

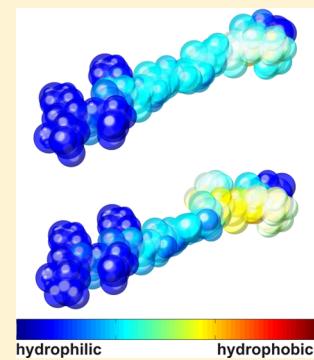
# Molecular Dynamics Simulations Approach for the Characterization of Peptides with Respect to Hydrophobicity

Sven Amrhein,<sup>†</sup> Stefan Alexander Oelmeier,<sup>†</sup> Florian Dismer,<sup>‡</sup> and Jürgen Hubbuch\*,<sup>†</sup>

<sup>†</sup>Institute of Process Engineering in Life Sciences, KIT, Engler-Bunte-Ring 1, D-76131 Karlsruhe, Germany

<sup>‡</sup>Novo Nordisk AS, Novo Allé, 2880 Bagsværd, Denmark

**ABSTRACT:** It has been shown that molecular dynamics (MD) simulations are a powerful tool to generate knowledge about complex interactions in the field of bioprocess technologies at the atomic level. In this field, one of the most important nonspecific interactions is the hydrophobic interaction, which is still not fully understood after nearly 30 years of research. To date established hydrophobicity scales, which base mostly on proteins' primary structure, are used to estimate the overall hydrophobicity. The structural complexity and the influence of the protein's environment cannot be accommodated with these scales. In this work, free solution molecular dynamics simulations were used to investigate the hydrophobic character of low molecular weight peptides. Therefore, local densities of a small hydrophobic tracer molecule and unprotonated triethylamin (TEA) in particular were used to localize and quantify hydrophobic patches among the peptide surface. Comparisons between local densities and the retention behavior in reversed phase chromatography showed significant correlations. Moreover, neighbor effects caused by charges could be identified. We were able to show that the developed *in silico* method is applicable to characterize peptides in respect to hydrophobicity in agreement with experimental data. We are confident to apply this method to larger protein structures.



## INTRODUCTION

Hydrophobicity is at the heart of our understanding how proteins behave in solutions in respect to solubility, folding, aggregation, or precipitation. In the field of biopharmaceuticals, hydrophobicity is a key factor from the down streaming processing over the formulation to the final drug application. In downstream processing and formulation, hydrophobicity dominates the retention behavior in hydrophobic interaction and reversed phase chromatography, the partitioning in extraction processes, exemplary aqueous two phase systems, and the tendency to forming oligomers in a preferred manner in crystallization steps or in undesirable manner of aggregation and precipitation, which leads to product loss and bears the risk of immune reactions upon applications. The bioavailability of a drug is influenced by solubility and membrane permeability, which depend on hydrophobicity in turn.

Thus, the quantification of hydrophobicity is highly relevant for the processability as well as the pharmacological property of a potential drug substance. In particular, with regard to time to market demands and reduction of time and investment for candidate selection, an *in silico* approach for measuring hydrophobicity is highly desirable. One method for the experimental determination of hydrophobicity of a protein is precipitation.<sup>1,2</sup> In this case proteins' solubility is influenced by the concentration of a precipitation species. The critical concentration at which protein precipitation begins is called  $m^*$ . The protein's hydrophobicity can be expressed by  $1/m^*$ . The precipitation or aggregation process respectively is highly complex and is influenced by a number of parameters such as pH, protein concentration, temperature or additives. Hence the

transferability of the hydrophobicity determined by the precipitation method to other environmental conditions is limited.

A second experimental approach to determine a surface hydrophobicity is the use of extrinsic fluorescent dyes such as ANS, Bis-ANS, Nile Red, or Tioflavin T.<sup>3</sup> This method focuses on the hydrophobic patches on the protein's surface. The sensitivity of the dyes' optical properties to their environment's polarity, basicity, or acidity, respectively, are used for protein characterization. The mechanisms are described elsewhere in detail.<sup>3</sup> These experimental procedures require high investments in time and material and are limited with regard to transferability to other systems because of the environmental influences mentioned above. This is why hydrophobicity scales based on structural data would be beneficial.

In the last few decades, a number of scales were developed to characterize proteins regarding hydrophobicity.<sup>4–9</sup> Partition coefficients between polar and apolar solvents, retention behavior in reversed phase chromatography or thermodynamic approaches were used to allocate a hydrophobicity value to every amino acid. Some of the scales, like the probably most widely used scales of Kyte and Doolittle, implicate the probability of exposure in a globular protein. Globular proteins prone to form a hydrophobic core to minimize the hydrophobic solvent accessible surface in polar solvents like water.<sup>10</sup> The probability of an amino acid being exposed on the surface is

Received: July 25, 2013

Revised: January 23, 2014

Published: January 27, 2014

accounted by Kyte and Doolittle by using the results from Chothia et al.<sup>11</sup> On the basis of the primary sequence of a protein, it is possible to estimate the overall hydrophobicity by summing up the hydropathy values for each residue and dividing them by the length of the sequence, according to Kyte and Doolittle.<sup>4</sup>

But estimating a protein's hydrophobicity solely based on its amino acid sequence cannot account for the complexity of its structure in solution. The structure of a protein is highly influenced by environmental properties like the nature of the solvent, temperature, cosolvents, additives or pH, which induces charge variations in turn. The allocation of single hydropathy values for each amino acid can give an impression of the proteins hydrophobicity but is not suitable to compare the protein properties depending on its environment.

Molecular dynamics (MD) simulations were shown to be an appropriate approach to accommodate for the proteins' complexity. For instance Cramer et al.<sup>12</sup> have used MD simulations to localize potential binding sites of a mixed mode ligand to various Ubiquitin mutants. In a previous work we have shown that MD simulations are suitable to simulate aqueous two phase systems (ATPS).<sup>13</sup> By a number of tools coming from bioinformatics such as homology modeling<sup>14,15</sup> it is possible to model a three-dimensional structure, without the actual need to crystallize that particular protein. Thereby, MD simulations can be performed in the very early stages of drug development, namely drug screening and candidate selection.

In the presented work, we developed an *in silico* method to characterize small peptides in aqueous solutions with respect to hydrophobicity and its distribution on the surface. The basic intention of the presented work is the development of an *in silico* approach to characterize proteins with respect to surface hydrophobicity. We are convinced, that a first step to reach these ambitious aim is the use of short and clearly defined amino acid structures which do not expose complex secondary or tertiary structures. Therefore, we have focused on small custom designed peptides for method development and validation in the presented work. The developed approach is potentially capable to be transferred to proteins. Explicit solvent molecular dynamic simulations containing hydrophobic tracer molecules were performed for eight different peptides. A modified radial distribution function (mRDF) was used to generate a hydrophobicity index. On the basis of the mRDF the peptides' overall hydrophobicity could be quantified and the distribution of hydrophobic areas could be identified. Experimentally the overall hydrophobicity could be confirmed by reversed phase ultra high performance liquid chromatography (RP-UHPLC). It could be shown that this *in silico* method yielded a better correlation to experimental data than the most commonly used scale of hydrophobicity presented by Kyte and Doolittle.

## MATERIALS AND METHODS

**Peptides.** To develop a method to describe hydrophobicity based on MD simulations, eight custom designed peptides, each with a length of 10 amino acids, were purchased from Thermo Fisher Scientific GmbH (Ulm, Germany). The primary sequences of these peptides are listed in one-letter-code in Table 1. All peptides include a nominally hydrophilic patch with duplicates of Lys and Asp and a nominally hydrophobic residue separated by neutral Gly. In case of the hydrophobic residue Leu three peptides with a different content

**Table 1. Primary Sequences of Used Model Peptides**

name	primary sequence
Pep <sub>1</sub> Leu	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-LEU-GLY-OH
Pep <sub>2</sub> Leu	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-LEU-LEU-GLY-OH
Pep <sub>3</sub> Leu	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-LEU-LEU-LEU-GLY-OH
Pep <sub>1</sub> Cys	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-CYS-GLY-OH
Pep <sub>1</sub> Ile	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-ILE-GLY-OH
Pep <sub>1</sub> Met	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-MET-GLY-OH
Pep <sub>1</sub> Phe	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-PHE-GLY-OH
Pep <sub>1</sub> Val	NH <sub>2</sub> -ASP-LYS-ASP-LYS-GLY-GLY-GLY-VAL-GLY-OH

of Leu were designed. The peptides were designed by following considerations:

- The peptide should not form complex secondary or tertiary structure like  $\beta$  sheets or  $\alpha$  helices.
- The hydrophobic substitution residue should not carry an additional charge, thus should not constitute the C-terminus.
- The hydrophobic substitution residue should be accessible for the C18 ligand of the reversed phase chromatography matrix. Thus, very small amino acids namely Gly were set next to the substitute to minimize steric hindrance to the C18 ligand.
- The peptide should have local charges to increase water solubility.

**Reversed Phase HPLC.** The retention factors of the peptides on a reversed phase UHPLC column were used as experimental measure of the peptides' hydrophobicity. These measurements were performed on a Dionex UltiMate 3000 RSLC x2 Dual system consisting of an HPG-3400RS binary rapid separation LC pump module, a WPS-3000TRS autosampler, a TCC-3000RS column compartment and a DAD3000(RS) diode array detector.

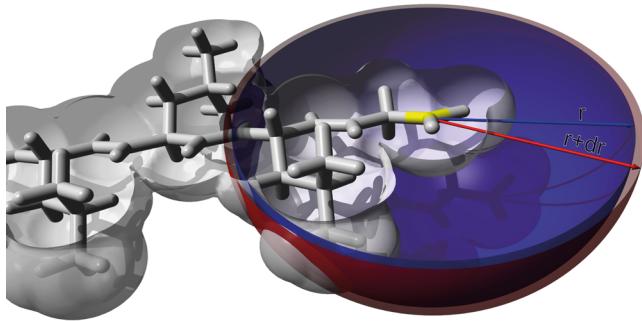
To avoid limitations by mass transfer the monolithic silica based HPLC column Chromolith Performance RP-18 end-capped (100 × 2 mm) from Merck KGaA (Darmstadt, Germany) was used. Each peptide was solubilized in ultra pure water to a concentration of 1 mg/mL. Samples were injected with a volume of 5  $\mu$ L. The column temperature was kept at 25 °C. The peptides were eluted with a gradient starting from 100% filtered, degassed, and ultrapure water to 40% degassed acetonitrile, purchased from Merck KGaA (Darmstadt, Germany) with a slope of 1 CV/min and a flow of 3 mL/min. The system dead time ( $t_{\text{Sys}}$ ) was measured by injecting 5  $\mu$ L of 50 mmol citric acid. The columns dead time ( $t_0$ ) was measured by injecting 5  $\mu$ L of 50 mmol citric acid onto the column. The retention factor  $k'_i$  was calculated by eq 1

$$k'_i = \frac{t'_{R,i} - t_0}{t_0} \quad (1)$$

$$t'_{R,i} = t_{R,i} - t_{\text{Sys}} \quad (2)$$

where  $t'_{R,i}$  is the absolute retention time of the peptide,  $t_{\text{Sys}}$  is the systems dead time, and  $t_0$  is the retention time of a nonretained analyte. All measurements were performed as quintets.

**MD Simulations.** The explicit MD simulations were performed on two distributed memory parallelized Intel Xeon Quad Core sockets with 2.66 GHz frequency with the software YASARA Structure (Version 10.10.29).<sup>16,17</sup> A cubic simulation cell with dimensions of about 57 Å and periodic boundaries was used. The temperature was kept at 298 K (isotherm) by



**Figure 1.** Schematic illustration of the procedure of calculating local distribution of each proteinogenic atom core using the example of the C-terminal carbon atom of Pep<sub>3</sub>LEU. The atom of interest is colored in red, the peptide's van der Waals surface is shown in gray and the balancing shells, limited by the radius  $r$  and  $r + dr$ , are shown in blue or red, respectively. The molecular graphic was created with YASARA<sup>28</sup> and POV-Ray.<sup>29</sup>

rescaling the atom velocities using a Berendsen thermostat.<sup>18</sup> Simulations were performed with constant atom number, constant volume and constant temperature.

The Amber03 force field<sup>19</sup> was applied to describe the interactions of the proteins, which has been adequate for the simulation of peptides and proteins. The TIP3P model was used for water. The particle mesh Ewald (PME) algorithm treatment<sup>20</sup> was used for longrange electrostatics with a 7.86 Å cutoff for nonbonded interactions. These cutoff of 7.86 Å was selected due to concerns regarding the performance/accuracy trade-off. As the software YASARA is optimized toward a cutoff that is a multiple of 2.62 Å and the chosen cutoff applies to the van der Waals forces only, we selected a cutoff of 7.86 Å. The simulation cell included one initially centered peptide of interest, 33 copies of the neutrally charged tracer molecule triethylamin (TEA) and over 6000 water molecules. The protonation states of residues corresponded to a pH of 7. At this pH the N-terminal Asp has a negatively charged carboxy group and a positively charged N-terminus, Asp 3 and C-terminal Gly are negatively charged and Lys 2 and Lys 4 are positively charged resulting in a neutrally charged peptide in total. Consequently, no counterions were added for cell neutralization.

The tracer TEA was parametrized using the AutoSMILES algorithm<sup>21</sup> implemented in the MD simulation software. In order to remove bumps and to correct the covalent geometry, every simulation was initially energy-minimized with the Amber03 force field.<sup>19</sup> After removal of conformational stress by a short steepest descent minimization, the procedure continued by simulated annealing until convergence was reached with a time step of 2 fs and scaled down atom velocities by 0.9 every 10th step. Convergence was reached when the energy improved by less than 0.05 kJ/mol per atom during 200 steps.

To gain information about the distribution of the hydrophobic tracer molecules or the local densities respectively, a

modified radial distribution function (mRDF) of each protein atom was calculated. In this work, the modified radial distribution function represents the concentration of the tracer atoms in dependency of the distance from an atomic core, which is illustrated in Figure 1. In contrast to the generally used radial distribution function the solvent accessible volume is considered in the mRDF. This modification is necessary in order to be able to distinguish between an atom with hydrophobic character with minor solvent accessibility and an atom with a neutral character which is exposed and completely solvent accessible. The mRDF corresponds to the probability of finding tracer molecules around a certain peptide atom. A high probability of finding a tracer molecule in proximity of a certain peptide surface region in turn corresponds to a high hydrophobicity of that region.

Data evaluation was done in MATLAB R2012a (7.14.0.739) and determined how many tracer atoms were within a shell limited by the inner radius  $r$  and the outer radius  $r + dr$  around each atom core of the protein. The solvent accessible volume (probe radius of 1.4 Å) of each shell was calculated by a numerical approximation as follows. Each shell was filled with 5000 ( $N_t$ ) evenly distributed grid points and the number of solvent accessible points ( $N_a$ ) was evaluated. The solvent accessible volume of the shells ( $V_a$ ) was determined by eq 3.

$$V_a = \frac{4}{3}\pi(r_a^3 - r_i^3)\frac{N_a}{N_t} \quad (3)$$

With knowledge about the number of tracer atoms and the solvent accessible volume of each shell, the concentration was calculated. For each atom core the concentration was determined within a distance of 10 Å by evaluating 40 shells with a width of 0.25 Å each. Simulations were evaluated every 25 ps within a time range from 10 to 150 ns. The results of all simulation snapshots were averaged. By this averaging over a relative long simulation time the modified radial distribution profiles correspond to the probability of a tracer atom being in a specific region of the respective atom.

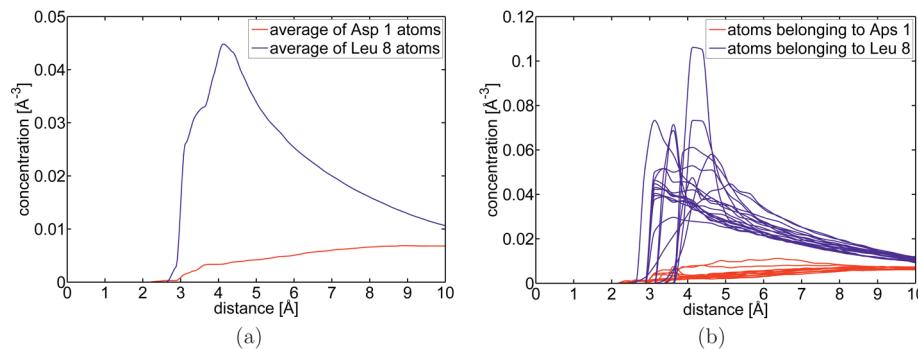
## RESULTS

**Experimental Results.** The retention factors on a monolithic reversed phase column of all peptides were measured to get a scale for their overall hydrophobicity. The peptides eluted in the following order and their retention factors are listed in Table 2: Pep<sub>1</sub>Cys < Pep<sub>1</sub>Val < Pep<sub>1</sub>Met < Pep<sub>1</sub>Ile < Pep<sub>1</sub>Leu < Pep<sub>1</sub>Phen < Pep<sub>2</sub>Leu < Pep<sub>3</sub>Leu. All retention factors could be determined with a standard deviation below 3.2 %.

**In Silico Results.** The evaluation of the modified radial distribution function of each atom was focused on the characterization of each atom with respect to hydrophobicity and the peptide in total, respectively. In general, the determined mRDF varied strongly within the primary sequence. The atoms of hydrophobic residues showed a peak shaped radial distribution which approximates the tracer's bulk concentration after reaching the maximal peak height. In contrast atoms of a

**Table 2.** Retention Factors  $k'$  and the Relative Standard Deviation  $\sigma_r$  of Peptide Pep<sub>Substitute</sub> Obtained by Reversed Phase Ultrahigh Performance Liquid Chromatography

substitute	1Cys	1Ile	1Leu	1Met	1Phen	1Val	2Leu	3Leu
$k' [-]$	0.52	4.73	4.80	2.27	6.11	1.86	11.97	16.22
$\sigma_r [\%]$	0.66	3.10	2.83	0.22	1.28	0.36	0.16	0.04



**Figure 2.** Comparison of density profiles of atoms belonging to Asp 1 and Leu 8. The distribution functions were averaged over a time range of 10 to 150 ns.

charged residue showed no or low peaks, which is illustrated by means of Figure 2. Figure 2a illustrates the averaged mRDF of all atoms belonging to Asp 1 and Leu 8. Figure 2b reveals the mRDF of every single atom of these amino acids. These simplified illustrations indicate, that the hydrophobicity of the peptides' atoms and amino acids correlates with the shape of the corresponding mRDF.

All determined mRDF on these atomic level of the investigated peptides are presented in Figure 3 to illustrate the basic observations. Within these subfigures the primary sequence of the respective peptide is located on the *x*-axis as one-letter-code. All atoms belonging to one residue are colored consistently.

As illustrated in Figure 3 the averaged modified radial distribution function (mRDF) of all peptides varied strongly within their primary sequence. All modified distribution density functions revealed no tracer atoms within a certain distance from the respective protein atom core due to steric hindrance. This distance was equal to the VdW radius of the respective atom plus the smallest VdW radius of the tracer atoms. The comparison of all peptides revealed a “fingerprint” character of the density profiles. Profiles were unique for every peptide. While the profiles correlated well to the nature of the residue, the atom belongs to, they surpassed the common amino acid based scale in resolution. For example, the charged and nominally hydrophilic Asp 4 and especially its backbone carbon atom revealed in all mRDF some peaks with a height of about 0.04 Å<sup>-3</sup>, equal to the hydrophobicity of the adjacent Gly and thus showing the influence of a hydrophobic residue on its surrounding. Comparing the nominally hydrophobic part of the sequence consisting of Gly and the substitute it was obvious, that the peak heights increased with the retention factor of the respective peptide. Comparing the Leu containing peptides, it was found that the peak heights of mRDF within the Leu residues increased with the number of Leu (compare parts e, g, and h of Figure 3) and the distance to the negatively charged C-terminal Gly (see Figure 3h). To determine the distribution or the localization of hydrophobic surfaces a descriptor  $H_{atom}$  was extracted from the mRDF ( $c(r)$ ) for each atom according to the following equations.

$$H_{atom} = A_{Core} \frac{A_{Core}}{\left( \sum_{i=1}^{n_{atoms}} A_{Core,i} \right)} \quad (4)$$

$$A_{Core} = \int_{r=r_{vdW}}^{r=r_{vdW}+6\text{\AA}} c(r) \, dr \quad (5)$$

The area  $A_{Core}$  of one core was calculated by integrating the mRDF of the respective atom. The integration of the mRDF corresponds to the possibility of tracer atoms being nearby the respective atom within the region limited by  $r_{vdW}$  and  $r_{vdW} + 6$  Å. The hydrophobicity value of an atom ( $H_{atom}$ ) was calculated by weighting the respective area ( $A_{Core}$ ) with its ratio to the sum of all areas within the peptide.

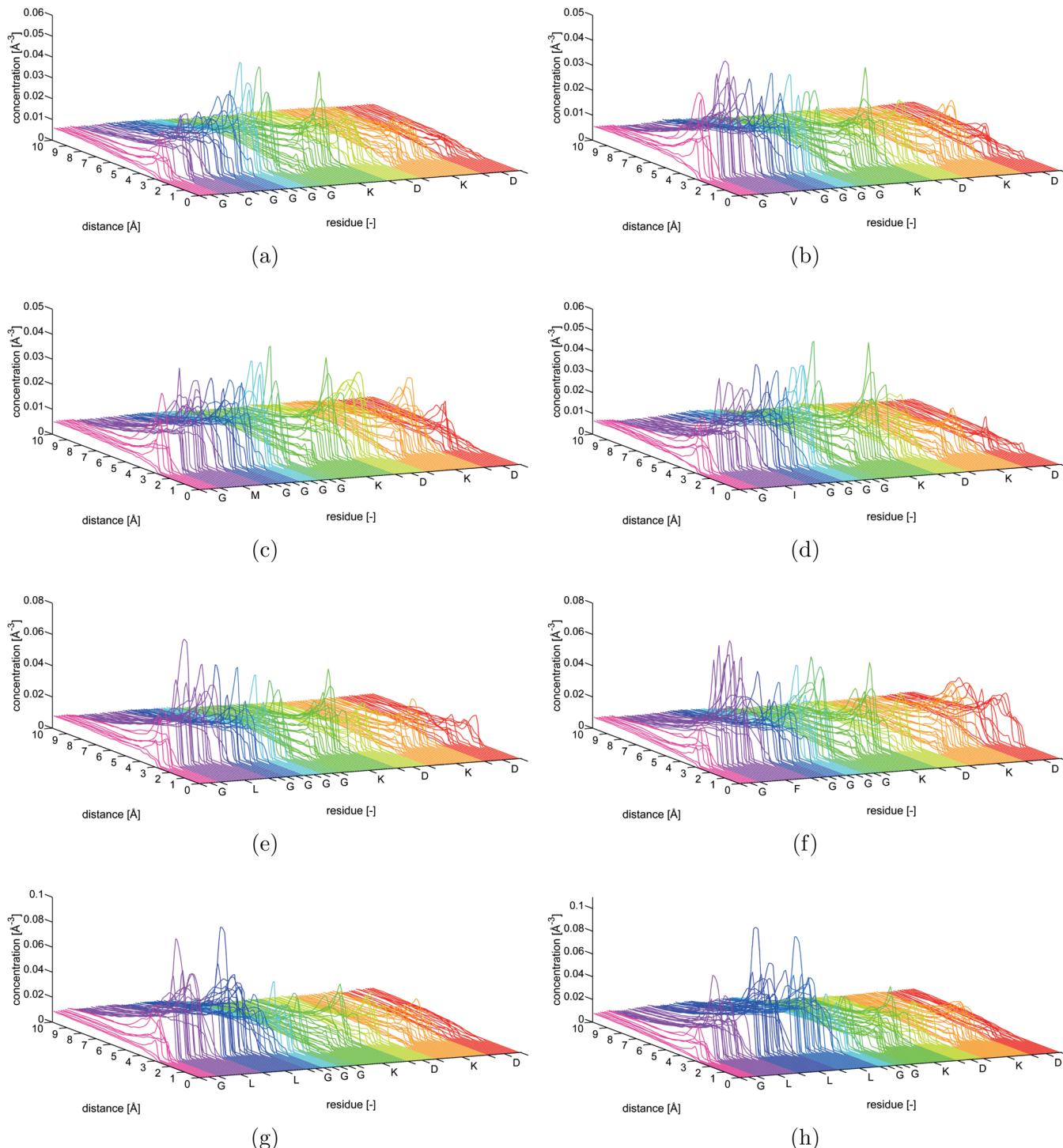
This atomic hydrophobicity descriptor  $H_{atom}$  was used for mapping the peptides' surfaces to characterize the distribution of hydrophobic areas. This mapping revealed a good correlation with expected hydrophobicity distribution and quantification. As an example, the peptides Pep<sub>1</sub>Val, Pep<sub>1</sub>Leu, and Pep<sub>2</sub>Leu mapped by  $H_{atom}$  are illustrated in Figure 4. Hydrophilic surfaces, colored in dark blue, can be precisely assigned to the charged residues. Hydrophobic surfaces, shown in warm colors, can be allocated to nominally neutral Gly and the nominally hydrophobic amino acids. However, the transitions within the sequence are smooth. The C-backbone in all peptides from Asp 3 to the C-terminus revealed a rather neutral character. Each peptide's overall hydrophobicity ( $H_{Pep}$ ) was calculated by summing up the hydrophobicity values  $H_{atom}$  of each of its atoms according to eq 6.

$$H_{Pep} = \sum_{i=1}^{n_{atoms}} H_{atom,i} \quad (6)$$

The overall peptides' hydrophobicity was tested experimentally by a RP-HPLC. As presented in Figure 5, there was a good linear correlation between the calculated hydrophobicity values ( $H_{Pep}$ ) and the retention factors, obtained from the reversed phase chromatography, with a coefficient of determination ( $R^2$ ) of 0.94. In comparison to the widely used GRAVY-value (grand average of hydropathy), which bases on the scale of Kyte and Doolittle, correlated with a  $R^2$  of 0.60. It is important to note that the order of elution and the order of the  $H_{Pep}$  values were equal which was in contrast to the GRAVY-value. However, the focus of the presented work was not to generate an universal hydrophobicity scale comparable to Kyte and Doolittle but to develop an *in silico* method to generate an impression of the hydrophobicity of the peptide of interest in its specific environment.

## ■ DISCUSSION

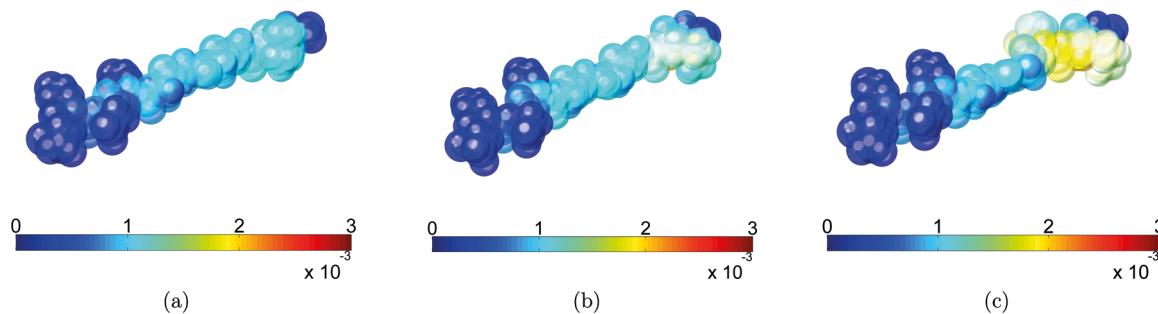
**RP HPLC.** To compare *in silico* generated hydrophobicity values, an experimental scale of this property was needed. We decided on using small peptides not forming stable secondary structures as initial benchmark for the *in silico* parameter generation. As experimental measure of the peptides' hydro-



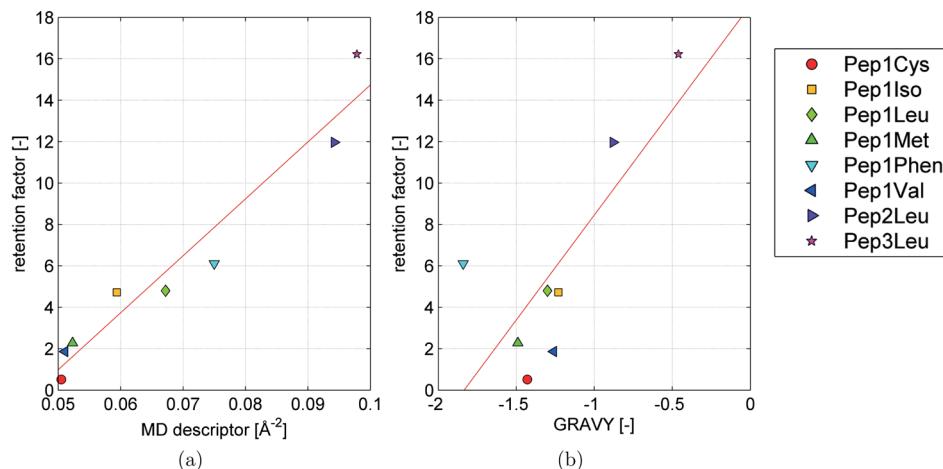
**Figure 3.** Compilation of all evaluated mRDF sorted by increasing retention factor (a) Pep<sub>1Cys</sub>, (b) Pep<sub>1Val</sub>, (c) Pep<sub>1Met</sub>, (d) Pep<sub>1Ile</sub>, (e) Pep<sub>1Leu</sub>, (f) Pep<sub>1Phen</sub>, (g) Pep<sub>2Leu</sub> and (h) Pep<sub>3Leu</sub>.

phobicities, a RP-HPLC assay was successfully established for one monolithic column. Retention in RP-HPLC is strongly dependent on the hydrophobicity of the analyte and the stationary phase.<sup>22</sup> However, we feel that the validation of a calculated overall hydrophobicity by retention factors in reversed phase chromatography is justified as the order of elution is a good experimental factor for the analyte's hydrophobicity, if size exclusion effects, diffusion limitations or steric hindrance of ligand-analyte binding can be considered insignificant. In our case we avoided any impact of size

exclusion and diffusion limitations by using a monolithic chromatography column. Small and highly flexible peptides were designed in order to provide fully accessible hydrophobic residues. Thus, steric hindrance of ligand interaction, which would be influenced by the ligand length, can be considered negligible. Charge effects were minimized by using an end-capped RP phase and by keeping the charge of each peptide constant. We are thus confident that the established assay is a valid representation of the peptides' hydrophobicity. The peptides did not elute according to the nominal hydrophobicity

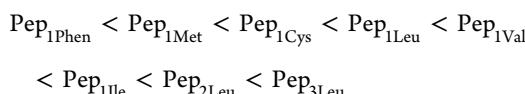


**Figure 4.** Characterization of the hydrophobicity distribution by mapping the peptide's surface according to the atomic descriptor  $H_{atom}$  by means of (a) Pep<sub>1Val</sub>, (b) Pep<sub>1Leu</sub> and (c) Pep<sub>2Leu</sub>. Colder colors refer to hydrophilic areas, whereas warm colors refer to hydrophobic areas.

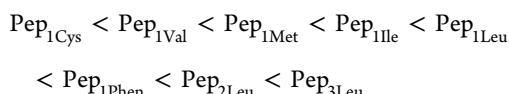


**Figure 5.** Correlation of the retention factor  $k'$  and (a) the descriptor obtained from MD simulations ( $R^2 = 0.94$ ) and (b) the GRAVY value ( $R^2 = 0.60$ ) for the eight model peptides.

of Kyte and Doolittle whereupon the elution order is supposed to be as follows:



This was contradictory to the experiments which revealed the following elution order:

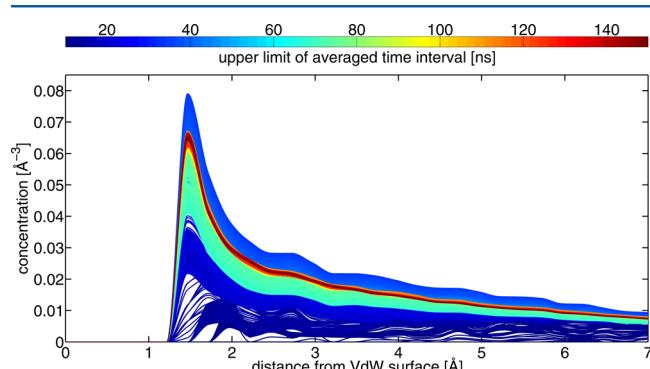


**MD.** The present work was designed to determine hydrophobic surfaces of peptides using molecular dynamics simulations. The evaluation of the MD simulation aims at the modified distribution function in respect of a hydrophobic tracer molecule.

The concentration of TEA was optimized toward the maximum number of tracer molecules to reduce the necessary simulation time and to ensure system stability. It was considered that the necessary time to get an impression of the peptides hydrophobicity can be reduced by increasing the tracer number, as it increases the probability of the peptide meeting a tracer molecule. In contrast, phase separation was observed in the presence of too many tracer atoms. Therefore, an optimization was performed. Experimental phase separation could be observed using a cloud point method at about 1.3 mol % or 0.06 g/mL respectively. A phase separation could be observed at 0.05 g/mL *in silico*, which equals to 1.0 mol %.

Even the basic character of TEA is not represented in MD simulations, this concentration was within the scope of the critical concentration where phase separation could be observed in MD simulation. As a comparison the concentration for the final MD simulations was 0.55 mol %.

The necessary simulation time to get a stable mRDF is illustrated in Figure 6 by means of a carbon atom (number 113) belonging to the phenol ring of the peptide Pep<sub>1Phen</sub>. Starting 10 ns into the simulation, the average mRDF was calculated for increasing simulation durations. The length of the period over which the average was calculated is color coded in



**Figure 6.** Stability of the mRDF illustrated by means of a carbon atom belonging to the phenol ring (atom 113) of the peptide Pep<sub>1Phen</sub>. The colorbar represents the time interval which was averaged to calculate the mRDF.

Figure 6. With increasing duration (with warmer color) the calculated average becomes increasingly stable. Figure 6 clearly shows that the chosen period of 10 to 150 ns is sufficient to yield a stable, representative average mRDF. As shown in the Results, the profiles of the mRDF indicate the hydrophobic or hydrophilic property of each protein atom. Charged atoms and atoms with nearby charges revealed low attractions to the hydrophobic tracer. This is due to the electrostatic interactions with the dipolar water molecules and hence the preference for forming hydration shells around charged atoms.<sup>23,24</sup> Charged or polar solvent molecules displace hydrophobic tracer molecules from charged peptide areas.

By contrast the mRDF profiles of atoms belonging to hydrophobic amino acids indicated an accumulation of hydrophobic tracer molecules as it is thermodynamically preferred to minimize the exposed hydrophobic area. The comparison of the mRDF profiles of the Leu containing peptides exhibit the influence of the nearby C-terminal charge and the cumulative effects of hydrophobic amino acids, as Leu 7 and Leu 8 accumulate more tracer molecules compared to Leu 9.

This shows, that this molecular dynamic approach is sensitive enough to identify hydrophobic areas of the peptides and can consider adjacency effects like decreasing hydrophobicity by nearby charges or increasing hydrophobicity by forming hydrophobic patches consisting of a number of hydrophobic amino acids. Potentially this approach could be used to identify aggregating prone areas for proteins that follow a native aggregation pattern.

A descriptor  $H_{Pep}$  was generated according to eq 6 on base of the evaluated mRDF of each atom as an overall characteristic value with respect to hydrophobicity. It is important to note that the choice of the descriptor to be correlated to experimental data was guided by the underlying physical principles. It was thus decided to calculate the mRDF to a distance of the VdW radius plus 6 Å as this represents the space in which the relevant forces predominantly act. While a slightly better correlation of MD data to experimental data could be derived by using other descriptors, we are confident that correlations based on reasoning rather than statistics are more likely to have predictive power. The descriptor correlates to the probability of finding a tracer molecule within a region limited by the van der Waals radius and the used cutoff where the van der Waals interactions are considered. It seems to be capable to characterize the peptides' surface with regard to hydrophobicity in terms of localization and strength. This descriptor seems to be sensitive enough to picture effects caused by adjacent charges on an atomic level. The overall descriptor  $H_{Pep}$  showed a satisfying agreement with the retention data.

Compared to the developed *in silico* descriptor, the GRAVY values correlated less well with the retention factors. This is not surprising considering the short length of the peptides. The terminal charges have a relatively high influence on the peptides' hydrophobicity, which is not accounted in the calculation of the GRAVY value.

Kim et al<sup>25</sup> have already observed, that there is no correlation between elution times and the GRAVY score. Although peptide retention is not supposed to correlate to the GRAVY score, this index appears as an useful scale for distinguishing hydrophobic from hydrophilic peptides.<sup>26</sup> A proteomics approach for predicting peptides retention was developed by Krokhin et al.<sup>27</sup> We are aware that the tracer substance triethylamine is not present only in neutral form due to the basic character.

However, this tracer molecule could meet our demands with respect to stability of the *in silico* system, accessibility due to small size and distinctive hydrophobic character.

The used peptides did not show a stable secondary structure, therefore the influence of the hydrophobic tracer molecule on the peptides' conformation was considered negligible. Molecular dynamics simulations of systems under identical conditions but without tracer molecules were performed to compare the conformations of the peptides. The root-mean-square deviation (RMSD) of the backbone of each peptide to its initial linear structure was calculated for all snapshots. Neither significant variations of the RSMD can be observed in the presence of tracer atoms, nor preferred conformations could be identified (data not shown). This is due to the high flexibility of the peptides. However, for larger proteins it must be considered, that the presence of the hydrophobic tracer molecules could lead to minor conformational changes. The simulation methodology used in the presented work is not adequate to simulate major structural conformational changes like folding or unfolding as the simulation time range is not appropriate to observe larger conformational changes. It cannot be excluded that minor structural conformational changes could still be induced by the presence of tracer molecules. For instance hydrophobic residue nearby the exposed surface could be induced to move slightly to the surface attracted by a tracer molecule. For the transfer to larger proteins, we thus recommend an equilibration without tracer molecules and constraining the backbone after adding the tracer molecules.

## CONCLUSION

This study has shown that molecular dynamic simulations can be used for the characterization of small peptides with respect to hydrophobicity with good experimental agreement. In this work an *in silico* method was developed, which can cope with peptides complexity. In particular the overall hydrophobicity of the peptides could be evaluated from molecular dynamics simulations and validated by reversed phase chromatography. Even on atomic level the hydrophobic distribution could be determined, which can be helpful to identify aggregation prone surfaces for proteins following a native aggregation pattern. Compared to the established hydrophobicity scales and Kyte and Doolittle scale (GRAVY) in particular, we found a better agreement of the *in silico* descriptor with the experiments. We are confident that this molecular dynamics approach can be transferred to larger proteins even when a reduction of computational effort is needed. This could be done by reducing the number of mRDF evaluations by calculating the mRDF of every amino acid in contrast to every atom.

Finally this approach could be used to characterize chromatographic ligands of HIC, RP, or mixed mode adsorber with respect to hydrophobicity.

## AUTHOR INFORMATION

### Corresponding Author

\*(J.H.) E-mail: Juergen.Hubbuch@kit.edu. Telephone: +49 721 608 42557. Fax: +49 721 608 46240.

### Notes

The authors declare no competing financial interest.

## REFERENCES

- (1) Hachem, F.; Andrews, B.; Asenjo, J. Hydrophobic Partitioning of Proteins in Aqueous Two-Phase Systems. *Enzyme Microbial Technol.* 1996, 19, 507–517.

- (2) Asenjo, J.; Schmidt, A.; Hachem, F.; Andrews, B. Model for Predicting the Partition Behaviour of Proteins in Aqueous Two-Phase Systems. *J. Chromatogr. A* **1994**, *668*, 47–54; 8th International Conference on Partitioning in Aqueous Two-Phase Systems.
- (3) Hawe, A.; Sutter, M.; Jiskoot, W. Extrinsic Fluorescent Dyes as Tools for Protein Characterization. *Pharm. Res.* **2008**, *25*, 1487–1499.
- (4) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (5) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, L.; White, S. H.; von Heijne, G. Recognition of Transmembrane Helices by the Endoplasmic Reticulum Translocon. *Nature* **2005**, *433*, 377–381.
- (6) Wimley, W. C.; White, S. H. Experimentally Determined Hydrophobicity Scale for Proteins at Membrane Interfaces. *Nat. Struct. Mol. Biol.* **1996**, *3*, 842–848.
- (7) Radzicka, A.; Wolfenden, R. Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous Solution. *Biochemistry* **1988**, *27*, 1664–1670.
- (8) Guy, H. Amino Acid Side-Chain Partition Energies and Distribution of Residues in Soluble Proteins. *Biophys. J.* **1985**, *47*, 61–70.
- (9) Bull, H. B.; Breese, K. Surface Tension of Amino Acid Solutions: A Hydrophobicity Scale of the Amino Acid Residues. *Arch. Biochem. Biophys.* **1974**, *161*, 665–70.
- (10) Moelbert, S.; Emberly, E.; Tang, C. Correlation Between Sequence Hydrophobicity and Surface-Exposure Pattern of Database Proteins. *Protein Sci.* **2004**, *13*, 752–762.
- (11) Chothia, C. The Nature of the Accessible and Buried Surfaces in Proteins. *J. Mol. Biol.* **1976**, *105*, 1–12.
- (12) Freed, A. S.; Garde, S.; Cramer, S. M. Molecular Simulations of Multimodal Ligand-Protein Binding: Elucidation of Binding Sites and Correlation With Experiments. *J. Phys. Chem. B* **2011**, *115*, 13320–13327.
- (13) Oelmeier, S.; Dismer, F.; Hubbuch, J. Gaining Mechanistic Understanding of Aqueous Two-Phase Systems for Bioseparation. *Chem. Ing. Tech.* **2012**, *84*, 1292–1292.
- (14) Bordoli, L.; Schwede, T. In *Homology Modeling*; Orry, A. J. W., Abagyan, R., Eds.; Methods in Molecular Biology 857; Humana Press: New York, 2012; pp 107–136.
- (15) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discovery Today* **2009**, *14*, 676–683.
- (16) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making Optimal Use of Empirical Energy Functions: Force-Field Parameterization in Crystal Space. *Proteins* **2004**, *57*, 678–83.
- (17) Krieger, E.; Koraimann, G.; Vriend, G. Increasing the Precision of Comparative Models With YASARA NOVA—a Self-Parameterizing Force Field. *Proteins* **2002**, *47*, 393–402.
- (18) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (19) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (20) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.
- (21) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–41.
- (22) Biswas, K. M.; DeVido, D. R.; Dorsey, J. G. Evaluation of Methods for Measuring Amino Acid Hydrophobicities and Interactions. *J. Chromatogr. A* **2003**, *1000*, 637–655, A Century of Chromatography 1903–2003.
- (23) Teeter, M. M. Water Structure of a Hydrophobic Protein at Atomic Resolution: Pentagon Rings of Water Molecules in Crystals of Crambin. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014–6018.
- (24) Némethy, G.; Scheraga, H. A. The Structure of Water and Hydrophobic Bonding in Proteins. III. The Thermodynamic Properties of Hydrophobic Bonds in Proteins. *J. Phys. Chem.* **1962**, *66*, 1773–1789.
- (25) Kim, J.; Petritis, K.; Shen, Y., II; Camp, D. G.; Moore, R. J.; Smith, R. D. Phosphopeptide Elution Times in Reversed-Phase Liquid Chromatography. *J. Chromatogr. A* **2007**, *1172*, 9–18.
- (26) Blonder, J.; Conrads, T. P.; Yu, L.-R.; Terunuma, A.; Janini, G. M.; Issaq, H. J.; Vogel, J. C.; Veenstra, T. D. A Detergent- and Cyanogen Bromide-Free Method for Integral Membrane Proteomics: Application to Halobacterium Purple Membranes and the Human Epidermal Membrane Proteome. *Proteomics* **2004**, *4*, 31–45.
- (27) Krokhin, O.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. An Improved Model for Prediction of Retention Times of Tryptic Peptides in Ion Pair Reversed-phase HPLC: Its Application to Protein Peptide Mapping by Off-Line HPLC-MALDI MS. *Mol. Cell. Proteomics* **2004**, *3*, 908–919.
- (28) YASARA: [www.yasara.org](http://www.yasara.org).
- (29) POVRay: [www.povray.org](http://www.povray.org).