∎ **Computational Chemistry** |*Hot Paper*|

Basal Histamine H$_4$ Receptor Activation: Agonist Mimicry by the Diphenylalanine Motif

David Wifling,[b] Christopher Pfleger,[c] Jonas Kaindl,[a] Passainte Ibrahim,[a] Ralf C. Kling,[a] Armin Buschauer+,[b] Holger Gohlke+,*[c, d] and Timothy Clark+*[a]

In memory of our friend and colleague Armin Buschauer

Abstract: Histamine H$_4$ receptor (H$_4$R) orthologues are G-protein-coupled receptors (GPCRs) that exhibit species-dependent basal activity. In contrast to the basally inactive mouse H$_4$R (mH$_4$R), human H$_4$R (hH$_4$R) shows a high degree of basal activity. We have performed long-timescale molecular dynamics simulations and rigidity analyses on wild-type hH$_4$R, the experimentally characterized hH$_4$R variants S179M, F169V, F169V+S179M, F168A, and on mH$_4$R to investigate the molecular nature of the differential basal activity. H$_4$R variant-dependent differences between essential motifs of GPCR activation and structural stabilities correlate with experimentally determined basal activities and provide a molecular explanation for the differences in basal activation. Strikingly, during the MD simulations, F169$^{45.55}$ dips into the orthosteric binding pocket only in the case of hH$_4$R, thus adopting the role of an agonist and contributing to the stabilization of the active state. The results shed new light on the molecular mechanism of basal H$_4$R activation that are of importance for other GPCRs.

Introduction

GPCR crystal structures,[1] coupled with molecular dynamics (MD) simulations, site-directed mutagenesis, and NMR studies, have led to increased knowledge about the activation mechanism of GPCRs.[2] Agonists that bind to the orthosteric binding site located between transmembrane helices (TM) III, V, VI and VII[3] favor the active receptor state, eliciting the maximal possible response and inducing G-protein- and/or β-arrestin-mediated signaling.[4] Inverse agonists exert the opposite effect to agonists and favor the inactive receptor state. Neutral antagonists do not affect the basal activity. However, little is known, both on the pharmacological and molecular levels, about basal (constitutive) GPCR activation, which describes the activity of a GPCR in the absence of a ligand.[5] The existence of basal activity suggests that conformational transitions leading to the active receptor state may occur spontaneously. Such dynamic oscillations between active and inactive conformations have also been posited as a mechanism for partial activation,[6] suggesting a mechanistic link between ligand-independent and -dependent GPCR activation. Alternatively, a partially active receptor may adopt a static conformation that is intermediate between active and inactive.[7] Furthermore, single-molecule studies have indicated that differential ligand dwell-times and/or influences on nucleotide exchange rates in the G-protein add further layers of complexity to GPCR activation.[7] Under-

[a] J. Kaindl, P. Ibrahim, Dr. R. C. Kling, Prof. Dr. T. Clark+
Computer Chemistry Center
Department of Chemistry and Pharmacy
University of Erlangen-Nürnberg
Nägelsbachstr. 25, 91052 Erlangen
(Germany)
E-mail: Tim.Clark@fau.de

[b] Dr. D. Wifling, Prof. Dr. A. Buschauer
Department of Pharmaceutical/Medicinal Chemistry II
Institute of Pharmacy, University of Regensburg
Universitätsstr. 31, 93053 Regensburg
(Germany)

[c] Dr. C. Pfleger, Prof. Dr. H. Gohlke+
Institute for Pharmaceutical and Medicinal Chemistry
Heinrich Heine University Düsseldorf
Universitätsstr. 1, 40225 Düsseldorf
(Germany)

[d] Prof. Dr. H. Gohlke+
John von Neumann Institute for Computing (NIC)
Jülich Supercomputing Centre (JSC) &
Institute for Complex Systems—Structural Biochemistry (ICS 6)
Forschungszentrum Jülich GmbH
Wilhelm-Johnen-Str., 52425 Jülich
(Germany)
E-mail: gohlke@hhu.de

[†] Deceased, July 18, 2017.

[+] These authors contributed equally to this work.

⊞ Supporting information and the ORCID identification number(s) for the
ⓘ author(s) of this article can be found under:
https://doi.org/10.1002/chem.201902801.

standing the mechanisms and pathways of GPCR activation will help design new GPCR-targeted drugs with tailored pharmacological responses and fewer side effects.

The histamine H$_4$ receptor (H$_4$R), a class A GPCR expressed mainly on immune cells, plays a fundamental role in processes such as cytokine release and chemotaxis.[8] H$_4$R offers a unique opportunity towards understanding the mechanisms and pathways of GPCR activation in a ligand-independent manner. H$_4$R orthologues represent ideal naturally occurring GPCRs with different degrees of basal activity. In contrast to mouse H$_4$R (mH$_4$R), human H$_4$R (hH$_4$R) shows a high level of basal activity (Figure 1).[3a, 9] In a previous molecular pharmacological study, we characterized the basal activity of a series of hH$_4$R variants, comprising hH$_4$R-S179M, hH$_4$R-F169V, hH$_4$R-F169V+S179M, and hH$_4$R-F168A (Figure 1).[3a, 10] While the hH$_4$R-S179M variant exhibits a basal activity similar to that of hH$_4$R, replacing F169$^{45.55}$, situated in extracellular loop 2 (ECL2, numbering scheme according to GPCRdb[11]), by valine significantly decreased the basal activity.[3a] The basal activity of the hH$_4$R-

F169V+S179M[3a] and hH$_4$R-F168A[10] variants is even comparable to that of mH$_4$R. These observations thus identified residues that account for the high basal activity of hH$_4$R (Figure 1). However, the molecular mechanisms of basal hH$_4$R activation, and how the equilibrium between inactive and active receptor states is shifted towards the inactive state in the other H$_4$R variants, are still unknown.

Here, we performed long-timescale MD simulations (2 μs) of each H$_4$R variant to monitor conformational transitions in the ECL2 region, the orthosteric binding pocket, the transmission region, and the G-protein binding site (Figure 1). Most notably, hH$_4$R showed the most pronounced binding-pocket contraction and TM VI outward movement, which reflects its highest basal activity. To complement the analyses of structural dynamics, we performed rigidity analyses to probe the structural stability of key H$_4$R elements involved in GPCR activation. Our results highlight differential activation patterns of the H$_4$R variants in different regions shown in Figure 1 and, hence, provide a molecular explanation for basal activity.
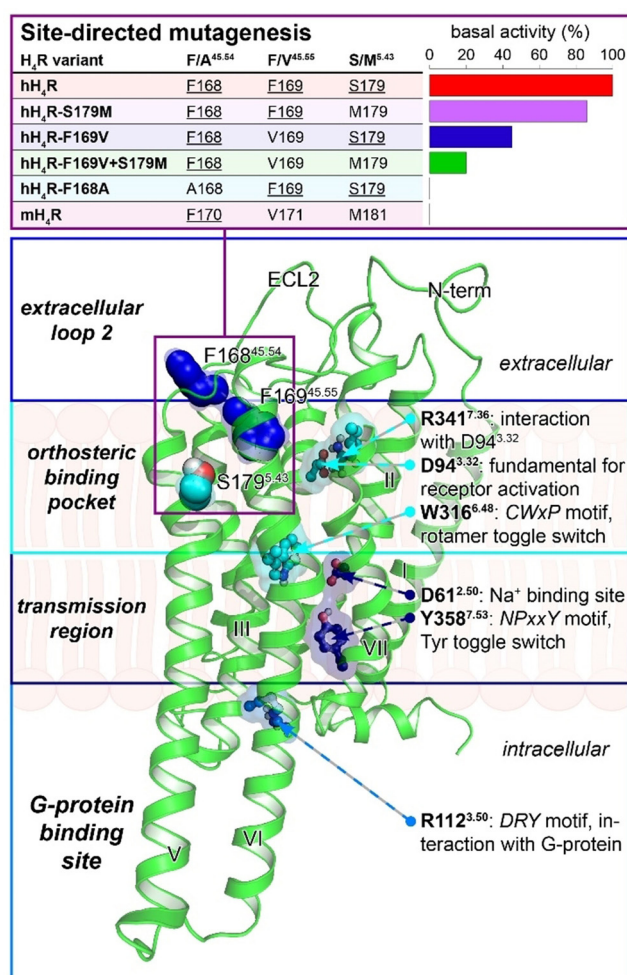
Results and Discussion

Dipping of F169$^{45.55}$ into the orthosteric binding pocket

ECL2 has been proposed to be involved in ligand binding, selectivity, recognition, and to act as a gatekeeper for ligand entry.[12] However, little is known about its contribution to basal GPCR activity. Among others, the M$_2$ muscarinic receptor (M$_2$R), the β$_2$-adrenergic receptor (β$_2$AR), the histamine H$_3$ receptor and hH$_4$R, which all show basal activity, share the same F$^{45.54}$–F$^{45.55}$ motif in ECL2 (F168–F169 for hH$_4$R). Furthermore, 15 other GPCRs feature analogous F–Y, Y–F or Y– motifs in place of F–F.[13] Remarkably, our MD simulations of hH$_4$R revealed that F169$^{45.55}$ dips into the orthosteric binding pocket and mainly interacts with the surrounding hydrophobic and aromatic residues W90$^{3.28}$, L91$^{3.29}$, Y95$^{3.33}$, P166$^{45.52}$, W172$^{5.36}$, L175$^{5.39}$, Y319$^{6.51}$, L326$^{6.58}$, Y340$^{7.35}$ and F344$^{7.39}$ (Figure 2, numbering of TMs I–VII according to the Ballesteros–Weinstein nomenclature[14]). The dipping of F169$^{45.55}$ was observed in four out of seven (approximately 60%) of our wild-type hH$_4$R simulations but not at all for other variants (simulations performed in singlet). Strikingly, F169$^{45.55}$ occupies the same area of the binding pocket as assumed[3a] for the imidazole moiety of histamine. It may thus act as a surrogate ligand in the highly basally active hH$_4$R. Interestingly, GPCR crystal structures show two different orientations of the diphenylalanine (F–Y in case of H$_1$R) motif: the two amino acids either point to opposite (β$_2$AR,[15] H$_1$R[16]) or the same (M$_2$R[17]) directions compared with hH$_4$R simulations. Hence, though our simulations suggest the dipping of F169$^{45.55}$ into the binding pocket and the direction of F168$^{45.54}$ towards the cytoplasm, the two amino acids could potentially swap their orientations on longer timescales.

Contraction of the orthosteric binding pocket

In all MD simulations, the binding-pocket dimension between TM III and VII, represented by the distances between the C$_α$

Figure 1. Amino acids that act as key players in basal H$_4$R activation. Fundamental amino acids for basal H$_4$R activation (shown as spheres) were previously determined by site-directed mutagenesis studies.[3a, 10] The respective basal activities of hH$_4$R, mH$_4$R, and the hH$_4$R variants containing mutations in either F168$^{45.54}$, F169$^{45.55}$, and/or S179$^{5.43}$ are depicted in the upper box. Further amino acids found to be involved in basal H$_4$R activation are shown as ball and sticks. Amino acid numbering corresponds to hH$_4$R.
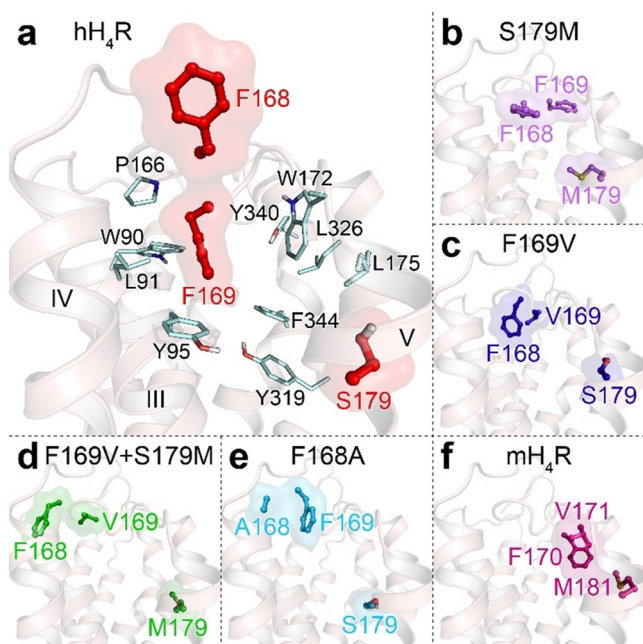
Figure 2. H4R variant-dependent locations and orientations of key amino acids for basal H4R activation: F/A168[45.54], F/V169[45.55], S/M179[5.43] for a) hH4R and the hH4R variants b) S179M, c) F169V, d) F169V+S179M, e) F168A, and f) F170[45.54], V171[45.55], M181[5.43] of mH4R. Wild-type and mutated residues of pre-aligned H4R variant structures (a–f), showing the respective predominating cluster, are illustrated as sticks and semi-transparent surface representation. To enhance comparability, the hH4R cartoon is highlighted in red (semi-transparent) for all H4R variants (a–f). Carbon atoms of aromatic and hydrophobic hH4R residues, capable of interacting with F169, are colored in gray (a).

atoms of D94[3.32] and Q347[7.42] (hH4R variants) or Q349[7.42] (mH4R), was initially approximately 11 Å, close to the corresponding distance[16] in the inactive hH1R structure used for homology modelling. This distance increases by a maximum of approximately 3 Å for the basally inactive hH4R variants F169V, F169V+S179M and F168A, and for mH4R (Figure 3b and Table 1). In contrast, it remained nearly constant for the hH4R-S179M variant and decreased by 2 to 3 Å for wild-type hH4R, both of which exhibit high basal activity. This distance is approximately 11 Å in aminergic GPCR crystal structures (Table S1 in the Supporting Information), both with bound inverse agonist and antagonist ((10.9±0.2) Å; $N=30$) and bound agonist ((10.9±0.5) Å; $N=20$). Thus, in our simulations, TM III approaches TM VII more closely in apo-hH4R (≈ 8 Å) than previously observed in any ligand-bound GPCR or found for the other H4R variants. The decreased distance between TM III and TM VII enables D94[3.32], which is involved in ligand binding and receptor activation,[3b,18] to form additional interactions to residues in TM VII. The simulations of hH4R, its variants S179M (basally active) and F169V+S179M (weakly basally active) exhibited hydrogen bonds between D94[3.32], R341[7.36] and W348[7.43] in addition to a water-mediated contact to Q347[7.42] (Figures 3a, 4, and S1a and c in the Supporting Information). By contrast, in all variants with lower basal activity and mH4R, D94[3.32] is involved in a hydrogen-bond network with S68[2.57] (and with W90[3.28] in hH4R-F169V+S179M) (Figures 3d, 4, and S1b–d in the Supporting Information). A network comparable to the latter has been described for the crystal structures of M1 and M4 receptor-antagonist complexes,[19] and may be an indicator of the inactive receptor conformation of H4R. In a further hy-
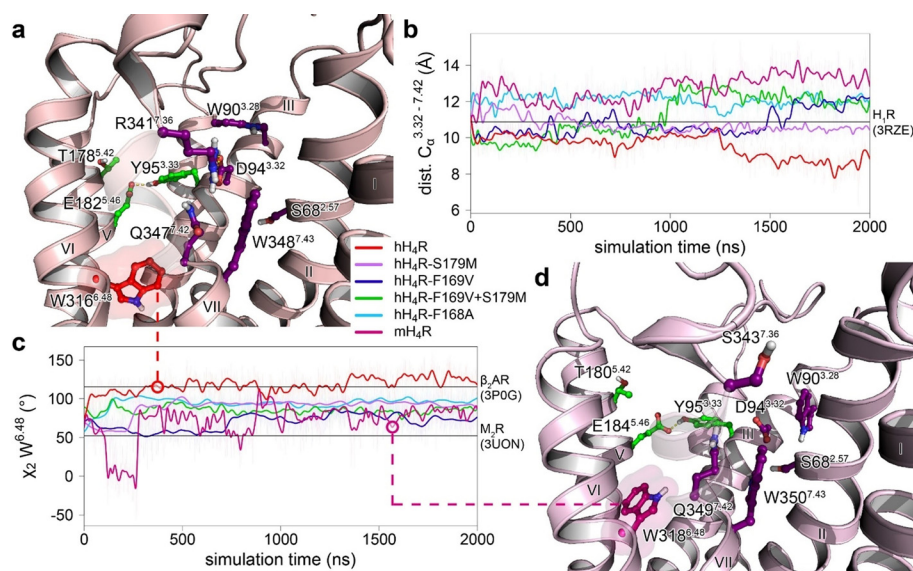

Figure 3. Structural differences within the orthosteric binding pocket that relate to differences in basal activity. Insights into a) hH4R and d) mH4R binding pockets, represented by the corresponding predominant cluster (for cluster sizes see Table S2 in the Supporting Information). Residues involved in a hydrogen-bond network with D94[3.32] (a,d) and E182[5.46] (a) or E184[5.46] (d) in at least one H4R variant are highlighted in dark purple and green, respectively. W316[6.48] (a) or W318[6.48] (d), and the semi-transparent surface are shaded in red (a) or magenta (d). b) Time-evolution of distances between Cα atoms of D94[3.32] and Q347[7.42] (hH4R variants) or Q349[7.42] (mH4R). c) Time course of W316[6.48] (hH4R variants) or W318[6.48] (mH4R) χ_2 rotational angles. For comparison, the corresponding distances of the inactive state hH1R[16] (b) and the χ_2 torsional angles of both the inactive state M2R[17a] and the active state β2AR[15b] (c) are illustrated with straight lines. Colors referring to the H4R variants in (b,c) are specified in the legend.

Table 1. Analysis of selected distances and torsional angles (mean ± standard deviation) of wild-type hH$_4$R, its four variants, and mH$_4$R.[a]						
Parameter	hH$_4$R	hH$_4$R-S179M	hH$_4$R-F169V	hH$_4$R-F169V+S179M	hH$_4$R-F168A	mH$_4$R
distance[b] (C$_\alpha$) D94$^{3.32}$–Q347$^{7.42}$ [Å]	9.5 ± 0.7	10.7 ± 0.5	10.8 ± 0.8	11.2 ± 1.2	12.1 ± 0.5	12.7 ± 0.8
torsional angle[c] for χ$_2$ W316$^{6.48}$ [°]	117.3 ± 13.6	89.5 ± 14.9	69.7 ± 14.7	85.3 ± 10.7	94.5 ± 11.4	71.5 ± 26.4
torsional angle[d] for χ$_2$ Y358$^{7.53}$ [°]	102.8 ± 21.1	87.3 ± 31.4	129.7 ± 30.3	126.9 ± 20.2	105.8 ± 20.4	139.1 ± 20.7
distance[e] (C$_\alpha$) R112$^{3.50}$–A298$^{6.30}$ [Å]	10.1 ± 0.6	9.1 ± 0.6	9.4 ± 0.4	9.3 ± 0.5	9.5 ± 0.4	8.5 ± 0.5

[a] Amino acid names and numbers correspond to hH$_4$R. [b] Inactive-state hH$_1$R[16]: 10.9, inactive-state β$_2$AR[15a]: 10.9, active-state β$_2$AR[15b]: 11.4 Å. [c] Inactive-state hH$_1$R[16]: 105.9, inactive-state M$_2$R[17a]: 52.2, active-state β$_2$AR[15b]: 115.4°. [d] Inactive-state hH$_1$R[16]: 128.7, inactive-state β$_2$AR[15a]: 138.4, active-state β$_2$AR[15b]: 86.3°. For torsion angle calculation, a periodicity of 180° was considered. [e] Inactive-state hH$_1$R[16]: 8.0, inactive-state β$_2$AR[15a]: 11.1, active-state β$_2$AR[15b]: 17.2 Å.

drogen-bonding network, all H$_4$R variants showed differential, highly variant-specific interactions between the key residue[3b, 18] E182$^{5.46}$ (E184$^{5.46}$ in mH$_4$R) and Y95$^{3.33}$ or T178$^{5.42}$ (Figures 3a, d, 4, and S1 in the Supporting Information). While E182$^{5.46}$ established hydrogen-bond contacts to both Y95$^{3.33}$ and T178$^{5.42}$ in hH$_4$R and its variants S179M and F169V+S179M, it only displayed a single interaction to T178$^{5.42}$ in hH$_4$R-F169V, and Y95$^{3.33}$ in hH$_4$R-F168A and mH$_4$R. Consequently, the results emphasize that the key amino acids D$^{3.32}$ and E$^{5.46}$, which are usually involved in ligand-dependent H$_4$R activation, formed intra-receptor interactions in the absence of a ligand. Depending on the amino acid composition, H$_4$R variant-specific active receptor conformations are stabilized to a greater or lesser extent.

The rotamer toggle switch contributes to basal H$_4$R activation

The binding pocket connects to structures closer to the cytoplasm, so that interactions formed within the binding pocket affect distal rearrangements of the receptor. In this context, the proposed W$^{6.48}$ rotamer toggle switch has been suggested[20] to drive the P$^{6.50}$ kink, and thus the outward movement of TM VI at the intracellular side. Analysis of rotameric W$^{6.48}$ states of aminergic GPCR crystal structures (Table S1 in the Supporting Information) resulted in χ$_2$ torsional angles ranging from 29.7 to 113.7° for antagonist/inverse agonist-bound structures ((94.7 ± 29.5)°, N = 30), from 80.1 to 86.1° for ligand-free structures ((82.2 ± 2.8)°, N = 3) and from 97.3 to 124.4° for agonist-bound structures ((112.4 ± 6.5)°, N = 20). In the hH$_4$R simulations, this angle ((117.3 ± 13.6)°) was comparable to that of active-state aminergic structures such as, for example, β$_2$AR[15b] (115.4°) (Table 1, Figure 3c). Simulations of both mH$_4$R and hH$_4$R-F169V gave values ((71.5 ± 26.4) and (69.7 ± 14.7)°) comparable to those of inactive-state aminergic structures such as, for example, M$_2$R[17a] (52.2°). The corresponding angles of the other hH$_4$R variants ranged between these two values. These findings thus support the hypothesis that particularly apo-hH$_4$R features more highly populated active states than the other H$_4$R variants. Hence, although the rotameric toggle switch has been discussed controversially in literature,[20, 21] our results suggest that it is involved in basal H$_4$R activation.

Structural rearrangements in the transmission region

D$^{2.50}$, whose fundamental role for agonist-dependent GPCR activation and allosteric sodium ion binding has been demonstrated,[22] is located in the central core of the receptor. Hydrogen bonds between D61$^{2.50}$ and S101$^{3.39}$ were more pronounced for both the highly basally active hH$_4$R and its hH$_4$R-S179M variant than for the other H$_4$R variants (see Figures 4, 5a, b, and S2 in the Supporting Information; for further interactions formed by D$^{2.50}$, please also see Figure S3a, Supporting Information). Consequently, the hydrogen-bond network of the
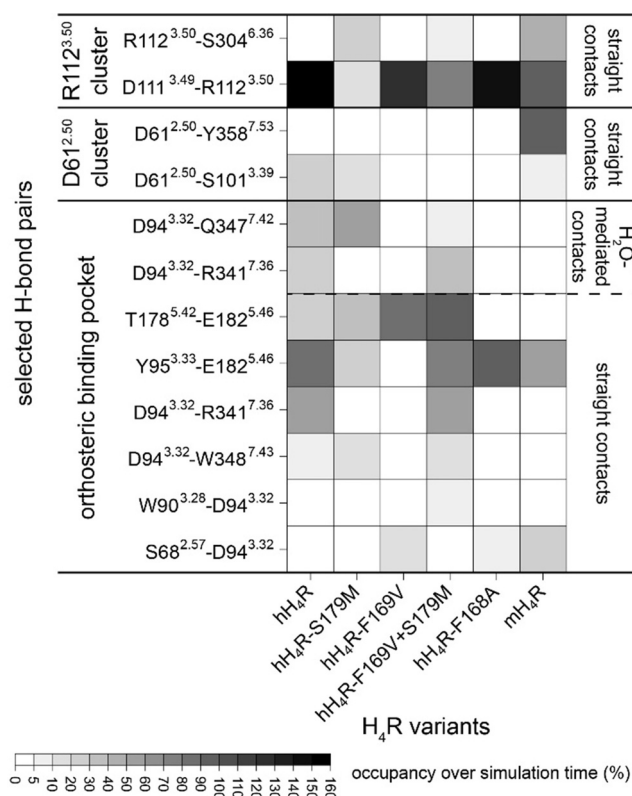

Figure 4. Hydrogen-bond (straight and water-mediated) analysis of the MD simulation trajectories performed for the orthosteric binding pocket, the D61$^{2.50}$ cluster (transmission region), and the R112$^{3.50}$ cluster (G-protein binding site). The % occupancy over the simulation time values represent cumulated values taking into account all possible hydrogen-bond donor and acceptor atom-to-atom combinations of the given amino acid pairs (excluding the backbone). The amino acid names and numbers given on the y axis correspond to hH$_4$R.
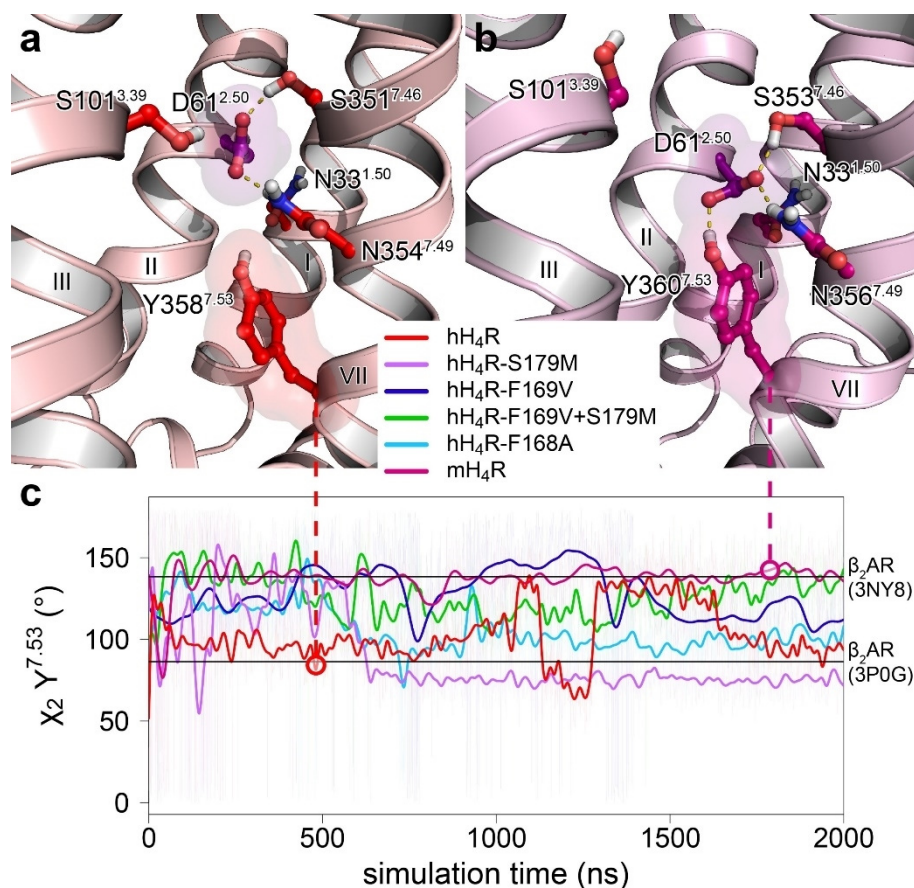
Figure 5. Structural differences within the transmission region that relate to differences in basal activity. Predominating cluster shown for a) hH_4R and b) mH_4R. Carbon atoms and the semi-transparent surface of $D61^{2.50}$ are shaded in dark purple, and carbon atoms of the other residues are colored red (a) or magenta (b) (semi-transparent surface shown for $Y358^{7.53}$ (a) and $Y360^{7.53}$ (b)). c) Time course of $Y358^{7.53}$ (hH_4R variants) and $Y360^{7.53}$ (mH_4R) χ_2 torsion angles (for color code, see legend). The χ_2 torsional angles of both inactive[15a]- and active[15b]-state β_2AR are shown with straight lines for comparison.

$D^{2.50}$ cluster is expanded, and thus, TM III could undergo concerted helix movements with TM I, II and VII. The fact that active-state crystal structures,[15b,17b,23] unlike inactive[15a,17a,24] ones, usually show this hydrogen bond confirms its important role for basal GPCR activation. $Y358^{7.53}$, which is part of the NPxxY motif and lies adjacent to $D^{2.50}$, acts as a further key player in GPCR activation.[20] It conducts the effect of the tyrosine toggle switch after the so-called hydrophobic barrier has been opened and thus, the water-mediated hydrogen bonding network is expanded towards the DRY motif.[20] Consistently with our interpretation of the simulations, $Y358^{7.53}$ ($Y360^{7.53}$ in mH_4R) revealed overall χ_2 dihedral angles comparable to those of the inactive[15a] and active[15b] state β_2AR (Table 1, Figure 5c) in mH_4R and hH_4R, respectively. The mean χ_2 angles increased in the order hH_4R-S179M < hH_4R ≈ hH_4R-F168A < hH_4R-F169V+S179M ≈ hH_4R-F169V < mH_4R (Table 1). Consequently, in hH_4R and its variants S179M and F168A, $Y^{7.53}$ prevalently occupies a conformation that corresponds to a (basally) activated receptor state. In contrast, $Y^{7.53}$ conformations of the hH_4R variants F169V and F169V+S179M are comparable to that of mH_4R and thus reflect an inactive receptor state. Driven by such conformational changes, $Y^{7.53}$ only formed a prevalent hydrogen bond with $D61^{2.50}$ in mH_4R (Figures 4, 5a, b and S2 in

the Supporting Information). Water-mediated interactions, however, were present both in the other H_4R variants (Figure S3a, Supporting Information) and identified in both inactive and active aminergic GPCR crystal structures.[24,25] This result further supports the hypothesis that, of the H_4R variants investigated, the inactive receptor state is most predominant in mH_4R.

Outward movement of TM VI at the intracellular side

$R^{3.50}$ has been proposed to play a key role in G-protein activation.[26] It forms the ionic lock with $E^{6.30}$ that has been found in many GPCRs such as, for example, β_2AR or H_1R, and was suggested to play a role in the stabilization of the inactive receptor state.[20] In contrast, H_4R orthologues cannot form the ionic lock because $E^{6.30}$ is replaced by non-acidic amino acids (hH_4R: Ala, mH_4R: Gly).[3a,27] The simulations suggest that $R112^{3.50}$ forms more prevalent hydrogen-bond contacts with $D111^{3.49}$ in the H_4R-$S^{5.43}$ variants (hH_4R, hH_4R-F169V, hH_4R-F168A) than H_4R-$M^{5.43}$ variants (hH_4R-S179M, hH_4R-F169V+S179M, mH_4R) (Figures 4, 6a, b, and S4 in the Supporting Information). However, only the latter showed hydrogen bonds between $R^{3.50}$ and $S^{6.36}$ that connect TM III and TM VI; these may replace the ionic-
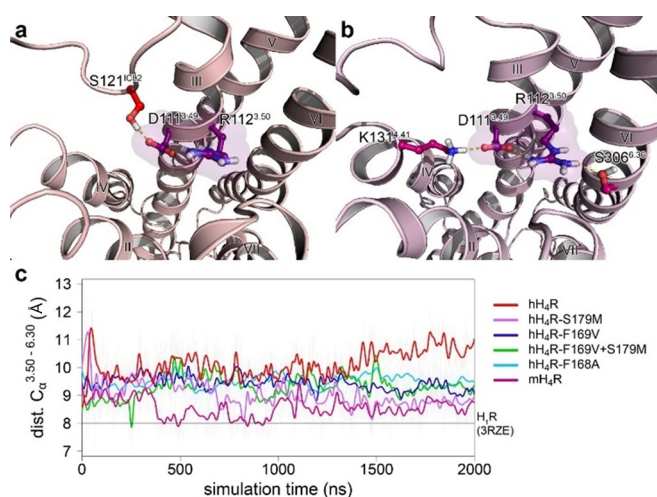
Figure 6. Structural differences within the G-protein binding site that relate to differences in basal activity. Predominating cluster shown for a) the hH$_4$R and b) the mH$_4$R. Carbon atoms and the semi-transparent surface of D111$^{3.49}$ and R112$^{3.50}$ are colored in dark purple. Carbon atoms of other residues and the respective cartoon are colored in red (a) or magenta (b). c) Time course of distances between R112$^{3.50}$ and A298$^{6.30}$ (hH$_4$R variants) or G300$^{6.30}$ (mH$_4$R) (for color code of H$_4$R variants, see legend). The corresponding distance of the inactive state hH$_1$R[16] is displayed with a straight line for comparison.

lock interaction and therefore stabilize the receptor in the inactive conformation. Consequently, the substitution of S$^{5.43}$ by Met leads to structural rearrangements that diminish the hydrogen-bond formation between D$^{3.49}$ and R$^{3.50}$ and simultaneously enhance contacts between R$^{3.50}$ and S$^{6.36}$ (see Figures 4, 6 a, b, and S4 in the Supporting Information; for further interactions formed by D$^{3.49}$, see Figure S3 b, Supporting Information). This result is in agreement with site-directed mutagenesis studies, which show that the missing E$^{6.30}$ alone is not the sole reason for the basal activity of H$_4$R.[27]

Depending on the amino acid composition of the H$_4$R variant, all aforementioned motifs in concert contribute more or less to helical rearrangements at the intracellular face. In particular, TM VI is bent around P$^{6.50}$ as part of the FxxCWxP[28] motif, and thus forms a kink at this position.[29] As a result, TM VI moves significantly outward upon receptor activation, enabling the binding of G-proteins into the gap thus formed, and downstream G-protein and/or β-arrestin-mediated signaling.[30] As a measure of GPCR activation,[31] initial mean C$_\alpha$–C$_\alpha$ distances of around 8.0 Å (as in the hH$_1$R[16] template) between R112$^{3.50}$ and A298$^{6.30}$ (G300$^{6.30}$ in mH$_4$R) increased by a maximum of approximately 3 Å for WT-hH$_4$R during the last quarter of the simulation (Figure 6 c). In all other H$_4$R variants the distance increase does not exceed 1.5 Å. This is consistent with the highest hH$_4$R basal activity of all the H$_4$R variants investigated, though not valid for hH$_4$R-S179M with the second highest basal activity. Comparison of the R$^{3.50}$–E$^{6.30}$ distance of inactive[15a] state β$_2$AR (11.1 Å) with that of active state[15b] β$_2$AR (17.2 Å) (see Table 1) revealed an outward movement of 6 Å during activation, suggesting that hH$_4$R ($\approx$3 Å) switched from an inactive to a partially active receptor conformation during MD simulation. Such an intermediate receptor conformation has also been suggested by Dror et al.[2b]

Mechanism of low-level activation

To both gain insights into the mechanism of low-level activation and to validate the reproducibility of our simulation results on basal hH$_4$R activation, we performed six additional replica MD simulations of wild-type hH$_4$R (Figure 7). During the last 500 ns of the in total seven hH$_4$R simulations, the dipping of F169$^{45.55}$ into the binding pocket was observed in four simulations (1–4). Strikingly, the two of those four simulations that showed the most predominant dipping of F169$^{45.55}$ into the binding pocket ($>$70%) and either a binding pocket contraction (1) or only a subtle dilatation (2) also showed the most pronounced TM VI outward movement ($>$2.5 Å). By contrast, the other two simulations that either revealed lower dipping occupancies of F169$^{45.55}$ (4) or higher binding pocket dilatations (3) showed a markedly reduced outward movement of TM VI ($<$2.0 Å). The three remaining simulations that showed no dipping of F169$^{45.55}$ (5–7) resembled the H$_4$R variants with a lower degree of basal activity (S179M, F169V, F169V+S179M or F168A) with respect to binding pocket contraction ($<$0.2 Å) and TM VI outward movement (0.9–1.3 Å). We thus observed two different activation states in the seven hH$_4$R MD simulations, suggesting that hH$_4$R switches between on (1, 2) and off (3–7) states in favor of adopting partially active conformations.

This is fully consistent with the concept of basal activity since basally active receptors transition spontaneously between inactive and active-like conformations in the absence of a ligand.[7] Note, however, that although the dipping and TM VI outward movement were reproducible and predominant for
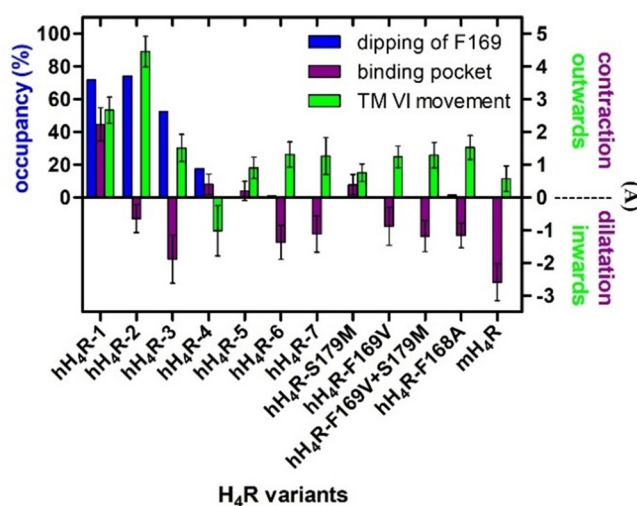

Figure 7. Activation patterns observed in MD simulations of hH$_4$R (seven replica) and other H$_4$R variants during the last 500 ns. The % dipping of F169$^{45.55}$ into the binding pocket (blue, left y axis), measured as the occupancy of z distances $<$3.5 Å between the CZ atom of F169$^{45.55}$ (CB in the case of Val) of the respective 500 simulation frames and the center of mass of the imidazole ring (non-hydrogen atoms) of histamine (aligned docking pose). Contraction (+)/ dilatation (−) of the orthosteric binding pocket (dark magenta, right y axis) and TM VI outward (+)/inward (−) movement (green, right y axis), respectively, measured as the mean difference ± standard deviation of the D$^{3.32}$–Q$^{7.42}$ or R$^{3.50}$–A/G$^{6.30}$ distances between hH$_1$R (PDB ID: 3RZE[16]) and the 500 simulation frames.

hH$_4$R, we cannot entirely exclude their occurrences for other variants on a statistical basis. Indeed, such structural changes would have been expected for the still highly basally active hH$_4$R-S179M variant.

Shift in structural stability occurs with changing basal activity

To complement the analyses of differences in structural dynamics, we used Constraint Network Analysis (CNA) to compare the structural stability of WT-hH$_4$R, its four variants S179M, F169V, F169V+S179M, F168A, and mH$_4$R. In CNA, biomolecules are modeled as a constraint network, in which atoms are represented as nodes, and covalent and non-covalent binding interactions as constraints. The constraint network is efficiently decomposed into rigid clusters and connecting flexible hinge regions by applying rigidity theory.[32] The inherent long-range aspect of rigidity percolation (i.e., whether a region is flexible or rigid depends on distant structural details) makes rigidity analysis an attractive tool for studying changes in structural stability due to distant influences.[33] Biomolecules display a hierarchy of structural stability that reflects the modularity of their structure.[34] To identify this hierarchy, a "constraint dilution trajectory" of network states, obtained from an initial network topology by successively removing non-covalent constraints, is analyzed (see Methods).[34,35] Figure 8a shows a succession of five network states of hH$_4$R and mH$_4$R along such a constraint dilution trajectory, starting from the homology models (see Methods). In particular, regions known to be important in GPCR activation such as helices TM V and TM VI, become flexible earlier in hH$_4$R than in mH$_4$R. This qualitative analysis suggests that the lower structural stability of hH$_4$R favors adaptations of these helices that allow the G-protein to bind and, hence, that hH$_4$R can be basally active, in contrast to mH$_4$R.

To obtain more detailed information about structural stability in hH$_4$R, the four hH$_4$R variants and mH$_4$R, we applied CNA to conformational ensembles of the last 500 ns of the 2 µs MD trajectories. We then computed the difference of the residue-wise chemical potential energy [Eq. (1)], which reflects changes in biomolecular stability between the system being considered and hH$_4$R (Figures 8b, S8, and S9 in the Supporting Information).

$$\Delta E_{i, CNA} = E_{i, CNA, hH_4R} - E_{i, CNA, hH_4R\ variants/mH_4R} \tag{1}$$

ΔE_{CNA} has been shown to be a robust local stability measure for predicting the thermal stability of proteins.[36] From the $\Delta E_{i,CNA}$ values, we identified regions in the receptors that correlate with the rank-ordering of the systems with respect to basal activity (Figure 8b): The structural stability of the systems increases with increasing levels of basal activity for the regions enclosing residues F54$^{2.43}$–P70$^{2.59}$ in helix TM II as well as W90$^{3.28}$–T92$^{3.30}$ and S109$^{3.47}$–D111$^{3.49}$ in helix TM III. In contrast, the structural stability of the systems decreases with increasing levels of basal activity for residues V184$^{5.48}$–V190$^{5.54}$ in TM V as well as A306$^{6.38}$–G310$^{6.42}$ and V314$^{6.46}$–A317$^{6.49}$ in TM VI. Notably,
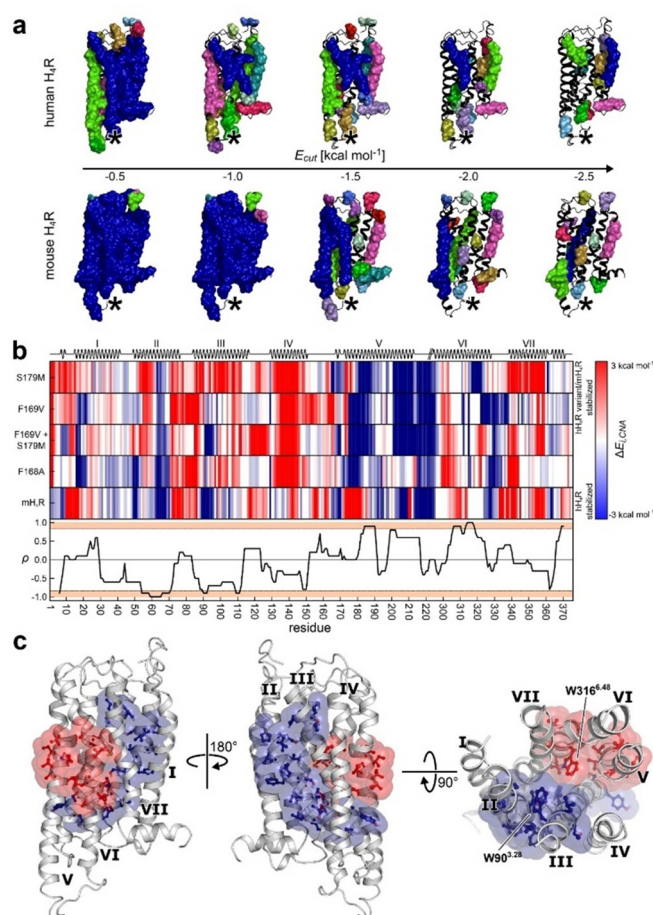

Figure 8. Comparison of structural stability between hH$_4$R and mH$_4$R as well as four hH$_4$R variants. a) Rigid cluster decomposition along the constraint dilution trajectory of hH$_4$R (top) and mH$_4$R (bottom). The five states show differences in the decay of structural stability in either species. At each state, all hydrogen bonds with an energy $E_{HB} > E_{cut}$ are removed from the network. The asterisks indicate the position of TM VI. b) The heat map depicts $\Delta E_{i,CNA}$ values [Eq. (5)] of the four hH$_4$R variants and mH$_4$R relative to hH$_4$R. A structural stabilization of either hH$_4$R or hH$_4$R variant/mH$_4$R is color coded in blue and red, respectively (see color scale). The line plot shows the Spearman's rank correlation ρ between the computed, relative $\Delta E_{i,CNA}$ values and the rank of the basal activity of each system. Orange regions depict significant rank correlations ρ ($p < 0.05$). c) Regions enclosing the residues F54$^{2.43}$–P70$^{2.59}$, W90$^{3.28}$–T92$^{3.30}$, and S109$^{3.47}$–D111$^{3.49}$ (V184$^{5.48}$–V190$^{5.54}$, A306$^{6.38}$–G310$^{6.42}$, and V314$^{6.46}$–A317$^{6.49}$) with a significant correlation between increasing (decreasing) structural stability and increasing levels of basal activity are mapped on the structure of H$_4$R in blue (red).

the analyses thus reveal focused, sequentially continuous but topologically distant residue clusters in the receptors in which changes in structural stability correlate with basal activity, rather than a broad distribution of stability changes across the structures.

Furthermore, the analyses show that, while regions critical for and known to move during GPCR activation[37] (TM V and TM VI) become less structurally stable with increasing levels of basal activity, regions that are rather immobile during GPCR activation[37] (TM II and TM III) become more structurally stable. Therefore, a local shift in structural stability within the GPCR occurs with changing basal activity, rather than a global, uni-

form change, so that overall structural stability is maintained in each case.

Interestingly, analyzing residues in TM VI adjacent to the G-protein binding site (L296[6.28]–F312[6.44]), which undergoes the largest structural changes during activation,[37] in terms of the sum of their $\Delta E_{i,CNA}$ values yields a measure that significantly and very well correlates ($R^2 = 0.94$, $p = 0.001$) with the level of basal activity (cf. Figure 9; a significant and good correlation is also found if the entire, 2 µs-long trajectories are used for analysis (Figure S10 in the Supporting Information); likewise, a significant and very good ($R^2 = 0.94$, $p = 0.001$) correlation is found if the last 500 ns of six additional replica MD simulations for hH₄R are evaluated (Figures 9a and S11 in the Supporting Information); both findings demonstrate the robustness of the CNA results): The hH₄R-S179M variant shows only a minor stabilization of -0.80 kcal mol^{-1} compared to hH₄R, which is in

agreement with the very similar basal activities of the two systems. A stronger stabilization is found for the hH₄R-F169V variant (-9.48 kcal mol^{-1}), and the strongest effect (-14.20, -20.94, and -17.57 kcal mol^{-1}, respectively) for hH₄R-F169V+S179M, hH₄R-F168A, and mH₄R. The findings are again in agreement with hH₄R-F169V+S179M and hH₄R-F168A showing a decrease in basal activity comparable to that of mH₄R.[3a, 10]

Predictive value of the parameters

Table 2 summarizes further geometrical features observed in the simulations in relation to the basal activity of the six H₄R variants. Although the binding pocket contraction is not as clear for the hH₄R-S179M mutant as for the wild-type receptor, the D[3.32]–Q[7.42] distance is quite predictive. Both direct and water-mediated hydrogen bonds between TM III and VII indicate activation, although they are also found for the inactive hH₄R-F169V+S179M variant. The hydrogen bond between D94[3.32] and S68[2.57] or W90[3.28], on the other hand, is only found in the inactive variants and mH₄R, making its absence indicative of activation. The remaining hydrogen-bond networks in the binding site and the W[6.48] toggle switch are not diagnostic. More distally located motifs such as the D[2.50]–S[3.39] hydrogen bond, the Y[7.53] toggle switch, and the R[3.50]–A[6.30] distance are quite predictive, but not the D[2.50]–Y[7.53] hydrogen bond. As an alternative measure, structural stabilities of the lower part of TM VI (close to the G-protein binding site) can accurately predict basal activities (Figure 9b).

Conclusion

Prior molecular pharmacological (site-directed mutagenesis) studies have demonstrated the importance of the diphenylalanine (F168[45.54]–F169[45.55]) motif for basal H₄R activation: whereas hH₄R-F169V showed intermediate basal activity, the basal activities of both hH₄R-F169V+S179M and hH₄R-F168A were even comparable to that of the basally inactive mH₄R. Therefore, the question arose, by which molecular mechanism the diphenylalanine motif contributes to basal H₄R activation.

Our simulations unveiled the dipping of F169[45.55] into the orthosteric binding pocket and its conformational stabilization by the neighboring F168[45.54] as the trigger for high basal activity. The dipping was observed in approximately 60% of the hH₄R simulations but not at all for other variants (Table 2). hH₄R, therefore, appears to be unique among the receptor variants studied. Strikingly, our hH₄R simulations showed an overlap of the pharmacophoric groups of F169[45.55] with those of hH₄R ligands such as histamine. F169[45.55] apparently resembles a ligand with an intrinsic activity ranging between that of a neutral antagonist and a full agonist. Depending on the ability of F169[45.55] to interact with the surrounding hydrophobic and aromatic amino acids and, thus, to contract the binding pocket, TM VI at the intracellular side moves outward to a greater or lesser extent (Table 2). In this respect, hH₄R prefers switching between on and off states rather than stabilizing partially active conformations. Basal activity thus increases to
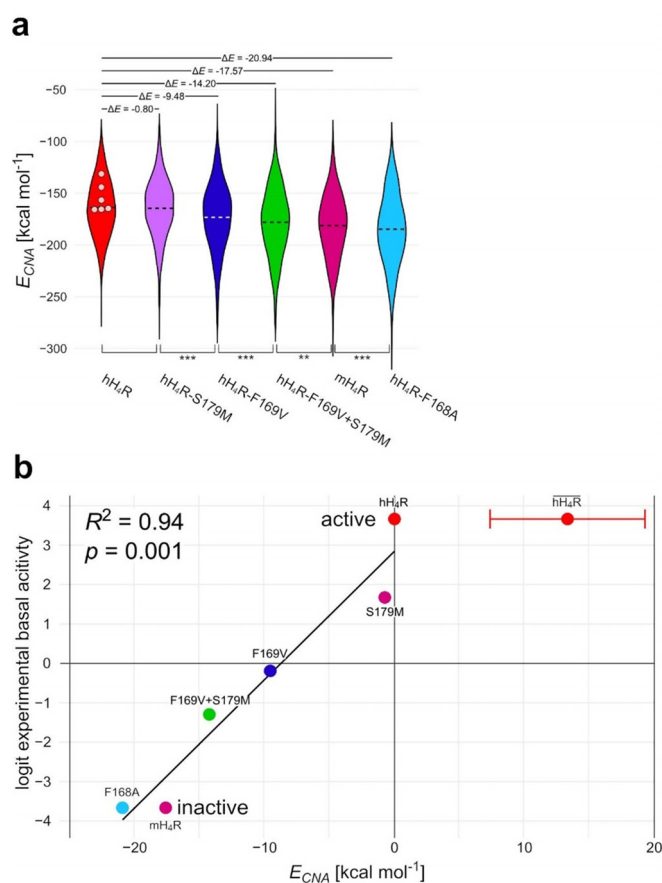

Figure 9. Correlation between structural stability of TM VI and experimental basal activity. a) The violin plot shows the distribution of structural stability in TM VI enclosing the residues L296[6.28]–F312[6.44]. The relative change in structural stability (ΔE_{CNA}) for each species compared to hH₄R is shown on the top. The statistical significance between pairs of mean E_{CNA} values (dashed line) in increasing order is *: $p < 0.01$, **: $p < 0.001$, ***: $p < 0.005$. For hH₄R, open circles depict the mean ΔE_{CNA} values of six additional replica MD simulations. b) The structural stability of TM VI correlates significantly and very well with experimental basal activity. Additionally, the mean ΔE_{CNA} value over all seven hH₄R replicas is shown, relative to the ΔE_{CNA} value of the one hH₄R trajectory shown as violin plot in Figure 9a; error bars denote the SEM. The basal activity in percentage (0 (basally inactive) to 100% (basally active)) was transformed into the *logit* scale.

Table 2. Summary[a] of the impact of wild-type hH$_4$R, its four variants, and mH$_4$R[b] on GPCR-activating motifs.

Parameter	hH$_4$R	S179M	F169V	F169V+S179M	F168A	mH$_4$R
basal activity[c] relative to hH$_4$R [%]	100	86	45	20	0	0
binding pocket						
dipping of F169$^{45.55}$ into the binding pocket	✓	×	×	×	×	×
distance (C$_\alpha$) D94$^{3.32}$–Q347$^{7.42}$ [%]	0	38	39	54	80	100
TM III–VII direct hydrogen bonds	✓	✓	×	✓	×	×
TM III–VII water-mediated hydrogen bonds	✓	✓	×	✓	×	×
hydrogen bond D94$^{3.32}$–S68$^{2.57}$ (or W90$^{3.28}$)	×	×	✓	✓	✓	✓
hydrogen bond Y95$^{3.33}$–E182$^{5.46}$ [%]	97	24	0	79	100	64
hydrogen bond T178$^{5.42}$–E182$^{5.46}$	24	37	90	100	0	4
Y95$^{3.33}$, T178$^{5.42}$, E182$^{5.46}$ hydrogen-bond network	✓	✓	×	✓	×	×
torsional χ_2 angle for W316$^{6.48}$ [%]	100	42	0	33	52	4
central region						
hydrogen bond D61$^{2.50}$–S101$^{3.39}$ [%]	100	58	0	2	1	19
hydrogen bond D61$^{2.50}$–Y358$^{7.53}$	×	×	×	×	×	✓
torsional χ_2 angle for Y358$^{7.53}$ [%]	30	0	82	77	36	100
G-protein binding site						
hydrogen bond R112$^{3.50}$–S304$^{6.36}$ [%]	0	45	2	13	0	100
distance (C$_\alpha$) R112$^{3.50}$–A298$^{6.30}$ [%]	100	32	57	48	61	0
structural stability L296$^{6.28}$–F312$^{6.44}$ [%]	0	4	45	68	100	84

[a] Effects of H$_4$R variants on GPCR activating motifs are either illustrated with ticks (applicable) and crosses (not applicable) or in a range between 0% (lowest value) and 100% (highest value). The percentage values were calculated from mean distances, torsional angles, or from occupancy over the simulation time values (hydrogen bonds) as outlined in the Computational methods section and Table S3 in the Supporting Information. [b] Amino acid naming and numbering correspond to hH$_4$R. [c] Basal activities (hH$_4$R = 100, mH$_4$R = 0%) were calculated from experimentally determined[3a, 10] intrinsic activities of thioperamide.

the extent that the basal equilibrium between inactive and active states is shifted towards the active state. An agonist binding to such a pre-activated receptor can subsequently only shift an intermediate to an active receptor state. In the absence of basal activity, by contrast, an agonist could shift an inactive to an active state.

Complementarily, results from rigidity analysis provide an excellent measure for describing the degree of basal activity in H$_4$R. First, we observed focused, sequentially continuous but topologically distant residue clusters in H$_4$R with structural stability characteristics that correlate significantly either with increasing or decreasing system's basal activity. Second, the structural stability of TM VI (L296$^{6.28}$–F312$^{6.44}$) alone, known to be important for GPCR activation, is an excellent indicator for the degree of a system's basal activity. Based on the almost perfect correlation between the computed structural stability and basal activity, a single-parameter model can be derived to efficiently probe the potential basal activity of new H$_4$R variants. Furthermore, this model might be transferable to probe for the efficacy of H$_4$R ligands.

This work and its results pose many questions such as whether the results are solely applicable for H$_4$R or more generally for class A GPCRs. Recently, the diphenylalanine motif gained high interest due to its central role for molecular recognition in peptide-based supramolecular assemblies and regulatory influences on bacterial morphology.[38] Notably, the class A GPCRs known to exhibit[39] the most pronounced basal activity (H$_4$R, H$_3$R and β_2AR) all contain a diphenylalanine motif in ECL2. Hence, it is tempting to speculate that the diphenylalanine motif exerts a general functional influence in which it drives the self-regulatory effect in transmembrane receptors, and thus, the system's basal activity. Furthermore, the exact nature of basal receptor activation, i.e., the structural comparability to agonist-induced activation, remains unknown. All these questions are the subject of further investigations.

Computational Methods

Homology modelling

A previous hH$_4$R homology model,[3a] based on the inactive state hH$_1$R[16] (PDB ID: 3RZE) was used. It includes all extracellular (ECL) and intracellular (ICL) loops.[3a] Models of the other H$_4$R variants (hH$_4$R variants S179M, F169V, F169V+S179M, F168A, and mH$_4$R) were prepared with the modeling suite SYBYL-X 1.3 (Tripos Inc., St. Louis, MO USA) using the hH$_4$R homology model as template. For the preparation of the mH$_4$R model, the ECL2 loop upstream of C164$^{45.50}$ was remodeled using the loop search module within SYBYL-X 1.3. All models showed the disulfide bond between C87$^{3.25}$ and C164$^{45.50}$ (hH$_4$R variants) or C166$^{45.50}$ (mH$_4$R). In place of ICL3, eight alanine residues were introduced. All residues were simulated in their dominant protonation state at pH 7. The N and C termini were positively (NH$_3$$^+$) and negatively (COO$^-$) charged.

Molecular dynamics simulations

Topologies and coordinates of the solvated (explicit water) and ionized (net charge of the entire system was zero) wild-type or mutant hH$_4$R models, and the mH$_4$R model, were prepared using the pdb2gmx module within GROMACS 4.5.[40] The systems were submitted to both steepest-descent and conjugated gradient energy minimization (without restraints) to remove bad van der Waals contacts of the amino acid side chains. The solvated GPCRs were then equilibrated in the NPT ensemble for 10 ns, applying harmonic restraints of 1000 kJ mol^{-1} nm^{-2} to protein main-chain atoms.

For membrane simulations, a pre-equilibrated dioleoylphosphatidylcholine (DOPC) lipid bilayer containing 72 DOPC lipids was used.[41] This system was enlarged in the x and y directions, resulting in a lipid bilayer consisting of 253 DOPC molecules.[42] This extended system was subjected to energy minimization (no restraints) and successively, equilibration runs were performed in the NVT (100 ps; weak harmonic restraints of $1000\ kJ\,mol^{-1}\,nm^{-2}$ applied on DOPC atoms) and NPT (10 ns; no restraints) ensembles.

Each receptor was inserted into this augmented, fully hydrated and equilibrated DOPC lipid bilayer using g_membed.[43] The net charge of the simulation systems was neutralized by addition of 5 (hH_4R variants) and 11 (mH_4R) Cl^- ions. The systems additionally contained around 217 DOPC and 28,000 water molecules, accounting for a total of around 120,000 atoms in a box spanning approximately $90\ Å \times 91\ Å \times 143\ Å$. Consecutively, these systems were energy minimized (without restraints) and equilibrated in the NPT ensemble for 10 ns (weak harmonic restraints of $1000\ kJ\,mol^{-1}\,nm^{-2}$ applied to protein main-chain atoms). Production molecular-dynamics (MD) simulations (2 μs) of each H_4R variant were started from the final frame of the equilibration runs. Data were collected every 100 ps. Although in other cases GPCR activation/deactivation typically demands much longer than 2 μs of simulation[2b, 44] or enhanced sampling,[45] we usually observe an induction period of approximately 500 ns with the AMBER force field, suggesting that 2 μs simulations are adequate for GPCR simulations in a membrane.[46] The nature[7] of basal activity might furthermore contribute to lower timescales needed for basal compared to ligand-dependent GPCR activation.

All minimization and MD simulation steps were carried out with GROMACS 4.5[40] (GROMACS 2016/2018 in the case of six replicate hH_4R simulations). Initial velocities of all MD simulation runs were randomly assigned to the atoms. The SPC/E water model,[47] the Amber ff99SB protein force field,[48] and for DOPC molecules the GAFF force field[41,49] were employed. A temperature of 310 K was maintained using a temperature bath[50] and a time constant of 0.1 ps. Protein, DOPC, and water (including Cl^- ions) atoms were coupled separately.[42] The Berendsen barostat[50] with a time constant of 5 ps and a surface tension of $22\ dyn\,cm^{-1}$ was applied to maintain a pressure of 1 bar and to ensure membrane properties in agreement with experiment.[42] The volume compressibility was chosen to be $4.5\times10^{-5}\ bar^{-1}$. The nonbonded cutoff was set to 10 Å and long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method[51] with an interpolation order of 4, an Ewald tolerance of 1×10^{-5} and FFT grid spacing of 1.2 Å. Periodic boundary conditions were applied in all directions. A cutoff of 14 Å was used for short-range van der Waals interactions. Bonds involving hydrogen atoms were constrained using LINCS,[52] enabling a time step of 2 fs.

Induced-fit docking

"Flexible" docking of histamine to the hH_4R model was essentially performed as described for muscarinic acetylcholine receptors in Pegoli et al.[53] by using the Protein Preparation Wizard, and the LigPrep and Induced-fit docking modules (Schrödinger LLC, Portland, OR, USA).

Data analysis

Data were analyzed by using GROMACS 4.5 (g_cluster, g_density, g_rms, g_rmsf) and GROMACS 5 (gmx distance) analysis tools every nanosecond. The *gromos* clustering method was applied for cluster analysis of the entire (2 μs) trajectories, setting an RMSD cutoff of 2.5 Å. Hydrogen bonds and torsional angles were analyzed using the CPPPTRAJ module of Amber 14 (University of California, San Francisco, CA USA). Hydrogen bond and xy plots (distance, torsion) were visualized by using the programing language R[54] and the packages devEM,F[55] plot3d,[56] Plotrix,[57] Peptides[58] and CircStats.[59] For the purpose of comparison, distances and torsional angles of aminergic GPCRs (Table S1 in the Supporting Information) were calculated using the UCSF Chimera[60] package, version 1.11.2, and are given as mean ± standard deviation. % values given in Table 2 were calculated as follows [Eq. (2)] from mean distances, torsional angles (Table 1), or occupancy over the simulation time values in the case of hydrogen bonds (Figure 4). For minimum and maximum values, please see Table S3 in the Supporting Information. Figure 7 was prepared with the Prism 5.01 software (GraphPad, San Diego, CA, USA), and all molecular figures with PyMOL Molecular Graphics system, version 1.8.2.1 (Schrödinger LLC, Portland, OR USA).

$$\% \text{ value} = \frac{\text{value} - \text{value\_min}(H_4R\ \text{variants})}{\text{value\_max}(H_4R\ \text{variants}) - \text{value\_min}(H_4R\ \text{variants})} \times 100 \tag{2}$$

Rigidity analysis

Rigidity analyzes were performed using the CNA software package.[32] CNA was applied on ensembles of network topologies generated from conformational ensembles. Each network of covalent and non-covalent (hydrogen bonds including salt bridges and hydrophobic tethers) interactions was constructed with the FIRST (Floppy Inclusions and Rigid Substructure Topography) software (version 6.2),[61] to which CNA is a front- and back-end. The strengths of hydrogen bonds (including salt bridges) were assigned by the energy E_{HB} computed by FIRST.[62] Hydrophobic interactions between carbon or sulfur atoms were taken into account if the distance between these atoms was less than the sum of their van der Waals radii (C: 1.7 Å, S: 1.8 Å) plus $D_{cut} = 0.25$ Å.[35d] In order to elucidate the hierarchy of structural stability in a biomolecule,[34] a "constraint dilution trajectory" of network states {σ} was analyzed, which was generated by successively removing hydrogen bond constraints (including salt bridges) in the order of increasing strength.[34,35] Thus, only those hydrogen bonds are retained in a network of state σ that have an energy $E_{HB} \leq E_{cut}(σ)$. Altered biomolecular stability along a constraint dilution trajectory was quantified based on neighbor stability maps ($rc_{ij,\text{neighbor}}$ with i, j being residue numbers) [Eq. (3)].[36]

$$rc_{ij} = \min\{E_{cut} | \exists c \in C^{E_{cut}} : R_i \Lambda R_j \in c\} \tag{3}$$

Here, only short-range rigid contacts were considered that have at least one pair of heavy atoms of the residue pair $R_{(i,j)}$, $A_{\{k \in i,\ l \in j\}}$, separated by a distance ≤ 4.5 Å.[63] A rigid contact rc_{ij} between pairs of residues ceases to exist when both residues stop sharing the same rigid cluster c of a set of rigid cluster $C^{E_{cut}}$ (Figure S7, Supporting Information). The double sum [Eq. (4)]

$$E_{CNA} = \sum_i^n \sum_{j>i}^n rc_{ij,\ \text{neighbor}} \tag{4}$$

then represents the chemical potential energy (E_{CNA}) due to non-covalent bonding, obtained from the coarse-grained, residue-wise network representation of the underlying biomolecular structure.[36,64] A per-residue decomposition of Equation (4), which is the chemical potential energy of residue i obtained by summation

over all *n* short-range rigid contacts the residue is involved in (Figure S7 in the Supporting Information), is computed according to Equation (5):

$$E_{i,\text{CNA}} = \frac{1}{2}\sum_{j\neq i}^{n} rc_{ij,\text{ neighbor}} \qquad (5)$$

Conformational ensembles used as input for CNA were extracted from MD trajectories of 2 μs length starting from hH$_4$R, the four hH$_4$R variants S179M, F169V, F169V+S179M, and F168A, as well as mH$_4$R. The first 1.5 μs of each trajectory were considered equilibration phase of the system, and conformational ensembles for CNA were extracted from the last 500 ns. To estimate the uncertainty in the CNA computations, we split the last 500 ns of each trajectory into five individual sets of 100 ns length. Because the correlation time for decay of fluctuations of E_{CNA} [Eq. (4)] for all trajectories is about 0.1 ns, the five individual sets are statistically independent, and the standard error of the mean (SEM) is calculated from the standard deviation (SD) of the five mean $E_{i,\text{CNA}}$ values [Eq. (5)] according to Equation (6):

$$\text{SEM} = \frac{\text{SD}}{\sqrt{5}} \qquad (6)$$

To probe for a drift in the $E_{i,\text{CNA}}$ values over the simulation time, we computed $E_{i,\text{CNA}}$ for intervals of 100 ns along the whole trajectory of 2 μs length for residues in TM VI, lying next to the orthosteric binding site and the G-protein binding site; these values show no correlation with simulation time, indicating that there is no drift (Figure S12, Supporting Information).

To probe for the robustness of the $E_{i,\text{CNA}}$ values for residues in TM VI, lying next to the orthosteric binding site and the G-protein binding site, we also analyzed entire, 2 μs-long MD trajectories (Figure S10 in the Supporting Information) as well as the last 500 ns of six additional replica MD simulations for hH$_4$R (Figures 9 and S11 in the Supporting Information).

Statistical analysis

For each system, the average $E_{i,\text{CNA}}$ values and the corresponding SEMs from Eq. 6 were calculated using NumPy.[65] According to the law of error propagation, the total SEM for differences in structural stability between wild-type hH$_4$R and the four hH$_4$R variants or mH$_4$R was computed according to Equation (7):

$$\text{SEM}_{\text{total}} = \sqrt{\text{SEM}^2_{\text{hH}_4\text{R}} + \text{SEM}^2_{\text{hH}_4\text{R variant/mH}_4\text{R}}} \qquad (7)$$

We used SciPy[66] for performing the following statistical analyses: (I) A two-sided Welch's t-test to probe if $\Delta E_{i,\text{CNA}}$ values are significantly different from zero ($p < 0.05$) and if independent samples of $E_{i,\text{CNA}}$ values have significantly different average values; (II) Pearson's and Spearman's rank correlation between $\Delta E_{i,\text{CNA}}$ values and (the rank of) the basal activity of each hH$_4$R variant and mH$_4$R. For computing the Pearson correlation, we summed $\Delta E_{i,\text{CNA}}$ values for the residues in TM VI (L296$^{6.28}$–F312$^{6.44}$) and then averaged over the last 500 ns for each species. The basal activity scale, ranging from 0% (inactive) to 100% (active), was transformed into the *logit* scale, with an adjustment of 2.5% for the lower and upper end of the scale. For computing the rank correlation, we calculated a running average over the $\Delta E_{i,\text{CNA}}$ values with respect to *i*, with a windows size of ten and a step size of one.

Conflict of interest

The authors declare no conflict of interest.

[1] a) A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, M. M. Babu, *Nature* **2013**, *494*, 185–194; b) K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739–745.

[2] a) B. K. Kobilka, *Biochim. Biophys. Acta Biomembr.* **2007**, *1768*, 794–807; b) R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18684–18689; c) R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser, B. K. Kobilka, *Cell* **2013**, *152*, 532–542; d) E. C. Hulme, *Trends Pharmacol. Sci.* **2013**, *34*, 67–84.

[3] a) D. Wifling, K. Loffel, U. Nordemann, A. Strasser, G. Bernhardt, S. Dove, R. Seifert, A. Buschauer, *Br. J. Pharmacol.* **2015**, *172*, 785–798; b) N. Shin, E. Coates, N. J. Murgolo, K. L. Morse, M. Bayne, C. D. Strader, F. J. Monsma, Jr., *Mol. Pharmacol.* **2002**, *62*, 38–47.

[4] T. Kenakin, *FASEB J.* **2001**, *15*, 598–611.

[5] R. J. Lefkowitz, S. Cotecchia, P. Samama, T. Costa, *Trends Pharmacol. Sci.* **1993**, *14*, 303–307.

[6] G. G. Gregorio, M. Masureel, D. Hilger, D. S. Terry, M. Juette, H. Zhao, Z. Zhou, J. M. Perez-Aguilar, M. Hauge, S. Mathiasen, J. A. Javitch, H. Weinstein, B. K. Kobilka, S. C. Blanchard, *Nature* **2017**, *547*, 68–73.

[7] R. Lamichhane, J. J. Liu, G. Pljevaljcic, K. L. White, E. van der Schans, V. Katritch, R. C. Stevens, K. Wüthrich, D. P. Millar, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14254–14259.

[8] E. Zampeli, E. Tiligada, *Br. J. Pharmacol.* **2009**, *157*, 24–33.

[9] D. Schnell, I. Brunskole, K. Ladova, E. H. Schneider, P. Igel, S. Dove, A. Buschauer, R. Seifert, *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2011**, *383*, 457–470.

[10] D. Wifling, G. Bernhardt, S. Dove, A. Buschauer, *PLoS One* **2015**, *10*, e0117185.

[11] V. Isberg, C. de Graaf, A. Bortolato, V. Cherezov, V. Katritch, F. H. Marshall, S. Mordalski, J.-P. Pin, R. C. Stevens, G. Vriend, D. E. Gloriam, *Trends Pharmacol. Sci.* **2015**, *36*, 22–31.

[12] a) M. Wheatley, D. Wootten, M. T. Conner, J. Simms, R. Kendrick, R. T. Logan, D. R. Poyner, J. Barwell, *Br. J. Pharmacol.* **2012**, *165*, 1688–1703; b) M. C. Peeters, G. J. van Westen, Q. Li, A. P. IJzerman, *Trends Pharmacol. Sci.* **2011**, *32*, 35–42.

[13] H. D. Lim, A. Jongejan, R. A. Bakker, E. Haaksma, I. J. de Esch, R. Leurs, *J. Pharmacol. Exp. Ther.* **2008**, *327*, 88–96.

[14] J. A. Ballesteros, H. Weinstein, *Methods Neurosci.* **1995**, *25*, 366–428.

[15] a) D. Wacker, G. Fenalti, M. A. Brown, V. Katritch, R. Abagyan, V. Cherezov, R. C. Stevens, *J. Am. Chem. Soc.* **2010**, *132*, 11443–11445; b) S. G. Rasmussen, H. J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. Devree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Ko-

netzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis, B. K. Kobilka, *Nature* **2011**, *469*, 175 – 180.

[16] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, S. Iwata, *Nature* **2011**, *475*, 65 – 70.

[17] a) K. Haga, A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga, T. Kobayashi, *Nature* **2012**, *482*, 547 – 551; b) A. C. Kruse, A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P. M. Sexton, A. Christopoulos, C. C. Felder, P. Gmeiner, J. Steyaert, W. I. Weis, K. C. Garcia, J. Wess, B. K. Kobilka, *Nature* **2013**, *504*, 101 – 106.

[18] A. Jongejan, H. D. Lim, R. A. Smits, I. J. de Esch, E. Haaksma, R. Leurs, *J. Chem. Inf. Model.* **2008**, *48*, 1455 – 1463.

[19] D. M. Thal, B. Sun, D. Feng, V. Nawaratne, K. Leach, C. C. Felder, M. G. Bures, D. A. Evans, W. I. Weis, P. Bachhawat, T. S. Kobilka, P. M. Sexton, B. K. Kobilka, A. Christopoulos, *Nature* **2016**, *531*, 335 – 340.

[20] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski, S. Filipek, *Curr. Med. Chem.* **2012**, *19*, 1090 – 1109.

[21] B. Taddese, L. M. Simpson, I. D. Wall, F. E. Blaney, C. A. Reynolds, *Methods Enzymol.* **2013**, *522*, 21 – 35.

[22] a) C. D. Strader, I. S. Sigal, M. R. Candelore, E. Rands, W. S. Hill, R. A. Dixon, *J. Biol. Chem.* **1988**, *263*, 10267 – 10271; b) B. P. Ceresa, L. E. Limbird, *J. Biol. Chem.* **1994**, *269*, 29557 – 29564; c) V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, V. Cherezov, R. C. Stevens, *Trends Biochem. Sci.* **2014**, *39*, 233 – 244.

[23] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, B. K. Kobilka, *Nature* **2011**, *477*, 549 – 555.

[24] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, R. C. Stevens, *Science* **2007**, *318*, 1258 – 1265.

[25] T. Warne, R. Moukhametzianov, J. G. Baker, R. Nehme, P. C. Edwards, A. G. Leslie, G. F. Schertler, C. G. Tate, *Nature* **2011**, *469*, 241 – 244.

[26] J. A. Ballesteros, A. D. Jensen, G. Liapakis, S. G. Rasmussen, L. Shi, U. Gether, J. A. Javitch, *J. Biol. Chem.* **2001**, *276*, 29171 – 29177.

[27] E. H. Schneider, D. Schnell, A. Strasser, S. Dove, R. Seifert, *J. Pharmacol. Exp. Ther.* **2010**, *333*, 382 – 392.

[28] C. Zhang, Y. Srinivasan, D. H. Arlow, J. J. Fung, D. Palmer, Y. Zheng, H. F. Green, A. Pandey, R. O. Dror, D. E. Shaw, W. I. Weis, S. R. Coughlin, B. K. Kobilka, *Nature* **2012**, *492*, 387 – 392.

[29] V. Katritch, V. Cherezov, R. C. Stevens, *Annu. Rev. Pharmacol. Toxicol.* **2013**, *53*, 531 – 556.

[30] D. M. Rosenbaum, S. G. Rasmussen, B. K. Kobilka, *Nature* **2009**, *459*, 356 – 363.

[31] R. O. Dror, D. H. Arlow, D. W. Borhani, M. O. Jensen, S. Piana, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4689 – 4694.

[32] C. Pfleger, P. C. Rathi, D. L. Klein, S. Radestock, H. Gohlke, *J. Chem. Inf. Model.* **2013**, *53*, 1007 – 1015.

[33] a) D. J. Jacobs, B. Hendrickson, *J. Comput. Phys.* **1997**, *137*, 346 – 365; b) C. F. Moukarzel, P. M. Duxbury in *Rigidity Theory and Applications* (Eds.: M. F. Thorpe, P. M. Duxbury), New York, Kluwer Academic/Plenum Publishers, **1999**, pp. 69 – 79; c) C. F. Moukarzel, P. M. Duxbury, *Phys. Rev. E* **1999**, *59*, 2614 – 2622.

[34] S. Radestock, H. Gohlke, *Eng. Life Sci.* **2008**, *8*, 507 – 522.

[35] a) A. J. Rader, *Phys. Biol.* **2009**, *7*, 016002; b) S. Radestock, H. Gohlke, *Proteins Struct. Funct. Bioinf.* **2011**, *79*, 1089 – 1108; c) B. M. Hespenheide, A. J. Rader, M. F. Thorpe, L. A. Kuhn, *J. Mol. Graphics Modell.* **2002**, *21*, 195 – 207; d) A. J. Rader, B. M. Hespenheide, L. A. Kuhn, M. F. Thorpe, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3540 – 3545.

[36] P. C. Rathi, K. E. Jaeger, H. Gohlke, *PLoS One* **2015**, *10*, e0130289.

[37] N. R. Latorraca, A. J. Venkatakrishnan, R. O. Dror, *Chem. Rev.* **2017**, *117*, 139 – 155.

[38] L. Schnaider, S. Brahmachari, N. W. Schmidt, B. Mensa, S. Shaham-Niv, D. Bychenko, L. Adler-Abramovich, L. J. W. Shimon, S. Kolusheva, W. F. De-Grado, E. Gazit, *Nat. Commun.* **2017**, *8*, 1365.

[39] R. Seifert, K. Wenzel-Seifert, *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2002**, *366*, 381 – 416.

[40] a) B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435 – 447; b) D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701 – 1718.

[41] M. C. Peeters, G. J. van Westen, Q. Li, A. P. IJzerman, *Trends Pharmacol. Sci.* **2011**, *32*, 35 – 42 > S. W. Siu, R. Vacha, P. Jungwirth, R. A. Bockmann, *J. Chem. Phys.* **2008**, *128*, 125103.

[42] A. Goetz, H. Lanig, P. Gmeiner, T. Clark, *J. Mol. Biol.* **2011**, *414*, 611 – 623.

[43] M. G. Wolf, M. Hoefling, C. Aponte-Santamaria, H. Grubmuller, G. Groenhof, *J. Comput. Chem.* **2010**, *31*, 2169 – 2174.

[44] K. J. Kohlhoff, D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman, V. S. Pande, *Nat. Chem.* **2014**, *6*, 15 – 21.

[45] Y. Miao, S. E. Nichols, P. M. Gasper, V. T. Metzger, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10982 – 10987.

[46] a) T. Clark, *Beilstein J. Org. Chem.* **2017**, *13*, 1071 – 1078; b) R. C. Kling, H. Lanig, T. Clark, P. Gmeiner, *PLoS One* **2013**, *8*, e67244.

[47] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269 – 6271.

[48] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712 – 725.

[49] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157 – 1174.

[50] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684 – 3690.

[51] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089 – 10092.

[52] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463 – 1472.

[53] A. Pegoli, D. Wifling, C. G. Gruber, X. She, H. Hübner, G. Bernhardt, P. Gmeiner, M. Keller, *J. Med. Chem.* **2019**, *62*, 5358 – 5369.

[54] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, **2016**, https://www.R-project.org/.

[55] P. Johnson, devEMF: EMF Graphics Output Device. R package version 3.3-1, **2017**, https://CRAN.R-project.org/package=devEMF.

[56] K. Soetaert, plot3D: Plotting Multi-Dimensional Data. R package version 1.1, **2016**, https://CRAN.R-project.org/package=plot3D.

[57] J. Lemon, *R-News* **2006**, *6*, 8 – 12.

[58] D. Osorio, P. Rondon-Villarreal, R. Torres, *C. R. Journ. Electron.* **2015**, *7*, 4 – 14.

[59] U. Lund, C. Agostinelli, CircStats: Circular Statistics, from "Topics in Circular Statistics" (2001). R package version 0.2-6, **2018**, https://CRAN.R-project.org/package=CircStats.

[60] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605 – 1612.

[61] D. J. Jacobs, A. J. Rader, L. A. Kuhn, M. F. Thorpe, *Proteins Struct. Funct. Bioinf.* **2001**, *44*, 150 – 165.

[62] B. I. Dahiyat, D. B. Gordon, S. L. Mayo, *Protein Sci.* **1997**, *6*, 1333 – 1337.

[63] J. Skolnick, L. Jaroszewski, A. Kolinski, A. Godzik, *Protein Sci.* **1997**, *6*, 676 – 688.

[64] C. Pfleger, A. Minges, M. Boehm, C. L. McClendon, R. Torella, H. Gohlke, *J. Chem. Theory Comput.* **2017**, *13*, 6343 – 6357.

[65] T. E. Oliphant, *A guide to NumPy, Vol. 1*, Trelgol Publishing USA, **2006**.

[66] E. Jones, T. Oliphant, P. Peterson, **2001**.