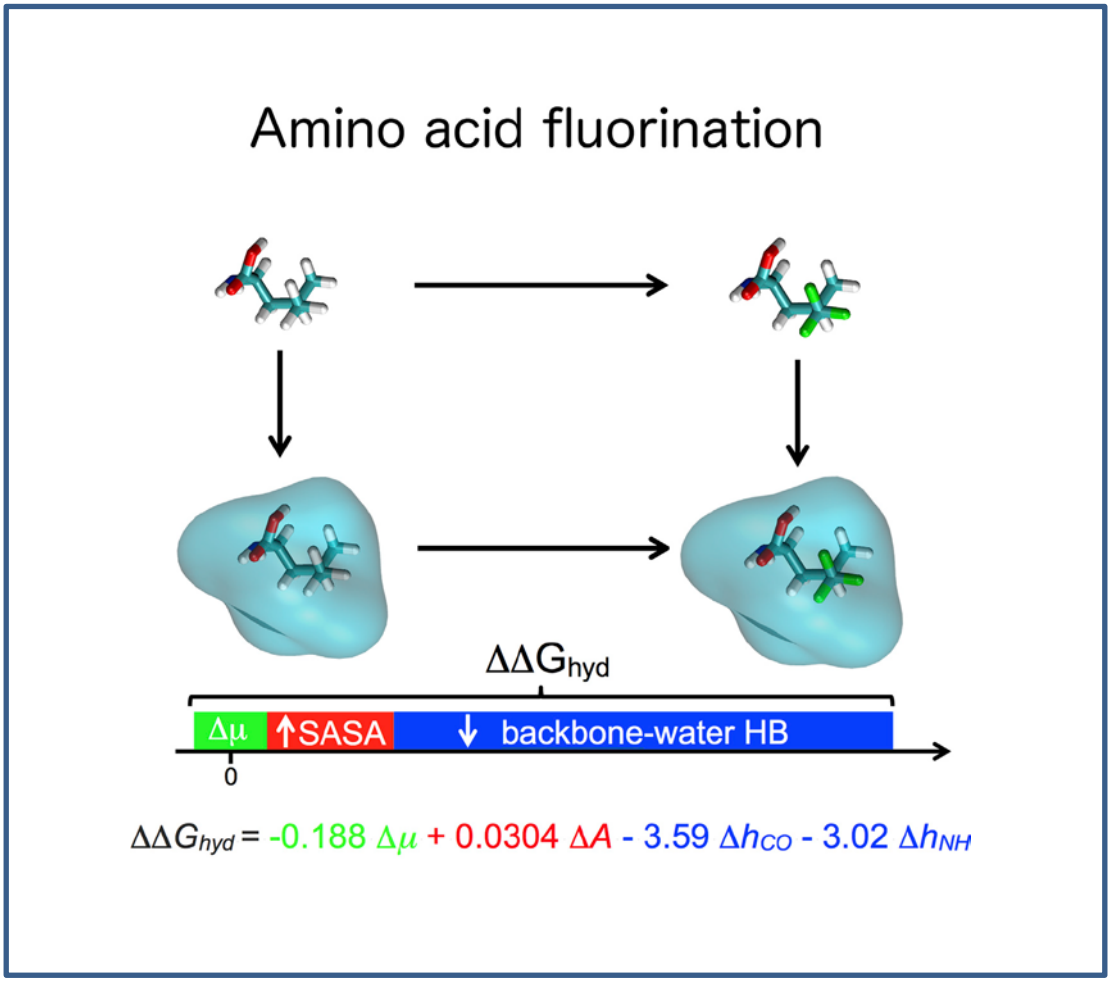
MPIKG Public Access
Author Manuscript

Published in final edited form as:
Robalo, J. R., Huhmann, S., Koksch, B., & Verde, A. V. (2017). The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids. Chem, 3(5), 881-897. doi:10.1016/j.chempr.2017.09.012.

The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids

João Ramiro Robalo, Susanne Huhmann, Beate Koksch,
Ana Vila Verde



This article may be used for non-commercial purposes in accordance with Cell Press Terms and Conditions for Self-Archiving.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids

João Ramiro Robalo,[1], Susanne Huhmann[2], Beate Koksch[2], Ana Vila Verde[1,3*]

[1] Department of Theory & Bio-systems, Max Planck Institute for Colloids and Interfaces, Science Park, Potsdam 14424, Germany.

[2] Department of Chemistry and Biochemistry, Freie Universität Berlin, Takustraße 3, 14195 Berlin, Germany

* Correspondence: ana.vilaverde@mpikg.mpg.de

[3] Lead contact

SUMMARY

The substitution of -CH_3 by -CF_3 groups in the side chain of hydrophobic amino acids often increases their hydrophobicity, but the impact of these substitutions on the thermal stability of proteins is system-specific. We investigate this issue using fixed-charge, all-atom molecular dynamics simulations and an AMBER-compatible library of fluorinated amino acids. We find that the changes in hydration free energy upon fluorination depend strongly on amino acid identity and location of the fluorinated site. We present a phenomenological model that quantitatively predicts the simulation results. The model demonstrates that changes in hydrophobicity upon fluorination largely arise from steric hindrance of backbone-water hydrogen bonds by -CF_3; changes in surface area, often invoked to explain experimental trends, have only a secondary contribution. The model and the force field presented here provide the indispensable molecular scale insight and simulation tools to understand and predict the impact of fluorination on protein properties.

Keywords: leucine, isoleucine, ethylglycine, valine, trifluoro amino acids, hexafluoro amino acids, TIP3P water, TIP4PEw water, free energy of solvation, protein thermal stability

Max Planck Institute of Colloids and Interfaces · Author Manuscript

INTRODUCTION

Fluorine, nearly unrepresented in biogenic compounds, has proven invaluable in the development of bioactive drugs: 20 to 25% of drugs contain at least one fluorine atom.[60] Fluorine substitutions are used frequently to improve the metabolic stability, bioavailability and interactions with relevant molecular partners of small organic molecules.[60] The success of fluorine substitutions to tune molecular properties is related to fluorine's unique physical-chemical properties. Fluorine's high electronegativity and low polarizability allows for inductive and hyper-conjugative effects to be modulated in biomolecules.[29,39,64] Simultaneously, fluorine's small atomic radius[15] suggests that substituting hydrogen by fluorine will be little disruptive in terms of molecular volume and geometry. Furthermore, by incorporating fluorine atoms it is possible to tune the solubility of a molecule in water, hydrocarbons or perfluorocarbons,[29,64] and to change its molecular dipolar moment.[40]

Given the enormous success that fluorine substitutions have had in tuning the properties of small molecules, it is natural to ask how this element can be used to tune the properties of proteins. Controlling protein thermal stability, resistance against proteolysis, affinity and specificity for ligands and efficiency of enzymatic activity is indispensable to bring promising protein-based drugs and devices to reality, *e.g.*, antimicrobial peptides as the next-generation antibiotics,[16,77] cell-penetrating peptides to carry drugs into the cell,[34] enzymes for industrial biocatalysis to produce biofuels, biomolecular photovoltaic devices to produce H_2 or soil nitrification under hypersaline conditions.[11,22,33] To explore the potential of fluorinated peptides and proteins, an extensive number of fluorinated amino acids has been investigated. Insertion of fluorinated amino acids into proteins has been accomplished both in synthetic constructs[38,40,51,69] and in naturally occurring peptides and proteins. [4,49,67,75] Those studies, focusing on characterizing the thermal and chemical stability of the fluorinated peptides, have demonstrated the potential of fluorine substitutions to tune protein properties. Introducing even a single fluorinated amino acid can change the melting temperature of proteins by tens of degrees,[38,40,41,61] alter proteolytic stability,[3] kinetics of peptide aggregation,[27] antimicrobial and hemolytic activity[54] and protein-ligand affinity.[53,75]

These experimental studies have also demonstrated that the effect of fluorination on protein properties cannot be easily understood and predicted. Fluorination of peptide ligands may increase or decrease their affinity for the protein substrate.[53] The resistance of peptides to proteolytic activity depends on the position of the fluorinated amino acid.[3,53] Substituting an amino acid, in the hydrophobic core of a protein, with a derivative which is fully-fluorinated at the side chain methyl groups, typically increases stability against temperature;[12,13,51,55,65,69] such an increase, however, is not always observed when

Max Planck Institute of Colloids and Interfaces · Author Manuscript

inserting a fluorinated derivative of a different amino acid.[38,40] Furthermore, the number of -CH_3 → -CF_3 substitutions has a measurable impact on the degree of thermal stabilization *per substitution*, and so has the localization of the substitution site.[17,44]

Achieving optimal protein thermal stability is critical for all potential applications of proteins. Controlling this property is best done by modifying the protein's hydrophobic core.[40,76] Control could be achieved by substituting one canonical, hydrophobic amino acid by another one, but this strategy has shortcomings: there are only 5 non-aromatic, hydrophobic amino acids, and each has different side chain geometry, which may alter packing in the hydrophobic core of the protein. Controlling protein thermal stability by -CH_3 → -CF_3 substitutions in the side chains of aliphatic, hydrophobic amino acids overcomes these shortcomings: many more fluorinated variants of amino acids exist, and if preserving the side chain geometry proves critical, a fluorinated variant of the same amino acid may be used.[18,51,76]

The difficulty in rationalizing the change in thermal stability induced by -CH_3 → -CF_3 substitutions in amino acids likely reflects multiple and interdependent mechanisms underlying fluorine–water and fluorine–protein interactions. Increases in thermal stability observed after replacing -CH_3 by -CF_3 have been attributed to the larger surface area of -CF_3: the larger area would result in a more positive (*i.e.*, less favorable) free energy of unfolding because of more water molecules being released from the hydration shell of the -CF_3 group upon burial in the hydrophobic core of the protein.[18,44,45] However, changes in side chain geometry and/or volume might also have the opposite effect on thermal stability, if they worsen packing in the protein's hydrophobic core.[19,30] Furthermore, for proteins and peptides with multiple fluorinated sites, the number of fluorine-fluorine contacts *versus* the number of hydrocarbon-fluorine contacts was also suggested to play a role in the stabilization of fluorinated proteins.[17] The possible existence of a "fluorous effect" – the preferential segregation of fluorinated groups from alkylated ones – in proteins remains controversial, being supported by some studies[13,14,40] and denied by others.[30,45,51] Finally, reports of strong slowdown of water molecules near fluorinated protein sites,[43] attributed to the larger dipole moment of the fluorinated side chains, suggest two other mechanisms by which fluorination might alter the thermal stability of proteins. If fluorination makes water–side chain interactions more favorable because of the larger dipole moment of the fluorinated side chain, then the thermal stability of fluorinated proteins would be reduced relative to their non-fluorinated parent via an enthalpic effect: it would be less favorable to bury the fluorinated side chain in the protein's hydrophobic core. However, if the larger dipole moment of the fluorinated side chain would in addition slow down nearby waters of hydration, then the burial of the fluorinated side chain in the hydrophobic core would result in a larger entropic gain. This entropic contribution would increase the stability of the fluorinated proteins relative to their non-fluorinated counterparts.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

At present, the possible contribution of all these mechanisms to changes in thermal stability upon fluorination remains unclear. This lack of knowledge severely hinders the efficient development of fluorinated proteins and peptides with the desired thermal stability for each application. It is therefore of interest to clarify the molecular mechanisms underlying changes in protein thermal stability upon fluorination. As highlighted above, this origin is intrinsically atomistic and multifold. In this context, molecular dynamics (MD) simulations are an invaluable tool because they can give simultaneous information on both the protein and the water, with atomistic resolution.

In this work we use molecular dynamics simulations to quantify the hydration free energy differences between tri- or hexa-fluorinated amino acid derivatives and their non-fluorinated variants, and to identify the mechanisms through which these differences arise. We focus on the hydration free energy because this observable is intrinsically related to the hydrophobic effect (*i.e.*, hydrophobic interaction). Hydrophobic molecules, defined as those with unfavorable (positive) hydration free energies, and hydrophilic molecules with hydrophobic patches typically experience net attraction in water – the hydrophobic effect.[5,6] This net attraction reflects the difference in hydration free energy between the separated molecules and that of the aggregate.[5,6] The hydrophobic effect is critical for protein folding and protein thermal stability,[70] so understanding it is indispensable for the usage of fluorination to tune the thermal stability of proteins. We analyze amino acid hydration free energies and we provide a quantitative phenomenological model to predict changes in free energy of hydration induced by $-CH_3 \rightarrow -CF_3$ substitutions. This model is general enough that it can be applied to any non-polar amino acid and requires as input data observables that can be easily calculated from short molecular dynamics simulations instead of time-consuming free energy calculations.

Performing this study requires reliable force fields for fluorinated amino acids. While several simulation studies of fluorinated peptides or proteins using classical, fixed charge, atomistic force fields already exist,[57,62] force fields for fluorinated amino acids have severe shortcomings. Firstly, these force fields are not fully compatible with commonly used protein force fields with which they are combined. Furthermore, Lennard-Jones parameters for fluorine, though abundant,[26,31,52,72] are inadequate for simulations of proteins in aqueous solution: they are either highly generic or parameterized to reproduce properties of pure fluorinated compounds or single molecules. To address these shortcomings and enable future studies of the impact of fluorination on protein structure and thermal stability, we present a force field for fluorinated compounds. This force field is specifically designed for fully fluorinated methyl or methylene groups (*i.e.*, $-CF_3$ or $-CF_2-$); a future paper will present accompanying parameters for hydrogen belonging to partially fluorinated groups (*e.g.*, $-CH_2F$, -CHF-). The Lennard-Jones (LJ) parameters for fluorine are optimized against the hydration free energy of CF_4 and the molar volume of an equimolar CH_4:CF_4 mixture.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

Parameterizing against these two properties is critical to capture the hydrophobic effect as well as packing constraints in the hydrophobic core of proteins. As mentioned above, both effects are expected to play a key role determining the structural stability of proteins. We use these parameters to create a library of amino acids fully compatible with the AMBER force field family, for the most commonly used hydrophobic fluorinated amino acids; see Figure 1. This library is given in AMBER format in the supplemental information file Robalo_et_al_force_field_parameters_amber.zip.

The quantitative insights acquired in this study provide the first-to-date comprehensive mechanism to the increased hydrophobicity of amino acids by the replacement of $-CH_3$ by $-CF_3$ groups. These insights, together with the force field we present, lay the foundations to understand the molecular mechanisms of changes in protein structure and thermal stability induced by fluorination.

RESULTS AND DISCUSSION

Hydration free energies of fluorinated amino acids strongly depend on amino acid identity and location of fluorination site

Experimental results from HPLC assays measuring retention times (references 24,27,62 and present work) indicate that $-CH_3 \rightarrow -CF_3$ substitutions increase retention times. This increase indicates that fluorination increases amino acid hydrophobicity, *i.e.*, leads to hydration free energies that increase in value, becoming more positive/less negative. This trend is also anticipated from our results, which indicate that CF_4 is more hydrophobic than CH_4 because of its larger surface area. To evaluate whether our force field for fluorinated amino acids correctly describes them as more hydrophobic than the parent canonical amino acid, we calculated the hydration free energies for each amino acid shown in Figure 1, using the force-field we developed (see Table 1, Figures S1, S2, Tables S1, S2, sections S1, S2.1 of the Supplemental Information (SI), and Procedures).

The differences in hydration free energy ($\Delta\Delta G_{Hyd}$) between the fluorinated amino acids and their non-fluorinated parent are shown in graphical form in Figure 2. The values of the free energy of hydration are also shown in tabular form in Table S3 of the SI. For all cases tested here, we find that $-CH_3 \rightarrow -CF_3$ substitutions make the hydration free energy less negative, *i.e.*, lead to an increase in amino acid hydrophobicity, in qualitative agreement with experiment. However, the increase in hydration free energy per $-CH_3 \rightarrow -CF_3$ substitution is not constant, but instead varies between 0.25 kcal/mol and 1.5 kcal/mol, depending on the identity of the amino acid and the stereochemistry of the fluorinated site. For both leucine and valine, the change in hydration free energy associated with the (R) isomer is three times larger

Max Planck Institute of Colloids and Interfaces · Author Manuscript

than that of the corresponding (S) isomer. For isoleucine, where fluorination of the gamma or delta carbons does not result in the formation of a chiral center, fluorination of either site leads to identical changes in hydration free energy. In all cases, the introduction of two -CF_3 groups has an additive effect: the $\Delta\Delta G_{Hyd}$ of hexafluorovaline and hexafluoroleucine is approximately equal to the sum of the $\Delta\Delta G_{Hyd}$ of the two corresponding trifluorinated variants.

The differences between amino acids, and between isomers of the same amino acid, we report here have not been observed in experiment. The likely reason is that the typical experimental observable used to compare hydrophobicity of amino acids – retention time measurements – is only a crude reporter of hydrophobicity, as we show below. Direct measurements of amino acid hydration free energies are not available.

The large variation we observe in $\Delta\Delta G_{Hyd}$ per -CH_3 → -CF_3 substitution is unexpected, in light of our results on the hydration free energy of CF_4 and CH_4 presented in Procedures and in section S2.1. Those results suggest that each -CH_3 → -CF_3 substitution in amino acids should lead to a change of approximately +0.5 kcal/mol in the hydration free energy. This estimate is based on the free energy difference between CF_4 and CH_4 (0.7 kcal/mol) and accounting for the difference in the number of hydrogen and fluorine atoms. The fact that in some cases, $\Delta\Delta G_{Hyd}$ can be as large as 1.5 kcal/mol/(- CF_3 group) suggests that factors other than the change in area also strongly influence how hydration free energies are affected by -CH_3 → -CF_3 substitutions.

In the following sections we examine the mechanisms by which fluorination changes the free energy of hydration of amino acids and quantify the magnitude of their contribution.

Non-constant changes in hydrophobicity per -CF_3 substitution are related to the Coulombic component of the hydration free energy

The hydration free energy of a solute can be described as the sum of three contributions.[68] In the context of our force field, these contributions are the Coulombic (or polar) contribution; the non-polar, repulsive contribution arising from the repulsive part of the Lennard-Jones potentials, in which the cost of forming a cavity in the solute is included; and the non-polar, attractive contribution, corresponding to the attractive component of the Lennard-Jones potentials.[68] To understand the origin of the variations in $\Delta\Delta G_{Hyd}$ per -CH_3 → -CF_3 substitution we observe, in Figure 3 and Tables S4 and S5 we separately show the Coulombic and the Lennard-Jones (attractive+repulsive) contributions to $\Delta\Delta G_{Hyd}$. We find that the Lennard-Jones component has an almost constant value per -CF_3 insertion. In contrast, the electrostatic contribution is site-dependent, being more favorable for the (S) rather than the (R) isomers of valine and leucine, thereby resulting in the overall differences between isomers.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

The increase in SASA due to -CF_3 substitutions does not fully explain the larger hydrophobicity of fluorinated amino acids

To understand the contribution of the solvent accessible surface area (SASA) to the change in free energy arising from -CH_3 → -CF_3 substitutions, we calculated the SASA of each amino acid. We found that fluorination changes the SASA in a non-constant manner (Figure 4 a and Table S6), with changes in area of 9 to 14 $Å^2$ being observed per -CH_3 → -CF_3 substitution, depending on the fluorination site. Even though the hydration free energies also change in a non-constant manner with fluorination, these energies correlate only weakly with SASA, as shown in Figure 4 a. This poor correlation is not surprising, in light of the importance of the Coulombic component for ΔG_{Hyd}, because the SASA is known to correlate very strongly only with the repulsive part of the Lennard-Jones component of ΔG_{Hyd}.[68] In Figure S3 we demonstrate that the correlation between SASA and the Lennard-Jones contribution of the hydration free energy is indeed much stronger than that shown in Figure 4 a. We emphasize that the free energy values correlated with the SASA in Figure S3 comprise the total LJ contribution. For that reason, the correlation between the two observables is weaker than those reported by others for the correlation between the cavity-formation contribution to the free energy and the SASA.[68]

The poor correlation between SASA and ΔG_{Hyd} and the much better correlation between SASA and $\Delta G_{Hyd,LJ}$ confirm that the different SASA of the various amino acids is not the main mechanism behind the changes in hydration free energy upon fluorination, as is often assumed.[18] Differences in SASA can also not explain the dependence of hydration free energies on the identity of the amino acid and the stereochemistry of the fluorinated site.

SASA and retention times A weak correlation between SASA and hydration free energy of amino acids has also been observed by others,[32] and does not reflect poor model quality. Evidence that there is no fundamental flaw in the models is presented in Figure 4 b, which demonstrates the excellent correlation between the experimental retention times of Fmoc-protected amino acids on a functionalized silica HPLC column (references 24,27,62 and present work) and the SASA calculated from our models. The strong correlation between SASA and experimental retention times (Figure 4 b), and the comparatively much weaker correlation between the retention times and our calculated hydration free energies (Figure 4 c) also suggests that, at least for molecules with comparable hydrophobicity, retention times are not determined by hydrophobicity itself, but by the contact area between a molecule and the hydrophobic chains in the HPLC substrate; this number is necessarily a function of the accessible surface area of the molecule. This conclusion is supported by the much stronger correlation

Max Planck Institute of Colloids and Interfaces · Author Manuscript

between measured retention times and the LJ component of the hydration free energy, shown in Figure S4, than with the total hydration free energy (Figure 4c).

-CF_3 substitutions increase amino acid hydrophobicity by decreasing the number of backbone-water hydrogen bonds

The formation of hydrogen bonds between water and the polar groups in the amino acid's backbone should be the predominant contribution of the backbone to changes in hydration free energy. The average number of hydrogen bonds between backbone amine and carbonyl groups and water is shown in Figure 5 and in Table S7. We used standard geometric criteria to identify hydrogen bonds: they were considered to form whenever the donor-acceptor distance is below 3.5 Å and the angle formed by the donor, hydrogen atom and acceptor ranges from 135° to 180°. We found that the total number of backbone-water hydrogen bonds decreases with fluorination, and correlates well with the amino acid hydration free energy (Figure 5). The decrease of backbone-water hydrogen bonds largely reflects the contribution of carbonyl-water hydrogen bonds (see Table S7 and Figure S5). Amine-water hydrogen bonds, in contrast, are much less abundant than carbonyl-water ones, and typically increase slightly upon fluorination, except for V3R and V6G. The same trends are observed when using more stringent criteria for hydrogen bond formation, see Table S8.

Close examination of the data in Table S7 and Figure 5 indicates that, in general, fluorination of the γ carbons leads to larger reduction of the number of backbone-water hydrogen bonds than fluorination of the δ carbons. This difference likely results from the closer proximity of the γ carbons to the backbone, that facilitates their direct interaction with the backbone and the consequent perturbation of backbone-water interactions.

-CF_3 substitutions modulate amino acid hydrophobicity by changing the electrical dipolar moment

The introduction of fluoromethyl groups in amino acids is expected to change their net dipole moment, and thus change amino acid interactions with water. To investigate whether changes in dipole moment also contribute to the observed trends, we calculated the global electrical dipolar moment of these molecules and correlated it with their hydration free energy; the results are compiled in Figure 6. Inspection of Figure 6 and of the distributions of dipolar moment magnitude shown in Figure S6 reveals that fluorination has a marked impact on the amino acid polarity, and that this impact is often non-intuitive. Even though fluorine is more electronegative than hydrogen, introduction of a -CF_3 group does not necessarily result in a molecule with a larger dipole moment. For example, fluorination of ETG

Max Planck Institute of Colloids and Interfaces · Author Manuscript

results in a lower average dipolar moment, derived from the appearance of a peak at low-magnitudes in the corresponding distribution; also, a lower average dipolar moment resulting from a broader band at small values of μ is observed for V3R and V6G when compared to valine; in contrast, V3S has a larger dipole moment than valine.

Figure 6 demonstrates that there is only a poor correlation between the average dipolar moment and the hydration free energies. The results that follow, however, show that this contribution is still important and should not be neglected.

A quantitative model to predict hydration free energies of fluorinated amino acids

Our results suggest the amino acid surface area, the number of backbone-water hydrogen bonds and the amino acid dipole moment all change upon fluorination. These changes must necessarily contribute to the observed non-intuitive changes in hydration free energy with fluorination (Figure 2), but the magnitude of each contribution is yet unclear. To quantitatively assess the contribution of each of these factors, we fitted the change in the free energies of hydration of the fluorinated amino acids with a linear multivariate model:

$$\Delta\Delta G_{Hyd} = k_1\Delta\mu + k_2\Delta A + k_3\Delta h_{CO} + k_4\Delta h_{NH} \quad (1)$$

where μ is the dipole moment, A is the solvent-accessible surface area, h_{CO} and h_{NH} are the number of hydrogen bonds between water and either carbonyl or amine and the differences are taken relative to the parent non-fluorinated amino acid; k_1 to k_4 are the fitting parameters. The first, third and fourth terms in the right hand side of Equation 1 principally reflect electrostatic contributions to $\Delta\Delta G_{Hyd}$; hydrogen bonds do have a Lennard-Jones component, but they are primarily of electrostatic nature. Electrostatic contributions associated with introducing CF_3 groups are included in the dipole term of Equation 1. In contrast, the second term in Equation 1 is associated with the change in area incurred on because of -CH_3 → -CF_3 substitutions. This term thus primarily contains two contributions, arising from changing both the repulsive and the attractive components of the Lennard-Jones potential when replacing hydrogen by fluorine. In the SI we show (section S2.7, Figures S8 and S9, Tables S10 and S11) that an analogous model can be used to predict the absolute free energies of hydration of the fluorinated and non-fluorinated amino acids studied here; there we also show P-values (Table S11), which indicate that the type of linear model we selected can be applied to our data. The P-values and the fact that the same type of linear model successfully predicts both ΔG_{Hyd} and $\Delta\Delta G_{Hyd}$ lends credence to our choice of model.

The value of k_2 can be estimated from the difference in areas and in hydration free energies between

Max Planck Institute of Colloids and Interfaces · Author Manuscript

CF_4 and CH_4: $k_2 = (\Delta G_{Hyd,CF_4} - \Delta G_{Hyd,CH_4})/(A_{CF_4} - A_{CH_4}) = 0.0304$ kcal/mol/Å^2; see Table 2 for the values of ΔG and section S2.1 for the values of the areas. The remaining parameter values are obtained by fitting to the $\Delta\Delta G_{Hyd}$ values calculated from free energy perturbation. The final fit parameters are shown in Table 3 together with the standard errors and P-values of the fitted coefficients. These parameters have reasonable physical values: the contribution of the water-carbonyl hydrogen bonds (k_3) is larger than that of the water-amine hydrogen bonds (k_4), and both are of the expected range for hydrogen bonds of this kind.[50] The correlation between model predictions and the results of the free energy perturbation simulations are shown in Figure 7.

Despite its simplicity, the model can quantitatively predict the change in amino acid free energies of hydration to within 0.3 kcal/mol for all cases except for L3S, for which it failed by >0.5 kcal/mol. Inspection of the quantities used as input for the model (μ, A, h_{CO} and h_{NH}) suggest L3S to be more hydrophobic than L3R, which opposes the free energy perturbation results. Given that the Coulombic contribution to $\Delta\Delta G_{Hyd}$ is the one that differs most between the two molecules (see Figure 3 and Table S4) and that the two isomers have the same atomic charges (allowing for the difference in chirality), we suspected that the unexpected $\Delta\Delta G_{Hyd,Coul}$ differences between L3S and L3R might have an entropic origin. Our calculations suggest this to be indeed the case. As described in section S2.6 of the SI, Figure S7 and Table S9, we find that the $-T\Delta S_{Hyd,Coul}$ term of the Coulombic component of the hydration free energy is 0.75±0.80 kcal/mol (at 298 K) larger for L3R than for L3S, very close to the 0.76 kcal/mol difference in total Coulombic hydration free energy between these two amino acids. This similarity suggests that the unexpected differences between L3R and L3S are entropic. Substantially reducing the uncertainty associated with the entropic term would be necessary to unequivocally establish this conclusion, but is at present computationally challenging. We emphasize that the $-T\Delta S_{Hyd,Coul}$ term is *not* related to the entropic cost of forming a cavity, which is accounted for by the Lennard-Jones component of the hydration free energy, *i.e.*, by the area term in Equation 1. By definition, the difference in $\Delta G_{Hyd,Coul}$ between L3S and L3R arises from different solute-solvent electrostatic interactions for these amino acids. We have not investigated the mechanism of this different interaction because existing methods to decompose entropic contributions, such as grid cell theory,[28] are approximate and have comparable uncertainty to the free energy difference in question.[28] Because the reason for the peculiar behavior of L3S is not known, the model (Equation 1) does not include a term to capture this effect, and hence it works less well for L3S. For this reason, the final fit shown in Figure 7, does not include L3S. The model may still be used to obtain coarse estimates for L3S provided that the extra entropic contribution is added *a posteriori*.

Figure 8 clarifies the origin of the non-intuitive variation in the hydration free energies induced by $-CH_3 \rightarrow -CF_3$ substitutions in amino acids. This figure contains the contribution of each term in Equation 1 to the

Max Planck Institute of Colloids and Interfaces · Author Manuscript

free energy difference between each fluorinated amino acid and its non-fluorinated parent. The various terms contribute to the observed free energies of hydration to a different extent, depending on the identity of the amino acid and the characteristics of the fluorinated site (*i.e.*, formation of a chiral center after fluorination and its configuration). For all cases, the largest contribution to $\Delta\Delta G_{Hyd}$ is the change in the number of backbone-water hydrogen bonds. The second largest contribution is the change in area in the amino acid side chain. The dipole has the smallest contribution of all to the hydration free energy, but it is nevertheless important. For example, ILE derivatives exhibit remarkably different dipolar moments, thereby justifying their comparable $\Delta\Delta G_{Hyd}$: even though I3G's backbone makes fewer hydrogen bonds with water than I3D, its average dipolar moment is *circa* 1.5 Debye greater and compensates for that difference. The 1 kcal/mol increase in free energy of hydration upon fluorination of ETG – much larger than that expected based on the area increase upon fluorination – is also consistent with the decrease in backbone-water hydrogen bonds shown in Table S7 of the SI, and has only a small contribution from a lower dipole moment. The (R) isomer of trifluorovaline is 1 kcal/mol more hydrophobic than the (S) isomer, despite having the same SASA. This trend is consistent with the fact that the R isomer forms fewer backbone-water hydrogen bonds than the S isomer, and also has a lower dipole moment.

The large variability in the weight of each contribution to the $\Delta\Delta G_{Hyd}$ of fluorinated amino acids illustrates the need for a quantitative model to understand how fluorination will alter the hydration free energy, and, consequently, the thermal stability of proteins. Although intuitive, the individual assessment of the dipole moment and the surface area – the typically invoked rules-of-thumb to explain the experimentally observed changes in thermal stability of proteins – is not enough to understand a multi-faceted observable such as the free energy of hydration.

The model given by Equation 1 may be used, together with short molecular dynamics simulations, to rapidly estimate how differences in the hydrophobicity of fluorinated and non-fluorinated amino acids contribute to changes in the stability of proteins upon fluorination. This contribution has two components: from the unfolded ensemble and from the folded state. The change in free energy arising from the unfolded state ensemble has the magnitude (per fluorinated amino acid) given in Figure 2. This estimate assumes that the unfolded state ensemble is largely hydrated and is similar for fluorinated and non-fluorinated proteins. These assumptions are reasonable, at least for peptides and small proteins with only a few fluorinated sites.[47] The change in free energy arising from the folded state can be quickly estimated by performing short molecular dynamics simulations of the folded fluorinated and non-fluorinated protein, to calculate the observables (number of backbone- water hydrogen bonds, SASA and dipole moment of the mutated residues) necessary as input for the analytical model (Equation 1).

Max Planck Institute of Colloids and Interfaces · Author Manuscript

CONCLUDING REMARKS

The structural thermal stability of a protein is determined by the free energy difference between the folded and the unfolded forms. A key aspect determining how fluorination alters the thermal stability of proteins where CH_3 groups are replaced by CF_3 groups is thus the difference in the hydration free energy of fluorinated and non-fluorinated amino acids. These differences have been explained solely in terms of qualitative rules-of-thumb invoking two competing effects: the larger dipole moment and larger surface area of the fluorinated amino acids versus their non-fluorinated parents. Using molecular dynamics simulations and a simple, analytical model, we show for the first time that this explanation is insufficient. Fluorination alters the hydration free energy of amino acids primarily by changing the number of backbone-water hydrogen bonds that the amino acid may form. The contribution of changes in surface area of the amino acid is sizeable, but always smaller than the contribution of the hydrogen bonds; the often-invoked contribution of changes in dipole moment is the smallest of the three.

The quantitative, phenomenological model we present predicts the difference in the hydration free energy between the most common tri-/hexa-fluorinated amino acids and their parent amino acids, to within 0.3 kcal/mol; the only exception is (S)-trifluoroleucine, for which the deviation is 0.7 kcal/mol because of entropic effects associated with this amino acid and not accounted for by this model. The accuracy of the model, and the AMBER-compatible force fields for fluorinated amino acids we present, enable researchers to easily estimate the contribution of differences in hydrophobicity between fluorinated and non-fluorinated amino acids to the changes in protein stability. These tools are indispensable to interpret experimental observations, and ultimately to guide the development of fluorinated peptides and proteins with the desired thermal and structural stability.

PROCEDURES

Force Fields

The TIP4P-Ew[37] water model was used in this work. For fluorine, we use parameters (ε_{FF} and σ_{FF}; see Table 2) optimized against the free energy of hydration of CF_4 and the molar volume of a 1:1 mixture of CH_4:CF_4, as described in detail in section S2.1 (Figure S2) of the SI. The parameters reproduce the target parameterization properties very well; in addition, they also reproduce the vaporization enthalpy of an equimolar CH_4:CF_4 mixture (see Table 3). The optimized parameters for fluorine were found to be valid also for the TIP3P[42] water model.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

Our optimized LJ parameters are used for the fluorine atoms of all fluorinated amino acids; for the remaining atom types, the AMBER FF14SB[48] LJ parameters are used. All fluorinated amino acids use AMBER FF14SB force field parameters for the bonds, angle and dihedral potentials of their canonical analogue; ethylglycine and its fluorinated versions were built out of a leucine backbone. Whenever bonded or non-bonded parameters were lacking, as in the case of side chain dihedrals where hydrogen was substituted by fluorine, the corresponding parameters for hydrogen were taken, thereby assuming the rotation of a -CF_3 group not to drastically differ from that of a -CH_3 group.

The partial charges for the fluorinated amino acids are obtained via a multi-configuration RESP fitting procedure, described in more detail in the supplemental information. Typically, an AMBER- compatible charge-fitting procedure is done using only two configurations (one α-helical and one β-sheet) per amino acid. We found, however, that this procedure is not optimal for fluorinated amino acids because the charge distribution strongly depends also on the particular side chain configuration used for the charge fit. To reduce the bias of the charge distribution to any particular configuration, we opted to perform the charge fit on 200 conformations (100 α-helical and 100 β-sheet) per amino acid.

Simulation details

We calculated the hydration free energy of each amino acid using Free Energy Perturbation (FEP) and Bennett Acceptance Ratio[8,63] (BAR). In FEP, the solute is progressively decoupled from the solvent in two steps. Initially, only the Coulombic interactions are turned off, by running multiple parallel simulations, each corresponding to a state of the coupling parameter λ ($0 \leq \lambda_C \leq 1$). After the electrostatic interactions between solute and solvent are disabled, Lennard-Jones interactions are turned off, this time by introducing a soft-core potential[10,58] which smooths the LJ potential progression to zero with increasing λ ($0 \leq \lambda_{LJ} \leq 1$). Coulomb-decoupling was done using 21 steps and LJ-decoupling used 59 steps; the values of the coupling parameters are given in section S1.4 of the SI .

The simulation box used is illustrated in Figure S1 d; periodic boundary conditions in all directions were used. The systems were assembled using the built-in tools in Gromacs.[1,9,36,46,56,59,71] The software ACPYPE[23] was used to create Gromacs-compatible structure and topology files from those in AMBER[20]-readable format. Simulations used a time-step of 2 fs and constraints (LINCS[35]) were applied to all bonds involving hydrogen atoms. Integration of the equations of motion was done using a leap-frog Langevin dynamics algorithm with a collision frequency of 1 ps^{-1} set at the desired temperature. Van der Waals interactions were modeled with a cutoff of 1.2 nm and long-range dispersion corrections were applied to

Max Planck Institute of Colloids and Interfaces · Author Manuscript

both pressure and energy. Long-range electrostatics were treated with the PME [25] scheme with a 1.2 nm cutoff, a grid spacing of 0.1 nm and a 4th order interpolation. Simulations consisted of a steepest descent minimization, a L-BFGS minimization, a 100 ps NVT heating from 0 K to 298 K, and a 100 ps NPT equilibration. The Monte Carlo barostat[2,20] was used with a relaxation time of 1 ps for an isotropic coupling of system pressure to 1 bar. The production step was run in the NPT ensemble and lasted 2 ns per λ value.

Energies were collected from the individual simulations using the BAR method as implemented in Gromacs, with the final free energy difference being the additive inverse of the sum of the free energy differences for each sequential simulation pair. The convergence of the calculation was assessed by inspecting the difference in entropy calculated for each state's configuration using the Hamiltonians of the adjacent states; the difference was considered to be sufficiently close to zero – thus indicating that both configurations were similar enough to allow for a correct sampling – when it did not exceed 0.3 kT (0.2 kcal/mol; k=Boltzmann constant, T = 298 K).

Retention times

The impact of fluorination on the amino acid hydrophobicity was determined with the help of an RP-HPLC assay established previously.[62] A LaChrom-ELITE-HPLC-system containing an organizer, two pumps L-2130 with solvent degaser, a diode array flow detector L-2455 and an autosampler L-2200 (VWR International Hitachi, Darmstadt, Germany) was used. The Nα-Fmoc protected amino acids were dissolved in a mixture of 40 % acetonitrile (ACN) in deionized water containing 0.1 % (v/v) trifluoroacetic acid (TFA) and their retention times were determined on a C18 column (Capcell PAK C18, 5 μm; Shiseido Co., Ltd., Tokio, Japan). As eluents water and ACN, both containing 0.1 % (v/v) TFA were utilized. A flow rate of 1.0 mL/min was used, and a linear gradient from 40 to 70 % ACN + 0.1 % (v/v) TFA in 20 min was applied at room temperature. All experiments were performed in triplicate. The data were analyzed with EZ Chrom ELITE software (version 3.3.2, Agilent Technologies, Santa Clara, CA, United States).

AUTHOR CONTRIBUTIONS

AVV: project conception, guidance for data analysis and interpretation, writing; JRR: simulations, data analysis and interpretation, writing; BK, SH: experimental data, writing.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

ACKNOWLEDGMENTS

A.V.V. and J.R.R. acknowledge the International Max Planck Research School (IMPRS) on Multiscale Bio-Systems for financial support to perform this work, and CECAM for financial support to disseminate it in CECAM workshops; we thank Marco Ehlert and René Genz from the IT team of the MPIKG for their help. S.H. and B.K. thank the Deutsche Forschungsgemeinschaft in the context of the Research Training Group 1582 “Fluorine as Key Element” for financial support.

References

1 Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1*, 19–25.

2 Allen, M. P. and Tildesley, D. J. (1989). Computer Simulation of Liquids. Oxford University Press.

3 Asante, V., Mortier, J., Schlüter, H. and Koksch, B. (2013). Impact of fluorination on proteolytic stability of peptides in human blood plasma. Bioorg. Med. Chem. *21*, 3542–3546.

4 Bae, J. H., Paramita Pal, P., Moroder, L., Huber, R. and Budisa, N. (2004). Crystallographic evidence for isomeric chromophores in 3-fluorotyrosyl-green fluorescent protein. ChemBioChem *5*, 720–722.

5 Ben-Amotz, D. (2015). Hydrophobic Ambivalence: Teetering on the Edge of Randomness. J. Phys. Chem. Lett. *6*, 1696–1701.

6 Ben-Amotz, D. (2016). Water-Mediated Hydrophobic Interactions. Annu. Rev. Phys. Chem. *67*, 617–638.

7 Ben-Naim, A. and Marcus, Y. (1984). Solvation thermodynamics of nonionic solutes. J. Chem. Phys. *81*, 2016–2027.

8 Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. J. Comput. Phys. *22*, 245–268.

9 Berendsen, H., van der Spoel, D. and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43 – 56.

10 Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. and van Gunsteren, W. F. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem. Phys. Lett. *222*, 529–539.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

11 Biava, H. and Budisa, N. (2014). Evolution of fluorinated enzymes: An emerging trend for biocatalyst stabilization. Eng. Life Sci. *14*, 340–351.

12 Bilgiçer, B., Fichera, A. and Kumar, K. (2001). A coiled coil with a fluorous core. J. Am. Chem. Soc. *123*, 4393–4399.

13 Bilgiçer, B. and Kumar, K. (2002). Synthesis and thermodynamic characterization of self-sorting coiled coils. Tetrahedron *58*, 4105–4112.

14 Bilgiçer, B., Xing, X. and Kumar, K. (2001). Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores. J. Am. Chem. Soc. *123*, 11815–11816.

15 Bondi, A. (1964). van der Waals volumes and radii. J. Phys. Chem. *68*, 441–451.

16 Brogden, K. A. (2005). Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. *3*, 238–250.

17 Buer, B. C., de la Salud-Bea, R., Al Hashimi, H. M. and Marsh, E. N. G. (2009). Engineering protein stability and specificity using fluorous amino acids: the importance of packing effects. Biochemistry *48*, 10810–10817.

18 Buer, B. C., Levin, B. J. and Marsh, E. N. G. (2012a). Influence of Fluorination on the Thermodynamics of Protein Folding. J. Am. Chem. Soc. *134*, 13027–13034.

19 Buer, B. C., Meagher, J. L., Stuckey, J. A. and Marsh, E. N. G. (2012b). Structural basis for the enhanced stability of highly fluorinated proteins. Proc. Natl. Acad. Sci. USA *109*, 4810–4815.

20 Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham, T., Darden, T., Duke, R., Gohlke, H., Goetz, A., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry I., Kovalenko, A., Lee, T., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K., Paesani, F., Roe, D., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C., Smith, W., Swails, J., Walker, Wang, J., Wolf, R., Wu, X. and Kollman, P. (2014). Amber 14.

21 Croll, I. M. and Scott, R. L. (1958). Fluorocarbon Solutions at Low Temperatures. III. Phase Equilibria and Volume Changes in the CH_4-CF_4System. J. Phys. Chem. *62*, 954–957.

22 Cui, Y.-W., Zhang, H.-Y., Ding, J.-R. and Peng, Y.-Z. (2016). The effects of salinity on nitrification using halophilic nitrifiers in a Sequencing Batch Reactor treating hypersaline wastewater. Sci. Rep. *6*, 24825.

23 da Silva, A. W. S. and Vranken, W. F. (2012). ACPYPE-Antechamber python parser interface. BMC Research Notes *5*, 1.

24 Erdbrink, H., Nyakatura, E. K., Huhmann, S., Gerling, U. I., Lentz, D., Koksch, B. and Czekelius, C. (2013). Synthesis of enantiomerically pure (2S, 3S)-5, 5, 5-trifluoroisoleucine and (2R, 3S)-5, 5, 5-trifluoro-allo-isoleucine. Beilstein J. Org. Chem. *9*, 2009–2014.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

25 Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577–8593.

26 Fioroni, M., Burger, K., Mark, A. E. and Roccatano, D. (2000). A new 2,2,2-trifluoroethanol model for molecular dynamics simulations. J. Phys. Chem. B *104*, 12347–12354.

27 Gerling, U. I., Salwiczek, M., Cadicamo, C. D., Erdbrink, H., Czekelius, C., Grage, S. L., Wadhwani, P., Ulrich, A. S., Behrends, M., Haufe, G. and Koksch, B. (2014). Fluorinated amino acids in amyloid formation: a symphony of size, hydrophobicity and *α*-helix propensity. Chem. Sci. *5*, 819–830.

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

29 Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J. and Meanwell, N. A. (2015). Applications of fluorine in medicinal chemistry. J. Med Chem. *58*, 8315–8359.

30 Gottler, L. M., de la Salud-Bea, R. and Marsh, E. N. G. (2008). The Fluorous Effect in Proteins: Properties of α_4F_6, a 4-*α*-Helix Bundle Protein with a Fluorocarbon Core. Biochemistry *47*, 4484– 4490.

31 Gough, C. A., Debolt, S. E. and Kollman, P. A. (1992). Derivation of fluorine and hydrogen atom parameters using liquid simulations. J. Comput. Chem. *13*, 963–970.

32 Hajari, T. and van der Vegt, N. F. (2015). Solvation thermodynamics of amino acid side chains on a short peptide backbone. J. Chem. Phys. *142*, 144502.

33 Harris, B. J., Cheng, X. and Frymier, P. (2016). Structure and Function of Photosystem I-[FeFe] Hydrogenase Protein Fusions: An All-Atom Molecular Dynamics Study. J. Phys. Chem. B *120*, 599– 609.

34 Henriques, S. T., Melo, M. N. and Castanho, M. A. R. B. (2006). Cell-penetrating peptides and antimicrobial peptides: how different are they? Biochem. J. *399*, 1–7.

35 Hess, B., Bekker, H., Berendsen, H. J. C. and Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463–1472.

36 Hess, B., Kutzner, C., Van Der Spoel, D. and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. *4*, 435–447.

37 Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L. and Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P- Ew. J. Chem. Phys. *120*, 9665–9678.

38 Huhmann, S., Nyakatura, E. K., Erdbrink, H., Gerling, U. I., Czekelius, C. and Koksch, B. (2015). Effects of single substitutions with hexafluoroleucine and trifluorovaline on the hydrophobic core formation of a heterodimeric coiled coil. J. Fluorine Chem. *175*, 32–35.

39 Jäckel, C. and Koksch, B. (2005). Fluorine in peptide design and protein engineering. Eur. J. Org.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

Chem. *2005*, 4483–4503.

40 Jäckel, C., Salwiczek, M. and Koksch, B. (2006). Fluorine in a native protein environment–how the spatial demand and polarity of fluoroalkyl groups affect protein folding. Angew. Chem. Int. Ed. *45*, 4198–4203.

41 Jäckel, C., Seufert, W., Thust, S. and Koksch, B. (2004). Evaluation of the molecular interactions of fluorinated amino acids with native polypepticles. ChemBioChem *5*, 717–720.

42 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

43 Kwon, O.-H., Yoo, T. H., Othon, C. M., Van Deventer, J. A., Tirrell, D. A. and Zewail, A. H. (2010). Hydration dynamics at fluorinated protein surfaces. Proc. Natl. Acad. Sci. USA *107*, 17101–17106.

44 Lee, H.-Y., Lee, K.-H., Al-Hashimi, H. M. and Marsh, E. N. G. (2006). Modulating protein structure with fluorous amino acids: increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. J. Am. Chem. Soc. *128*, 337–343.

45 Lee, K.-H., Lee, H.-Y., Slutsky, M. M., Anderson, J. T. and Marsh, E. N. G. (2004). Fluorous effect in proteins: de novo design and characterization of a four-*α*-helix bundle protein containing hexafluoroleucine. Biochemistry *43*, 16277–16284.

46 Lindahl, E., Hess, B. and van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. *7*, 306–317.

47 Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. and Shaw, D. E. (2012). Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. J. Am. Chem. Soc. *134*, 3787–3791.

48 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696–3713.

49 Maisch, D., Wadhwani, P., Afonin, S., Böttcher, C., Koksch, B. and Ulrich, A. S. (2009). Chemical labeling strategy with (R)-and (S)-trifluoromethylalanine for solid state 19F NMR analysis of peptaibols in membranes. J. Am. Chem. Soc. *131*, 15596–15597.

50 Maréchal, Y. (2006). The hydrogen bond and the water molecule: The physics and chemistry of water, aqueous and bio-media. Elsevier.

51 Marsh, E. N. G. (2014). Fluorinated proteins: from design and synthesis to structure and stability. Accounts Chem. Res. *47*, 2878–2886.

52 Mayo, S. L., Olafson, B. D. and Goddard, W. A. (1990). DREIDING: a generic force field for molecular simulations. J. Phys. Chem. *94*, 8897–8909.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

53 Meng, H., Krishnaji, S. T., Beinborn, M. and Kumar, K. (2008). Influence of Selective Fluorination on the Biological Activity and Proteolytic Stability of Glucagon-like Peptide-1. J. Med. Chem. *51*, 7303–7307.

54 Meng, H. and Kumar, K. (2007). Antimicrobial Activity and Protease Stability of Peptides Containing Fluorinated Amino Acids. J. Am. Chem. Soc. *129*, 15615–15622.

55 Montclare, J. K., Son, S., Clark, G. A., Kumar, K. and Tirrell, D. A. (2009). Biosynthesis and Stability of Coiled-Coil Peptides Containing (2S,4R)-5,5,5-Trifluoroleucine and (2S,4S)-5,5,5-Trifluoroleucine. ChemBioChem *10*, 84–86.

56 Páll, S., Abraham, M. J., Kutzner, C., Hess, B. and Lindahl, E. (2014). Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In International Conference on Exascale Applications and Software pp. 3–27, Springer International Publishing.

57 Pendley, S. S., Yu, Y. B. and Cheatham, T. E. (2009). Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. Proteins: Struct., Funct., Bioinf. *74*, 612–629.

58 Pitera, J. W. and van Gunsteren, W. F. (2002). A comparison of non-bonded scaling approaches for free energy calculations. Mol. Sim. *28*, 45–65.

59 Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics *29*, 845–854.

60 Purser, S., Moore, P. R., Swallow, S. and Gouverneur, V. (2008). Fluorine in medicinal chemistry. Chem. Soc. Rev. *37*, 320–330.

61 Salwiczek, M., Samsonov, S., Vagt, T., Nyakatura, E., Fleige, E., Numata, J., Cölfen, H., Pisabarro, M. T. and Koksch, B. (2009). Position-Dependent Effects of Fluorinated Amino Acids on the Hydrophobic Core Formation of a Heterodimeric Coiled Coil. Chem. Eur. J. *15*, 7628–7636.

62 Samsonov, S. A., Salwiczek, M., Anders, G., Koksch, B. and Pisabarro, M. T. (2009). Fluorine in protein environments: a QM and MD study. J. Phys. Chem. B *113*, 16400–16408.

63 Shirts, M. R., Bair, E., Hooker, G. and Pande, V. S. (2003). Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. Phys. Rev. Lett. *91*, 140601.

64 Smart, B. E. (2001). Fluorine substituent effects (on bioactivity). J. Fluorine Chem. *109*, 3–11.

65 Son, S., Tanrikulu, I. C. and Tirrell, D. A. (2006). Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. ChemBioChem *7*, 1251–1257.

66 Song, W., Rossky, P. J. and Maroncelli, M. (2003). Modeling alkane+ perfluoroalkane interactions using all-atom potentials: Failure of the usual combining rules. J. Chem. Phys. *119*, 9145–9162.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

67 Steiner, T., Hess, P., Bae, J. H., Wiltschi, B., Moroder, L. and Budisa, N. (2008). Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. PLoS One *3*, e1680.

68 Tan, C., Tan, Y.-H. and Luo, R. (2007). Implicit Nonpolar Solvent Models. J. Phys. Chem. B *111*, 12263–12274.

69 Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, W. A., DeGrado, W. F. and Tirrell, D. A. (2001). Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine. Biochemistry *40*, 2790–2796.

70 Thirumalai, D., O'Brien, E. P., Morrison, G. and Hyeon, C. (2010). Theoretical Perspectives on Protein Folding. Annu. Rev. Biophys. *39*, 159–183.

71 Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. J. Comput. Chem. *26*, 1701–1718.

72 Vymětal, J. and Vondrášek, J. (2014). Parametrization of 2,2,2-trifluoroethanol based on the generalized AMBER force field provides realistic agreement between experimental and calculated properties of pure liquid as well as water-mixed solutions. J. Phys. Chem. B *118*, 10390–10404.

73 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. and Case, D. A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157–1174.

74 Wilhelm, E., Battino, R. and Wilcock, R. J. (1977). Low-pressure solubility of gases in liquid water. Chem. Rev. *77*, 219–262.

75 Ye, S., Loll, B., Berger, A. A., Mülow, U., Alings, C., Wahl, M. C. and Koksch, B. (2015). Fluorine teams up with water to restore inhibitor activity to mutant BPTI. Chem. Sci. *6*, 5246–5254.

76 Yoder, N. C. and Kumar, K. (2002). Fluorinated amino acids in protein design and engineering. Chem. Soc. Rev. *31*, 335–341.

77 Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. Nature *415*, 389–395.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

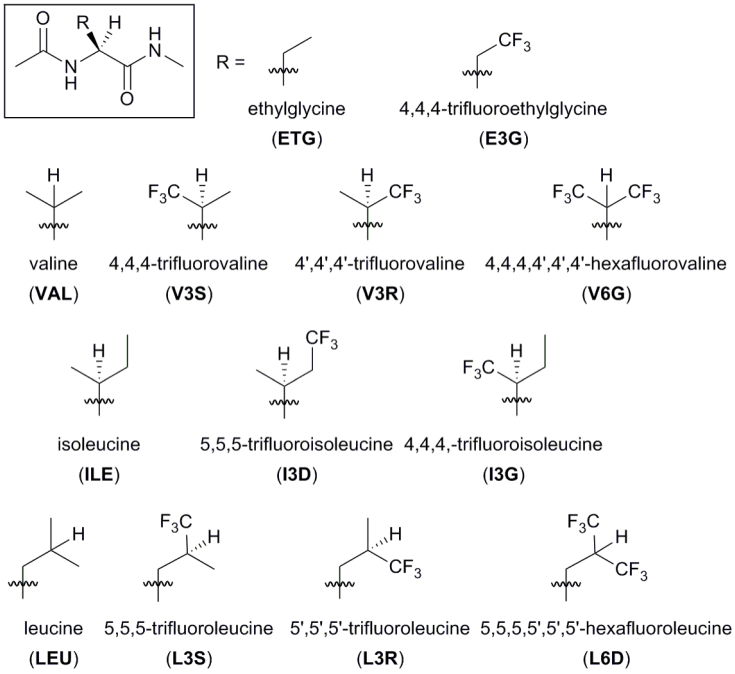


Fig. 1 Structure and nomenclature of the amino acids studied in this work. Amino acids parameterized and studied here. Each amino acid residue is capped at the N-terminus with an acetate group (ACE, -$COCH_3$, left-hand side of the inset structure) and capped at the C-terminus with an N-methyl group (NME, -$NHCH_3$, right-hand side of the inset structure). Abbreviations for fluorinated amino acids follow a three-character nomenclature: initial character of parent (non-fluorinated) amino acid name (E, V, I, L); number of fluorine atoms (3, 6); fluorination site (delta carbon as D, gamma carbon as G or, in the case of chiral center formation following fluorination, R or S).

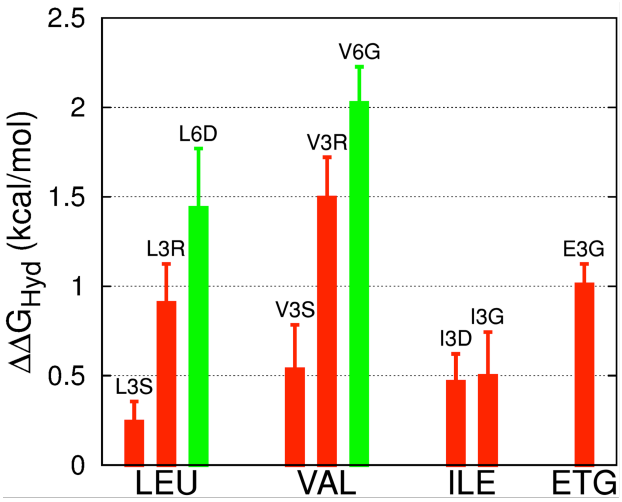


Fig. 2 Difference between the hydration free energy of fluorinated and non-fluorinated amino acids. Differences in hydration free energy ($\Delta\Delta G_{Hyd}$) for each series of amino acids (LEU, VAL, ILE, ETG); in each series the non-fluorinated amino acid is taken as reference; tri-fluorinated amino acids are shown in red, hexa-fluorinated ones are shown in green. Data are represented as mean ± standard deviation. See also Table S3.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

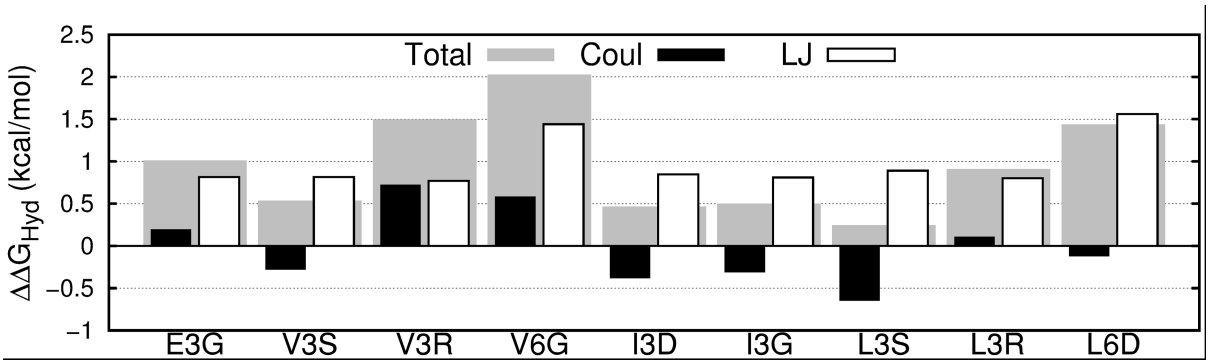


Fig. 3 Coulombic and Lennard-Jones contributions to the hydration free energy. Coulombic contribution (Coul), Lennard-Jones contribution (LJ) and sum of the contributions (total) to the differences in hydration free energy ($\Delta\Delta G_{Hyd}$) between the fluorinated and the non-fluorinated amino acids. See also Tables S4 and S5.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

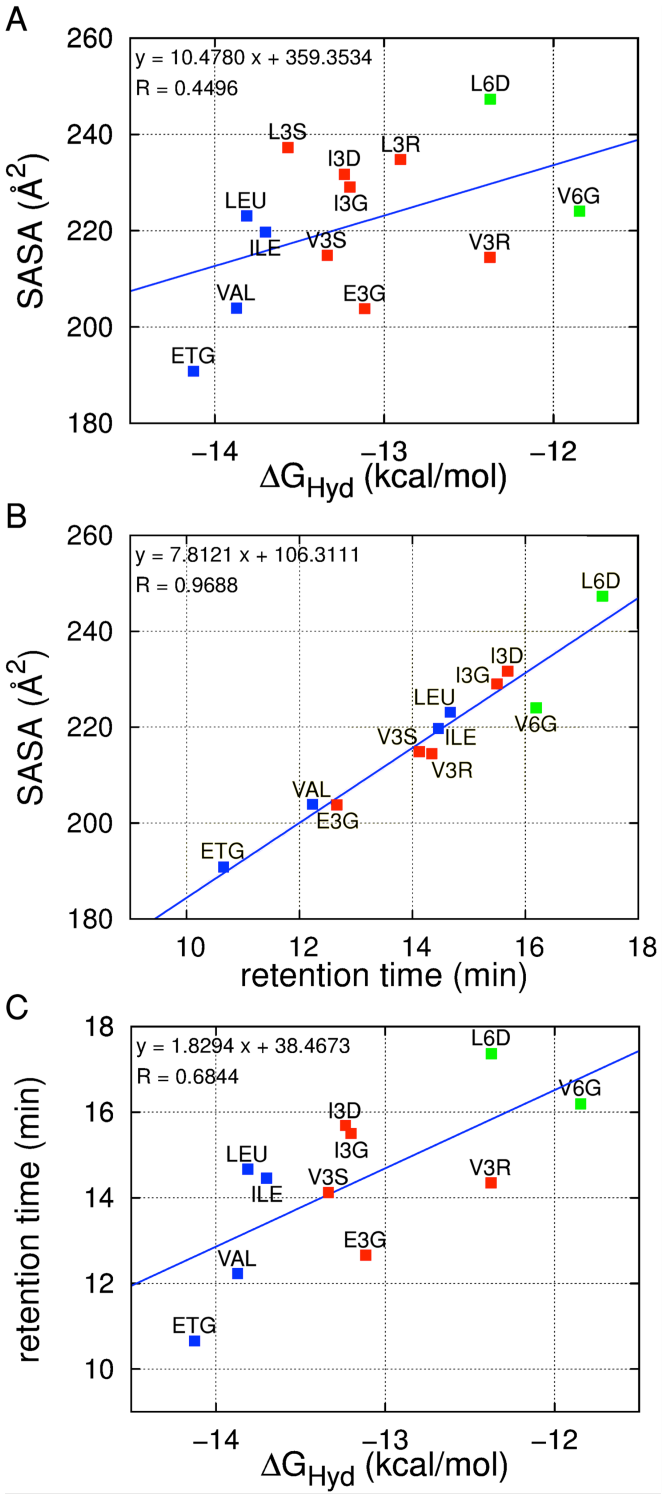


Fig. 4 Correlation between the hydration free energy and either the surface area or the retention time of the amino acids; correlation between the surface area and the retention time. A) Solvent accessible surface area (SASA) *vs* the calculated hydration free energy (ΔG_{Hyd}); B) SASA *vs* experimental retention times for each amino acid and C) experimental retention time *vs* calculated ΔG_{Hyd}. Blue=non-fluorinated, red=tri- fluorinated, green=hexa-fluorinated amino acids. The blue line is a linear fit to the data, with the corresponding equation and regression coefficient shown in each panel. Retention times from references 24,27,62 and present work. See also Figure S3 and S4 and Table S6.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

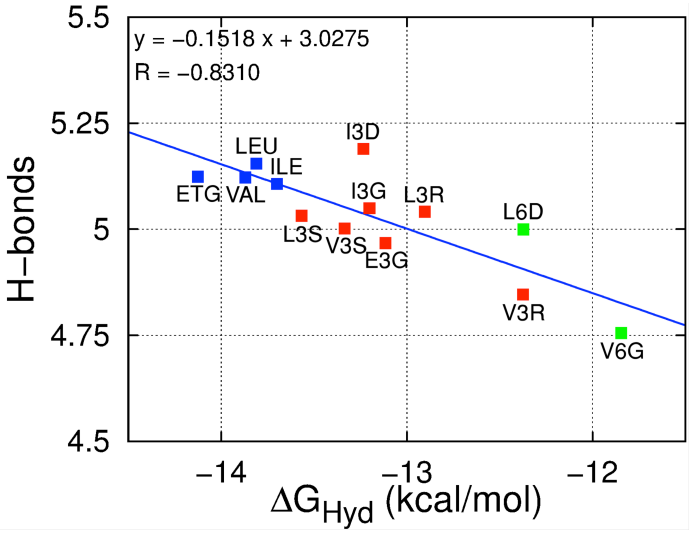


Fig. 5 Average number of backbone – water hydrogen bonds *vs.* the hydration free energy. Number of hydrogen bonds (H-bonds) between water and both the backbone's amine groups (donor and acceptor) and carbonyl groups (acceptor) *vs* the calculated hydration free energy (ΔG_{Hyd}) for each amino acid. Blue=non-fluorinated, red=tri- fluorinated, green=hexa-fluorinated amino acids. The blue line is a linear fit to the data, with the corresponding equation and regression coefficient shown in the graph. See also Figure S5 and Tables S7 and S8.

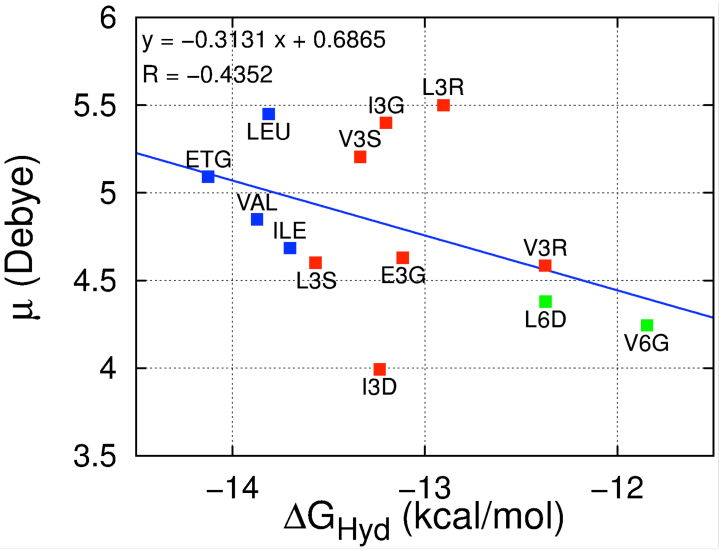


Fig. 6 Electrical dipolar moment *vs.* free energy of hydration. Average electrical dipolar moment (μ) *vs* the calculated hydration free energy (ΔG_{Hyd}) for each molecule; the dipolar moment is averaged from 25 ns simulations, with the corresponding distributions depicted in Figure S4. Blue=non-fluorinated, red=tri-fluorinated, green=hexa-fluorinated amino acids. The blue line is a linear fit to the data, with the corresponding equation and regression coefficient shown in the plot. See also Figure S6.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

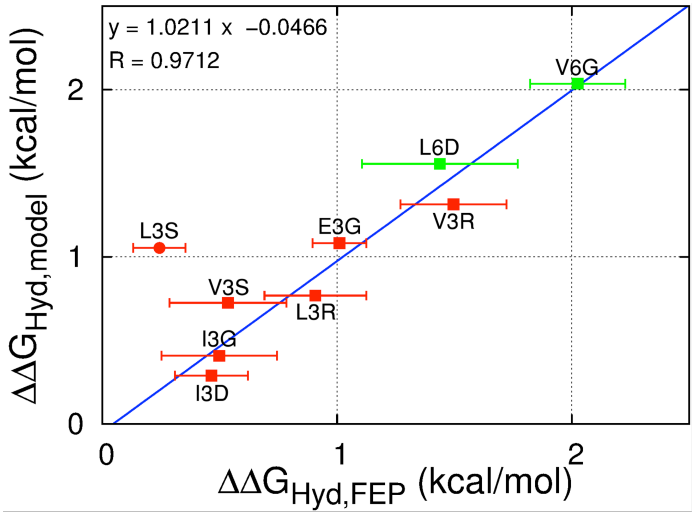


Fig. 7 Correlation between the differences in hydration free energies calculated with free energy perturbation and those calculated with Equation 1. Calculated (FEP) differences in amino acid hydration free energies against those predicted (model) by fitting equation 1 to the measured SASA, μ and number of backbone-water hydration bonds. L3S was not included in the fit. The blue line is a linear fit to the data (excluding L3S), with the corresponding equation and regression coefficient shown in the plot. Data from the FEP simulations are represented as mean ± standard deviation. See also Figure S8 and Tables S10 and S11.

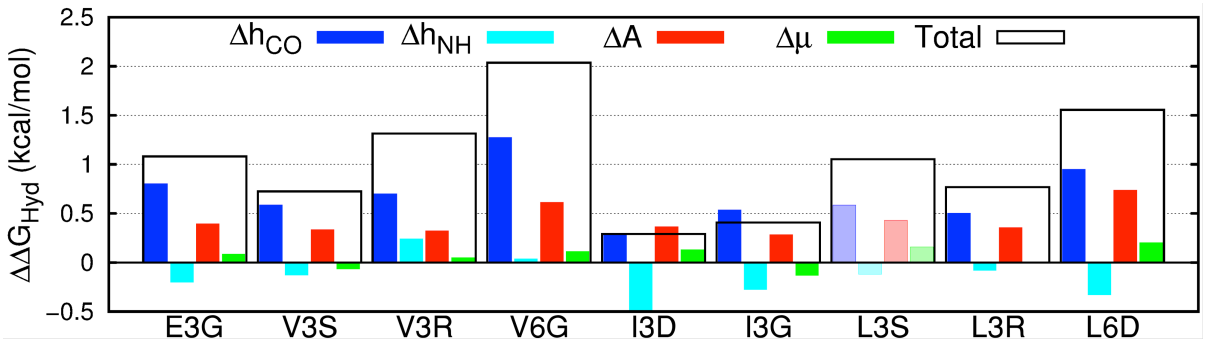


Fig. 8 Contribution of the backbone – water hydrogen bonds, surface area and dipolar moment to the free energy differences. Differences in hydration free energies ($\Delta\Delta G_{Hyd}$) between fluorinated and non-fluorinated amino acids predicted by equation 1. The narrow bars show the contribution of each term, the broad bars show the sum of all contributions. See also Figure S9.

Max Planck Institute of Colloids and Interfaces · Author Manuscript

Table 1 GAFF and optimized parameters for fluorinated groups. Lennard-Jones parameters ε and σ for the fluorine atom, from GAFF [31,73] and optimized in this work. See also Table S1.

	$\varepsilon_{F,F}$ (kJ/mol)	$\sigma_{F,F}$ (Å)
GAFF	0.255224	3.11815
This Work	0.090820	2.80000

Table 2 Calculated and target observables from the optimization of the Lennard-Jones parameters for fluorine. Calculated hydration free energy (ΔG_{Hyd}), enthalpy of vaporization (ΔH_{Vap}) and molar volume (V_{Mol}) for CF_4, CH_4 and an equimolar CF_4:CH_4 mixture using our optimized parameters; target values and their sources are shown within parenthesis.

* ΔG_{Hyd} of CF_4 with the TIP3P water model: 3.026 kcal/mol. ‡ Ref. 31. § Ref. 66. # Ref. 21. †Solubilities are converted to the Ostwald coefficient, [74] from which ΔG_{Hyd} can be calculated.[7] See also Table S2 and Figure S1.

	ΔG_{Hyd} (kcal/mol)	ΔH_{Vap} (kcal/mol)	V_{Mol} (cm^3/mol)
CF4	3.25* (3.12†)	0.45 (2.86‡)	2173.1 (54.8‡)
CH4	2.520 (2.004†)	-	-
CF4 : CH4	-	2.973 (2.476§)	44.0 (43.1#)

Max Planck Institute of Colloids and Interfaces · Author Manuscript

Table 3 Fitting parameters, parameter errors and parameter P-values associated with Equation 1. Values of the fitting parameters from Equation 1, and associated standard errors and P-values. NA: not applicable; *calculated via error propagation. See also Figure S8 and Tables S10 and S11.

	k_1 (μ) kcal/mol/Debye	k_2 (A) kcal/mol/Å^2	k_3 (h_{CO}) kcal/mol/H-bond	k_4 (h_{NH}) kcal/mol/H-bond
Pre-Factor	-0.188	0.0304	-3.59	-3.02
Error	0.119	0.001*	0.350	0.780
P-value	0.177	NA	0.000	0.012

SUPPLEMENTAL INFORMATION

Robalo_et_al_supplemental_Information.pdf: **Supplemental computational details, figures and tables.** Force field details, charge-fitting protocol, simulation details for the calculation of hydration free energies, vaporization enthalpy and molar volume; optimized Lennard-Jones parameters for fluorine, hydration free energies, solvent accessible surface area, backbone-water hydrogen bonds, dipole moment distributions, validation of the use of a linear model and entropies of solvation for L3S and L3R.

Robalo_et_al_force_field_parameters_amber.zip **Force field parameters.** Parameters in AMBER format for all fluorinated amino acids used in this work, and for ethylglycine, related to Procedures.