Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces

Johannes Kraml,[†] Anna S. Kamenik,[†] Franz Waibl,[†] Michael Schauperl,[‡] and Klaus R. Liedl*,[†]

[†]Institute of General, Inorganic and Theoretical Chemistry and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innsbruck 6020, Austria

[‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92039-0736, United States

Ⓢ *Supporting Information*

ABSTRACT: Solvation and hydrophobicity play a key role in a variety of biological mechanisms. In substrate binding, but also in structure-based drug design, the thermodynamic properties of water molecules surrounding a given protein are of high interest. One of the main algorithms devised in recent years to quantify thermodynamic properties of water is the grid inhomogeneous solvation theory (GIST), which calculates these features on a grid surrounding the protein. Despite the inherent advantages of GIST, the computational demand is a major drawback, as calculations for larger systems can take days or even weeks. Here, we present a GPU accelerated version of the GIST algorithm, which facilitates efficient estimates of solvation free energy even of large biomolecular interfaces. Furthermore, we show that GIST can be used as a reliable tool to evaluate protein surface hydrophobicity. We apply the approach on a set of nine different proteases calculating localized solvation free energies on the surface of the binding interfaces as a measure of their hydrophobicity. We find a compelling agreement with the hydrophobicity of their substrates, i.e., peptides, binding into the binding cleft, and thus our approach provides a reliable description of hydrophobicity characteristics of these biological interfaces.
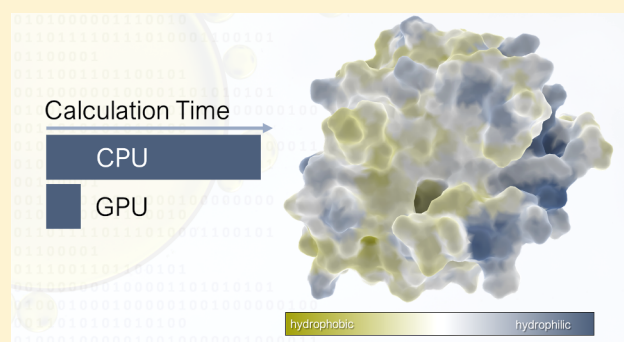
INTRODUCTION

The hydrophobic effect is a major driving force for chemical interactions in aqueous solution and thus a heavily studied phenomenon.[1−4] However, it is discussed to be the least understood intermolecular interaction,[5] which is particularly true in the complex environment of biomolecules.[6,7] In spite of the conceptual challenges, a detailed profile of the hydrophobic surface area is critical in state-of-the-art drug design.[8,9] Hydrophobic areas, where water is easily replaced and released into the bulk, indicate possible interaction sites at the protein surface. On the one hand, this information can be exploited to optimize binding properties of small molecules and biopharmaceuticals alike.[10−12] On the other hand, surface hydrophobicity is believed to be linked to protein aggregation, a crucial factor in the developability of biopharmaceutical products.[13,14] Hence, there is an acute demand for methods to estimate the hydrophobicity of biomolecular interfaces.

A broad range of concepts and algorithms have already been devised to calculate the surface hydrophobicity of proteins. Most common approaches describe hydrophobicity residue-wise, relying on an average hydrophobicity for each amino acid, which is extracted from premeasured or precalculated hydrophobicity scales.[15−20] These traditional hydrophobicity scales found a broad spectrum of applications, ranging from purely sequence-based secondary structure prediction tools to identification of hydrophobic patches on the protein surfaces.[13−15,21−23] More fine-grained scales have been developed, where distinctions are made between different atom types in the amino acids.[8,24] Despite their applicability for a broad range of problems, hydrophobicity scales share a common limitation—they only consider the average hydrophobicity of each building block independently. Yet, in the context of a protein, the solvent interactions of each residue strongly depend on its surrounding, i.e., on nonadditive effects. Furthermore, it is interesting to study hydrophobicity in terms of entropic and enthalpic contributions to gain insight into the underlying mechanisms of hydrophobicity and protein hydration. However, none of the aforementioned methods can quantify the entropic contribution to hydrophobicity, or hydration, directly. Free energy perturbation and thermodynamic integration are often used to calculate the hydration free energy of various molecules.[25,26] But, to calculate the entropic contribution to hydration these models use either the temperature dependence of the free energy or the difference of the free energy and enthalpy.[27,28]

As hydration is a key feature for a variety of different processes, such as biomolecular recognition, a plethora of methods has been developed in recent years to model and analyze water molecules and interactions thereof (AcquaAlta,[29] AQUARIUS2,[30,31] Biki Hydra,[32] Consolv,[33] Dowser++,[34] GRID,[35−37] HINT toolkit,[38] MCSS,[39−41] PyWATER,[42] SPAM,[43] SZMAP,[44] WatCH,[45] WATCLUST,[46] WaterDock,[47] WaterFLAP,[48] WaterScore,[49] WATGEN,[50] wPMF[51]). Methods like AQUARIUS2, MCSS, PyWATER, or WatCH were developed to suggest the placement of water molecules, which can be extremely useful either for further molecular dynamics simulations or molecular docking studies. Another group of methods, such as GRID, WaterDock, or SPAM, calculate the interaction energies of water molecules around a protein. But, most of these cannot calculate the entropic contributions directly or can only analyze certain parts of the protein. Another approach to estimate the solvation free energy is presented by methods based on the Inhomogeneous Solvation Theory (IST). Multiple implementations exist, some of which only analyze the free energy of water on specific, high density water sites, i.e., solvation thermodynamic of ordered water (STOW)[52] or WaterMap.[53,54] Grid inhomogeneous solvation theory (GIST),[12,55−57] on the other hand, calculates the thermodynamic properties of the water molecules for each voxel on a three-dimensional grid. A highly versatile alternative to GIST is the 3D reference site interaction model (3D-RISM).[58,59] This approach relies on the Ornstein−Zernike equation to calculate the enthalpy and entropy directly. However, due to the estimations necessary to calculate the closure relationship, the values obtained by the 3D-RISM theory are approximate. In a recent paper, Nguyen et al.[60] analyzed the differences of GIST and 3D-RISM and have shown that, in general, 3D-RISM yields lower enthalpic and entropic penalties for moving a water molecule from the bulk to the protein's surface. This leads to lower free energies than are calculated by GIST. Another method similar to GIST is the grid cell theory (GCT),[61] in which the thermodynamic values of the water molecules are calculated on a grid according to the cell theory. However, compared to GIST this approach is extremely computationally expensive.

We recently suggested an approach based on GIST to calculate hydrophobicity as solvation free energy.[62] The strengths of this approach include the possibility to dissect solvation free energy into its contributions—enthalpy and entropy. In this method, the individual contributions to the free energy of water surrounding a solute are analyzed on a grid. This allows a detailed analysis of the different contributions to the solvation free energy. Furthermore, GIST allows areas contributing to the solvation free energy of the solute to be identified, therefore providing a method to estimate local contributions to the ensemble property. However, a major drawback we encountered is the extensive computational demand of the approach, especially when it comes to large biomolecular systems. In recent years, the GPU has been harnessed more and more to accelerate simulations and analyses thereof. Here, we reimplement the GIST algorithm to make use of the GPU's calculation power. Thereby, the approach becomes feasible for the analysis of extensive biomolecular interfaces.

We apply our algorithm to characterize the hydrophobicity of the biomolecular interfaces of nine Chymotrypsin-like serine proteases. Proteases in general accelerate the hydrolysis of peptide bonds, via nucleophilic attack on the otherwise kinetically stable amide group.[63] Due to the versatility of this reaction, proteases fulfill a broad range of functions in the human body.[64] Thus, they are highly appealing drug targets and a detailed understanding of their substrate recognition mechanism is paramount.[65,66] For numerous proteases, their substrate preference has been extremely well characterized by massive amounts of cleavage data accumulated in the public MEROPS database.[67] A key feature of protease binding sites captured by the MEROPS is that binding data can be localized to so-called subpockets.[68,69] The concept of subpockets goes back to Schechter and Berger,[70] who suggested to number the binding interface of proteases with S4−S4′ and to do the same on the substrate side, where the substrate positions are named P4−P4′. This partitioning of the binding interface facilitates a localized analysis of the contributions to biomolecular recognition.

Due to their role in various diseases, the Chymotrypsin-like serine protease family has been extensively studied experimentally as well as computationally.[69,71,72] In one of our recent studies, we exploited the experimental data on local substrate preferences to show that similarity of Chymotrypsin-like proteases in terms of their electrostatic substrate recognition correlates with the similarity of their electrostatic molecular interaction fields.[73] This finding suggests that electrostatic recognition and shape-based recognition are orthogonal aspects of substrate binding. Our results so far are thus well in-line with a two-step mechanism of protein−protein binding, where in a first step, an initial preordering of the substrate takes place, which relies on long-range electrostatic interactions between the binding partners.[74−76] To allow for the second step, where the formation of the complex structure is partly driven by shape complementarity, water in the binding interface needs to be displaced. Displacement of energetically unfavorable water molecules in the binding interface contributes beneficially to the binding process. This effect is strongest when both the protease binding interface and substrate are strongly hydrophobic, as water molecules are then more easily stripped from both surfaces. We thus surmise that, especially during this last step of protein−protein complex formation, the hydrophobicity of the binding interface plays a key role.

In this study, we implement and employ the GIST algorithm on the GPU, for an accelerated estimation of solvation free energy from molecular dynamics simulations. Focusing on a set of nine Chymotrypsin-like serine proteases, we use the GIST results to characterize and localize the hydrophobicity in the respective binding interfaces. Furthermore, we compare this analysis to the proteases' substrate data, by introducing a straightforward methodology to calculate the average hydrophobicity of the substrate binding to each of the proteases' subpockets. Combining both analyses we then show that the substrate hydrophobicity correlates well with the local solvation patterns of the binding site.

■ METHODS

Grid Inhomogeneous Solvation Theory. In this section, a brief overview of the grid inhomogeneous solvation theory, devised by Gilson and co-workers, will be given. For a more detailed introduction, the reader is referred to their original studies.[12,55−57] GIST uses a grid to replace the spatial integrals from IST with discrete sums over the voxels on the constructed grid, and only the discrete case will be explained here. For a more general discussion of IST itself, we again refer the reader

to the original publication.[77] In GIST, the contributions to a localized, density-weighted free energy of solvation is split into energy and entropy contributions according to eq 1. The energy contribution is then split up further into the solvent–solvent interaction energy and the solute–solvent interaction energy, as shown in eq 2. For the entropy, only the two-body term is considered and split into translational and orientational entropy (eq 3).

$$\Delta A(r_k) = \Delta E_{total}(r_k) - T\Delta S_{uv}^{total}(r_k) \tag{1}$$

$$\Delta E_{total}(r_k) = \Delta E_{vv}(r_k) + \Delta E_{uv}(r_k) \tag{2}$$

$$\Delta S_{uv}^{total}(r_k) = \Delta S_{uv}^{trans}(r_k) + \Delta S_{uv}^{orient}(r_k) \tag{3}$$

The total equation for the localized free energy of solvation can be derived by inserting eqs 2 and 3 into eq 1. The calculation of E_{vv} and E_{uv} is straightforward, as this is the averaged interaction energy between the solvent molecules in the current voxel with the solute (E_{uv}) or the other solvent molecules (E_{vv}). Algorithmically, these interaction energies directly derive from the applied force field and are simply averaged over all frames of the trajectory. To avoid double counting of the solvent–solvent interactions, this contribution is divided by 2. For the entropic contribution to solvation free energy, solvent–solvent correlations are neglected. The entropy is then calculated via the distribution of the solvent molecules in each voxel. This distribution is a function of the position of the solvent molecule in three-dimensional space and of the orientation of the solvent molecule. This can be partitioned into two different functions, one considering only the position of the solvent molecule, which is calculated as the translational entropy, and a second one depending on the orientation of the solvent molecule in a certain grid voxel, which is the orientational entropy. The translational entropy is then calculated as a Shannon entropy of the distribution of the position of the solvent molecules, as can be seen in eq 4.

$$\Delta S_{uv}^{trans}(r_k) = -k_B\rho^0 V_k g(r_k) \ln g(r_k) \tag{4}$$

where k_B is Boltzmann's constant, ρ^0 is the reference density of the solvent model, V_k is the volume of a voxel, and $g(r_k)$ is the positional distribution of the solvent molecule.

Similarly, the orientational entropy is also calculated as a Shannon entropy of the orientational distribution of the solvent molecules at a certain position, following eq 5.

$$\Delta S_{uv}^{orient}(r_k) = \rho^0 V_k g(r_k) S^\omega(r_k)$$
$$S^\omega = -\frac{k_B}{V_k 8\pi^2} \int_k dr \int g(\omega|r) \ln g(\omega|r) \, d\omega \tag{5}$$

As proposed by the original GIST implementation,[57] we use a nearest neighbor method[78] (eqs 6 and 7) to compute the orientational and translational entropy for voxel k, from the simulation data.

$$S_k^{trans} = R\left(\gamma + \frac{1}{N_k}\sum_{i=1}^{N_k} \ln \frac{N_f \rho^0 4\pi \cdot d_{trans}^3}{3}\right) \tag{6}$$

$$S_k^{orient} = R\left(\gamma + \frac{1}{N_k}\sum_{i=1}^{N_k} \ln \frac{N_k(\Delta\omega_i)^3}{6\pi}\right) \tag{7}$$

where N_k is the total number of solvent molecules in voxel k and γ is Euler's constant that corrects for an asymptotic bias,

d_{trans} is the closest distance between two solvent molecules, $\Delta\omega_i$ is the closest angular distance between two solvent molecules, and N_f is the total number of frames. Algorithmically, d_{trans} is calculated as the Euclidean distance between two solvent molecules. $\Delta\omega_i$ is calculated as the distance between two quaternions, which represent the orientation of each solvent molecule in space (eq 8).[79,80]

$$\Delta\omega_i = 2\cdot\arccos(|\langle q_1, q_2\rangle|) \tag{8}$$

where |.| denotes the absolute value and $\langle q_1, q_2\rangle$ denotes the dot product between the two quaternions.

Both contributions to the entropy can also be estimated together via the same nearest neighbor method, estimating the integral of sixth order.[57]

For all discussed metrics, the reference state is a pure solvent box without solute. Since the solvent–solvent correlations are neglected for the entropic contributions, the reference value for both entropic contributions is zero, as there is no preferred position or orientation resulting in a uniform distribution. Furthermore, the reference for the solute–solvent interaction energy is also zero, with only the solvent–solvent interaction energy differing from zero. In the case of TIP3P water this reference energy is −9.533 kcal/mol for the water–water interaction energy per water molecule.

GIST Implementation. The GPU-accelerated GIST algorithm was implemented using C++ and CUDA. Cpptraj,[57,81] the analysis tool chain included in the AMBER software package, served as backbone for the code. The presented GPU-accelerated GIST algorithm will be provided to be included to the official version of cpptraj. The calculation of the interaction energies was reimplemented on the GPU, since this is the most time-consuming part. The calculation of the entropic contributions is still performed on the CPU, as this calculation is very fast in comparison. The GPU implementation of GIST was analyzed with respect to speedup by performing GIST calculations on pure truncated octahedron water boxes of various sizes. The entire GIST calculation was considered for the analysis of the speedup. These water boxes ranged from a minimum wall distance from the origin of 20 Å to 50 Å, which we will denote as a box radius for readability purposes. For each water box 5000 frames were analyzed, and the grid was created in a way to cover the entire simulation box, using a grid size of 125 Å × 125 Å × 125 Å and a grid spacing of 0.5 Å, resulting in a total number of 250 × 250 × 250 or 15,625,000 grid points.

Simulation Setup. We performed MD simulations starting from the crystal structures of Chymotrypsin (4CHA),[82] Elastase (1QNJ),[83] Factor VIIa (1KLI),[84] Factor Xa (1C5M),[85] Granzyme B (1IAU),[86] Granzyme M (2ZGH),[87] Kallikrein-1 (1SPJ),[88] Thrombin (4AYY),[89] and Trypsin (1PQ7)[90] for subsequent GIST calculations. The structures were prepared using the MOE software package (2019.01),[91] and ligands were deleted where present. We used the implemented Protonate3D tool[92] to resemble a pH of 7.0. Special attention was devoted to the catalytic histidine, to make sure it was not protonated. A TIP3P[93] water box was added with a minimal wall distance of 12 Å. All simulations were performed using the AMBER simulation engine,[94] with the GPU implementation of the particle mesh Ewald algorithm.[95]

We employed the Langevin thermostat for temperature control of the system and simulated at a temperature of 300 K. Since simulations were performed in the NpT ensemble, the
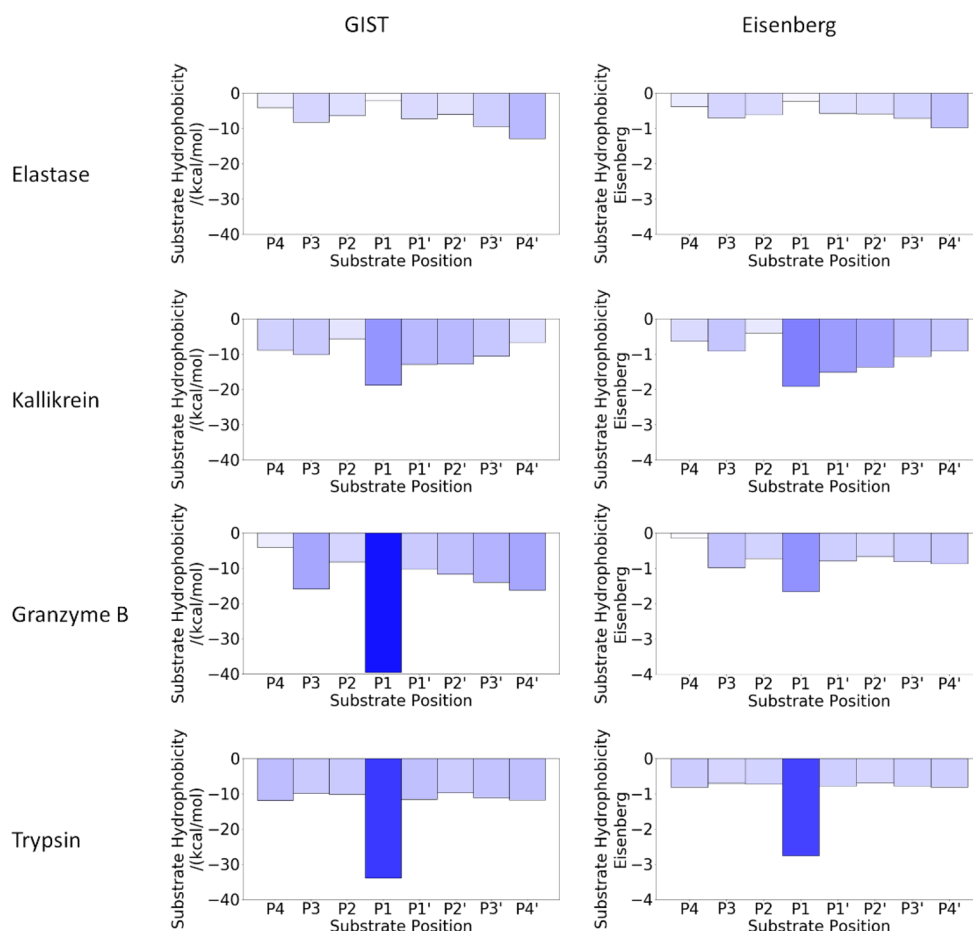
Figure 1. Average hydrophobicity at each substrate position (P4 to P4′). Based on the MEROPS cleavage data, we calculate substrate hydrophobicity using GIST (left) and the Eisenberg scale (right). Substrate positions range from more hydrophobic, i.e., higher score on the hydrophobicity scale (white), to more hydrophilic, i.e., lower score on the hydrophobicity scale (blue).

Berendsen barostat was used with a relaxation time of 2.0 ps. Each structure was simulated for 100 ns, and 5000 frames were used, to ensure convergence. The solute was restrained using a restraint weight of 50 kcal/(mol Å²).

The GIST analysis for the proteases was carried out, using the newly implemented GPU version of the GIST algorithm. A grid of size 100 Å × 100 Å × 100 Å was used for all protease structures, with grid spacing of 0.5 Å, resulting in a grid of size 200 × 200 × 200, or 8×10^6 grid points. The temperature was set to 300 K, and the reference density of water was chosen according to the AMBER manual (TIP3P: 0.0329 molecules/Å³). For the analysis, 5000 frames of the simulation were used. The reference water−water interaction energy was set to the same as stated in the AMBER manual (TIP3P: −9.533 kcal/mol).

Amino Acid Solvation Free Energy. To calculate the solvation free energy of each amino acid in the substrates, the same procedure was followed as published previously.[62] Three conformations of each of the naturally occurring 20 amino acids were considered to calculate their solvation free energy. In a distance of 5 Å, integration of the solvation free energy was performed to calculate the total solvation free energy for each conformation of each amino acid. These values were then averaged, and for histidine, the average of the two possible uncharged tautomeric states was used.

Substrate Hydrophobicity. The substrate hydrophobicity is calculated as the weighted average hydrophobicity at each substrate position, based on the MEROPS cleavage data.

Here, we used the hydrophobicity scale from Eisenberg[15] as a reference to calculate the average hydrophobicity of an amino acid at each substrate position. We normalized the values of that scale by shifting the hydrophobicity value of the most hydrophobic amino acid to zero. This enables an easy comparison between different methods.

Additionally, we calculated the solvation free energy profile of each substrate position using GIST. Also, for this scale, the most hydrophobic amino acid was shifted to have a score of 0. Here, we used the solvation free energy of the different amino acids as a scale, as suggested previously.[62] Then, the hydrophobicity for each substrate position was calculated via eq 9, which is the weighted average of the hydrophobicity score of each individual amino acid. This step was performed for both scales to compare the solvation free energy ansatz with the experimental concept of Eisenberg's scale.

$$H_{GIST}^{P_i} = \frac{\sum_{a \in AA} n_{a,i} \cdot \Delta G_a}{\sum_{a \in AA} n_{a,i}} \tag{9}$$

where P_i is the substrate position i, AA is the set of all 20 naturally occurring amino acids, $n_{a,i}$ is the number of substrates containing amino acid a in substrate position i, and ΔG_a is the solvation free energy of amino acid a. The substrate binding
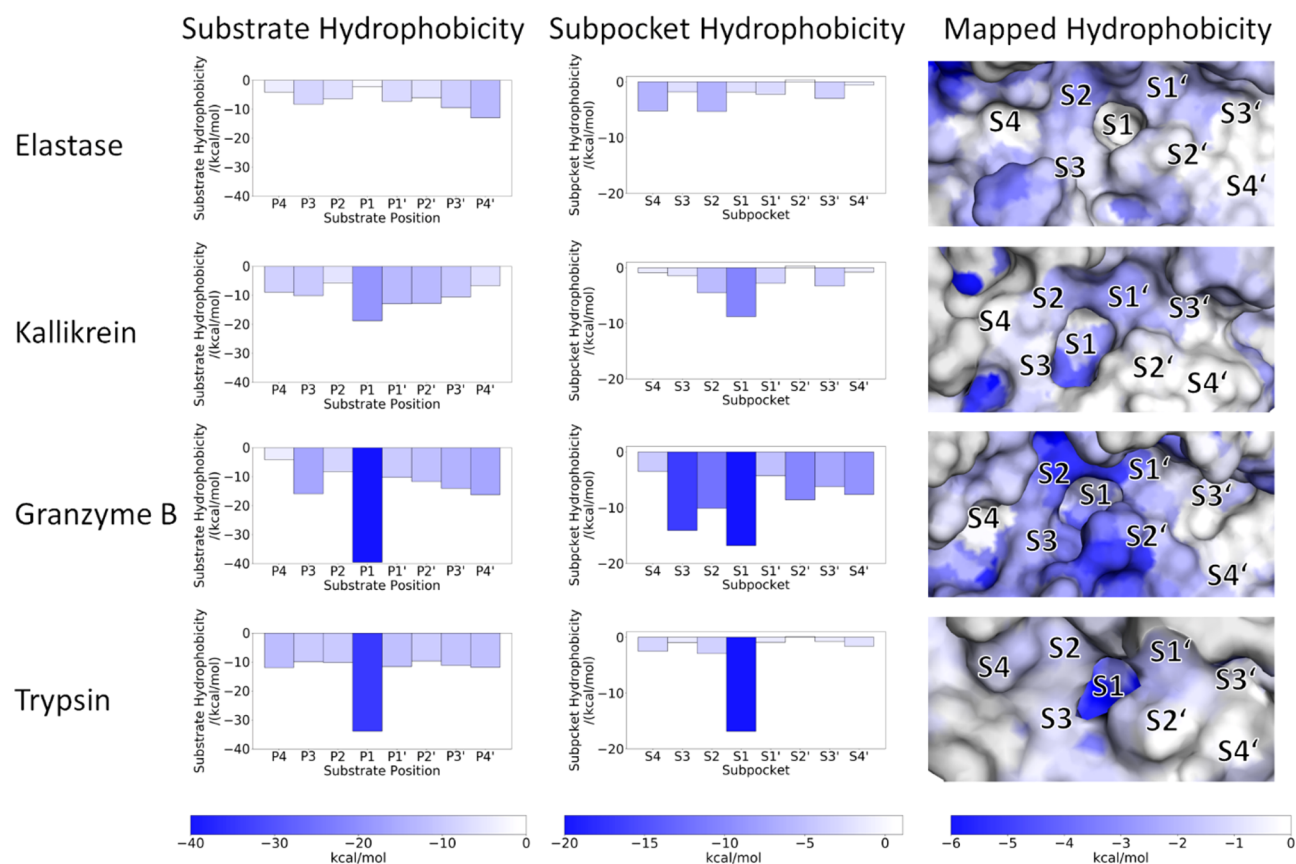
Figure 2. Comparison of the substrate hydrophobicity (left), the localized subpocket hydrophobicity (middle), and the protease binding interface colored by the localized hydrophobicity (right). The surface is colored from white—more hydrophobic, higher solvation free energy (cutoff: 0 kcal/mol)—to blue—more hydrophilic, more negative solvation free energy (−6 kcal/mol). The substrate hydrophobicity from Figure 1 is depicted on the far left, and the localized subpocket hydrophobicity is depicted in the middle. For both, the values range from more hydrophobic (white) to more hydrophilic (blue).

data needed for each protease was gathered from the MEROPS database.

Protein Surface Hydrophobicity Mapping. The surface hydrophobicity mapping projects the values of the solvation free energy around a molecule onto its surface. In this case, the mapping was performed on an atomistic level, but also more coarse- and more fine-grained levels could be used, e.g., considering the entire residue or calculating points on the surface and projecting the values there.

The hydrophobicity was calculated by a density-weighted average around each atom. Additionally, a Gaussian cutoff function with a σ of 1 Å was applied between 4 and 7 Å.

Localized Subpocket Hydrophobicity. The localized subpocket hydrophobicity is a metric to calculate the solvation free energy that would need to be replaced by a ligand binding into a certain pocket of a protease.

The localized subpocket hydrophobicity was calculated, using four different ligands to define the binding interface of each protease. The four different ligands where chosen from different X-ray structures to cover the binding interface of the different proteases from the S4 subpocket to the S4′ subpocket (PDB Codes: 1DE7,[96] 3LU9,[97] 2ZGH,[87] and 3TJV[98]). The ligands were positioned by aligning the protease crystal structures to the simulated structures.

Grid voxels were assigned to each of the amino acids of these ligands to calculate the solvation free energy of the water molecules replaced by the ligands. The grid points were weighted by a Gaussian function with respect to the distance of the atoms of the ligand residue. For the distance in the Gaussian function, the nearest grid point to any side chain atom of any ligand amino acid at the specific substrate position was used. A σ of 1 Å was used to define the Gaussian, and a hard cutoff was defined after 3 σ (3 Å away from any of the ligand's atoms).

■ **RESULTS**

The solvation free energy was calculated for all 20 naturally occurring amino acids as described in the Methods section. The resulting values for these calculations are reported in the SI (Table S1).

We then calculated the hydrophobicity in substrate space via eq 9. For each substrate position, a value was assigned according to the average hydrophobicity of an amino acid in the substrate at that position. In this metric, small values (less negative) imply that the pocket is prone to bind hydrophobic residues. Large values, i.e., favorable solvation free energies, indicate that mostly hydrophilic residues were determined to bind in this pocket. As a reference for our GIST-based approach, we calculated the same values based on the widely established Eisenberg scale. This scale has been created as a consensus scale of five different hydrophobicity scales. The scales considered to generate this consensus scale ranged from scales based on the free energies for the transition from water to ethanol[99,100] and for the transfer from water to vapor,[101] as

well as two scales based on calculated free energies for the transfer of amino acid side chains from the surface into the interior of a protein[102,103] and a last semiempirical scale.[104] Figure 1 compares the substrate hydrophobicity from GIST and the Eisenberg scale for four representative proteases. The results for the remaining five Chymotrypsin-like serine proteases can be found in the SI (Figure S1). Comparing the two metrices for all studied proteases and subpockets, we find a striking Pearson correlation of 0.88.

Furthermore, we used GIST to quantify and localize the hydrophobicity within the binding site of each studied protease. To illustrate the resulting free energy grid, we mapped it onto the protein surfaces. Figure 2 compares localized hydrophobicity within the protease binding interface with the respective substrate hydrophobicity. The same analysis was carried out for the remaining five proteases and can be found in the Supporting Information (Figure S2).

To quantify the local hydrophobicity in each subpocket, we calculated the localized subpocket hydrophobicity using four different ligands, as described in the Methods section. This approach allowed us to calculate the correlation between the localized subpocket hydrophobicity and the substrate hydrophobicity. Including all nine proteases in our set, we find a Pearson correlation coefficient of 0.70.

For a more qualitative visualization we also mapped the values for the hydrophobicity onto the surface, as mentioned in the Methods section, and for visualization we used PyMOL.[105] Here again, a good agreement between the substrate hydrophobicity and our projection becomes apparent. In this way, Figure 2 intuitively reflects that the studied proteases are not only distinct in their substrate preferences but also in their respective surface hydrophobicities.

Elastase, for example, is known to bind a broad range of hydrophobic substrates. First, the substrate hydrophobicity of Elastase is significantly higher than for the remaining proteases, with only minor signals in the hydrophilic range. Second, also the solvation free energy within the binding site of Elastase clearly points toward an overall hydrophobic site. In particular, its S1 subpocket is considerably more hydrophobic than in other proteases, but also the outer subpockets of Elastase show barely any indication of hydrophilicity.

Kallikrein is a protease that recognizes a higher variety of amino acids at the P1 substrate position, is more hydrophilic than, for example, Elastase, but is rather hydrophobic compared to the strongly hydrophilic proteases, e.g., Trypsin. Using the described metrics, we can clearly identify Kallikrein as amphiphilic in the S1 subpocket. This is interesting as also the surface mapping shows different parts in the S1 subpocket, where a very hydrophilic and a more hydrophobic part can be seen.

Granzyme B shows a fairly different pattern as it accepts almost exclusively hydrophilic substrates at the S1 position, recognizing primarily aspartate and glutamate. This is well in line with the surface mapping of solvation free energy, where we find a pronounced hydrophilic signal at that position. Interestingly, Granzyme B is rather hydrophobic at the P4 substrate position, which also agrees well with the binding site's solvation free energy at the S4 subpocket.

Trypsin, compared to the other proteases, is strongly hydrophilic at the P1 substrate position but rather hydrophobic in the remaining substrate positions. We observe the same trend in the surface mapping, where the subpocket S1 is depicted as extraordinarily hydrophilic while the other

subpockets are much more hydrophobic. Also, our localized substrate hydrophobicity metric clearly quantifies that the S1 subpocket is significantly more hydrophilic than the other subpockets.

GPU Implementation. We tested the efficiency of our parallelization efforts on water boxes of increasing size and on the example of Chymotrypsin. The computational time of the calculations for the different systems is shown in Figure 3. For
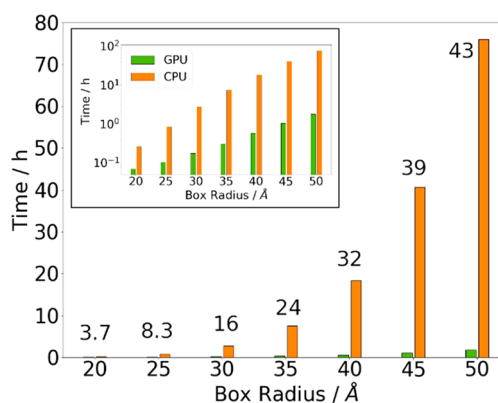

Figure 3. Calculation time of the GIST calculation on the GPU (green) compared to the original implementation on the CPU (orange) and the respective speedup, for the entire analysis. The same is shown as the inlet, using a logarithmic scale for the time. Calculations were performed on the same machine, containing an AMD Ryzen 7 1800X CPU and a NVIDIA GeForce GTX 1080 GPU. GIST calculations were performed on a truncated octahedron water box, with increasing radius.

the water boxes an almost linear increase in the acceleration can be seen even to the 50 Å radius. When looking at the speedup with respect to the number of atoms, we still observe a significant acceleration up to our biggest system with almost 90 000 atoms. For the considered systems, we achieved four times faster calculations on the smaller systems, which increase to even 45 times faster calculations for larger systems. So far, these trends are not really showing indications of leveling off. We did not test the speedup for systems larger than 50 Å, because of the substantial computational demand of the CPU implemented version of GIST. All benchmarks were carried out using a machine consisting of an AMD Ryzen 7 1800X and a NVIDIA GeForce GTX 1080. The exact values are given in the SI (Tables S2 and S3) both for the speedup of the entire cpptraj run and for only the parallelized part of the code.

■ DISCUSSION

We use the physics-based method GIST to profile the hydrophobicity of protein−protein interaction sites. We test the reliability of the approach on the thoroughly studied family of Chymotrypsin-like serine proteases. The virtues of this set of model systems include structurally well-characterized binding sites as well as compelling statistics on their substrate preferences. Furthermore, protease binding interfaces are designed to allow significant differences in their local physicochemical properties.

To evaluate the binding preferences of these nine different proteases, we performed a simple analysis of binding data found in the MEROPS database. We benchmark the plausibility of the GIST-based results against the well-established Eisenberg hydrophobicity scale. Reassuringly, we

find that both methods perform very similar in respect to the calculation of substrate hydrophobicity, despite major differences in their underlying concepts. This consistency between the two metrics is further highlighted by a Pearson correlation coefficient of 0.88 over all substrate positions in all nine proteases.

Interestingly, the scale devised by Eisenberg rates the hydrophilic substrate positions less strongly than our GIST-based approach. This trend probably originates from the generally broader spread of the data calculated via GIST. The values calculated by GIST are more negative for the hydrophilic residues than is the case in the Eisenberg scale. Additionally, the GIST calculation typically predicts a lower hydrophobicity for the negatively charged amino acids than other hydrophobicity scales. This effect has been observed before[62] and is again clearly visible in Figure 1, when comparing substrate hydrophilicities of Granzyme B, which mainly binds aspartate and glutamate in the S1. In the GIST calculation, the P1 substrate position of Granzyme B is depicted as strongly hydrophilic. However, using the Eisenberg scale, this position is substantially more hydrophobic than in the GIST approach. This discrepancy stems from the fact that the GIST based method ranks the two negatively charged amino acids, i.e., aspartate and glutamate, more hydrophilic than the Eisenberg scale. In the Eisenberg scale these amino acids are considered even less hydrophilic than some of the uncharged amino acids. However, in our scale, these amino acids are ranked at the hydrophilic end of the scale, with aspartate and glutamate being the most hydrophilic amino acids.

In addition to the hydrophobicity of each substrate position we use GIST to establish a localized quantification of hydrophobicity within the protease binding sites. Already visually, the qualitative agreement between hydrophobicity mapped onto the protease surface and the substrate hydrophobicity is evident. We further calculate the localized subpocket hydrophobicity to quantitively compare substrate and binding site hydrophobicity. The resulting Pearson correlation coefficient of 0.70 indicates that a hydrophilic subpocket prefers hydrophilic substrates. This is quite intuitive, as the favorable interactions between the water and the hydrophilic surface need to be replaced by other favorable interactions, e.g., with hydrophilic residues. Hydrophobic surfaces, on the other hand, are drawn to each other as thereby unfavorable interactions of water molecules are then removed and replaced. Water molecules can be released into the bulk, where it has more favorable interactions than with the biomolecule. Furthermore, a high gain in free energy can be achieved when releasing fixed, i.e., entropically unfavorable, water molecules into the bulk, where these molecules can move freely. Thus, a high gain in entropy can compensate for a smaller loss in enthalpy. This is also understood as enthalpy–entropy compensation, where a favorable enthalpic interaction comes at the price of an unfavorable entropic interaction.[106,107]

The projection of the hydrophobicity onto the surface shows quite interesting features on the protease binding sites. A generally reassuring observation we made is that more hydrophilic substrates correspond to more hydrophilic subpockets. This can be visualized for the two proteases that accept primarily charged amino acids, i.e., Trypsin and Granzyme B. However, for Elastase, which favors hydrophobic amino acids, all subpockets are hydrophobic. In this protease, the S1 subpocket seems to be the most hydrophobic subpocket

of the entire set. Kallikrein binds both hydrophilic and hydrophobic residues, and strikingly its S1 subpocket is split into two distinct sections, which could explain its amphipathic binding behavior.

Interestingly, such a partitioning into different areas is also found for proteases favoring solely hydrophilic amino acids in the S1. In most of the cases where a protease favors a positively charged amino acid in the S1 subpocket, this subpocket is hydrophilic at the bottom but is rather hydrophobic around its edges. At first glance, this seems a little counterintuitive, but the positively charged amino acids lysine and arginine consist of a charged group on a long hydrophobic chain. The hydrophobic area at the S1 pocket outlines might thus facilitate replacing water molecules at these positions. This hypothesis is further strengthened by the surface hydrophobicity of Granzyme B's S1 subpocket. Neither residue favored, i.e., aspartate and glutamate, contains a long hydrophobic chain but is primarily characterized by its negatively charged group.

A coloring based solely on the existing hydrophobicity scales is not able to capture the differences in hydrophobicity present in the different areas of the binding interface. When using the Eisenberg scale to analyze the binding interfaces in terms of hydrophobicity, the projected values in most cases do not fit to the values calculated from the substrate distribution (Figure S3). This is not surprising, as this method uses predefined values that cannot consider the surroundings of the different amino acids, as discussed previously. Interestingly, for many different proteases, the Eisenberg scale actually suggests the inverse trend, for example, Chymotrypsin, which is very hydrophobic in the S1 and is colored as slightly hydrophilic. However, Thrombin and Factor Xa, which both heavily favor hydrophilic substrates in the S1 pocket, are colored as almost completely hydrophobic. But both are considered hydrophilic in our presented method.

This finding emphasizes one of the most intriguing qualities of the presented metric, which is that it captures nonadditive effects to hydrophobicity, and it can also analyze cooperative contributions to hydrophobicity. Cooperative effects might play a significant role in different aspects of hydrophobicity. A large amount of hydrophobic amino acids, building a cavity, induce order on the encapsulated water molecules, thus leading to an unfavorable entropic contribution. This fixed water then induces order to the other water molecules in its surrounding. This can even go so far that the encapsulated water molecules are forced into a fixed position due to the arrangement of the other water molecules in that cavity. Most prominently, this has been shown to be the case in the streptavidin system by Young et al.,[53] where a network of water molecules is constrained into a similar pattern as the ligand, meaning that the hydration of amino acids is quite different in the complex environment of a protein compared to the free amino acids. When considering only the free amino acids, it is not possible to capture these cooperative effects of the solvent molecules. However, the GIST-based approach allows for a localized analysis of hydrophobicity while considering the surrounding of each residue. This facilitates a reliable and in-depth analysis of the protein surface.

Another quite interesting feature of this metric is the potential to split the hydrophobicity into enthalpic and entropic contributions (Figures S3 and S4). Hydration of proteins is still a field of ongoing research with a lot of controversy about the actual mechanism. We believe that molecular dynamics simulation and the analysis of the

hydration on an atomistic level is of high interest. Our method can directly measure the entropic and enthalpic contributions to hydration around the protein's surface. This is particularly interesting for the entropic contributions, as these are directly calculated from the phase space of the water molecules. Thus, from one simulation all major contributions to hydrophobicity of a particular conformation can be studied separately.

In their recent work, Nguyen et al.[60] compared the thermodynamic quantities generated by GIST with 3D-RISM. In their study, they used Factor Xa, a protein also investigated here. As mentioned previously, they found that 3D-RISM generally underestimates the enthalpic penalty for moving a water molecule out of the bulk onto a surface. The entropic penalty showed the same behavior when compared to GIST. However, both methods use different approximations for the calculation of the entropy and thus no conclusions were drawn in that regard. For this study, their findings suggest that 3D-RISM will generally yield more hydrophilic properties on the surface.

A key aspect of this work was the reimplementation of the GIST algorithm on the GPU. We already briefly discussed the speedup of the GPU implemented GIST calculation, showing a speedup of up to 30 for the protease systems and up to 45 for the simulated water boxes. We believe that the higher amount of speedup for the largest system is an indication that the significant gain in speed still has not started to level off with system size. A few points limit the speedup in smaller systems; first the copying of the data to and from the GPU's memory might be fast but is, for very small systems, also a significant factor slowing down the calculation on the GPU. Furthermore, the limited number of atoms in small systems might not fully utilize the GPU's potential, thus not allowing for very high speedup, compared to the larger systems. For our largest system, we reached a speed of about 1 frame per second, while the CPU implementation of GIST required almost 1 min for each frame, showing the highest speedup in our test systems of nearly 45. This significant acceleration of GIST, coupled with cpptraj, allowed us to analyze our test systems (ca. 30000 atoms) in roughly 10 min, compared to the hours it would have taken otherwise. This is most apparent when looking at the calculation of the largest water box, where the calculation with our GPU implementation took about 1.5 h. However, when calculating it on the CPU, the identical analysis took over three full days.

We surmise that the solvation free energy as a metric for protein surface hydrophobicity will be particularly valuable for drug discovery and optimization. The GPU implementation facilitates a fast detection of trapped water molecules on the surface, which are easily replaced and thus increase binding affinity of putative ligands. Furthermore, surface hydrophobicity is discussed as a determinant for protein aggregation. The presented metric is fast enough to reliably identify hydrophobic areas on extensive biomolecular surfaces and thus estimate the developability of biopharmaceuticals.

CONCLUSION

Hydrophobicity scales have been around for over 50 years now and are still an integral part in an enormous amount of theoretical methods. They are of crucial importance for analyzing membrane-bound proteins and in various other fields. The major drawback of these methods is that hydrophobicity is considered residue-wise, neglecting all cooperative effects, which is particularly problematic in the complex environment of biomolecules. The presented method is based on the solvation free energy of a solute and inherently includes cooperative effects for protein surfaces.

We show that the approach can localize hydrophobicity in the binding interfaces of nine Chymotrypsin-like serine proteases. The resulting solvation patterns are well in line with the respective substrate preferences deposited in the MEROPS database. Thus, our physics-based approach to analyze hydrophobicity is able to reliably characterize the binding interface of the studied proteases.

Reimplementation of the GIST algorithm on the GPU led to a significant speedup. For our tested systems, we reached a maximum speedup of almost 45 times faster than calculation on the CPU. With the presented GPU implementation of GIST, solvation free energy analysis of large biomolecular systems at reasonable calculation times is now publicly accessible.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00742.

> The GPU accelerated code of GIST is available on https://github.com/liedllab/gigist. Detailed Information on the speed of the GPU implementation of GIST, the data for the remaining five Chymotrypsin-like proteases, and the contributions, i.e., enthalpy and entropy, to the hydrophobicity (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*(K.R.L.) E-mail: klaus.liedl@uibk.ac.at.
ORCID ⓘ
Michael Schauperl: 0000-0001-5648-8170
Klaus R. Liedl: 0000-0002-0985-2299

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

■ ABBREVIATIONS

GIST, grid inhomogeneous solvation theory; IST, inhomogeneous solvation theory; STOW, solvation thermodynamic of ordered water; 3D-RISM, 3D reference interaction site model; GCT, grid cell theory

■ REFERENCES

(1) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437* (7059), 640−647.

(2) Young, L.; Jernigan, R. L.; Covell, D. G. A Role for Surface Hydrophobicity in Protein-Protein Recognition. *Protein Sci.* **1994**, *3* (5), 717−729.

(3) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116* (13), 7673−7697.

(4) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* **1945**, *13* (11), 507−532.

(5) Biedermann, F.; Nau, W. M.; Schneider, H.-J. The Hydrophobic Effect Revisited—Studies with Supramolecular Complexes Imply High-Energy Water as a Noncovalent Driving Force. *Angew. Chem., Int. Ed.* **2014**, *53* (42), 11158−11171.

(6) Bagchi, B. Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* **2005**, *105* (9), 3197−3219.

(7) Brooks, C. L.; Karplus, M. Solvent Effects on Protein Motion and Protein Effects on Solvent Motion: Dynamics of the Active Site Region of Lysozyme. *J. Mol. Biol.* **1989**, *208* (1), 159−181.

(8) Lijnzaad, P.; Berendsen, H. J. C.; Argos, P. A Method for Detecting Hydrophobic Patches on Protein Surfaces. *Proteins: Struct., Funct., Genet.* **1996**, *26* (2), 192−203.

(9) Sarkar, A.; Kellogg, G. E. Hydrophobicity–Shake Flasks, Protein Folding and Drug Discovery. *Curr. Top. Med. Chem.* **2010**, *10* (1), 67−83.

(10) Uehara, S.; Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **2016**, *21* (11), 1604.

(11) Balius, T. E.; Fischer, M.; Stein, R. M.; Adler, T. B.; Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T.; Shoichet, B. K. Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (33), E6839−E6846.

(12) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10* (7), 2769−2780.

(13) Liu, D. R.; Phillips, K. J.; Lawrence, M. S. Protein Surface Remodeling. EP3045532A1, 2016.

(14) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Design of Therapeutic Proteins with Enhanced Stability. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (29), 11937.

(15) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* **1984**, *179* (1), 125−142.

(16) Kovacs, J. M.; Mant, C. T.; Hodges, R. S. Determination of Intrinsic Hydrophilicity/Hydrophobicity of Amino Acid Side Chains in Peptides in the Absence of Nearest-Neighbor or Conformational Effects. *Biopolymers* **2006**, *84* (3), 283−297.

(17) Zamora, W. J.; Campanera, J. M.; Luque, F. J. Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations. *J. Phys. Chem. Lett.* **2019**, *10* (4), 883−889.

(18) Black, S. D.; Mould, D. R. Development of Hydrophobicity Parameters to Analyze Proteins which Bear Post- or Cotranslational Modifications. *Anal. Biochem.* **1991**, *193* (1), 72−82.

(19) Mazzé, F. M.; Fuzo, C. A.; Degrève, L. A new Amphipathy Scale: I. Determination of the Scale from Molecular Dynamics Data. *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, *1747* (1), 35−46.

(20) Simm, S.; Einloft, J.; Mirus, O.; Schleiff, E. 50 Years of Amino Acid Hydrophobicity Scales: Revisiting the Capacity for Peptide Classification. *Biol. Res.* **2016**, *49* (1), 31.

(21) MacCallum, J. L.; Tieleman, D. P. Hydrophobicity Scales: A Thermodynamic Looking Glass into Lipid-Protein Interactions. *Trends Biochem. Sci.* **2011**, *36* (12), 653−662.

(22) von Heijne, G. Membrane Protein Structure Prediction: Hydrophobicity Analysis and the Positive-Inside Rule. *J. Mol. Biol.* **1992**, *225* (2), 487−494.

(23) Perunov, N.; England, J. L. Quantitative Theory of Hydrophobic Effect as a Driving Force of Protein Structure. *Protein Sci.* **2014**, *23* (4), 387−399.

(24) Nicolau Jr, D. V.; Paszek, E.; Fulga, F.; Nicolau, D. V. Mapping Hydrophobicity on the Protein Molecular Surface at Atom-Level Resolution. *PLoS One* **2014**, *9* (12), No. e114042.

(25) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. A Molecular Dynamics Study of Thermodynamic and Structural Aspects of the Hydration of Cavities in Proteins. *Biopolymers* **1991**, *31* (8), 919−931.

(26) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509−1519.

(27) Hajari, T.; van der Vegt, N. F. A. Solvation Thermodynamics of Amino Acid Side Chains on a Short Peptide Backbone. *J. Chem. Phys.* **2015**, *142* (14), 144502.

(28) Peter, C.; Oostenbrink, C.; Van Dorp, A.; Van Gunsteren, W. F. Estimating Entropies from Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *120* (6), 2652−2661.

(29) Rossato, G.; Ernst, B.; Vedani, A.; Smieško, M. AcquaAlta: A Directional Approach to the Solvation of Ligand-Protein Complexes. *J. Chem. Inf. Model.* **2011**, *51* (8), 1867−1881.

(30) Pitt, W. R.; Goodfellow, J. M. Modelling of Solvent Positions around Polar Groups in Proteins. *Protein Eng., Des. Sel.* **1991**, *4* (5), 531−537.

(31) Pitt, W. R.; Murray-Rust, J.; Goodfellow, J. M. AQUARIUS2: Knowledge-Based Modeling of Solvent Sites around Proteins. *J. Comput. Chem.* **1993**, *14* (9), 1007−1018.

(32) Zia, S. R.; Gaspari, R.; Decherchi, S.; Rocchia, W. Probing Hydration Patterns in Class-A GPCRs via Biased MD: The A2A Receptor. *J. Chem. Theory Comput.* **2016**, *12* (12), 6049−6061.

(33) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting Conserved Water-Mediated and Polar Ligand Interactions in Proteins using a K-Nearest-Neighbors Genetic Algorithm. *J. Mol. Biol.* **1997**, *265* (4), 445−464.

(34) Morozenko, A.; Stuchebrukhov, A. A. Dowser++, A New Method of Hydrating Protein Structures. *Proteins: Struct., Funct., Genet.* **2016**, *84* (10), 1347−1357.

(35) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849−857.

(36) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups with the Ability to Form Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36* (1), 140−147.

(37) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability to Form More than Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36* (1), 148−156.

(38) Kellogg, G. E.; Fornabaio, M.; Chen, D. L.; Abraham, D. J. New Application Design for a 3D Hydropathic Map-Based Search for Potential Water Molecules Bridging between Protein and Ligand. *Internet Electron J. Mol. Des* **2005**, *4*, 194−209.

(39) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Genet.* **1991**, *11* (1), 29−34.

(40) Bitetti-Putzer, R.; Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Functional Group Placement in Protein Binding Sites: A Comparison of GRID and MCSS. *J. Comput.-Aided Mol. Des.* **2001**, *15* (10), 935−960.

(41) Evensen, E.; Joseph-McCarthy, D.; Weiss, G. A.; Schreiber, S. L.; Karplus, M. Ligand Design by a Combinatorial Approach Based on Modeling and Experiment: Application to HLA-DR4. *J. Comput.-Aided Mol. Des.* **2007**, *21* (7), 395−418.

(42) Patel, H.; Grüning, B. A.; Günther, S.; Merfort, I. PyWATER: A PyMOL Plug-in to Find Conserved Water Molecules in Proteins by Clustering. *Bioinformatics* **2014**, *30* (20), 2978−2980.

(43) Cui, G.; Swails, J. M.; Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J. Chem. Theory Comput.* **2013**, *9* (12), 5539−5549.

(44) Bayden, A. S.; Moustakas, D. T.; Joseph-McCarthy, D.; Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J. Chem. Inf. Model.* **2015**, *55* (8), 1552−1565.

(45) Sanschagrin, P. C.; Kuhn, L. A. Cluster Analysis of Consensus Water Sites in Thrombin and Trypsin Shows Conservation between Serine Proteases and Contributions to Ligand Specificity. *Protein Sci.* **1998**, *7* (10), 2054−2064.

(46) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Marti, M. A.; Turjanski, A. G. WATCLUST: A Tool for Improving the Design of Drugs Based on Protein-Water Interactions. *Bioinformatics* **2015**, *31* (22), 3697−3699.

(47) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS One* **2012**, *7* (3), No. e32036.

(48) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47* (2), 279−294.

(49) García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A Novel Method for Distinguishing between Bound and Displaceable Water Molecules in the Crystal Structure of the Binding Site of Protein-Ligand Complexes. *J. Mol. Model.* **2003**, *9* (3), 172−182.

(50) Bui, H.-H.; Schiewe, A. J.; Haworth, I. S. WATGEN: An Algorithm for Modeling Water Networks at Protein-Protein Interfaces. *J. Comput. Chem.* **2007**, *28* (14), 2241−2251.

(51) Zheng, M.; Li, Y.; Xiong, B.; Jiang, H.; Shen, J. Water PMF for Predicting the Properties of Water Molecules in Protein Binding Site. *J. Comput. Chem.* **2013**, *34* (7), 583−592.

(52) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer New York: New York, NY, 2012; pp 393−404.

(53) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 808.

(54) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817−2831.

(55) Nguyen, C.; Gilson, M. K.; Young, T., Structure and Thermodynamics of Molecular Hydration via Grid Inhomogeneous Solvation Theory. arXiv preprint. arXiv:1108.4876 2011.

(56) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137* (4), 044101.

(57) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37* (21), 2029−2037.

(58) Kovalenko, A.; Hirata, F. Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chem. Phys. Lett.* **1998**, *290* (1), 237−244.

(59) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal-Water Interface by the Kohn-Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **1999**, *110* (20), 10095−10112.

(60) Nguyen, C.; Yamazaki, T.; Kovalenko, A.; Case, D. A.; Gilson, M. K.; Kurtzman, T.; Luchko, T. A Molecular Reconstruction Approach to Site-Based 3D-RISM and Comparison to GIST Hydration Thermodynamic Maps in an Enzyme Active Site. *PLoS One* **2019**, *14* (7), No. e0219473.

(61) Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J.; Michel, J. Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *J. Chem. Theory Comput.* **2014**, *10* (1), 35−48.

(62) Schauperl, M.; Podewitz, M.; Waldner, B. J.; Liedl, K. R. Enthalpic and Entropic Contributions to Hydrophobicity. *J. Chem. Theory Comput.* **2016**, *12* (9), 4600−4610.

(63) Radzicka, A.; Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.* **1996**, *118* (26), 6105−6109.

(64) López-Otín, C.; Overall, C. M. Protease Degradomics: A New Challenge for Proteomics. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (7), 509−519.

(65) Mackman, N. Triggers, Targets and Treatments for Thrombosis. *Nature* **2008**, *451*, 914.

(66) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discovery* **2010**, *9* (9), 690−701.

(67) Rawlings, N. D.; Barrett, A. J.; Finn, R. Twenty Years of the MEROPS Database of Proteolytic Enzymes, their Substrates and Inhibitors. *Nucleic Acids Res.* **2016**, *44* (D1), D343−D350.

(68) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Margreiter, M. A.; Spitzer, G. M.; Wallnoefer, H. G.; Liedl, K. R. Cleavage Entropy as Quantitative Measure of Protease Specificity. *PLoS Comput. Biol.* **2013**, *9* (4), No. e1003007.

(69) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102* (12), 4501−4524.

(70) Schechter, I.; Berger, A. On the Size of the Acitive Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157−162.

(71) Coughlin, S. R. Thrombin Signalling and Protease-Activated Receptors. *Nature* **2000**, *407* (6801), 258−264.

(72) Johnson, D. E. Noncaspase Proteases in Apoptosis. *Leukemia* **2000**, *14* (9), 1695−1703.

(73) Waldner, B. J.; Kraml, J.; Kahler, U.; Spinn, A.; Schauperl, M.; Podewitz, M.; Fuchs, J. E.; Cruciani, G.; Liedl, K. R. Electrostatic Recognition in Substrate Binding to Serine Proteases. *J. Mol. Recognit.* **2018**, *31* (10), No. e2727.

(74) Campbell, B.; Petukh, M.; Alexov, E.; Li, C. On the Electrostatic Properties of Homodimeric Proteins. *J. Theor. Comput. Chem.* **2014**, *13* (03), 1440007.

(75) Tworowski, D.; Feldman, A. V.; Safro, M. G. Electrostatic Potential of Aminoacyl-tRNA Synthetase Navigates tRNA on its Pathway to the Binding Site. *J. Mol. Biol.* **2005**, *350* (5), 866−882.

(76) Wade, R. C.; Gabdoulline, R. R.; Lüdemann, S. K.; Lounnas, V. Electrostatic Steering and Ionic Tethering in Enzyme-Ligand Binding: Insights from Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (11), 5942.

(77) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102* (18), 3531−3541.

(78) Singh, H.; Misra, N.; Hnizdo, V.; Fedorowicz, A.; Demchuk, E. Nearest Neighbor Estimates of Entropy. *American Journal of Mathematical and Management Sciences* **2003**, *23* (3−4), 301−321.

(79) Huynh, D. Q. Metrics for 3D Rotations: Comparison and Analysis. *Journal of Mathematical Imaging and Vision* **2009**, *35* (2), 155−164.

(80) Huggins, D. J. Estimating Translational and Orientational Entropies Using the k-Nearest Neighbors Algorithm. *J. Chem. Theory Comput.* **2014**, *10* (9), 3617−3625.

(81) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084−3095.

(82) Tsukada, H.; Blow, D. M. Structure of Alpha-Chymotrypsin Refined at 1.68 A Resolution. *J. Mol. Biol.* **1985**, *184* (4), 703−11.

(83) Wurtele, M.; Hahn, M.; Hilpert, K.; Hohne, W. Atomic Resolution Structure of Native Porcine Pancreatic Elastase at 1.1 A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56* (4), 520−523.

(84) Sichler, K.; Banner, D. W.; D'Arcy, A.; Hopfner, K. P.; Huber, R.; Bode, W.; Kresse, G. B.; Kopetzki, E.; Brandstetter, H. Crystal Structures of Uninhibited Factor VIIa Link its Cofactor and Substrate-Assisted Activation to Specific Interactions. *J. Mol. Biol.* **2002**, *322* (3), 591−603.

(85) Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengeler, P. A.; Wang, J.; Chan, H.; Wong, L. Structural Basis for Selectivity of a Small Molecule, S1-binding, Submicromolar Inhibitor of Urokinase-Type Plasminogen Activator. *Chem. Biol.* **2000**, *7* (4), 299−312.

(86) Rotonda, J.; Garcia-Calvo, M.; Bull, H. G.; Geissler, W. M.; McKeever, B. M.; Willoughby, C. A.; Thornberry, N. A.; Becker, J. W. The Three-Dimensional Structure of Human Granzyme B Compared to Caspase-3, Key Mediators of Cell Death With Cleavage Specificity for Aspartic Acid in P1. *Chem. Biol.* **2001**, *8* (4), 357−68.

(87) Wu, L.; Wang, L.; Hua, G.; Liu, K.; Yang, X.; Zhai, Y.; Bartlam, M.; Sun, F.; Fan, Z. Structural Basis for Proteolytic Specificity of the Human Apoptosis-Inducing Granzyme M. *J. Immunol.* **2009**, *183* (1), 421−9.

(88) Laxmikanthan, G.; Blaber, S. I.; Bernett, M. J.; Scarisbrick, I. A.; Juliano, M. A.; Blaber, M. 1.70 A X-Ray Structure of Human Apo Kallikrein 1: Structural Changes upon Peptide Inhibitor/Substrate Binding. *Proteins: Struct., Funct., Genet.* **2005**, *58* (4), 802−14.

(89) Hilpert, K.; Ackermann, J.; Banner, D. W.; Gast, A.; Gubernator, K.; Hadvary, P.; Labler, L.; Muller, K.; Schmid, G.; Tschopp, T. B.; et al. Design and Synthesis of Potent and Highly Selective Thrombin Inhibitors. *J. Med. Chem.* **1994**, *37* (23), 3889−3901.

(90) Schmidt, A.; Jelsch, C.; Ostergaard, P.; Rypniewski, W.; Lamzin, V. S. Trypsin Revisited: Crystallography at (sub) Atomic Resolution and Quantum Chemistry Revealing Details of Catalysis. *J. Biol. Chem.* **2003**, *278* (44), 43357−62.

(91) *MOE (Molecular Operating Environment), 2019.1*; Chemical Computing Group (CCG): 2019.

(92) Labute, P. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins: Struct., Funct., Genet.* **2009**, *75* (1), 187−205.

(93) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(94) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T., III; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M. *AMBER 2018*, 2018.

(95) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878−3888.

(96) Sadasivan, C.; Yee, V. C. Interaction of the Factor XIII Activation Peptide with Alpha -Thrombin. Crystal Structure of its Enzyme-Substrate Analog Complex. *J. Biol. Chem.* **2000**, *275* (47), 36942−8.

(97) Gandhi, P. S.; Chen, Z.; Di Cera, E. Crystal Structure of Thrombin Bound to the Uncleaved Extracellular Fragment of PAR1. *J. Biol. Chem.* **2010**, *285* (20), 15393−8.

(98) Wang, L.; Zhang, K.; Wu, L.; Liu, S.; Zhang, H.; Zhou, Q.; Tong, L.; Sun, F.; Fan, Z. Structural Insights into the Substrate Specificity of Human Granzyme H: The Functional Roles of a Novel RKR Motif. *J. Immunol.* **2012**, *188* (2), 765−73.

(99) Segrest, J. P.; Feldmann, R. J. Membrane Proteins: Amino Acid Sequence and Membrane Penetration. *J. Mol. Biol.* **1974**, *87* (4), 853−858.

(100) Nozaki, Y.; Tanford, C. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions Establishment of a Hydrophobicity Scale. *J. Biol. Chem.* **1971**, *246* (7), 2211−2217.

(101) Wolfenden, R.; Andersson, L.; Cullis, P.; Southgate, C. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20* (4), 849−855.

(102) Chothia, C. The Nature of The Accessible and Buried Surfaces in Proteins. *J. Mol. Biol.* **1976**, *105* (1), 1−12.

(103) Janin, J. Surface and Inside Volumes in Globular Proteins. *Nature* **1979**, *277* (5696), 491.

(104) Von Heijne, G.; Blomberg, C. Trans-Membrane Translocation of Proteins: The Direct Transfer Model. *Eur. J. Biochem.* **1979**, *97* (1), 175−181.

(105) DeLano, W. L. *PyMOL*, 2002.

(106) Lumry, R.; Rajender, S. Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Properly of Water. *Biopolymers* **1970**, *9* (10), 1125−1227.

(107) Dunitz, J. D. Win Some, Lose Some: Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. *Chem. Biol.* **1995**, *2* (11), 709−712.