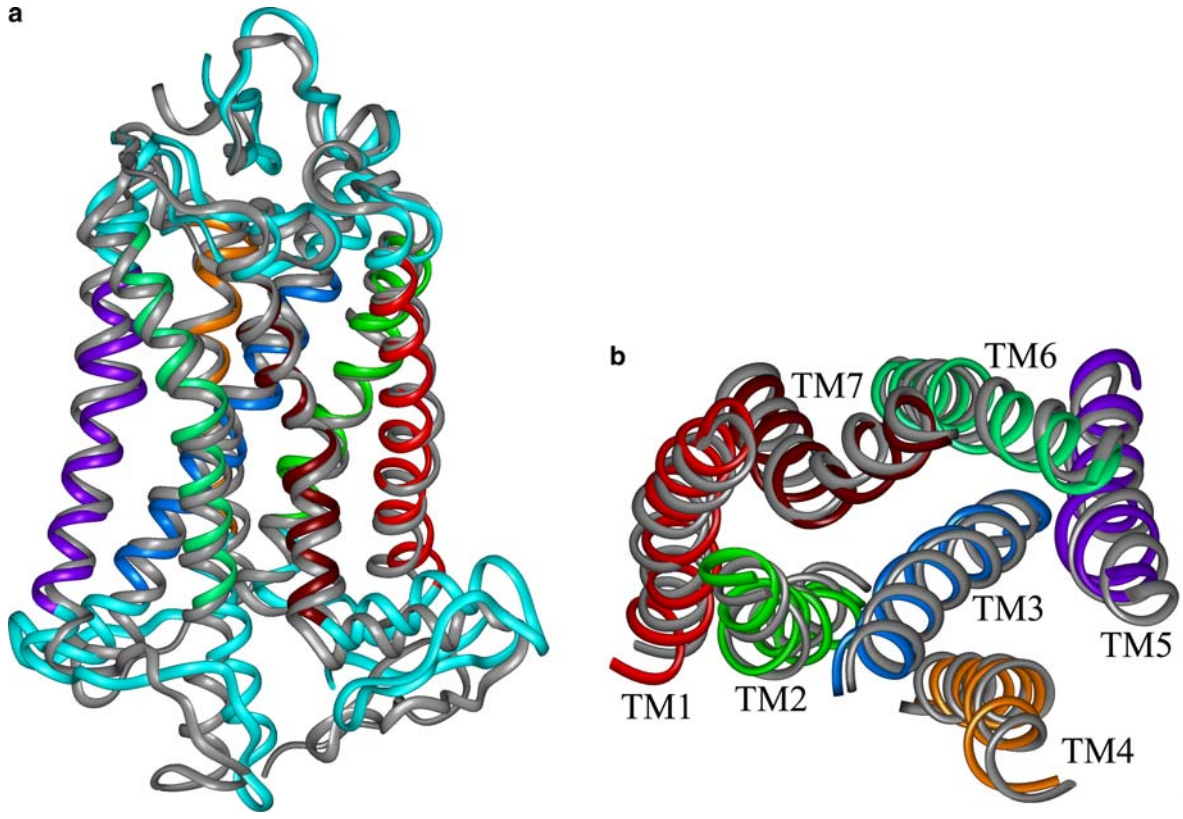

Fig. 3 Comparison of the backbone region of the crystal structure of bovine rhodopsin 1HZX (*light gray*) and a model that was simulated for 5 ns without constraints. **a** Side view; **b** helix bundle viewed from the extracellular space (loops are neglected for reasons of clarity). Apparently deviations from the crystal structure are small and predominantly limited to the intracellular loop region

the quality of a structure [25] or the end of the equilibration period required.

Figure 5 shows all interhelical hydrogen bonds described by Teller et al. [27] and additional ones that occurred at frequencies above 1% of the analyzed frames, which were written out for every 2 ps. Table 3 shows the corresponding frequencies of occurrence.

The calculations imply that strong interhelical contacts exist between helices TM3-TM7 that are on average linked by 1.9 hydrogen bonds. These interactions are produced mainly through the hydrogen bond reinforced ionic interaction of residue E113/3.28 and K296/7.43. Although in the crystal structure only a lateral interaction of one carbonyl oxygen atom with the nitrogen atom is observed, two simultaneous hydrogen bonds are found in most frames of the MD-simulations due to the fact that force-field methods favor a chelate-type interaction that is, however, rarely found for this kind of interaction in natural protein-ligand complexes. TM1-TM7 (1.7), TM2-TM4 (1.5), TM3-TM5 (1.3), TM4-TM5 (1.0), TM2-3 (0.8), TM6-TM7 (0.7) are linked by the average number of interactions indicated in brackets (see Fig. 5 for detailed contacts). In this simulation setup, an occasional hydrogen bond was also observed between helices 2 and 7, which seemed, however, artificial due to the neglect of internal water molecules. Consideration of an internal water molecule at a position corresponding to site 1a [29] would especially affect the hydrogen-bonding pattern around D83/2.50, as this water molecule forms contacts with D83/2.50, N302/7.49 and potentially G120/3.35, thus, cross-linking helices 2, 3, and 7. The corresponding hydrogen-bond pattern is described in a later section. Consideration of internal water molecules also significantly reduced the interactions between helices 6 and 7.

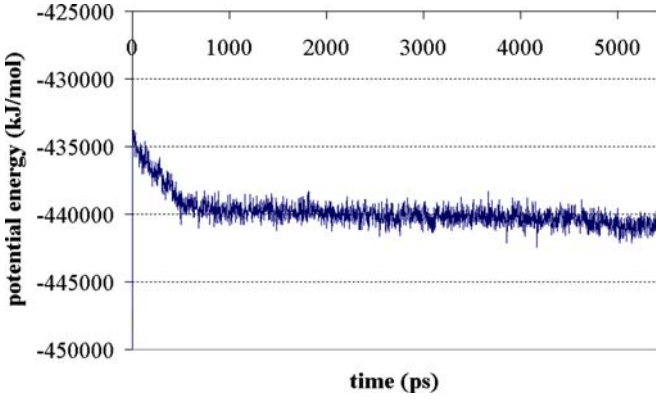

Fig. 4 Lapse of potential energy of the entire system during the simulation of a model of bovine rhodopsin, indicating a temporal equilibration period from approximately 1–4.8 ns. The eventual decrease of potential energy at the end of the simulation results from an artificial hydrogen bond described in the text

Further important hydrogen bonds within the extracellular and intracellular loop regions were established

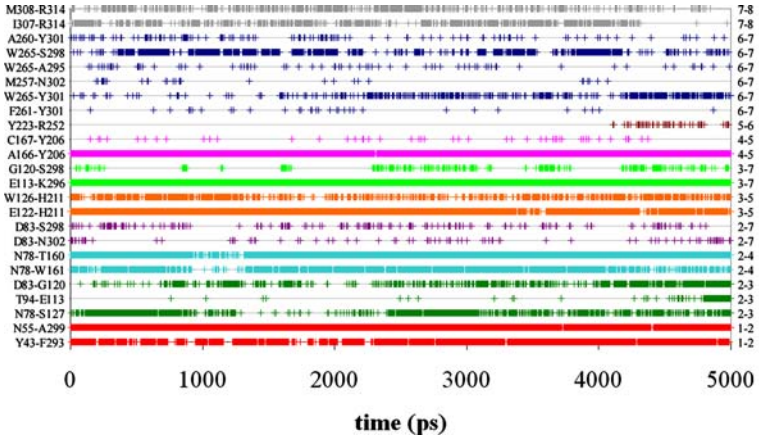

Fig. 5 Interhelical hydrogen bonds observed during a stable simulation of a model of bovine rhodopsin. Frames were written out every 2 ps. Each time the given hydrogen-bond interaction was found, a + was set. See Table 3 for frequencies of occurrence

between residues W175/4.64-S202/5.37 (92%) and R135/3.50-E247/6.30 (47%) at the frequencies given in brackets. Obviously, the ionic lock between helices 3 and 6, which was described previously by Shapiro et al., [37] was not present continuously. However, it was never lost during the course of the simulation.

Table 3 Frequencies of interhelical hydrogen bonds observed during the simulation of a model of bovine rhodopsin[a]

Interhelical contact	Residues involved	Frequency of occurrence (%)
TM1–TM2	N55-A80 1.50–2.47	< 1
TM1–TM2	N55-D83 1.50–2.50	< 1
TM1–TM7	Y43-F293 1.38–7.40	75
TM1–TM7	Y43-F294 1.38–7.41	< 1
TM1–TM7	N55-A299 1.50–7.46	95
TM1–TM8	Q64-T320 1.59–7.67	6
TM2–TM3	N78-S127 2.45–3.42	39
TM2–TM3	T94-E113 2.61–3.28	5
TM2–TM3	D83-G120 2.50–3.35	27*
TM2–TM4	N78-W161 2.45–4.50	61
TM2–TM4	N78-T160 2.45–4.49	92
TM2–TM7	D83-S298 2.50–7.45	10*
TM2–TM7	D83-A299 2.50–7.46	< 1
TM2–TM7	D83-V300 2.50–7.47	< 1
TM2–TM7	D83-N302 2.50–7.49	4*
TM3–TM4	W126-M163 3.41–4.52	< 1
TM3–TM5	E122-H211 3.37–5.46	91
TM3–TM5	W126-H211 3.41–5.46	37
TM3–TM7	E113-K296 3.28–7.43	99
TM3–TM7	G120-S298 3.35–7.45	13*
TM4–TM5	M163-H211 4.52–5.46	< 1
TM4–TM5	A166-H211 4.55–5.46	< 1
TM4–TM5	A166-Y206 4.55–5.41	97
TM4–TM5	C167-Y206 4.56–5.41	2
TM4–TM5	C167-H211 4.56–5.46	< 1
TM5–TM6	Y223-R252 5.58–6.35	2
TM6–TM7	E249-M309 6.32–7.56	< 1
TM6–TM7	F261-Y301 6.44–7.48	1
TM6–TM7	W265-Y301 6.48–7.48	18
TM6–TM7	M257-N302 6.40–7.49	1
TM6–TM7	W265-A295 6.48–7.42	3
TM6–TM7	W265-S298 6.48–7.45	40
TM6–TM7	A260-Y301 6.43–7.48	5
TM7–TM8	I307-R314 7.54–7.61	40
TM7–TM8	M308-R314 7.55–7.61	26

[a] Frames were written out every 2 ps. Not all of the spacially possible interactions as derived from the crystal structure by Teller et al. were observed during the simulation and seemed necessary for upholding the protein conformation. An asterisk was added to hydrogen-bond interactions that occur due to the neglection of internal water molecules that would otherwise prevent these hydrogen bonds

Truncated versus entire protein models

In order to test the influence of the explicit presence of the N-terminal ending on the overall protein integrity, a model of bovine rhodopsin lacking the initial stretch of 33 residues was simulated. In bovine rhodopsin, helix 1 interacts via two hydrogen bonds with helix 7, the highly conserved interaction of N55/1.50-A299/7.46 and the hydrogen bond formed between Y43/1.38 and the backbone oxygen of F293/7.40. Thus, it seems that the N-terminal ending folding over the entire protein is not necessary to anchor helix 1 to the remaining helix bundle. However, N-terminal truncation introduced undesirable side effects, as residues originally buried were brought to the protein surface facing the aqueous environment.

In the simulation of the truncated rhodopsin model (D83/2.50 and E122/3.37 protonated, internal water molecules considered, DPPC/water environment), deviations in the backbone region of the second extracellular loop and deviations of the position of helix 1 compared to the crystal structure were observed. While distortions of the E2 loop were due to the direct contact with the aqueous environment, artifacts in helix 1 were caused by the intrusion of extracellular water molecules and disruption of the interaction Y43/1.38-F293/7.40 after 3500 ps of unconstrained simulation. As a consequence, helix 1 separated with its extracellular part from helix 7, leaving, however, the interaction N55/1.50-A299/7.46 unaffected. At the height of L1.38, the shift of helix 1 amounted to 1.83 Å (measured from backbone L1.38:C to L1.38:C′). When the simulation was repeated in a CCl_4/water environment, the shift increased to 3.78 Å. The interaction N55/1.50-A299/7.46 was again unaffected, whereas the interaction Y43/1.38-F293/7.40 was lost after 2700 ps. In this simulation setup, the disruption of this hydrogen bond was, however, not caused by the intrusion of water molecules but rather owing to the

increased flexibility of the N-terminal ending of helix 1 starting at residue G1.46. The significantly larger structural deviations in this simulation are a consequence of the environment used, as will be discussed in more detail in a later section (CCl_4/water versus DPPC/water environment). Briefly, the DPPC/water environment has a greater capability to preserve this imperfect (because truncated) structure, whereas in the CCl_4/water environment deviations from the crystal structure were more pronounced as structural motions are less impeded by this environment.

In conclusion, truncation of the N-terminal region resulted in the deviations within the extracellular loop region and of the position of helix 1, compared to the crystal structure. Although helix 1 is anchored to helix 7 by hydrogen-bond interactions that were observed to be highly stable when the entire protein structure was simulated (Fig. 5), these interactions were disrupted in the truncated model either by the intrusion of water molecules or by the increased flexibility of helix 1 lacking the N-terminal ending.

Rhodopsin-like GPCRs usually lack the proline kink in helix 1 [38] and no hydrogen-bond linkage anchors the extracellular part of helix 1 to the helix bundle. During the simulation of homology models of bovine rhodopsin, the bend in helix 1 will quite likely disappear. In order to still anchor helix 1 to the helix bundle, an explicit consideration of the N-terminal ending will be essential.

Consideration of internal water molecules

In the crystal structure 1L9H of bovine rhodopsin (resolution 2.6 Å), 11 internal water molecules have been described by Okada et al. [29], of which seven are found within the transmembrane region. In order to analyze the dynamic behavior of these water molecules, we set up a 25 ns MD-simulation of an entire bovine rhodopsin model in a DPPC/water environment with D83/2.50, E122/3.37, and E181/4.70 considered in their protonated form using a time step of 2 fs. The 11 water molecules were positioned corresponding to coordinates resolved in 1L9H. In the structure 1U19, which became available only recently, the position of most water molecules at the transmembrane level is almost identical. A main difference is, however, the presence of an additional water molecule in the water cluster near residue D83. The consequences of an additional water molecule at this position on the output of the MD-simulation is difficult to judge. In the paper of Crozier et al., the simulation was carried out with structure 1F88, which was the only available experimental structure at that time. Although it had been argued that the simulation should not depend on possible alternative starting points—referring to the structure 1L9H with more water molecules resolved—several factors analyzed from their dynamics simulation, such as the interaction energy of Glu181 with retinal or the amount of transition observed for this residue, could well be influenced by a water molecule considered (1L9H) or neglected (1F88) interacting with Glu181. Although the intrusion and correct positioning of water molecules to the protein structure can occur during an MD-simulation, the required timescales to ascertain correct placement of all water molecules may be impracticably high.

In the simulation described here, the explicit consideration or neglect of certain water molecules did have effects on the hydrogen-bond pattern. This can also be seen from an inspection of the RMSD within the helical region as a function of time (Fig. 6, red curve). In Fig. 6, three zones can be roughly discriminated. An initial short plateau at 0.15 nm was formed during the first 2.5 ns, then the RMSD remained stable at 0.17 nm within the next 10 ns and eventually rose to a value of 0.2 nm, which was observed until the end of this simulation. When the RMSD curve is related to the interhelical hydrogen pattern, a potential reason for the increase of the RMSD value from 0.15 to 0.17 nm could be the loss of the interaction between Q64/1.59 and T320/7.67. More obvious is the coincidence of the increase of RMSD after 12 ns with the switch from hydrogen bond N55/1.50-A299/7.46 to the linkage between D83/2.50-A299/7.46 (Fig. 7).

The consideration of internal water molecules thus did not necessarily improve the quality of simulation. Internal water molecules were able to provoke disruption of conserved hydrogen bonds due to their high flexibility in regions rich in polar residues where changes in the hydrogen-bond pattern could occur easily. Rather than stabilizing the existing hydrogen-bond pattern, internal water molecules tended to catalyze the switch to alternatively possible hydrogen bonds. This is, of course, quite likely the intended natural function. However, the

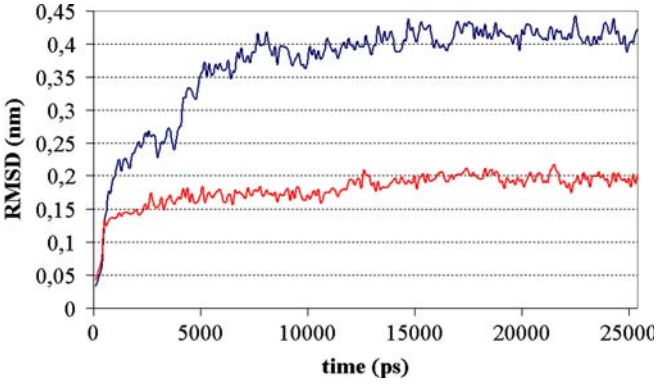

Fig. 6 Course of RMSD within the helical region (*red curve*) and the entire backbone region (*blue curve*) during the MD-simulation of a model of bovin rhodopsin comprising the entire sequence. D83/2.50, E122/3.35, and E181/4.70 were considered in their protonated state and 11 internal water molecules were placed according to Ref. [29]. Simulation was carried out in a DPPC/water environment with a time step of 2 fs. Although the simulation setup is more realistic, artifact production still cannot be preverted due to the high flexibility of the internal water molecules.Thus, an artificial linkage between D83/2.50 and A299/7.46 is formed resulting in a rise of RMSD after 12 ns

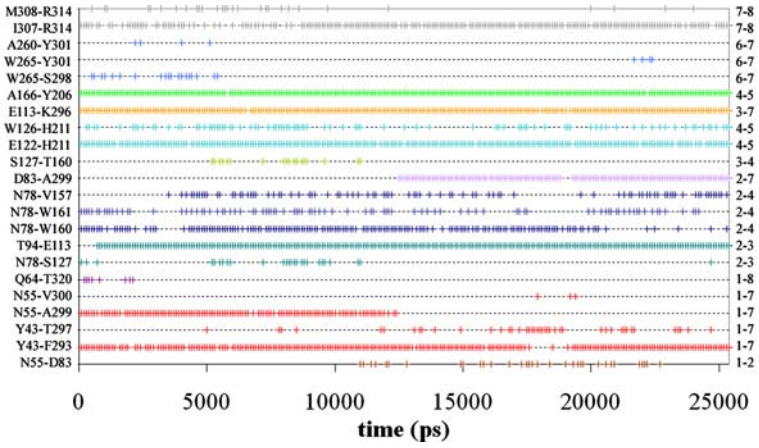

Fig. 7 Interhelical hydrogen bonds observed during a stable simulation of a model of bovine rhodopsin with internal water molecules considered. Frames were written out every 100 ps. Each time the given hydrogen-bond interaction was found, a + was set

original hydrogen-bond interaction pattern as observed in the crystal structure was preserved better when the disruption of a hydrogen bond could not be compensated by a temporary interaction with a flexible water molecule before a new interaction was established. However, this holds true only for water molecules located in protein-regions that were rich in polar residues. Water molecule 2021 at site 2b (numbering scheme referring to Ref. [29]) formed a stable interaction with E113/3.28, maintaining the exact side-chain conformation for this residue as observed in the crystal structure during the entire simulation. Molecule 964 (site 3) likewise did not substantially deviate from its original position. The reason for this persistence was that these water molecules were trapped with their interaction partner in an otherwise apolar environment.

For water molecules 2015, 2017, and 2020 at sites 1a, 1b, and 1c, respectively (cross linking helices 2, 3, and 7), and 2024 at site 4 (linking helices 2, 3, and 7) on the other hand, a great flexibility was observed. The best example for this behavior was molecule 2020, which hopped during the course of the simulation from its original position 1c to the proximal position 1a on to position 4 more than 12 Å away from the original site. Thus, there seems to exist a passage from the water cluster near D83/2.50 to the intracellular space. As mentioned above, an additional water molecule has been resolved in the structure 1U19 [19] that forms part of the water cluster 1a, 1b, and 1c. It will be interesting to analyze how the consideration of this additional molecule will influence the simulation outcome and the flexibility observed for the water molecules in this region.

Comparison of simulations in the presence and absence of internal water molecules showed that neglect of water molecules 2027, 2028, 2000, and 2014 did not alter side-chain conformation at the corresponding sites substantially. When these water molecules were, however, considered, they moved away from their original position, interfering with the existing hydrogen bonds located nearby.

As a subsumption, only two (2021, 964) of the 11 water molecules resolved in Ref. [29] remained stable at their original position hydrogen bonding to E113/3.28 and Y268/6.51-C264/2.47, respectively. Three water molecules (2015, 2017, and 2020) are described to form a cluster in the proximity to the highly conserved residues D83/2.50 and N302/7.49. These solvent molecules and molecule 2024 close to the intracellular loop were highly flexible, indicating a possible passage to the intracellular space. Four water molecules (2027, 2000, 2014, and 2028) did not seem to be essential.

Apparently, internal water molecules play an essential role in stabilizing the local hydrogen-bond networks. Unfortunately, the exact localization of internal water molecules is extremely difficult and no reliable methods exist to solve this task satisfactorily. Rarey et al. [39] have, for example, described a method for water-molecule placement during ligand docking. However, the success of this strategy was limited; of the 200 PDB–protein-ligand complexes in which water molecules were involved in the active site, only 35% of the water locations could be predicted correctly. Alternatively, a crystal structure lacking internal water molecules could be scanned with a water probe, for example, as implemented in the program GRID [40] that maps favorable interaction sites for H_2O molecules. However, here the results are often quite difficult to interpret because of the large number of indicated sites with small interaction fields and depend strongly on the contour level used. Using this program, the strongest interaction fields, indicating favorable sites for water molecules, were generated for the sites corresponding to the location of water molecules 2020 (in contact with Y301/7.48 and N302/7.49) and 2015 (in contact with D83/2.50, G120/3.35, and N302/7.49).

In homology models, the exact localization of water molecules will be even harder because of the uncertainties regarding the side-chain conformation that are usually added by programs such as SCWRL [41]. When using the SCWRL algorithm on the rhodopsin backbone, many residues are predicted correctly. Deviations were observed mostly for side chains facing the lipid or aqueous environment, for which, however, a high flexibility was observed under MD-conditions. Variations in

the side-chain conformations of residues facing the helix bundle interior were observed for Y306/7.53, E113/3.28, E181/4.70, and R135/3.50. Interestingly, the first three residues mentioned are in contact with internal water molecules; thus, a potential function for these water molecules could be to restrain an otherwise suboptimal side-chain conformation.

Although the results of applying the SCWRL algorithm to the rhodopsin backbone are quite promising, it must be held in mind that SCWRL places the side chains in a backbone-dependent manner so that in the case of backbone deviations occurring in the homology model a correct placement will be problematic.

Even when the exact position of the H_2O–oxygen is known, the exact orientation of the water molecules is still not, and anyway seems to be highly flexible within regions of polar character, at least under MD-simulation conditions. In order to avoid the described flexibility that can cause the disruption of conserved hydrogen bonds, a possible strategy would be the introduction of distance restraints fixing the position with respect to the interacting residues. In the case of homology models, this would be an applicable strategy for water molecules cross linking highly conserved residues such as solvent molecule 2015 of the water cluster near D83/2.50.

A stable simulation can be achieved even in the absence of internal water molecules. However, the corresponding sites will then turn metastable. It is thus possible that long-lasting unconstrained simulations of GPCR homology models lacking internal water molecules will introduce artifacts in the models simulated.

CCl_4/water versus DPPC/water environment

A typical phospholipid membrane such as DPPC/water spans about 4.5 nm [42] of which the hydrocarbon interior accounts for approximately 3.0 nm. Although the lecithin headgroup is of zwitterionic nature, the resulting charge distribution is almost completely cancelled since the charge distribution of choline and the phosphate atom overlap to a large extent and the remaining charge is neutralized further by the distribution of water dipoles [43].

In a CCl_4/water simulation environment, the CCl_4 layer is adjusted to 3 nm, thus mimicking only the hydrophobic core of the membrane. This simulation system obviously lacks the DPPC–headgroup interface. However, since this region is rich in water molecules, replacement by an aqueous environment should not introduce large errors.

In order to compare the suitability of the CCl_4/water environments with the DPPC/water environments for a GPCR simulation, we compared the simulation described in ''Introduction'' section to a simulation in a CCl_4/water environment.

Figure 8 shows the RMSD curve within the transmembrane region for both the DPPC/water (red curve) and a CCl_4/water (blue curve) environments. During the simulation in the DPPC environment, a lower RMSD value was maintained for about 4800 ps until the hydrogen bond Y223/5.58-R252/6.35 previously described occurred. In a CCl_4/water environment, the same hydrogen bond was established at an earlier stage (after 500 ps), resulting in a 0.02 nm upward shift of the RMSD curve during the first 4800 ps of simulation until the two curves eventually converged. Apart from the interaction between helices 5 and 6, the interhelical hydrogen-bond pattern was highly comparable within the two simulation environments (Fig. 9). An exception is the interaction between helices 2 and 7, which is an artifact of simulation setups lacking internal water molecules that would otherwise impede this interaction, as described in the section ''Consideration of internal water molecules.'' Thus, the only relevant differences in the hydrogen-bonding pattern rather take place due to local metastabilities resulting from the neglect of the G-protein (that could be a potential reason for the formation of the hydrogen bond between residues Y223/5.58 and R252/6.35) or internal water molecules than due to differences of the solvent environment used. In both cases a stable simulation (resulting in a plateau of the RMSD value) was possible and eventually the same errors occurred.

The earlier onset of deviations from the original structure in the CCl_4/water environment lets the DPPC/water environment appear more suitable for maintaining the model close to the experimental structure at first sight. This result is, however, consistent with two possible interpretations. Either a DPPC/water simulation system temporarily maintains smaller deviations due to the more natural environment or due to a generally decreased flexibility in a more viscous medium. In order to test these hypotheses, we repeated the simulation with

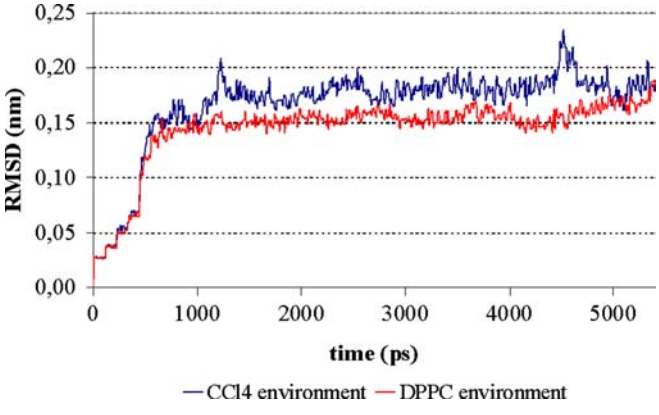

Fig. 8 RMSD curves within the transmembrane regions for a simulation of a model of bovine rhodopsin in a CCl_4/water (*blue curve*) and DPPC/water (*red curve*) environment. The upward shift of the curve corresponding to the CCl_4/water environment is caused by an artificial hydrogen bond. In the DPPC/water environment, formation of this interaction is postponed but still occurs resulting in the convergence of the curves towards the end of the simulations

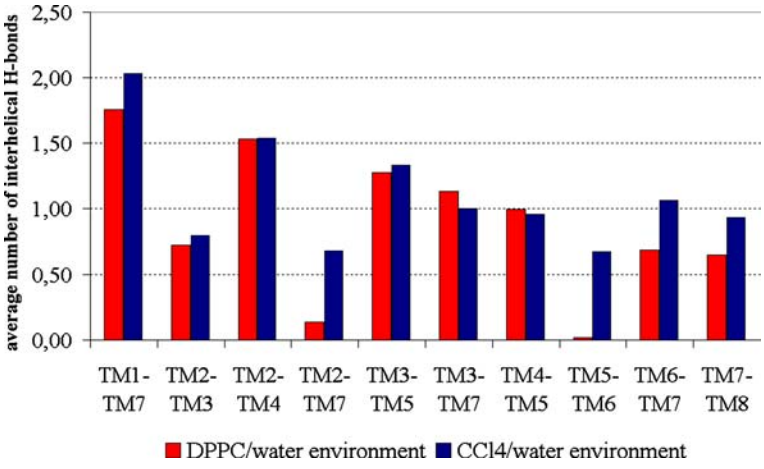

Fig. 9 Comparison of the average number of interhelical interactions for the two environments probed. Apart from the differences in the interaction of TM2–TM7 and TM5–TM6 (see text), comparable results were obtained

a truncated model of bovine rhodopsin with D83/2.50 and E122/3.37 in the deprotonated state—for which we observed large deviations from the starting structure when simulated in a CCl_4/water environment (Fig. 10, blue curve)—in the DPPC/water environment (Fig. 10, red curve). Apparently, a DPPC/water environment seems to slow down adaptation of the protein structure, even when they seem justified. This must be considered as a drawback of this simulation system when used to simulate homology models, where structural adaptations are expected to take place during the course of a simulation (e.g., in helix 1). Visual inspection of the resulting protein structures reveals that local distortions of the backbone regions occurred in both simulation setups. However, these modifications were accompanied by translocations of entire helices in the CCl_4/water environment that seem to be impeded in the DPPC/water environment.

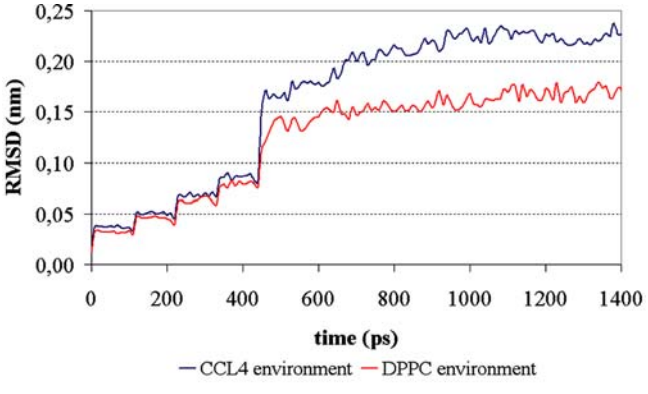

Fig. 10 Comparison of RMSD within transmembrane regions of a model of bovine rhodopsin simulated in a CCl_4/water (*blue curve*) and DPPC/water (*red curve*) environment. Choice of an incorrect protonation state for D83/2.50 and E122/3.37 resulted in a pronounced increase of RMSD in the CCl_4/water environment, whereas in the DPPC/water environment the slope is apparently much smaller. This indicates that a DPPC/water environment slows down adaptations of protein structures much more than the CCl_4/water environment

Choice of the correct state of protonation

Calculation of pKa-shifts using the UHBD program

At physiological pH of 7.4, acidic amino acids such as aspartate and glutamate (pKa = 4.4) are deprotonated/negatively charged, while basic amino acids such as lysine (pKa = 10.0) and arginine (pKa = 12.0) are protonated, thus positively charged. Histidine (pKa = 6.5) will be present in an equilibrium of protonated and deprotonated states, while cysteins (pKa = 8.5), and tyrosines (pKa = 10.0) are to a large extent neutral. These pKa-values are, however, only true for an aqueous environment with a dielectric constant of $\varepsilon_r \approx 80$. In a protein environment with an estimated dielectric constant of 2–20, the pKa values can considerably shift from the values given. This shift is mainly due to two factors: a desolvation of the residue (that will favor the corresponding neutral forms of all titratable sites) and stabilizing interactions between close titratable residues within the protein (that can stabilize a charge even in an apolar environment). The UHBD program calculates pKa-shifts in proteins by determining the differences in the electrostatic work of altering the charge of a titratable group from the unprotonated to the protonated state in the protein and the work of making the same alteration for the residue in aqueous solution (for details see the UHBD manual). The program output consists of the calculated state of protonation at various pH-values and pKa-shifts for all titratable sites in the protein given.

Experimental findings obtained by Fourier-transform infrared difference spectra (FTIR) analysis of rhodopsin and metarhodopsin II [44] indicate the protonation of D83 and E122, while Raman vibrational spectra described in [45] support the view that E181 is also present in its protonated state. However, for other rhodopsin-like GPCRs, no such data exist and it was thus tried to evaluate if analogous results could be implied from an application of the UHBD program, MD-simulations and multiple sequence alignments (MSA).

In Table 4, the state of protonation and the calculated pKa-shifts, as suggested by the UHBD program, are listed for those residues that deviate significantly from the pKa in aqueous solution. The calculations were carried out twice, once for the Opsin–Retinal complex, then also taking into account the water molecules as resolved in the crystal structure 1L9H [29]. However, pKa-shifts were identical so that only the results of one calculation are given. The most striking pKa-shifts were observed for the acidic residues D83/2.50, E122/3.37, and E181/4.70 in the E2 loop, which were all predicted to be in their neutral states.

D83/2.50 is a highly conserved residue throughout the GPCR family and mutation studies reveal that this residue is involved in the activation process. As described in Ref. [29], D83/2.50 is stabilized by an interaction with water molecule 1a linking helices 2 (D83/2.50), 3 (G120/3.35), and 7 (S298/7.45). Protonation of this residue is also supported by the observation of the existence of a sodium-binding site [46]. The presence of a sodium ion would result in a compensation of the potential negative charge of D83/2.50; thus, the existence of the neutral form of D83/2.50 in the absence of ions seems convincing.

If the protonated neutral forms of E122/3.37 and E181/4.70 were considered, E122 could form a hydrogen bond with the backbone carbonyl group of H211/5.46, and E181 could be stabilized by an additional hydrogen bond to Y268/6.51. The existence of E122/3.37 in its protonated neutral form is also supported by a MSA (ClustalW, default settings, identity matrix) of biogen aminergic GPCRs around this site (Fig. 11). In most sequences the corresponding position is occupied by a threonine, serine or asparagine; thus, the hydrogen donor ability seems to be the key function of residues at this location, rather than the introduction of a charge. Finally, for residue E181/4.70, a potential implication as a counterion switch for residue E113/3.28 has been discussed [45].

Table 4 pKa-shifts for the titratable sites in the Opsin–Retinal complex showing significant deviations from the expected protonation state in aqueous solution

Residue	Aqueous pKa	Intrinsic pKa	Charge at pH = 7.4
D83/2.50	4.0	9.8	0.0
E122/3.37	4.4	12.2	0.0
E181/4.70	4.4	12.1	0.0

MD-simulations comparing different states of protonation for residues D83/2.50, E122/3.37 and E181/4.70

In order to analyze the effect of different states of protonation of residues D83/2.50, E122/3.37, and E181/4.70 on the overall protein stability, MD-simulations with models varying in these parameters were carried out.

If residue D83/2.50 was considered negatively charged and no internal water molecules were considered, this residue would disrupt the highly conserved hydrogen-bond interaction between N55/1.50 and A299/7.46 (Fig. 12a). Residue D83, which was not involved in any stabilizing interaction in the crystal structure, formed a hydrogen bond with the backbone nitrogen of residue V300/7.47 (Fig. 12b). Formation of this interaction was accompanied with a pronounced relocation of the backbone region of helix 7, resulting in deviations from the ideal helix structure. Whereas consideration of D83 in its protonated state resulted in a conformation analogous to that observed in the crystal structure, where D83 was not involved in any hydrogen-bond interaction in simulations neglecting internal water molecules (Fig. 12c).

In one simulation setup with the consideration of D83 in its charged state, we observed the intrusion of an intracellular water molecule that moved up to the location of the water cluster described in Ref. [29] during the equilibration period. Under these conditions, residue D83/2.50 interacted permanently with G120/3.35 and S298/7.45 bridged by this water molecule, which prevented the distortions previously described.

Of the four possible setups for residue D83/2.50, i.e., D83 protonated or unprotonated ± internal water molecules, the alternative D83(−) without any water molecules consistently resulted in the distortion of the protein backbone. The variant D83H in the absence of internal water molecules temporarily reproduced the

α1-AA	C	N	I	W	A	A	V	D	V	L	C	C	**T**	A	S
β1-A	C	E	L	W	T	S	V	D	V	L	C	V	**T**	A	S
D1	C	N	I	W	V	A	F	D	I	M	C	S	**T**	A	S
M2	C	D	L	W	L	A	L	D	Y	V	V	S	**N**	A	S
α2-AA	C	E	I	Y	L	A	L	D	V	L	F	C	**T**	S	S
D2	C	D	I	F	V	T	L	D	V	M	M	C	**T**	A	S
5-HT2A	C	A	V	W	I	Y	L	D	V	L	F	S	**T**	A	S
5-HT1A	C	D	L	F	I	A	L	D	V	L	C	C	**T**	S	S
5-HT4	C	L	V	R	T	S	L	D	V	L	L	T	**T**	A	S
H3	C	K	L	W	L	V	V	D	Y	L	L	C	**T**	S	S
rhodopsin	C	N	L	E	G	F	F	A	T	L	G	G	**E**	I	A

Fig. 11 MSA of representatives of the family of human biogene aminergic GPCRs and bovine rhodopsin starting with the highly conserved cysteins residue in the E1 loop involved in the disulfide linkage. The corresponding position to E122 in bovine rhodopsin is highlighted and aligns with residues characterized by their hydrogen-donor functions

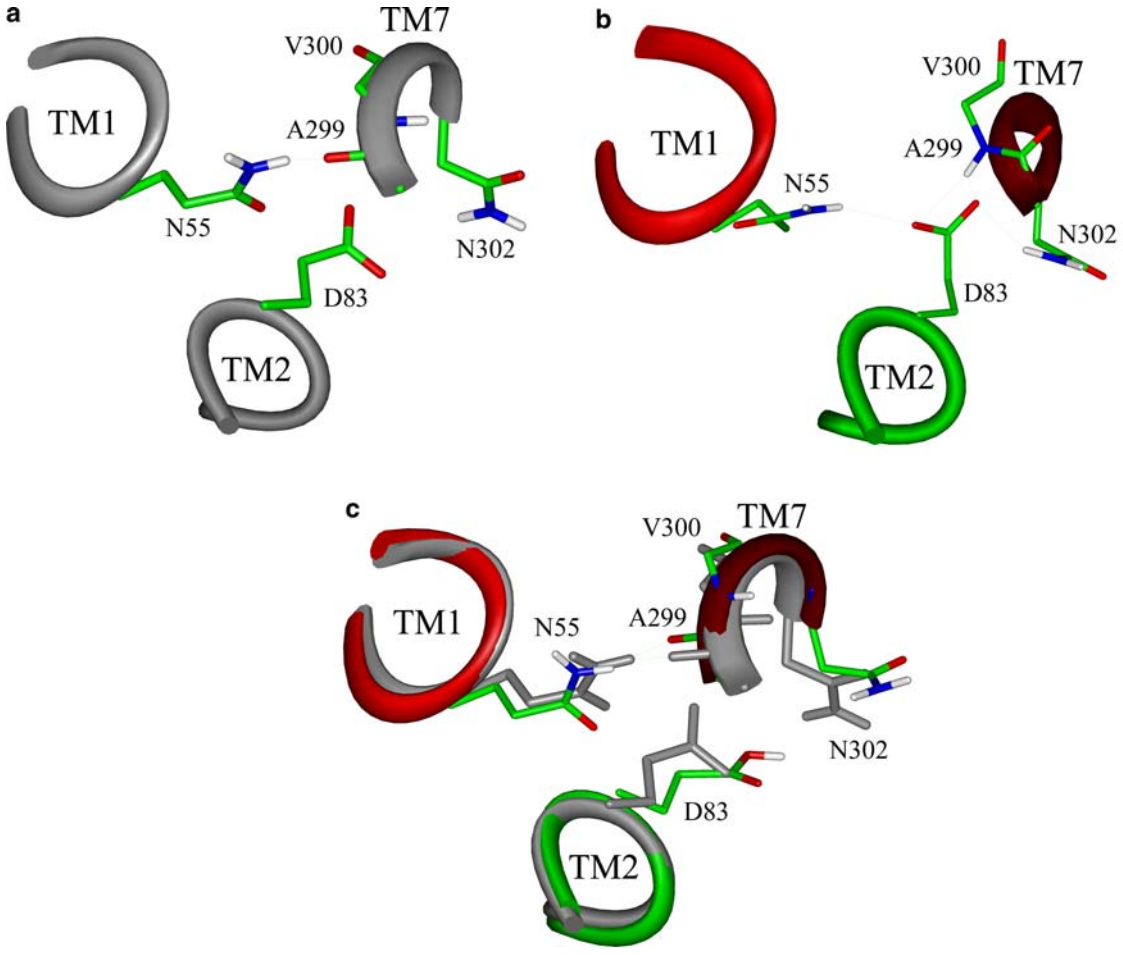

Fig. 12 **a** Hydrogen-bond pattern around residue D83/2.50 as observed in the crystal structure 1HZX. **b** Hydrogen-bond interactions obtained after 1 ns unconstrained MD-simulation of a model of bovine rhodopsin with the consideration D83 in its deprotonated state. **c** Hydrogen-bond pattern observed after 5 ns unconstrained MD-simulation of a model of bovine rhodopsin with D83 in its protonated state compared to the reference structure (*light gray*)

circumstances in the crystal structure. It will, however, represent a metastable system, as D83 is not involved in any hydrogen-bond interaction. In the presence of internal water molecules, both protonation states of D83 are consistent with structural preservation. However, due to the high flexibility of internal water molecules in an unconstrained dynamics simulation, the hydrogen-bonding pattern that was observed in the crystal structure was only temporarily preserved (Figs. 6, 7).

The situation for residue E122/3.37 was rather straightforward. A MSA (Fig. 11) suggested that at this position a hydrogen-donor function might be important. This was supported by the MD-simulations that were carried out. In the setup where E122 was considered negatively charged, a distortion of the backbone region of helix 4 could be observed, and residues in this region deviated significantly from the conformation observed in the crystal structure (Fig. 13a, b), whereas consideration of E122 in its protonated state led to negligible deviations from the crystal structure (Fig. 13c).

Similarly, protonation of residue E181 resulted in smaller deviations from the crystal structure.

As mentioned before, experimental data available for bovine rhodopsin has suggested that D83, E122, and E181 are protonated within the natural structure. This was also the conclusion that could be drawn from the predicted pKa-shifts by the UHBD program, MD-simulations and the analysis of MSA. This coincidence of calculated values with experimental findings stresses the usefulness of studying homology models of GPCRs by means of MD-simulations and chemoinformatic tools such as the program UHBD or MSA as these methods have been shown to give valuable information on the protonation states of buried residues [47, 48]. In order to further assess the ability of the UHBD program predicting pKa-values of buried residues in GPCR proteins correctly, we calculated the protonation state of D3.32 in a model of the human histamine H_3 receptor. Residue D3.32 is conserved within the family of biogen aminergic receptors and known to interact with protonated moieties of numerous ligands. In the case of a ligand-occupied or solvent-occupied binding pocket, D3.32 was predicted to be in its deprotonated form. However, although this result coincides with the experimental observations, it must be

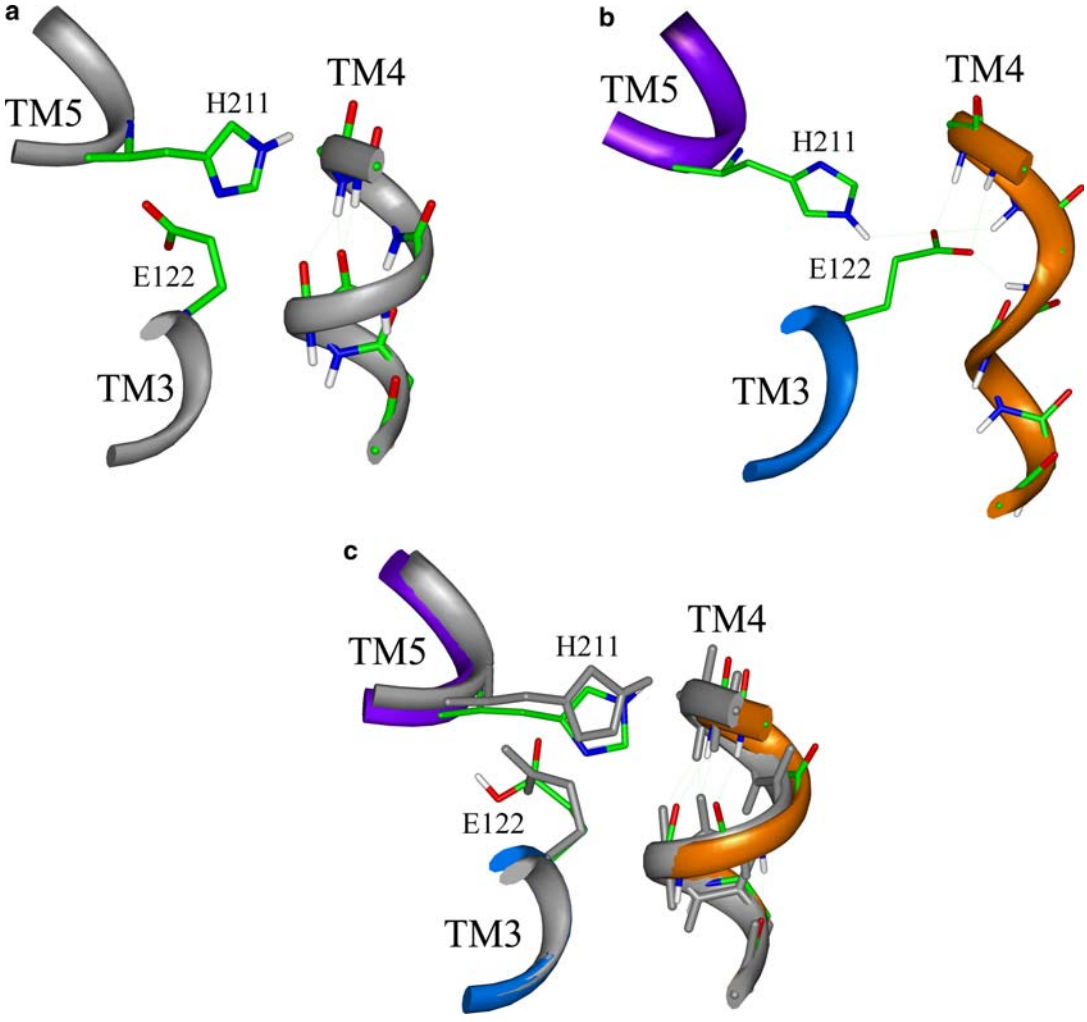

Fig. 13 **a** Hydrogen-bond pattern around residue E122/3.37 as observed in the crystal structure 1HZX. **b** Hydrogen-bond interactions obtained after 1 ns unconstrained MD-simulation of a model of bovine rhodopsin with the consideration of E122 in its deprotonated state. **c** Hydrogen-bond pattern observed after 5 ns unconstrained MD-simulation of a model of bovine rhodopsin with E122 in its protonated state compared to the reference structure (*light gray*)

considered that the calculated pKa shift will strongly depend upon the exact side-chain conformation, which is not provided by experimental data in the case of a homology model.

Conclusions

The rational design of new drugs targeting GPCRs is often hampered because no crystal structures are available for these target proteins. A potential workaround is the approach of comparative modeling, and homology models of GPCRs have been shown to be valuable tools in the drug-design process, e.g., through their use in database-screening or virtual de novo ligand-design techniques. After the initial model generation, a subsequent MD-simulation is often carried out in order to study the dynamic behavior of the protein-ligand complex and allow for structural adaptations. Unconstrained MD-simulation as a tool to refine homology models has, however, been reported to lead to models that resemble the experimental structure less [49]. We therefore analyzed an MD-simulation of bovine rhodopsin, which is used as template structure for most GPCR homology models in order to identify contacts that could serve as constraints in the simulation of homology models. Additionally, the effect of an N-terminal truncation, the nature of the membrane mimic, the influence of varying protonation states of buried residues and the importance of internal water molecules has been analyzed. N-terminal truncation had a negative effect on the structural preservation during MD-simulations and although currently no reliable computational procedures exist for determining the secondary structure of this region, its explicit consideration seems to be reasonable. Concerning the simulation environment, we observed that protein flexibility in a DPPC/water environment is significantly decreased compared to a CCl_4/water environment. Although the DPPC/water environment emulates the natural circumstances more clo-

sely, the increased viscosity of the medium will require much longer simulation periods in order to observe comparable deviations—when required. If the goal is to rapidly compare alternative GPCR homology models with respect to their structural stability, application of the more artificial CCl_4/water environment might be justified because of its significantly lower hindrance of even large conformational changes. Regarding the choice of protonation states for buried residues, calculated pKa-shifts by the program UHBD together with the results from MD-simulations and the analysis of MSA were able to reproduce experimental data available for the bovine rhodopsin structure, suggesting its application for the analysis of rhodopsin-like GPCR homology models.

Acknowledgements HDH acknowledges financial support from the HPC-Europa transnational access program. We wish to thank Peter Tieleman for making available the DPPC/SOL box to the scientific community via Ref. [33].

References

1. Klabunde T, Hessler G (2002) Chem Bio Chem 3:928–944
2. Rang H, Dale M, Ritter J, Moore P (2003) Pharmacology, 5th edn. Churchill Livingstone, London,
3. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima A, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Science 289:739–745
4. Caffrey M (2003) J Struct Biol 142:108–132
5. Baldwin J, Schertler G, Unger V (1997) J Mol Biol 272:144–164
6. Bissantz C, Bernard P, Hibert M, Rognan D (2003) Proteins: Struct Funct Genet 50:5–25
7. Voigtländer U, Raasch A, Tränkle C, Buller S, Ellis J, Mohr K, Jöhren K, Höltje HD (2003) Mol Pharmacol 64:21–31
8. Bröer B, Gurrath M, Höltje HD (2003) J Comput Aided Mol Des 17:739–754
9. Anezo C, de Vries AH, Höltje HD, Tieleman DP, Marrink SJ (2003) J Phys Chem 107:9424–9433
10. Schuler LD, Daura X, van Gunsteren WF (2001) J Comput Chem 22:1205–1218
11. Chandrasekhar I, Kastenholz M, Lins RD, Oostenbrink C, Schuler LD, Tieleman DP, van Gunsteren WF (2003) Eur Biophys J 32:67–77
12. Berendsen HJ, van der Spoel D, van Drunen R (1995) Comput Phys Commun 91:43–56
13. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7:306–317
14. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G (2003) Nucleic Acids Res 31:294–297
15. Piascik MT, Perez DM (2001) J Pharmacol Exp Ther 298:403–410
16. Zhao MM, Hwa J, Perez DM (1996) Mol Pharm 50:1118–1126
17. Wurch T, Pauwels PJ (2000) J Neurochem 75:1180–1189
18. Mehler EL, Periole X, Hassan SA, Weinstein H (2002) J Comput Aided Mol Des 16:841–853
19. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V (2004) J Mol Biol 342:571–583
20. Wymore T, Wong T (1999) Biophys J 76:1199–1212
21. Tieleman DP, Berendsen HJ, Sansom MS (2001) Biophys J 80:331–346
22. Trent JO, Wang ZX, Murray JL, Shao W, Tamamura H, Fujll N, Peiper SC (2003) J Biol Chem 278:47136–47144
23. Crozier PS, Stevens MJ, Forrest LR, Woolf TB (2003) J Mol Biol 333:493–514
24. Huber T, Botelho AV, Beyer K, Brown MF (2004) Biophys J 86(4):2078–2100
25. Flohil JA, Vriend G, Berendsen HJC (2002) Proteins: Struct Funct Genet 48:593–604
26. Fan H, Mark AE (2004) Protein Sci 13:211–220
27. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Biochemistry 40:7761–7772
28. INSIGHT II (2000) FDISCOVER/HOMOLOGY, Biosym/MSI, San Diego, USA
29. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) PNAS 99:5982–5987
30. Essman U, Perela L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577–8592
31. Ibragimova GT, Wade RC (1998) Biophys J 74:2906–2911
32. Tieleman DP, Berendsen HJ (1996) J Chem Phys 105:4871–4880
33. Department of Biological Sciences, Calgary/Canada, http://moose.bio.ucalgary.ca
34. Faraldo-Gómez JD, Smith GR, Sansom MS (2002) Eur Biophys J 31:217–227
35. Briggs JM, Madura JD, Davis ME, Gilson MK, Antosiewicz J, Luty BA, Wade RC, Bagheri B, Ilin A, Tan RC, McCammon JA (1989) UHBD (University of Houston Brownian Dynamics) Release 5.1
36. Schutz CN, Warshel A (2001) Proteins: Struct Funct Genet 44:400–417
37. Shapiro DA, Kristiansen K, Kroeze WK, Roth BL (2000) Mol Pharm 58:877–886
38. Shi L, Simpson MM, Ballesteros JA, Javitch JA (2001) Biochemistry 40/41:12339–12348
39. Rarey M, Kramer B, Lengauer T (1999) Proteins: Struct Funct Genet 34:17–28
40. von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, van Phan T, Smythe ML, White HF, Oliver SW (1993) Nature 363:418–423
41. Canutescu AA, Shelenkov AA, Dunbrack RL (2003) Protein Sci 12:2001–2014
42. Braganza LF, Worcester DL (1986) Biochemistry 25:2591–2596
43. Egberts E, Marrink SJ, Berendsen HJ (1994) Eur Phys J 22:423–436
44. Fahmy K, Jager F, Beck M, Zvyaga TA, Sakamar TP, Siebertm F (1993) PNAS 90:10206–10210
45. Yan EC, Kazmi MA, Ganim Y, Hou JM, Pan D, Chang BS, Sakmar TP, Mathies RA (2003) PNAS 100:9262–9267
46. Neve KA, Cumbay MG, Thompson KR, Yang R, Buck DC, Watts VJ, Durand CJ, Teeter MM (2001) Mol Pharm 60:373–381
47. Khare D, Alexander P, Antosiewicz J, Bryan P, Gilson M, Orban J (1997) Biochemistry 36:3580–3589
48. Noble MA, Gul S, Verma CS, Brocklehurst K (2000) Biochem J 351:723–733
49. Koehl P, Levitt M (1999) Nat Struct Biol 6:108–111