Water at the Hydrated Membrane-Protein Interface

Author: Natalia Martínez Torres.
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Giancarlo Franzese

LRP1 is a 600 kDa membrane protein with more than 4500 amino acids and many repeated units, crucial for blood-brain barrier function and possibly implicated in Alzheimer's disease. We analyze its hydration properties starting from molecular dynamics simulations data produced for all-atom models of the separate units suspended in water. We focus on six amino acids that preserve their position in the units. We show evidence that their average interaction with water depends on the local environment and the 3D protein structure, at least for one of the amino acids. This observation highlights that the hydration property is, in principle, context-dependent and does not depend solely on the hydrophilicity of the amino acid. Therefore, in developing coarse-grained models for large-scale simulations based on effective interactions, our result calls for capturing this feature that could be relevant in large-scale numerical studies of LRP1 within the general context of finding new treatments for neurodegenerative diseases.

I. INTRODUCTION

Low-density lipoprotein receptor-related protein 1 (LRP1) is a large glycoprotein found on the plasma membrane of cells that plays a crucial role in the endocytosis and transcytosis of molecules and macromolecules. It is widely expressed in various tissues, including the endothelial cells of the blood-brain barrier (BBB). LRP1 facilitates a rapid and efficient transcytosis mechanism, essential for maintaining homeostasis and protecting the brain from potentially harmful substances [1].

LRP1 is significant due to its involvement in neurodegenerative diseases. The receptor's malfunctioning in removing damaged or excess proteins is associated with pathologies such as Alzheimer's and Parkinson's diseases [2]. Therefore, LRP1 is a key target for nanoparticle-mediated drug delivery, offering effective treatment pathways for central nervous system disorders. This targeted approach aims to develop more effective drugs with higher specificity and greater cell selectivity, thereby reducing side effects and minimizing damage to healthy tissues.

To achieve these goals, it is essential to understand how LRP1 interacts with its environment. Given the large size and complexity of the LRP1 protein, studying its behavior at the atomistic level in conjunction with the receptor is beyond the computational capacity of our available resources. Therefore, developing a multiscale approach to investigate this protein is reasonable. Starting from all-atoms molecular dynamics simulations of small parts of the LRP1, we aim to define a coarse-grained model that can reproduce the atomistic system's essential features at a larger scale. In this process, an important step is to find consistent interaction potentials between the LRP1 amino acids and the surrounding water, as well as the nanoparticle components with water, in such a way that it will be possible to describe accurately the water-mediated interactions between receptor and nanoparticle. Here, we focus on the hydration properties of the LRP1 amino acids.

II. METHODOLOGY

In this analysis, we use unpublished data from molecular dynamics simulations with the authors' authorization. These data were generated through a collaboration between Franzese's group at UB and Battaglia's group at the IBEC [3]. Due to the large size of the protein sequence, the authors concentrated on simulating different parts of the protein individually.

LRP1 has repeated motifs and, based on a detailed analysis reported in Ref. [3], it has a stable dimer structure (Fig. 1). There are three main types of motifs: calcium-binding (CB) domains, epidermal growth factor (EGF)-like domains found in many extracellular proteins, and beta-propeller domains, called in this way for their specific form.

In the present analysis, we focus on CB domains since they are the structures that present the most significant conservation in their amino acid sequence across the entire LRP1 structure. The CB motifs are crucial for the structural integrity and function of LRP1, stabilizing the protein through interactions with calcium ions. They cluster into four groups separated by EGF-like and beta-propeller domains (Fig. 1). We refer to the four sequences of CB domains as clusters I, II, III, and IV.

For each of the 31 CB domains distributed in the four clusters, we select for our analysis of the hydration properties six of the amino acids that repeat their position more frequently within each CB sequence: phenylalanine (PHE), cysteine (CYS), serine (SER), glutamic acid (GLU), aspartic acid (ASP), and lysine (LYS), ordered here from more to less hydrophobic as in the Kyte and

*Electronic address: nataliamartinez1310@gmail.com

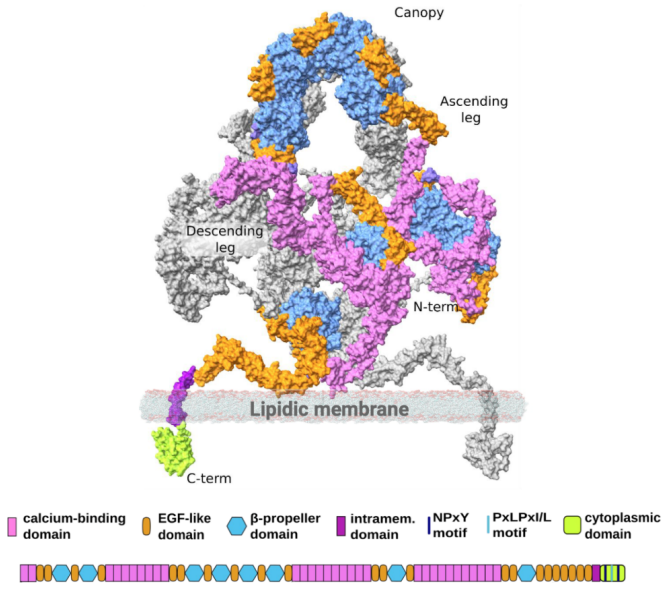

FIG. 1: LRP1 dimer structure and its motifs [3]. Top: One (back) monomer is represented in gray, while the other (front) monomer is colored following the legend shown on the bottom. Bottom: The sequence of repeated units in one monomer, with their names indicated in the legend.

Doolittle hydrophobicity scales [4]. In the CB motifs, we consider 6 ASP, 6 CYS, 3 LYS, 1 PHE, 1 SER, and 1 GLU are repeated in (almost) all clusters. For each occurrence of these six amino acids in the 31 CB domains, we analyze molecular dynamics time series over 1000 ns, saved every 0.1 ns, for 10^4 configurations.

In our analysis, we use several tools. To obtain the amino acid sequences of the calcium-binding units, we use **UniProt**, an online database offering extensive information on protein sequences.

For the analysis of molecular dynamics, we employ **GROMACS** (GROningen MAchine for Chemical Simulations). This powerful software allows us to select the desired amino acids and utilize its integrated functions to calculate the radial distribution function between selected molecules at specified distances.

To visualize the structure of the clusters and assess the exposure of the studied amino acids to water, we use **VMD** (Visual Molecular Dynamics), a tool designed for displaying and analyzing molecular assemblies. VMD enables us to gain a clear, visual understanding of the spatial arrangements and interactions within the protein.

Finally, we processed and plotted the data using **Python**. By integrating these tools and resources, we conducted a comprehensive and detailed study of the protein's behavior and interactions at the molecular level.

The following section defines the quantity used to characterize hydration properties.

A. Radial Distribution Function and Effective Interaction Potential

To characterize hydration around each conserved amino acid in LRP1, we calculate the Radial Distribution Function (RDF) between the oxygen atoms of water molecules and the carbon alpha (CA) atoms of the selected amino acids. Furthermore, to define an effective interaction potential between the amino acid and water, we adopt a simple strategy to evaluate it from the RDF.

The RDF describes how particle density varies as a function of distance from a reference particle. In the context of protein hydration, the RDF provides insight into the spatial distribution of water molecules around the CAs of the LRP1 protein. By definition, the RDF is

$$g(r) = \frac{\langle \rho(r) \rangle}{\rho_{\text{bulk}}} \tag{1}$$

where $\langle \rho(r) \rangle$ is the average local number of water molecules at a distance r from the considered CA, and ρ_{bulk} is the bulk density of water.

To estimate the effective pair-interaction between the CAs and water, we follow the strategy suggested by Stillinger and Head-Gordon [5], based on the hyper-netted chain approximation (HNC),

$$U(r) \cong k_B T[g(r) - 1 - \ln(g(r) - C(r))] \tag{2}$$

where $C(r)$ is the direct correlation function between the two interaction points at a distance r, defined in terms of the $g(r)$ by the Ornstein-Zernike integral equation [6]. Here, we make a further approximation assuming that $C(r)$ is negligible because the CAs and the water molecules are separated by the amino acid residues with average sizes of approximately 0.5 nm, which is of the order of one water layer.

III. RESULTS

A. Radial Distribution Function Analysis

Our initial hypothesis is that the RDF will adjust to each amino acid's hydrophobic or hydrophilic behavior. Conversely, results show this hypothesis needs to be more complex, and an effective interaction would need more information.

First, we calculated the average RDF across every calcium-binding unit and every cluster for each amino acid (Fig. 2). The peak in every $g(r)$ represents the most probable distance at which water molecules are found around the CA atoms of the amino acids, i.e., the first hydration shell. The height of the peak is indicative of the strength of the interaction. A higher peak suggests stronger interaction, typical for hydrophilic amino acids, given their ability to form hydrogen bonds between their residues and water molecules. Conversely, a lower peak

indicates weaker interactions, as seen in more hydrophobic amino acids, as their interactions with water are primarily driven by van der Waals forces rather than hydrogen bonding.

We observe several interesting trends when comparing the obtained radial distribution function values to the expected behaviors. Because we focus on CAs, separated from water by their residues, all the peaks are smaller than one. However, we expected that the first peak of the $g(r)$ would be smaller for the most hydrophobic amino acid and higher for the less hydrophobic. However, the overall results are only partially consistent with this expectation.

In particular, phenylalanine (PHE) and cysteine (CYS) present the lowest peaks in the RDF, with values lower than 0.4, which aligns with their hydrophobic nature. This low peak indicates fewer water molecules surrounding these amino acids, consistent with their tendency to avoid aqueous environments.

In the intermediate range, with values around 0.5, we find serine (SER), glutamic acid (GLU), and lysine (LYS) in this order. While it is expected that glutamic acid and serine would fall into this middle category due to their moderate hydrophilic properties, the position of lysine is surprising. Given lysine's strong hydrophilic nature, it should exhibit a higher peak in the RDF, indicating a greater interaction with water. Additionally, we would expect serine to be lower than glutamic acid due to its lower hydrophobicity value, but the results show serine having a higher peak than glutamic acid.

Finally, as anticipated, aspartic acid (ASP) appears highly hydrated, with a peak around 0.6, which matches its known hydrophilic properties. However, lysine should be even more hydrated, yet the results do not reflect this, suggesting other factors may influence its hydration behavior in the protein environment.

To further elucidate these influences, we discuss in the next section a focused investigation on how the structural context and local environment affect the hydration of the amino acids studied. This additional analysis aims to provide a more comprehensive understanding of the factors governing amino acid hydration in the LRP1 protein.

B. Structural context

We understand that an averaged RDF for each amino acid does not accurately describe the hydration properties because the specific position of an amino acid within a protein structure significantly affects its exposure and interaction with the surrounding environment. As we can see in Fig.3, proteins naturally fold into specific three-dimensional shapes to achieve a state of equilibrium, minimizing their free energy. This folding process leads to the formation of secondary structures, such as alpha helices and beta sheets, which influence the amino acids' exposure to the solvent environment.

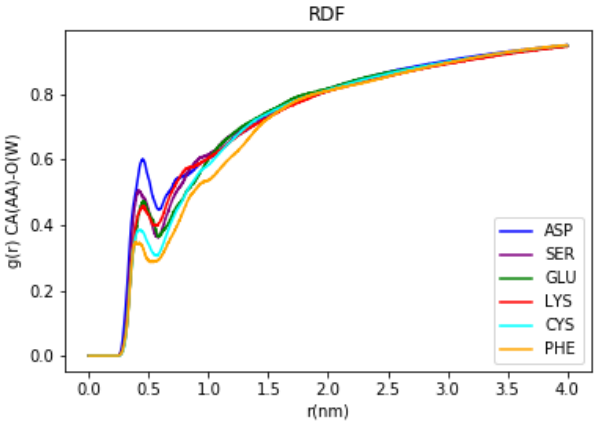

FIG. 2: The averaged radial distribution function (g(r)), calculated between the carbon alpha (CA) of each amino acid (AA) and the oxygen in water (O(W)), as a function of the distance to the AA in nanometers (r). Each color represents a different amino acid, as shown in the legend.

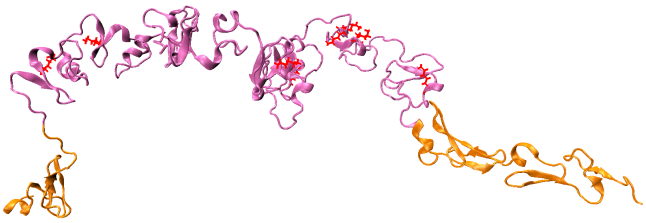

FIG. 3: Cartoon representation of cluster II, with the same color lagend as in Fig.1: we show the calcium-binding units (CBU3-CBU10) in pink, the EGF-like domains in orange.

However, for amino acids appearing multiple times in each calcium-binding unit across the clusters, we could expect a higher degree of homogeneity in the RDFs when they have similar spatial arrangements. Therefore, we compare these results and group all the cases with similar profiles. Following this characterization method, the appendix illustrates the various characteristic hydration behaviors for each amino acid.

Here, we describe the case of ASP in detail, which results in the one with the largest variation of hydration properties depending on the specific position occupied within the CB sequence. By examining ASP in different structural contexts within the protein, we aim to understand better the factors influencing its hydration and to shed light on the discrepancies observed in our initial results.

We found that ASP5, categorized as the most hydrophobic, is consistently located in alpha-helices, which results in low water exposure. In contrast, other ASP residues are in non-coiled structures, leading to higher water exposure. ASP residues in categories 1 and 2 have similar structural contexts, so their different hydration

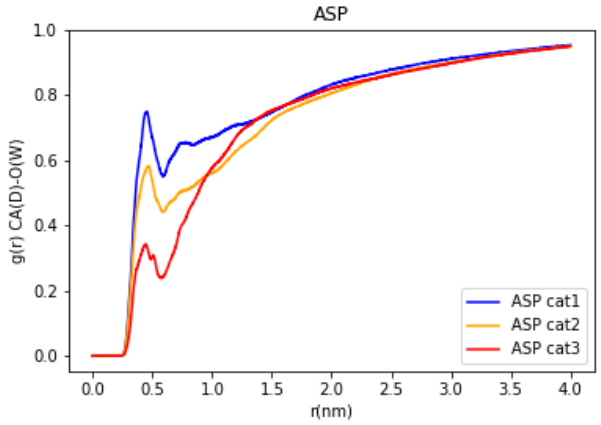

FIG. 4: RDF for the 3 aspartic acid's hydration categories. In blue, cat1, the most hydrophilic, in orange, cat2, an intermediate behavior and in red, cat3, the most hydrophobic.

levels can't be attributed to structure alone. These findings suggest that factors beyond structural context influence hydration behavior, warranting further investigation to fully understand the hydration properties of ASP residues in the LRP1 protein.

IV. DISCUSSION

Nearby amino acids can create specific microenvironments through electrostatic interactions, hydrogen bonding, and steric effects. These interactions influence the folding and stability of the protein, thereby affecting the exposure of individual amino acids to water molecules. Understanding the impact of these neighboring residues is crucial for insights into protein behavior, stability, and function.

In a first approximation, we aimed to study the amino acids in groups of three to determine if there was any correlation between the exposure to water of the central amino acid and the hydrophilicity of its nearest neighbors. However, the results indicated that studying amino acids in triplets did not reveal a clear correlation. The same three amino acids sometimes exhibited different hydration behaviors depending on their position within the protein. Although two neighboring hydrophobic amino acids occasionally resulted in a more hydrophobic profile for the central amino acid, this was not consistent enough to attribute it solely to the neighbors' nature.

These observations suggest that if we want to characterize an amino acid's behavior based on its neighbors, we must consider more of them. Understanding how many neighboring amino acids are necessary to accurately describe a given amino acid's hydration and exposure properties could provide deeper insights into protein behavior and stability. These findings open the doors for more sophisticated studies exploring the range of influence neighboring amino acids could have.

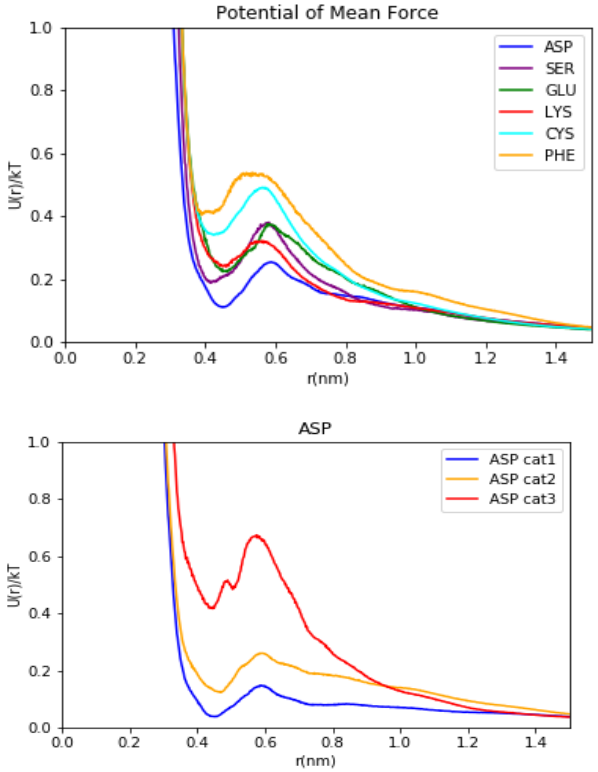

FIG. 5: 1.The average effective interaction potential $U(r)/k_BT$ for the six different amino acids in different colors, as shown in the legend, as a function of the distance in nanometers. 2.The average effective interaction potential $U(r)/k_BT$ for the three different hydration categories of ASP.

Although recognizing the complexities in isolating specific amino acid-water interactions, we first study LRP1 protein hydration by deriving the effective interaction from Eq.(2) averaging over all the same amino acids regardless of their position. The effective potential offers insights into the system's free energy profile, highlighting regions of higher and lower binding affinity. This analysis helps us better comprehend the hydration behavior and the underlying energetic landscape.

A minimum in the effective interaction indicates a favorable, stable interaction distance with high water binding affinity. At the same time, a subsequent local maximum signifies an energy barrier that water molecules must overcome to move away from the favorable interaction site. The effective interaction energy at short distances goes to infinity due to the strong repulsive forces due to the excluded volumes. It tends to zero at long distances as interactions diminish and molecules behave independently.

Following the initial idea discussed in this work, hydrophilic amino acids (ASP, LYS, GLU, and SER) should have minima at shorter distances, indicating strong interactions with water, and narrower maxima, suggesting stable hydration shells. Conversely, hydrophobic PHE

and CYS should show a minimum at longer distances and a broader, more pronounced maximum, reflecting weaker, less specific interactions. However, this approximation needs improvement. Therefore, we derive the effective interactions for the different characteristic behaviors of each amino acid obtained from the RDF. These figures are shown in the appendix and provide a better approximation for characterizing LRP1 hydration.

In contrast to the previous Fig5.1, Fig5.2 of the PMF for aspartic acid (ASP) across different categories reveals distinct behaviors, emphasizing the limitations of a first-order approximation. The PMF curves for ASP show significant variations: Category 1 (blue) indicates strong water interactions with a sharp peak and low PMF values, suggesting high hydration. Category 2 (orange) exhibits moderate hydration with intermediate PMF values and partial structural shielding. Category 3 (red) displays the weakest hydration, with high PMF values and a broad peak, reflecting its position in an alpha-helix. These differences highlight the importance of structural context and local environment for accurate characterization.

V. CONCLUSIONS

In summary, while most of the results align with the hydration expectations based on the amino acids' intrinsic hydrophobic or hydrophilic nature, some discrepancies suggest that intrinsic properties are not the sole determinants of hydration behavior. These inconsistencies indicate that additional factors, such as the position of the amino acid within the protein structure and the surrounding environment, could significantly influence hydration patterns.

Therefore, we identify distinct hydration behaviors of the amino acids in LRP1 depending on each amino acid environment. This approach better approximates the diverse hydration patterns, allowing us to characterize the protein's interaction with water more effectively.

Our results can be summarized as follows.

- We have successfully obtained the averaged RDF and effective interaction potential of six characteristic amino acids from the LRP1 protein, and we have categorized the different behaviors each present. Results approximately align with the expected according to their hydrophilic nature, but some inconsistencies show the possible influence of other factors.

- We have studied factors influencing amino acid exposure to water, such as the structural context and the local environment. We see a correlation between the structure an amino acid is found on and its hydration. Still, our first approach to characterizing amino acids according to their nearest neighbors needs to be completed.
- Categorization of the different shown behaviors for each amino acid sets the possibility to characterize the protein's interaction not only on 20 different interactions (one for every amino acid) but on more.

- Characterizing a protein as a simple sum of individual amino acid behaviors is insufficient. Proteins are complex systems that interact on many levels; first-order approximations cannot capture the nuanced interactions and influences present.

- This study sets the basis for more sophisticated studies considering the full spectrum of interactions within the protein to characterize its interaction and hydration dynamics.

Acknowledgments

I would like to thank Gian Marco Tuveri at IBEC, who helped me understand how to manage the molecular simulation data he produced and for his patience in teaching me how all the programs work. I am also thankful to my advisor, Giancarlo Franzese, for introducing me to this fascinating world of studying biology through the lens of physics, allowing me to combine my two passions. Finally, I want to thank the IBEC for welcoming me during this period and allowing me to use their infrastructure.

[1] A. P. Lillis, I. Mikhailenko, D. K. Strickland, Beyond endocytosis: LRP function in cell migration, proliferation and vascular permeability, J Thromb Haemost. 3, 1884 (2005). DOI:10.1111/j.1538-7836.2005.01371.x

[2] A. P. Lillis, L B. Van Duyn, J. E. Murphy-Ullrich, D. K. Strickland, LDL Receptor-Related Protein 1: Unique Tissue-Specific Functions Revealed by Selective Gene Knockout Studies, Physiol. Rev. 88, 899 (2008). DOI: 10.1152/physrev.00033.2007

[3] G. M. Tuveri, S. Acosta Gutiérez, G. Franzese, L. Ruiz Pérez, G. Battaglia, Computational reconstruction of the LDL-receptor-related protein 1 (LRP1) atomistic structure evolution from supertertiary to quaternary, Byophys. J., 123 (2024). DOI:https://doi.org/10.1016/j.bpj.2023.11.1327

[4] J. Kyte, and R. F. Doolittle A simple method for displaying the hydropathic character of a protein, J Mol Biol, 157, 105 (1982). DOI: 10.1016/0022-2836(82)90515-0

[5] F. H. Stillinger and T. Head-Gordon, Perturbational view of inherent structures in water. Phys. Rev. E **47**, 4 (1993).

[6] G. Franzese, G. Malescio, A. Skibinsky, S. V. Buldyrev, and H. E. Stanley, Metastable liquid-liquid phase transition in a single-component system with only one crystal phase and no density anomaly. Phys. Rev. E, **66**, 051206 (2002).

VI. APPENDIX

A. Categorization Hydration Behaviors

1. Radial Distrubution Function

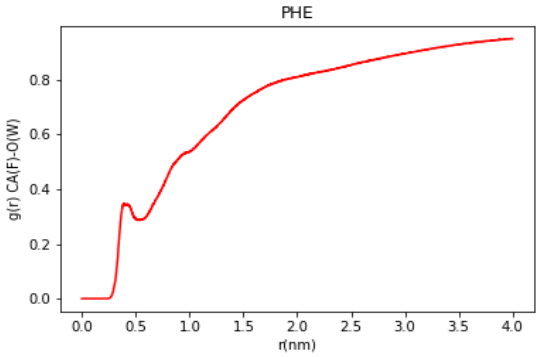

FIG. 6: RDF(g(r)) of the different characteristic hydration behaviors of Phenylalanine (PHE) in LRP1.

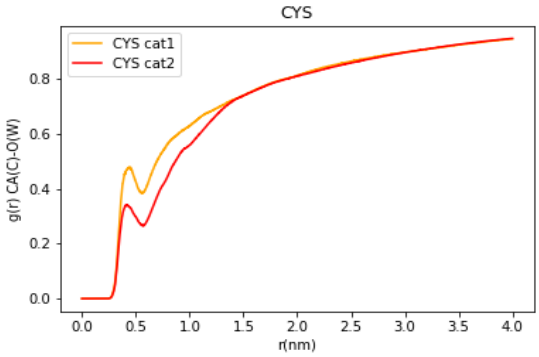

FIG. 7: RDF(g(r)) of the characteristic hydration behavior of Cysteine (CYS) in LRP1.

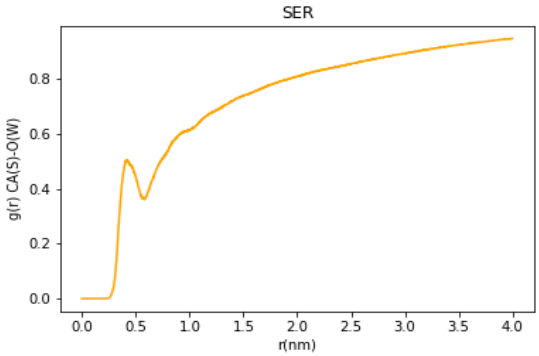

FIG. 8: RDF(g(r)) of the characteristic hydration behavior of Serine (SER) in LRP1.

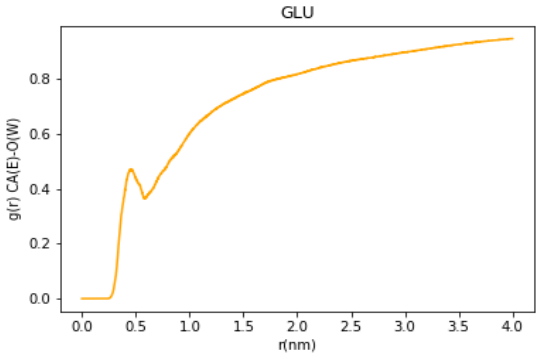

FIG. 9: RDF(g(r)) of the different characteristic hydration behaviors of Glutamic Acid (GLU) in LRP1.

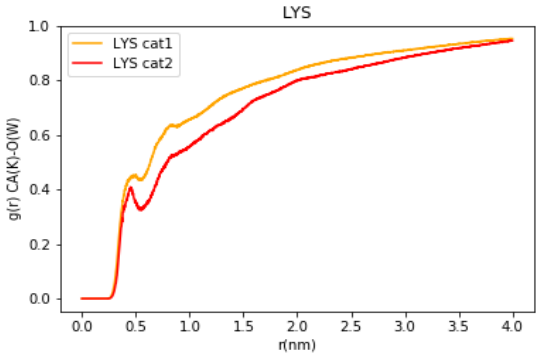

FIG. 10: RDF(g(r)) of the different characteristic hydration behaviors of Lysine (LYS) in LRP1.

2. Effective interaction potentials

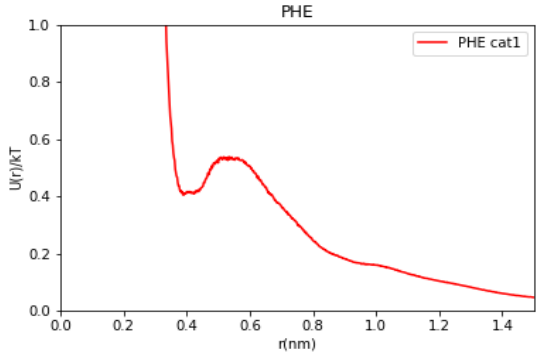

FIG. 11: effective interaction potential (U/kT) of the different characteristic behaviors of Phenyalanine (PHE) in LRP1.

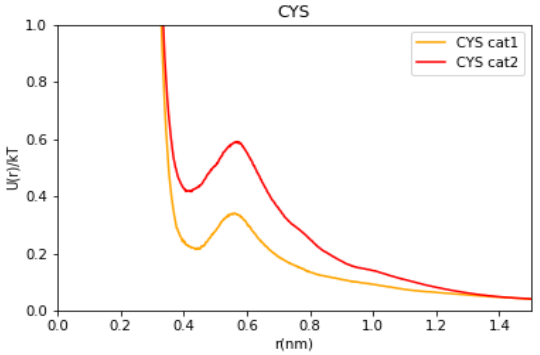

FIG. 12: effective interaction potential (U/kT) of the characteristic behavior of Cysteine (CYS) in LRP1.

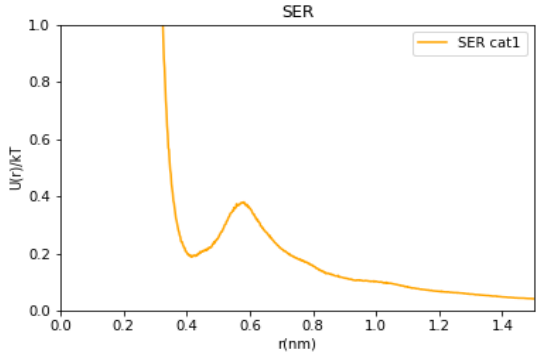

FIG. 13: effective interaction potential (U/kT) of the characteristic behavior of Serine (SER) in LRP1.

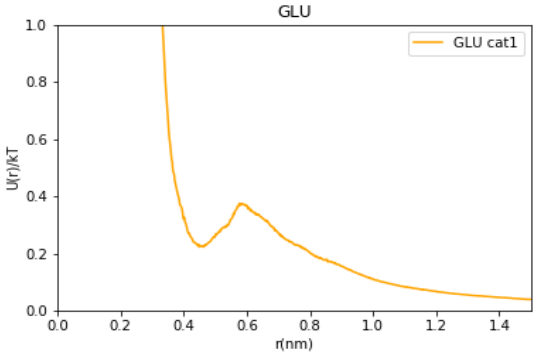

FIG. 14: effective interaction potential (U/kT) of the different characteristic behaviors of Glutamic Acid (GLU) in LRP1.

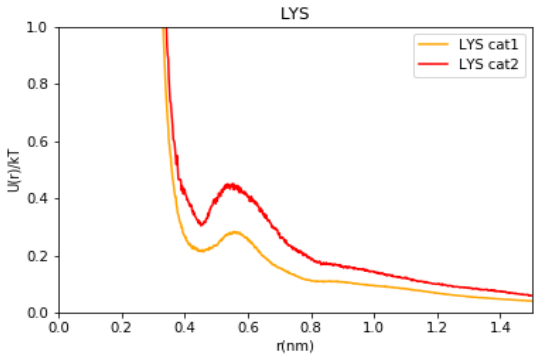

FIG. 15: effective interaction potential (U/kT) of the different characteristic behaviors of Lysine (LYS) in LRP1.