Leucine tunes hydropathy of class A GPCRs

Christian Baumann[1] and Oliver Zerbe[1]

[1]Universitat Zurich Institut fur Chemie

March 17, 2023

Abstract

Leucine and Isoleucine are two amino acids that differ only by the positioning of one methyl group. This small difference has however important consequences in α-helices, as the β-branching of Ile results in helix destabilization. We set out to investigate whether there are general trends for the occurrences of Leu and Ile residues in structures and sequences of class A GPCRs (G protein-coupled receptors). GPCRs are integral membrane proteins in which α-helices span the plasma membrane seven times and which play a crucial role in signal transmission into the cell. We found that Leu side chains are generally present in less densely packed regions and are more protein-surface exposed than Ile side chains. We explored whether this difference might be attributed to different functions of the two amino acids and tested if Leu adjusts the hydrophobicity of the transmembrane domain based on the Wimley-White whole-residue hydrophobicity scales. In class A GPCRs, Leu decreases the variation in hydropathy between receptors and Leu content correlates positively with hydropathy calculated without Leu. Both measures indicate that hydropathy is tuned by Leu. To test this idea further, we generated protein sequences with random amino acid compositions using a simple numerical model, in which hydropathy was tuned by adjusting the number of Leu residues. The model was able to replicate the observations made with class A GPCR sequences. We speculate that Leu tunes the hydropathy of the transmembrane domain of class A GPCRs to facilitate correct insertion into membranes and/or for stability within them.

Leucine tunes hydropathy of class A GPCRs

Christian Baumann*, Oliver Zerbe

Department of Chemistry, University of Zurich, Zurich, Switzerland

*Corresponding author: christian.baumann@chem.uzh.ch

Co-author: oliver.zerbe@chem.uzh.ch

Running title: Leu tunes GPCR hydropathy

Declaration of interest

The authors declare no conflict of interest.

Abstract

Leucine and Isoleucine are two amino acids that differ only by the positioning of one methyl group. This small difference has however important consequences in α-helices, as the β-branching of Ile results in helix destabilization. We set out to investigate whether there are general trends for the occurrences of Leu and Ile residues in structures and sequences of class A GPCRs (G protein-coupled receptors). GPCRs are integral membrane proteins in which α-helices span the plasma membrane seven times and which play a crucial role in signal transmission into the cell. We found that Leu side chains are generally present in less densely packed regions and are more protein-surface exposed than Ile side chains. We explored whether this difference might be attributed to different functions of the two amino acids and tested if Leu adjusts the hydrophobicity of the transmembrane domain based on the Wimley-White whole-residue hydrophobicity scales. In class A

GPCRs, Leu decreases the variation in hydropathy between receptors and Leu content correlates positively with hydropathy calculated without Leu. Both measures indicate that hydropathy is tuned by Leu. To test this idea further, we generated protein sequences with random amino acid compositions using a simple numerical model, in which hydropathy was tuned by adjusting the number of Leu residues. The model was able to replicate the observations made with class A GPCR sequences. We speculate that Leu tunes the hydropathy of the transmembrane domain of class A GPCRs to facilitate correct insertion into membranes and/or for stability within them.

Keywords

Hydropathy, Hydrophobicity, G protein-coupled receptor, GPCR, membrane, leucine, isoleucine

Introduction

Leucine and isoleucine are two amino acids that are identical except for the position of one methyl group, which is attached at the γ-carbon in Leu and the β-carbon in Ile. The high similarity raises the question to which degree these two amino acids are used differently in proteins by nature. Most well-known is the fact that Ile has a lower propensity to be within α-helices, due to steric clashes caused by the β-branching.[1,2] However, little is known if there are additional general trends that distinguish the two amino acids within proteins. We were interested in whether differences exist between these two amino acids within the structures and sequences of class A G protein-coupled receptors (GPCRs).

GPCRs are eukaryotic membrane proteins that possess a transmembrane domain (TMD) consisting of seven plasma membrane spanning α-helices (TM1-7). These proteins are receptors that detect a variety of GPCR subtype-specific extracellular signals, ranging from photons over small organic molecules to proteins. Absorption of a photon or binding of a molecule leads to conformational rearrangements that activate the receptor. Active GPCRs transmit the received signal further to cellular transducers such as G proteins and β-arrestins, which in turn initiate specific signaling cascades. GPCRs are commonly divided into different classes (A to F) based on sequence homology.[3,4] Class A (or rhodopsin-like) GPCRs are the most abundant and diverse receptors and include the most thoroughly studied GPCRs.[5] Since GPCRs span the hydrophobic environment of the plasma membrane, there is no partitioning into hydrophobic core and hydrophilic shell as present in soluble proteins. Rather, TMDs of GPCRs need to maintain favorable interactions with lipids and between TMs, and they need to enable the correct insertion into the membrane during protein translation.[6-9] These factors add different restraints on the primary sequence and lead to a general increase in hydrophobicity of GPCRs and other membrane proteins in comparison to soluble proteins.

One way to determine the overall hydrophobicity of an entire protein or a stretch of an amino acid sequence is by applying the Wimley-White whole-residue hydrophobicity scales.[10,11] These scales are based on the change in free energy (ΔG) for the transfer of amino acids from water to a bilayer interface (ΔG_{wif}) and from water to octanol (ΔG_{woct}). Negative values for either indicate that an amino acid is hydrophobic in the sense that it energetically disfavors to be in water. The difference between both values (ΔG_{woct}-ΔG_{wif}) captures the change in free energy for the insertion into a membrane. Negative values indicate that an amino acid favors the aliphatic environment of octanol over the membrane interface and thus favors the insertion into a membrane. For conciseness, we refer to the difference in octanol and interface scales (ΔG_{woct}-ΔG_{wif}) as hydropathy. Amino acid sequence stretches with negative hydropathy typically indicate transmembrane elements. This is used to predict membrane-spanning elements within membrane protein sequences based on hydropathy plots.[10]

We found differences between Leu and Ile within class A GPCR structures with respect to packing density and protein-surface exposure of the side chains. Leu residues are more commonly found at the receptor surface and in less densely packed areas of the receptor. We explored the idea that Leu adopts a role in adjusting TMD hydropathy and shows thus differences in these structural properties compared to Ile. Leu appears in specific patterns within the amino acid compositions of these GPCR TMDs that would match this putative role. We further assessed to which extent the observed patterns could be expected based on a simple numerical model for amino acid frequencies within TMDs.

Methods

A total of 216 GPCR structures (119 annotated as active and 97 as inactive) were downloaded from the GPCRdb[12], selecting all TMD helices, helix 8 and the loops between these elements. Protons were added using PyMol v2.4.2[13]. PDB identification codes of the structures are listed in table S1 in the supplementary information (SI). Packing densities were calculated by counting the number of atoms within a 5 Å radius around the δ-methyl carbons of Leu and Ile. Atoms that belong to the same residue as the probed methyl group were not included in the count. The area of side chains at the protein surface was quantified by determining the solvent-accessible surface areas using GETAREA[14] with default settings including a water probe radius of 1.4 Å. The relative area of side chains at the protein surfaces was calculated based on mean values obtained for side chains of the free amino acids (174.2 $Å^2$ for Ile and 174.0 $Å^2$ for Leu).

Class A GPCR sequences (1580 in total from 325 targets) were downloaded from GPCRdb selecting only sequences of TMD helices. Hydropathies were calculated based on the differences of Wimley-White whole-residue hydrophobicity scales for the transfer of an amino acid from water to a bilayer interface and from water to octanol (ΔG_{woct}-ΔG_{wif}).[10,11]The sum of all amino acid hydropathies was taken as the hydropathy of the TMDs. Only the hydropathies for protonation states of amino acids at pH 7 were considered. Spearman's rank correlation coefficients (ρ) were calculated based on amino acid content, i.e. the number of residues of an amino acid within the TMD divided by the sequence length of the TMD. All calculations, statistical analyses, and plots were done using R v4.0.3[15] with RStudio v1.4.1103[16]and the packages bio3d[17] and stringr[18].

Results

Leu & Ile in class A GPCRs

We compared Leu and Ile residues based on how densely packed their side chains are and how strongly these side chains are exposed on the protein surface. Since GPCRs are membrane proteins, protein-surface exposure captures contact with the lipid bilayer or solvent, depending on the location of the side chain. Further, side-chain packing density and protein-surface exposure measure overlapping properties of residues within protein structures, i.e. a high level of protein-surface exposure will lead to a small packing density for a given residue. A sample of 216 experimental structures from 95 unique GPCRs indicates that Ile tends to occur in more densely packed regions than Leu (Fig. 1A) and that Leu tends to be more protein-surface exposed than Ile (Fig. 1B). This is true for the majority of receptor structures, with a total of 88.4 % of them displaying more densely packed Ile residues and 81.0 % more protein-surface exposed Leu. This general difference between the two amino acids suggests that Leu and Ile residues tend to occur within different structural contexts within class A GPCRs, with Leu being more prone to interact with the lipids in the membrane bilayer.

Interestingly, significant differences in side-chain packing between active- and inactive-state GPCRs are present for Leu and Ile (Fig. 1C). In both cases, packing density is smaller in active- than in inactive-state structures (Ile: - 8.6 %, Leu: - 4.9 %). This is further accompanied by a less pronounced and non-significant increase in protein-surface exposure of the two amino acids (Ile: + 4.6 %, Leu: + 2.8 %) (SI Fig. S1). GPCR activation leads to conformational changes that allow G proteins to bind. These conformational changes include the outward movement of TM6 and the subsequent opening of a cytosolic crevice that accommodates the C-terminal helix of a G protein.[5,19] Active-state GPCR structures are generally solved in presence of G proteins or G protein-alike substitutions[20,21], which were not included in the calculation of packing densities. It is therefore likely that the decrease in packing density upon activation reflects the opening-up of the G protein binding pocket. This further matches the more pronounced decrease in packing density for Ile than Leu since this conformational change can be expected to have a stronger impact on buried residues.

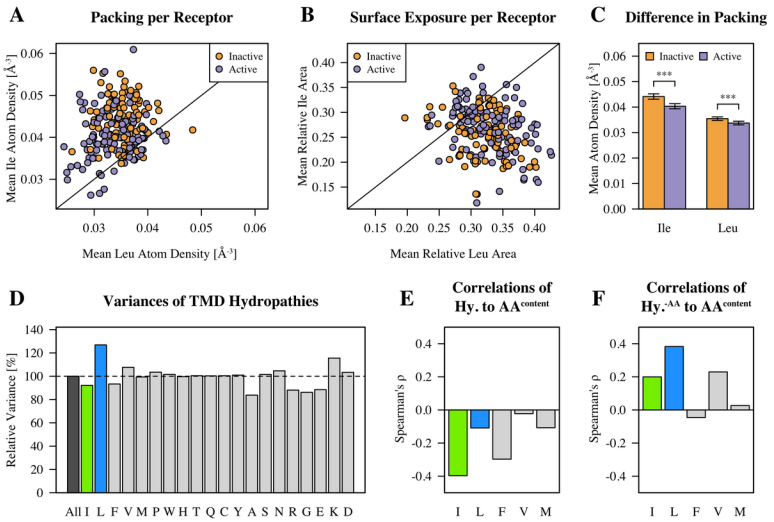

Fig. 1. Differences between Leu and Ile in class A GPCRs.(A) Packing density and **(B)** protein-surface exposure of Leu and Ile side chains in 216 GPCR structures. The diagonals indicate positions where Leu and Ile side chains are equally packed (A) or protein-surface exposed (B). Values above the diagonal indicate that Ile side chains are more densely packed (in A) and that Ile side chains are more protein surface exposed (in B) than Leu side chains. Structures of GPCRs in active and inactive states are highlighted in purple and orange, respectively. **(C)** Differences in side-chain packing density between inactive- and active-state structures. Error bars indicate 95 % confidence intervals. P-values were determined using Welch's t-tests and were < 0.001 in both cases. **(D)**Relative variances of TMD hydropathies. Variances were calculated for hydropathies of complete TMD sequences (≪All≫) and for hydropathies of TMD sequences from which the indicated amino acid was removed. Amino acids are ordered from low to high hydropathy. The dashed line indicates the variance of hydropathies of complete sequences as a visual reference. **(E)** Correlations between amino acid content and TMD hydropathy based on Spearman's rank correlation coefficient (ρ). Spearman's ρ ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation) with 0 indicating the absence of any correlation. Negative correlations show that TMD hydropathy decreases (i.e. TMD is more hydrophobic) when the amino acid content increases.**(F)** Correlations between amino acid content and TMD hydropathy calculated without given amino acid. Positive correlations show that TMD hydropathy (calculated without given amino acid) increases (i.e. TMD is more hydrophilic) when the amino acid content increases. Only correlations for the five amino acids with the lowest hydropathy are shown in E and F.

We are not aware of an obvious reason that accounts for the differences in packing densities and protein-surface exposure between Leu and Ile side chains in class A GPCRs. It is however possible to formulate several hypotheses that could explain the observed differences, and which are not necessarily mutually exclusive. For example, the destabilizing effects due to the β-branching of Ile might require additional structural restraints to be present for adequately accommodating Ile within α-helices, preventing it from being too protein surface exposed. Another underlying rationale could be that Leu might form better interactions with lipids than Ile and thus occurs more often on the protein surface. The hypothesis that intrigued us the most was that Leu is more protein surface exposed because it adjusts the hydropathy of class A GPCRs for optimal insertion into membranes and/or for stability within them.

We hypothesized that whether or not Leu is important for optimizing TMD hydropathy in class A GPCRs

could be detected by two patterns with which this amino acid occurs in the overall amino acid composition. The first pattern is based on the spread of values of a property within a population: A property that needs to adopt a defined optimal value will display little variation between different members of the population. If mainly one factor optimizes the value of such a property, then the removal of that factor will lead to a larger variation in the resulting values, since they are no longer optimized. Hence, if Leu is responsible for optimizing TMD hydropathy, then hydropathies calculated without Leu should show larger variations than when calculated including Leu. This was indeed the case for the sequences of 1580 class A GPCR TMDs (Fig. 1D). Among all amino acids, Leu displays the strongest impact on TMD hydropathy variation.

The second expected pattern is related to correlations between Leu content and TMD hydropathy when calculated with and without Leu residues. If Leu tunes hydropathy, then a positive correlation between Leu content and TMD hydropathy (calculated without Leu) is expected because more Leu residues are required to compensate for a more hydrophilic sequence, i.e. the more hydrophilic the TMD of a GPCR is (without Leu), the more Leu residues are required to make this TMD sufficiently hydrophobic (Fig. 1F). This, however, means that the Leu content and the overall TMD hydropathies should be uncorrelated, i.e. the TMD of a GPCR does not become more hydrophobic the more Leu it contains (Fig. 1E). Overall, the correlations between Leu content and TMD hydropathy in the sequence sample match these predictions (Fig. 1E & 1F).

Interestingly, the patterns that are observed with Leu are absent or weaker with Ile, suggesting that Ile is not (or much less) involved in adjusting the hydropathy of the TMDs. This is insofar surprising as both amino acid share similar hydropathies, with Ile (-0.81 kcal/mol) being even slightly more hydrophobic than Leu (-0.69 kcal/mol). However, even if Leu would be the main driving force in adjusting TMD hydropathy in GPCRs, this will not be the only function of Leu. To quantify the extent to which the above-described effects are present when only a part of all Leu residues is involved in hydropathy tuning, we performed numerical simulations based on a simplified model for amino acid compositions.

A numerical model for hydropathy tuning

To investigate hydropathy tuning by Leu, we used a simple model with sequences composed of only 4 ≪amino acids≫: A, B, C and D. These form sequences of the type $A_aB_bC_cD_d$, with small letters indicating the number of the amino acids. A and B were modeled according to Ile and Leu, with ≪a≫ and ≪b≫ corresponding to the occurrences of the two amino acids within the TMD sequences of class A GPCRs. The hydropathy of Ile was assigned to A (h_A = -0.81 kcal/mol) and the one of Leu was assigned to B (h_B= -0.69 kcal/mol). C and D, and their counts ¡¡c¿¿ and ¡¡d¿¿, were modeled to reflect all other amino acids with hydropathies smaller than zero and amino acids with hydropathies larger than zero. C and D are thus generic amino acids that represent the averages of all hydrophobic (except Ile and Leu) and all hydrophilic amino acids, respectively. For C and D, the average hydropathies of the amino acids they represent were used (h_C: -0.36 kcal/mol, h_D: 0.8175 kcal/mol).

Amino acid compositions for simulated sequences were created by generating Gaussian distributed numbers for a-d based on the amino acid occurrences in the class A GPCR TMD sequences (A: 8.8 % +- 3.0 % (SD), B: 15.2 % +- 3.4 %, C: 28.0 % +- 3.0 %, D: 48.0 % +- 2.8 %). The generated random numbers a-d were then multiplied by 220 and rounded to the nearest integer to obtain sequence lengths that are comparable to the lengths of the TMD sequences. To test for statistical features, a total of 1'500 sequences were generated in each of the 10'000 runs.

Driver residues were introduced to drive hydropathies towards a defined optimum value h_{opt}, which was set to -1.5 kcal/mol to resemble the mean hydropathy of the TMD sequences (-1.47 kcal/mol). With B as the driver, ¡¡a¿¿, ¡¡c¿¿ and ¡¡d¿¿ were randomly determined by a Gaussian distribution as described above. Then ¡¡b¿¿ was determined as shown by the equation below, with g(B) being a randomly Gaussian distributed number and h_B being the hydropathy of B. The first term calculates the difference between the optimal and the already present hydropathy, and divides it through the hydropathy of B, yielding the value of ¡¡b¿¿ needed to get to the optimal hydropathy. A defined degree of noise was introduced using f_{drive}, which determines the

amount of drive towards the optimum value h_{opt}, with the rest (1- f_{drive}) being determined randomly by the Gaussian distribution g(B). The value of f_{drive} used was 0.25, which, however, does not mean that 25 % of the final number of ¡¡b¿¿ is driving the hydropathy towards the desired value since this fraction additionally depends on the value of h_{opt}. Interestingly, the variances and correlations were identical between runs with different values for h_{opt}, indicating that the actual value of h_{opt} is not important to observe the effects of tuning towards it.

$$b = f_{\text{drive}} \times \frac{h_{\text{opt}} - (a \times h_A + c \times h_C + d \times h_D)}{h_B} + (1 - f_{\text{drive}}) \times g(B)$$

Two different models were tested (Fig. 2). In the first model, all amino acids were modeled independently from the resulting hydropathies by generating a-d based on Gaussian distributions alone. This simulates a case in which TMD hydropathy is not optimized (Fig. 2A-2C). In the second model, ≪a≫, ≪c≫ and ≪d≫ were generated based on Gaussian distributions, whereas ≪b≫ was chosen based on the equation shown above. This simulates the case in which Leu would be the driving force for adjusting the hydropathy of the TMDs (Fig. 2D-2F).

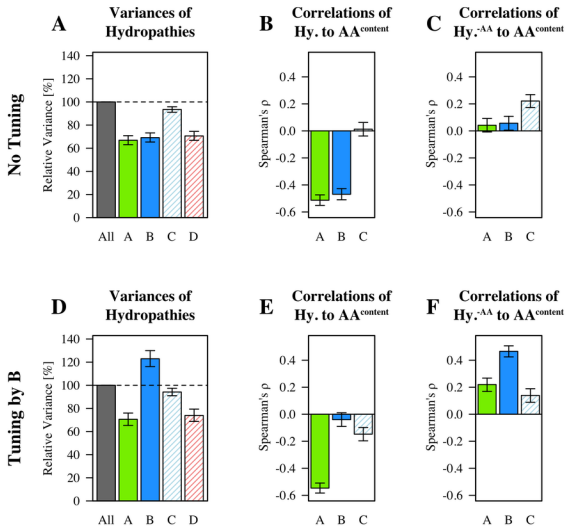

Fig. 2. Numerical simulation of hydropathy tuning. (A, B, C) Simulated sequences in the absence of hydropathy tuning and**(D, E, F)** when hydropathy is tuned by B. Error bars indicate the 95 % confidence interval within which values were obtained among all runs. **(A, D)** Relative variances of hydropathies. Variances were calculated for hydropathies of complete sequences (≪All≫) and for hydropathies of sequences from which the indicated amino acid was removed. The dashed line indicates the hydropathy variance of complete sequences as a visual reference. **(B, E)** Correlations between amino acid content and hydropathy based on Spearman's rank correlation coefficient (ϱ). **(C, F)** Correlations between amino acid content and hydropathy calculated without given amino acid. Only correlations for the hydrophobic amino acids A, B and C are shown in bar plots B, C, E and F. The generic hydrophilic amino acid D displayed strong positive correlations in each case.

The results confirm the anticipated effects for hydropathy tuning: the variation between hydropathies increases when calculated without the tuning amino acid (Fig. 2D) and a positive correlation exists between the content of the tuning amino acid and the hydropathy calculated without this amino acid (Fig. 2F). Further, the simulation shows the degree to which the effects are present when only a fraction of the tuning amino acid is driving the hydropathy towards the optimum value. The f_{drive} of 0.25 leads to similar patterns as were observed within the sequences of class A GPCR TMDs, supporting the idea that Leu is responsible for adjusting the hydropathy of these TMDs. Interestingly, the numerical model also captures the overall patterns of Ile within the TMD sequences. In the simulation, A was modeled after Ile and its content was determined by a Gaussian distribution alone, suggesting that Ile is not involved in tuning TMD hydropathy in class A GPCRs.

Discussion

To summarize, Leu side chains generally occur in less densely packed regions and are more protein-surface exposed than Ile side chains in structures of class A GPCRs, indicating that Leu interacts generally more with lipids. Within the TMD sequences of class A GPCRs, Leu decreases the variation in hydropathy between receptors and Leu content correlates with hydropathies calculated without Leu. A simple numerical model was able to reproduce the overall magnitudes of these two patterns when the number of Leu was adjusted to drive the hydropathy toward an optimal value. Taken together, these observations suggest that the hydropathy of class A GPCR TMDs is tuned by Leu. Since hydropathy is a measure for the energetics of membrane insertion, an appropriate Leu content appears to ensure that Class A GPCRs are inserted into membranes and/or are stable within them. The sequence patterns observed with Leu are absent with Ile, indicating that Ile is not involved in adjusting TMD hydropathy.

Leu content and protein hydrophobicity have previously been linked in proteins of thermophiles. In thermophilic organisms, an increased hydrophobicity in the protein core improves thermostability, which keeps these proteins functional at elevated temperatures.[22]The comparison between 110 pairs of homologous proteins from thermophilic and mesophilic organisms indicated that the Leu content is significantly higher in thermophilic proteins and accounts for a significant change in the aliphatic index.[23] The aliphatic index quantifies hydrophobicity based on the Ala, Val, Ile and Leu content of a protein.[24] Interestingly, the authors of that study used the correlation between aliphatic index and Leu content to question the validity of the aliphatic index, whereas we would interpret it in a way that the increase in Leu content is the reason for the increased hydrophobicity of these proteins.

One underlying rationale could be that a mutation of any α-helical residue to Leu is less destabilizing than a mutation to the β-branched Ile. Therefore, if an increased protein hydrophobicity is beneficial, then a mutation to Leu might be preserved more commonly than a mutation to Ile, despite their comparable hydrophobicity. In the case of GPCRs, such a stability-driven effect could be further amplified due to the lower intrinsic stability of GPCRs compared to other proteins.[25] However, it is unclear to what degree such an effect exists in a membrane environment since α-helix-destabilization by β-branching appears to be absent within membranes, at least for single-span α-helices.[26]

It is unclear how generalizable the observations made on Class A GPCRs are, particularly because the patterns with Leu are completely absent in GPCRs outside of class A (SI Fig. S2). For these receptors, Leu resembles Ile, whereas Val shows slightly more pronounced correlations that are indicative of hydropathy tuning. Additionally, we considered the entire TMDs as being important for membrane insertion or stability and neglected that residues buried within the TMD are unlikely to contribute to the overall hydropathy. Still, the patterns we observed in the TMD sequences of class A GPCRs remain highly suggestive of Leu tuning the hydropathies of at least this group of proteins. So far, we could not come up with alternative explanations that would produce similar statistical patterns without connecting Leu to hydropathy tuning. To further support the hypothesis that hydropathy is indeed tuned by Leu, and to rule out potential statistical anomalies and alternative explanations, more sophisticated models and alternative approaches need to be explored.

CRediT authorship contribution statement

Christian Baumann: Conceptualization, Methodology, Formal analysis, Visualization, Writing – Original Draft, Funding acquisition**Oliver Zerbe:** Writing – Original Draft, Supervision, Funding acquisition

Acknowledgements

We thank Wan-Chin Chiang for translating the publication of Lu et al.[23], Dr. Mattia Deluigi for insightful discussions, and Prof. Dr. Alexandria Deliz Liang and Dr. Matthias Schuster for helpful feedback and suggestions on the manuscript. This project was supported by research grants from the Swiss National Science Foundation (grants No. 310030-197679 and 310030-159453) and the Forschungskredit of the University of Zurich (FK-18-083).

References:

1. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 1994;3(5):843-852.

2. Heinz DW, Baase WA, Matthews BW. Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc Natl Acad Sci U S A.*1992;89(9):3751-3755.

3. Davies MN, Secker A, Freitas AA, Mendao M, Timmis J, Flower DR. On the hierarchical classification of G protein-coupled receptors.*Bioinformatics.* 2007;23(23):3113-3118.

4. Isberg V, de Graaf C, Bortolato A, et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol Sci.* 2015;36(1):22-31.

5. Zhou Q, Yang D, Wu M, et al. Common activation mechanism of class A GPCRs. *Elife.* 2019;8.

6. Cymer F, von Heijne G, White SH. Mechanisms of integral membrane protein insertion and folding. *J Mol Biol.* 2015;427(5):999-1022.

7. Skach WR. Cellular mechanisms of membrane protein folding. *Nat Struct Mol Biol.* 2009;16(6):606-612.

8. Popot JL, Engelman DM. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry.* 1990;29(17):4031-4037.

9. Yamane K, Akiyama Y, Ito K, Mizushima S. A positively charged region is a determinant of the orientation of cytoplasmic membrane proteins in Escherichia coli. *J Biol Chem.* 1990;265(34):21166-21171.

10. White SH, Wimley WC. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct.*1999;28:319-365.

11. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol.*1996;3(10):842-848.

12. Pándy-Szekeres G, Munk C, Tsonkov TM, et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.*2018;46(D1):D440-D446.

13. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. In.

14. Fraczkiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. In. Vol 19. J Comp Chem1998:319-333.

15. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. In:2020.

16. Team R. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/. In:2021.

17. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures.*Bioinformatics.* 2006;22(21):2695-2696.

18. Wickham H. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr. In:2019.

19. Ahn D, Ham D, Chung KY. The conformational transition during G protein-coupled receptor (GPCR) and G protein interaction. *Curr Opin Struct Biol.* 2021;69:117-123.

20. Wang J, Hua T, Liu ZJ. Structural features of activated GPCR signaling complexes. *Curr Opin Struct Biol.* 2020;63:82-89.

21. Congreve M, de Graaf C, Swain NA, Tate CG. Impact of GPCR Structures on Drug Discovery. *Cell.* 2020;181(1):81-91.

22. Haney P, Konisky J, Koretke KK, Luthey-Schulten Z, Wolynes PG. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus Methanococcus. *Proteins.* 1997;28(1):117-130.

23. Lu B, Wang G, Huang P. [A comparison of amino acid composition of proteins from thermophiles and mesophiles]. *Wei Sheng Wu Xue Bao.* 1998;38(1):20-25.

24. Ikai A. Thermostability and aliphatic index of globular proteins.*J Biochem.* 1980;88(6):1895-1898.

25. Maeda S, Schertler GF. Production of GPCR and GPCR complexes for structure determination. *Curr Opin Struct Biol.*2013;23(3):381-392.

26. Li SC, Deber CM. A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol.* 1994;1(8):558.