SUPPLEMENTAL INFORMATION

João R. Robalo,[1] Ana Vila Verde[1*]

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

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

1 Force Fields

In the TIP4P-Ew (ref. 1) water model, used in this work, the hydrogen has a charge $q_H = +0.52422\ |e|$, oxygen carries no charge, and a negative charge of magnitude $q_M = -1.04844\ |e|$ is located along the direction bisecting the region between the two hydrogens, 0.1250 Å away from the oxygen atom. Lennard-Jones (LJ) parameters are zero for hydrogen atoms; oxygen atoms have $\varepsilon_{OO} = 0.680946$ kJ mol^{-1} and $\sigma_{OO} = 3.16435$ Å; see SI Equation 1 for the functional form of the LJ potential.

We use the AMBER14 (ref. 2) force field for the canonical amino acids, and a force field developed by us (previous own work[3] and present work) compatible with AMBER14 for the remaining ones. The force field for ethylglycine and for tri- and hexa-fluorinated amino acids, together with a detailed description of the procedure used to develop it, is given elsewhere[3]. Here and in SI section 2 we give only the relevant details for the parameterization of mono- and difluorinated amino acids. Bonded parameters for the fluorinated amino acids were taken from the parent non-fluorinated amino acid; non-bonded, LJ, parameters are from AMBER14 (ref. 2), save for ε_{FF}, σ_{FF}, $\varepsilon_{H_FH_F}$ and $\sigma_{H_FH_F}$ LJ parameters, where F indicates a fluorine atom covalently bound to a carbon, and H_F indicates a hydrogen atom covalently bound to a fluorinated carbon; see SI Equation 1 for the definition of these parameters. LJ parameters used for methyl and fluoromethyl groups are presented in SI Table 1. For fluorine we use LJ parameters (ε_{FF} and σ_{FF}) optimized against the free energy of hydration of CF_4 and the molar volume of a liquid 1:1 mixture of CH_4:CF_4, as described elsewhere[3]. LJ parameters for fluorocarbon-bound hydrogen, H_F, were optimized to describe the free energy of hydration of CHF_3 and the molar volume of liquid CHF_3, as described in SI Section 2.

i	ε_{ii} (kJ mol^{-1})	σ_{ii} (Å)
C (aliphatic, GAFF[4])	0.4577300	3.3996700
F (ref. 3)	0.0908200	2.8000000
H (aliphatic, GAFF)	0.0656888	2.6495300
H_F (optimized, this work)	0.1746000	2.0000000

Table 1 LJ parameters ε_{ii} and σ_{ii} for the carbon, fluorine and hydrogen atoms in methane/methyl/methylene and fluorinated methane/methyl/methylene groups used in this work. All final results with partially fluorinated alkyl groups use the H_F parameters optimized in this work.

Non-bonded Van der Waals interactions between any two particles, i and j, over an inter-particle distance r_{ij} are calculated via the LJ potential as

$$V_{ij}^{LJ} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{1}$$

with

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \text{ and } \sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2} \tag{2}$$

following the Lorentz-Berthelot combination rules[5–11].

Atomic partial charges for the fluorinated methane derivatives were obtained with the Merz-Singh-Kollman Restrained Electrostatic Potential (RESP) methodology on a single conformation for each molecule, following the standard GAFF protocol. The charges were calculated at the RHF/6-31G* level of theory from a two-step geometry optimization (MP2/6-31G* optimization followed by RHF/6-31G* optimization). Quantum mechanical calculations were performed with the Gaussian 03 software[12], and RESP fitting of the partial charges was performed with the Antechamber package[13].

The atomic partial charges for all non-canonical amino acids are obtained via a multi-configuration RESP fitting procedure, described in greater detail in ref. 3, performed over 200 conformations (100 α-helical and 100 β-strand) per amino acid.

2 Simulation Details

The results discussed in the main text are obtained from Free Energy Perturbation (FEP) simulations – to calculate the hydration free energies of the amino acids – and molecular dynamics (MD) simulations – all other observables, *e.g.* solvent accessible surface area, dipole moment, numbers of hydrogen bonds – of single copies of the amino acids in water. In addition, FEP simulations of a single CHF_3 in water were used to calculate hydration free energies in the parameterization of fluorocarbon-bound hydrogen (H_F). Also used in the parameterization were simulations of liquid CHF_3, to calculate its molar volume, and additionally simulations of gas phase CHF_3, to calculate its vaporization enthalpy. Obtaining the partial charges for the partially fluorinated amino acids required initial simulations of single copies of the amino acids (with non-optimized charges) in water, to obtain multiple configurations that were then used in the multi-configuration RESP fit, as described in ref. 3. The simulation boxes used for each type of simulation are illustrated in SI Figure 1; periodic boundary conditions in all directions were used in all cases. All systems were assembled using the built-in tools of the software package used to perform the simulations. A summary of the most relevant parameters used during the production runs is given in SI Table 2. Simulations used a time-step of 2 fs and constraints (LINCS[15] in Gromacs, SHAKE[16] in Amber) were applied to all bonds involving hydrogen atoms. Integration of the equations of motion was done using a leap-frog Langevin algorithm. Van der Waals interactions were shifted to zero between 1.0 and 1.2 nm, and long-range dispersion corrections were applied to both pressure and energy.

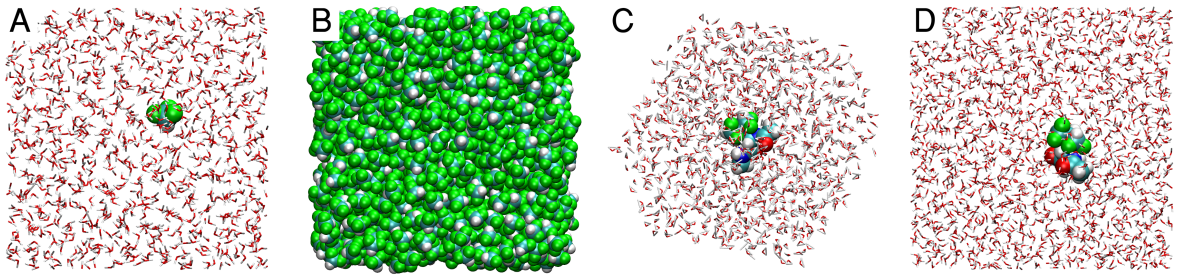

Fig. 1 Liquid phase simulation boxes of A) one CHF_3 molecule in water; B) pure CHF_3; C,D) L4D, capped with ACE and NME, in an octahedral (C) or cubic (D) box. Fluorinated molecules are shown as van der Waals surfaces, water as narrow tubes.

	$CH_3F/CH_2F_2/CHF_3$ (aq)	CHF_3 (l)	AA (aq)	AA (aq)	AA (aq)
Obs	ΔG_{Hyd}	ΔH_{Vap}, V_{Mol}	RESP fit	ΔG_{Hyd}	Analysis
Box	cubic (A)	cubic (B)	trunc. oct (C)	cubic (D)	trunc. oct. (C)
# mol	1070 (H_2O)	1150 (CHF_3)	1000 (H_2O)	1750 (H_2O)	1000 (H_2O)
p (bar)	1	1	1	1	1
T (K)	298	173	298	298	298
t (ns)	1*	10	50	2*	25
# rep	3	3	1	5	1
Package	Gromacs 5.0[5–11]	Gromacs 5.0	AMBER 14[14]	Gromacs 5.0	AMBER 14

* per λ value

Table 2 Simulation details for the production runs of the systems: C_nH_m (aq) = one C_nH_m molecule in water; CHF_3 (l) = pure CHF_3; AA (aq) = one capped amino acid in water. Obs = calculated observable; Box = box type (corresponding panel in Figure 1); # mol = number of molecules; p = pressure; T = temperature; t = production run length; # rep = number of independent runs (with different velocity assignments in the NVT equilibration step); Package = software package used in the simulations.

For the parameterization of the Lennard-Jones $\sigma_{H_FH_F}$ and $\varepsilon_{H_FH_F}$, each FEP and liquid phase simulation (molar volume and enthalpy of vaporization) consisted of a steepest-descent minimization, BFGS minimization, 100 ps of NVT equilibration, 100 ps of NpT equilibration and production run; gas phase simulations (enthalpy of vaporization) consisted only of the production run. FEP simulation of the amino acids also used the same equilibration procedure. Production runs were done in the NpT ensemble (liquid simulations) or the NVT ensemble (for gas phase). Long-range electrostatics were treated with the PME[17] scheme with a 1.2 nm cutoff, a grid spacing of 0.12 nm and a 6th order interpolation. The Berendsen barostat[18] was used with a time constant of 1 ps for coupling pressure to 1 bar; temperature coupling was enforced in the leap-frog Langevin integrator with a time constant of 1 ps.

Simulations for the second iteration of the RESP fit (extracting multiple configurations of the amino acids) consisted of a steepest descent minimization, a conjugated gradient minimization, a 200 ps NVT heating from 0 K to 298 K, and a 1 ns NpT equilibration; in the minimization, heating and equilibration steps the coordinates of the backbone atoms were restrained with a 20 kcal mol^{-1} potential. The production step was run in the NpT ensemble, keeping the same restraints on the backbone atoms. Long-range electrostatics were treated with the PME scheme with a 1.2 nm cutoff, a grid spacing of 0.1 nm and a 4th order interpolation. The Monte Carlo barostat[14,19] was used with a relaxation time of 1 ps for an isotropic coupling of system pressure to 1 bar; temperature coupling was handled by the leap-frog Langevin integrator with a collision frequency of 1 ps^{-1}. Simulations for the calculation of the input parameters in Equation 1 in the main text followed the same protocol except that no

restraints were imposed on the backbone atoms; these trajectories were used as input for the APBS[20] software to estimate the electrostatic hydration free energy of the amino acids using the linearized Poisson-Boltzmann equation, as described in more detail below.

2.1 Free Energy Perturbation

The potential form of the scaled intermolecular non-bonded interactions follows SI Equation 3, with the soft-core parameter α_{LJ} set to 0.5 for any $\lambda \neq \{0,1\}$ and zero otherwise.

$$V_{NB}(\lambda_C,\lambda_{LJ}) = \sum_i \sum_j (1-\lambda_C)\frac{q_i q_j}{r_{ij}} + \frac{(1-\lambda_{LJ})4\varepsilon_{ij}}{\left[\alpha_{LJ}\lambda_{LJ} + (r_{ij}/\sigma_{ij})^6\right]^2} - \frac{(1-\lambda_{LJ})4\varepsilon_{ij}}{\alpha_{LJ}\lambda_{LJ} + (r_{ij}/\sigma_{ij})^6} \tag{3}$$

Hydration free energies were calculated using Free Energy Perturbation (FEP) and Bennett Acceptance Ratio[21,22] (BAR), following the protocol we have previously adopted[3]. In each case, simulations consisting of a single solute molecule in water were conducted, first decoupling Coulombic interactions and then LJ interactions. For the parameterization of hydrogen LJ coefficients, the coupling parameter λ_C for the Coulombic interactions adopted the value of 0.00 (fully coupled), 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95 and 1.00 (fully decoupled); the resulting Coulombic potential is scaled linearly with λ. λ_{LJ} values were 0.00 (fully coupled), 0.06, 0.12, 0.18, 0.24, 0.30, 0.36, 0.42, 0.46, 0.50, 0.52, 0.54, 0.56, 0.58, 0.60, 0.64, 0.68, 0.72, 0.76, 0.80, and 1.00 (fully decoupled), for a total of 42 simulations; the hydration free energies for methane and fluorinated methane derivatives presented in SI Table 4 were calculated using 41 λ_{LJ} states: 0.00 (fully coupled), 0.03, 0.06, 0.09, 0.12, 0.15, 0.18, 0.21, 0.24, 0.27, 0.30, 0.33, 0.36, 0.39, 0.42, 0.44, 0.46, 0.48, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.62, 0.64, 0.66, 0.68, 0.70, 0.72, 0.74, 0.76, 0.78, 0.80, 0.90 and 1.00 (fully decoupled), adding to a total of 62 simulations. For the calculation of amino acid hydration free energies, the λ_C were kept from the methane simulations. λ_{LJ} adopted the values 0.00 (fully coupled), 0.03, 0.06, 0.09, 0.12, 0.15, 0.18, 0.21, 0.24, 0.27, 0.30, 0.33, 0.36, 0.39, 0.42, 0.44, 0.46, 0.48, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.80, 0.82, 0.84, 0.86, 0.88, 0.90, 0.92, 0.94, 0.96, 0.98 and 1.00 (fully decoupled), for a total of 80 Coulomb- and LJ-decoupling steps. An increase in both the number of states and the simulation time (compared to the simulations for methane derivatives, refer to SI Table 2) was necessary due to poor convergence for λ_{LJ} values above 0.6. Convergence was assumed when the entropy difference between adjacent states, calculated by analyzing the trajectory at each λ state with the adjacent states' Hamiltonians, did not exceed 0.2 kcal mol^{-1}.

2.2 Enthalpy of Vaporization and Molar Volume

Following the protocol of Gough and co-workers[23], the enthalpy of vaporization per mole, ΔH, of pure CHF_3 was calculated as

$$\Delta H = \Delta E + \Delta(pV) \approx \Delta E + RT \tag{4}$$

assuming that the gas phase systems behave ideally. In this expression, $\Delta E = E_g - E_l$, p is the pressure, V is the volume and R is the ideal gas constant. E_l is the potential energy per mole calculated from a simulation of a cubic box of pure CHF_3 in the liquid phase, and E_g is the equivalent quantity calculated using a simulation box containing only one molecule of CHF_3 (gas phase) and otherwise the same simulation parameters. For each system, the molar volume was calculated by dividing the average volume of the liquid phase simulation box by the total number of molecules, then multiplying by Avogadro's constant.

2.3 Parameterizing the H_F Atom

Given the known impact of the highly electronegative fluorine element in adjacent atoms, parameters for a fluorocarbon-bound hydrogen atom are required. To this end, we calculated the hydration free energy and molar volume of the CHF_3 molecule for all combinations of $\varepsilon_{H_FH_F}$ equal to (0.0656888*,0.15, 0.20, 0.25,0.30) kJ mol^{-1} and $\sigma_{H_FH_F}$ equal to (1.0, 1.5, 2.0, 2.64953†, 3.0) Å, totaling 25 ($\varepsilon_{H_FH_F}$, $\sigma_{H_FH_F}$) pairs. We then fit each of the calculated data sets and intersected the resulting surfaces with the experimental value of the hydration free energy or molar volume of CHF_3, given in SI Table 3. The intersection of both data sets yields the optimized parameters given in SI Table 1.

	ΔG_{Hyd} (kcal mol^{-1})			ΔH_{Vap} (kcal mol^{-1})		V_{Mol} (cm^3 mol^{-1})	
	this work (TIP4P-Ew)	this work (TIP3P)	exp.	this work	exp.	this work	exp.
CH_3F	-0.952 ± 0.026	-1.023 ± 0.012	-0.22‡				
CH_2F_2	-0.461 ± 0.001	-0.747 ± 0.033					
CHF_3	1.012 ± 0.029	0.687 ± 0.044	0.659‡	3.470 ± 0.010	4.25♯	44.453 ± 0.188	46.1♯

‡ Ref. 24,25.
♯ Ref. 23.

Table 3 Calculated hydration free energy (ΔG_{Hyd}), enthalpy of vaporization (ΔH_{Vap}) and molar volume (V_{Mol}) for CHF_3 and hydration free energy for CH_2F_2 and CH_3F, using our optimized parameters, as well as available reference experimental (exp.) values. For the hydration free energy, solubilities are converted to the Ostwald coefficient[24], from which the standard free energy of solvation can be calculated[25]; for the vaporization enthalpy and molar volume of CHF_3, experimental values are taken from Gough and colleagues[23]. Results are shown as mean ± standard deviation of five independent simulations in all cases except for ΔG_{Hyd} in TIP3P water, for which only 3 independent simulations were performed.

It would also be of interest to have an estimate of the optimal parameters for hydrogen in mono- and di-fluorinated carbons, which would be more relevant considering the fluorination patterns in the amino acids we are studying. Unfortunately, relevant experimental data on these species is sparse: for CHF_3, used in our parameterization, there are experimentally obtained values of its hydration free

* Original value in the GAFF[4] force field for carbon-bound hydrogen.
† Original value in the GAFF[4] force field for carbon-bound hydrogen.

energy, molar volume and enthalpy of vaporization; for CH_3F there is only the hydration free energy, for CH_2F_2 there is no data. Hydration free energies, enthalpies of vaporization and molar volumes, calculated with our parameters for CHF_3, CH_2F_2 and CH_3F, are given in SI Table 3. Our parameters follow the expected trend of increasingly positive hydration free energy with increasing degree of fluorination, though the hydration free energy for CH_3F is off by 0.7 kcal mol^{-1}. Because of the lack of experimental data, we cannot comment on the quality of our parameters for CH_2F_2.

It is worthwhile noting that the polar hydrogen atom we have parameterized is a main contributor to the more hydrophilic character of partially fluorinated methane relative to either CF_4 or CH_4, as becomes obvious by inspecting SI Tables 3 and 4. This effect of the polar hydrogen partially explains why the proportionality between surface area and hydration free energy no longer stands for partially fluorinated methane derivatives: each of the methane derivatives in SI Table 3 has a higher surface area than CH_4 while having a much lower hydration free energy (the hydration free energy of methane is 2.549 kcal mol^{-1}; see SI Table 4). In addition, the non-uniform charge distribution in the partially fluorinated molecules also increases the hydrophilicity of these molecules relative to CH_4 or CF_4.

3 Hydration free energies of fluorinated and non-fluorinated amino acids

The hydration free energy and differences in hydration free energy between fluorinated and non-fluorinated amino acids are presented in SI Table 4; also presented are the Coulombic and LJ components of the hydration free energy. Values for ethylglycine (ETG), valine (VAL), isoleucine (ILE), leucine (LEU) and all fully fluorinated amino acids (E3G, V3S, V3R, V6G, I3D, I3G, L3S, L3R, L6D) were retrieved from ref. 3 and are repeated here for the convenience of the reader.

3.1 Estimating electrostatic hydration free energies using Poisson-Boltzmann theory

We tested whether it was possible to estimate the electrostatic hydration free energy of the amino acids by solving the linearized Poisson-Boltzmann (PB) equation via a multigrid, finite difference, method[26,27], as implemented in the APBS[20] software. For the aqueous phase, default options for the APBS package were used, namely a multigrid level of four, with 129 grid points *per* processor in each direction (x, y, z; 0.1 Å grid spacing) and a multiple Debye-Hückel boundary condition. The solute dielectric constant was set to one, the solvent dielectric constant to 63 (corresponding to the TIP4P-Ew water model at 298 K[1]), with a cubic B-spline discretization of the charge mapping. The solvent-excluded volume was defined from the molecular surface, with atomic radii adopted from ref. 28; the solvent probe radius was taken as 1.4 Å (corresponding to water). The temperature was set at 298.15 K. Calculations in vacuum were performed by setting the solvent dielectric constant to one and the hydration free energy was calculated as the difference in free energy from the gas state to the solvated state. For each solvation free energy calculation, 25 configurations were extracted from a 25 ns simulation of the solvated amino acid. SI Table 5 holds the results of the PB calculations. SI Figure 2 A makes the comparison between PB and FEP hydration free energies. This figure clearly demonstrates

	ΔG_{Hyd}	$\Delta G_{Hyd,Coul}$	$\Delta G_{Hyd,LJ}$	$\Delta\Delta G_{Hyd}$	$\Delta\Delta G_{Hyd,Coul}$	$\Delta\Delta G_{Hyd,LJ}$
CH_4	2.549 ± 0.022	0.003 ± 0.001	2.546 ± 0.022			
CH_3F	-0.952 ± 0.026	-2.862 ± 0.009	1.910 ± 0.022	-3.501 ± 0.049	-2.866 ± 0.011	-0.636 ± 0.044
CH_2F_2	-0.461 ± 0.001	-2.956 ± 0.012	2.495 ± 0.017	-3.010 ± 0.024	-2.959 ± 0.013	-0.051 ± 0.039
CHF_3	1.012 ± 0.029	-2.069 ± 0.007	3.082 ± 0.019	-1.537 ± 0.052	-2.073 ± 0.008	0.536 ± 0.041
CF_4	3.258 ± 0.033	-0.379 ± 0.003	3.637 ± 0.028	0.709 ± 0.055	-0.382 ± 0.004	1.091 ± 0.050
CH_3CH_3	2.621 ± 0.021	0.0024 ± 0.000	2.618 ± 0.021			
CHF_2CH_3	0.018 ± 0.010	-2.823 ± 0.010	2.841 ± 0.015			
CF_3CF_3	3.871 ± 0.037	-0.386 ± 0.002	4.257 ± 0.038			
$CH_3CH_2CH_3$	2.759 ± 0.047	-0.002 ± 0.000	2.761 ± 0.047			
$CF_3CF_2CF_3$	4.394 ± 0.055	-0.378 ± 0.002	4.772 ± 0.055			
ETG*	-14.125 ± 0.054	-15.976 ± 0.037	1.851 ± 0.024			
E1G	-13.722 ± 0.128	-15.260 ± 0.106	1.538 ± 0.050	0.403 ± 0.182	0.716 ± 0.143	-0.313 ± 0.075
E2G	-13.121 ± 0.096	-15.248 ± 0.114	2.127 ± 0.023	1.004 ± 0.150	0.728 ± 0.151	0.276 ± 0.047
E3G*	-13.115 ± 0.060	-15.781 ± 0.059	2.666 ± 0.022	1.010 ± 0.114	0.195 ± 0.096	0.815 ± 0.046
PRG	-14.025 ± 0.067	-16.098 ± 0.050	2.074 ± 0.030			
P2G	-14.056 ± 0.159	-16.665 ± 0.126	2.608 ± 0.066	-0.032 ± 0.226	-0.566 ± 0.176	0.535 ± 0.096
VAL*	-13.871 ± 0.120	-15.965 ± 0.122	2.094 ± 0.025			
V3S*	-13.335 ± 0.129	-16.244 ± 0.111	2.909 ± 0.057	0.536 ± 0.249	-0.279 ± 0.233	0.815 ± 0.082
V3R*	-12.375 ± 0.105	-15.240 ± 0.153	2.865 ± 0.052	1.496 ± 0.226	0.725 ± 0.275	0.771 ± 0.078
V6G*	-11.846 ± 0.082	-15.381 ± 0.105	3.535 ± 0.030	2.025 ± 0.202	0.584 ± 0.227	1.441 ± 0.055
ILE*	-13.700 ± 0.100	-16.066 ± 0.101	2.366 ± 0.043			
I1G	-13.823 ± 0.209	-15.963 ± 0.204	2.140 ± 0.025	-0.123 ± 0.309	0.103 ± 0.306	-0.226 ± 0.068
I3D*	-13.235 ± 0.056	-16.448 ± 0.077	3.213 ± 0.048	0.465 ± 0.156	-0.382 ± 0.178	0.847 ± 0.091
I3G*	-13.202 ± 0.146	-16.379 ± 0.094	3.177 ± 0.075	0.498 ± 0.246	-0.313 ± 0.196	0.811 ± 0.118
LEU*	-13.811 ± 0.070	-16.109 ± 0.050	2.298 ± 0.030			
L1S	-15.296 ± 0.046	-17.335 ± 0.056	2.040 ± 0.017	-1.485 ± 0.116	-1.226 ± 0.106	-0.259 ± 0.047
L1R	-14.797 ± 0.121	-16.838 ± 0.062	2.041 ± 0.068	-0.987 ± 0.191	-0.729 ± 0.112	-0.258 ± 0.098
L4D	-14.670 ± 0.087	-17.580 ± 0.064	2.910 ± 0.045	-0.859 ± 0.157	-1.471 ± 0.114	0.612 ± 0.075
L3S*	-13.568 ± 0.042	-16.757 ± 0.063	3.190 ± 0.047	0.243 ± 0.112	-0.648 ± 0.113	0.892 ± 0.077
L3R*	-12.903 ± 0.147	-16.002 ± 0.129	3.099 ± 0.033	0.908 ± 0.217	0.107 ± 0.180	0.801 ± 0.063
L6D*	-12.373 ± 0.262	-16.231 ± 0.175	3.858 ± 0.089	1.438 ± 0.332	-0.122 ± 0.226	1.560 ± 0.119

Table 4 Free energy of hydration (ΔG_{Hyd}), Coulombic ($\Delta G_{Hyd,Coul}$) and Lennard-Jones ($\Delta G_{Hyd,LJ}$) components of the free energy of hydration for small molecules and amino acids; differences in the free energy of hydration ($\Delta\Delta G_{Hyd}$), Coulombic ($\Delta\Delta G_{Hyd,Coul}$) and Lennard-Jones components of the free energy of hydration ($\Delta\Delta G_{Hyd,LJ}$) between fluorinated and non-fluorinated molecules. All quantities were calculated using free energy perturbation and are shown as mean ± standard deviation, of five independent simulations, in kcal mol^{-1}. * retrieved from ref. 3. Related to Figures 2 and 3 in the main text.

that the PB approach does not capture the relative differences in the electrostatic hydration free energy between the amino acids, and does not give insight into these systems.

4 Calculating the electrostatic potential in the vicinity of the side chains

The electrostatic potential was calculated on a grid (0.1 Å of spacing) around the side chain of each amino acid, with conformations extracted from a 25 ns simulation of each amino acid in water, using only the partial charges of the side chain atoms. Then, the electrostatic potential was interpolated to the coordinates of oxygen atoms belonging to water molecules within a radius of 5.5 Å of the side chain carbon atoms; this radius corresponds to the first minimum of the radial distribution function of water oxygen atoms near methane's carbon atom, that is, it corresponds to the farther limit of the first hydration shell of methane. The probability density distributions of the electrostatic potential, Φ, at these positions is shown in SI Figure 2 B-F. For each fluorinated amino acid, the $\Delta\Phi^+$ term in Equation 1 in the main text corresponds to the number of water molecules, within 5.5 Å of the side chain carbon atoms, that experience a potential which is more positive than that of the waters at the 90^{th} percentile of the Φ probability distribution of the corresponding non-fluorinated amino acid. SI Figure 3 shows the electrostatic potential Φ around the side chain of ETG, E1G, E2G and E3G. Compared to ETG, the negative potential on the fluorinated derivatives arises from the fluorine atoms,

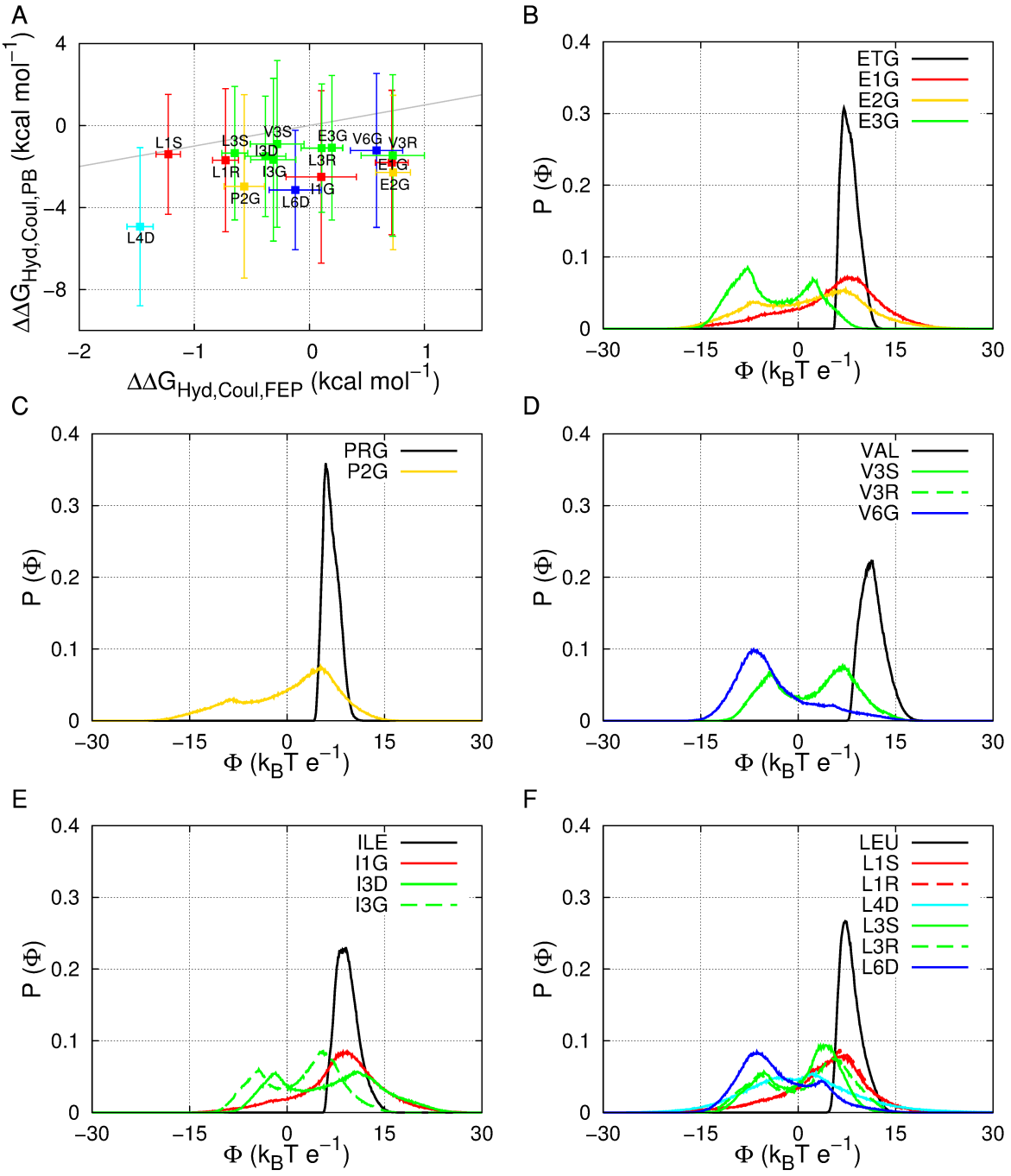

Fig. 2 A) Free energy of hydration calculated by the linearized Poisson-Boltzmann equation ($\Delta\Delta G_{Hyd,PB}$) *versus* the free energy of hydration calculated by free energy perturbation method ($\Delta\Delta G_{Hyd,FEP}$); data shown as mean $\pm$ standard deviation of 25 independent calculations (PB) or five independent simulations (FEP); the gray line indicates perfect correlation; B-F) probability density associated with the electrostatic potential induced by the side chain atoms only, at the position of the oxygen atoms of water molecules in the hydration shell of the side chains of: B) ethylglycine and derivatives, C) propylglycine and derivatives, D) valine and derivatives, E) isoleucine and derivatives and F) leucine and derivatives. The color code indicates the number of fluorine atoms in the side chain: red = one; yellow = two; green = three; cyan = four; blue = six. The hydration shell is composed of all water molecules within 5.5 Å of the side chain carbon atoms.

	$\Delta G_{Hyd,Coul,PB}$	$\Delta\Delta G_{Hyd,Coul,PB}$
ETG	-18.281 ± 1.497	
E1G	-20.086 ± 2.027	-1.805 ± 3.525
E2G	-20.572 ± 2.270	-2.290 ± 3.767
E3G	-19.365 ± 2.030	-1.083 ± 3.528
PRG	-17.612 ± 1.865	
P2G	-20.583 ± 2.608	-2.971 ± 4.474
VAL	-17.568 ± 1.617	
V3S	-18.466 ± 2.448	-0.898 ± 4.066
V3R	-19.031 ± 2.323	-1.463 ± 3.941
V6G	-18.782 ± 2.139	-1.214 ± 3.756
ILE	-17.277 ± 1.664	
I1G	-19.785 ± 2.543	-2.508 ± 4.207
I3D	-18.777 ± 1.272	-1.500 ± 2.935
I3G	-18.945 ± 2.305	-1.668 ± 3.968
LEU	-18.241 ± 1.444	
L1S	-19.643 ± 1.483	-1.403 ± 2.927
L1R	-19.933 ± 2.050	-1.692 ± 3.494
L4D	-23.175 ± 2.411	-4.934 ± 3.855
L3S	-19.586 ± 1.815	-1.345 ± 3.259
L3R	-19.343 ± 1.689	-1.102 ± 3.134
L6D	-21.385 ± 1.467	-3.144 ± 2.911

Table 5 Coulombic component of the free energy of hydration calculated using the linearized form of the Poisson-Boltzmann equation ($\Delta G_{Hyd,Coul,PB}$) and differences in the Coulombic component of the free energy of hydration calculated using the linearized form of the Poisson-Boltzmann equation ($\Delta\Delta G_{Hyd,Coul,PB}$). All quantities are shown as mean ± standard deviation in kcal mol^{-1}. Calculations were performed over amino acid conformations obtained from 25 equally spaced frames extracted from a 25 ns simulation.

the positive potential from the exposed carbon skeleton, especially from the fluorinated carbon (the hydrogen atoms in $-CH_3$ and, *e.g.*, $-CH_2F$ have a charge of similar magnitude; the carbon atoms have charges of opposite sign, negative in $-CH_3$ and positive in $-CH_2F$). The number of water molecules in the hydration shell was calculated with the Amber software package[14], the electrostatic potential was calculated using the VMD software package[29].

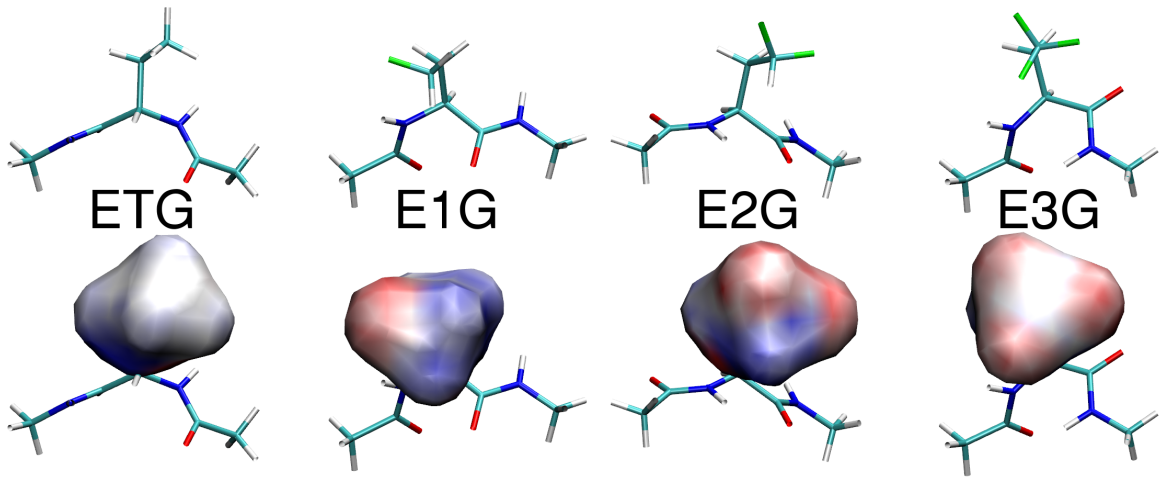

Fig. 3 Molecular structure (upper row) and electrostatic potential (lower row) for a single conformation of the side chain of ETG, E1G, E2G and E3G. The surfaces where the electrostatic potential is mapped were calculated using a probe of radius 5.5 Å. The color code for the electrostatic potential map is the same for all molecules, scaling from red (negative) to blue (positive).

5 Modeling the hydration free energy of fluorinated amino acids

5.1 Two linear, multivariate models that fail to explain how fluorination alters hydration free energies of amino acids

We attempted to fit the $\Delta\Delta G_{Hyd}$ for all amino acids using SI Equation 5:

$$\Delta\Delta G_{Hyd} = k_1\Delta A + k_2\Delta h_{CO} + k_3\Delta h_{NH} + k_4\Delta\mu \tag{5}$$

In this equation, the coefficient k_1 of the ΔA term is the LJ component of the hydration free energy *per* surface area unit of methane and the coefficients k_{2-4} for the water–carbonyl hydrogen bond (Δh_{CO}), the water–amine hydrogen bond (Δh_{NH}) and the dipole moment ($\Delta\mu$) terms are freely optimized in the fitting procedure. SI Table 6 holds the values of the fitting parameters, errors and P-values resulting from this attempt. As described in the main text, this model proved unsuccessful: the contributions of the water-backbone hydrogen bonds are unphysical, with hydrogen bonds with amines being much stronger than with carbonyls, and increases in dipole moment increasing amino acid hydrophobicity.

Fitting Parameter	Coefficient	Error	P-value
k_1 (ΔA; kcal mol^{-1} Å^{-2}))	0.053	0.001*	NA
k_2 (Δh_{CO}; kcal mol^{-1} H-Bond^{-1})	-1.121	1.484	0.463
k_3 (Δh_{NH}; kcal mol^{-1} H-Bond^{-1})	-5.732	3.863	0.161
k_4 ($\Delta\mu$; kcal mol^{-1} Debye^{-1})	0.173	0.461	0.714

Table 6 Values of the fitting parameters from SI Equation 5, and associated standard errors and P-values. NA: not applicable; * calculated via error propagation.

In a second attempt to understand our data, we turn to a model (SI Equation 6) we have previously successfully fitted to the changes in hydration free energy associated with tri- and hexa-fluorinated amino acids only[3]. This model has similar terms to those in SI Equation 5; the main difference is that the coefficient of ΔA is estimated from the *difference* in areas and in hydration free energies between CF_4 and CH_4.

$$\Delta\Delta G_{Hyd} = 0.0304\Delta A - 3.59\Delta h_{CO} - 3.02\Delta h_{NH} - 0.188\Delta\mu \tag{6}$$

SI Figure 4 illustrates the correlation between the model predictions and the hydration free energies calculated using FEP for all the amino acids under consideration here. The model largely fails for the mono- and difluorinated amino acids.

5.2 A successful linear, multivariate model to understand how fluorination alters hydration free energies of amino acids

This model, described by Equation 1 in the main text, includes a surface area contribution, captured by the $k_1\Delta A$ term, which is identical to that in SI Equation 5. This contribution increases linearly with the number of fluorine atoms,as shown in SI Figure 5. The model also includes two terms that reflect the contributions of carbonyl-water and amine-water hydrogen bonds. The main difference between this model and SI Equations 5 and 6 is that it includes three separate terms for water–fluorine

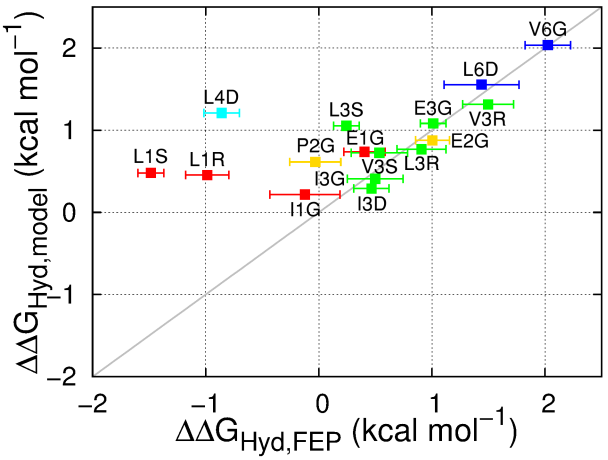

Fig. 4 Correlation between the hydration free energy differences between fluorinated and non-fluorinated amino acids calculated with FEP ($\Delta\Delta G_{Hyd,FEP}$) or as a result of the multivariate linear model ($\Delta\Delta G_{Hyd,model}$) given by SI equation 6, originally reported in ref. 3. This model takes as input the differences in surface area, water–carbonyl hydrogen bonds, water–amine hydrogen bonds and molecular dipole moment. FEP data points are presented as mean ± standard deviation of five independent simulations. The color code corresponds to the number of fluorine atoms in the side chain: red = one; yellow = two; green = three; cyan = four; blue = six. The gray line indicates perfect correlation.

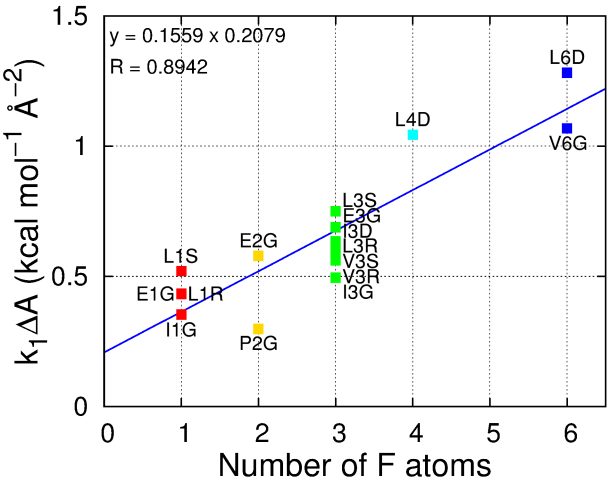

Fig. 5 Correlation between the contribution of the surface area to the hydration free energy and the number of fluorine atoms *per* fluorinated group. The color code corresponds to the number of fluorine atoms in the side chain: red = one; yellow = two; green = three; cyan = four; blue = six. The blue line is a linear fit to the data points with the corresponding equation and regression coefficient at the top left. Related to Equation 1 in the main text.

hydrogen bonds because their strength depends on the number of fluorine atoms in each group. Our simulations show that the donor–acceptor distance for these hydrogen bonds (SI Table 7) decreases with decreasing degree of fluorination of each alkyl group, indicating stronger interactions. Ab initio calculations, and the strong correlation between ^{19}F NMR isotropic chemical shifts and the type of fluorine-protein interactions observed in the Protein Data Bank also suggest that hydrogen bonds to fluorinated alkyl groups are strongest for groups with low degrees of fluorination[30]. The substantially different values of the fitting parameters k_{5-7} (see Table 1 in the main text), which translate the strength of water–fluorine hydrogen bonds, indicate this choice is necessary to understand interactions between fluorinated amino acids and water.

	distance (Å)		
	$-CH_2F$	$-CHF_2/-CF_2-$	$-CF_3$
E1G	2.711		
E2G		2.835	
E3G			2.951
P2G		2.813	
V3S			2.966
V3R			2.959
V6G			3.005
I1G	2.740		
I3D			2.925
I3G			2.971
L1S	2.705		
L1R	2.701		
L4D		2.847	
L3S			2.928
L3R			2.939
L6D			2.987

Table 7 Average distance between the water oxygen atom and the fluorine atom when a water–fluorine hydrogen bond is formed in a monofluoromethyl group ($-CH_2F$), a difluoromethyl/difluoromethylene group ($-CHF_2/-CF_2-$) or a trifluoromethyl group ($-CF_3$). Related to Equation 1 in the main text.

5.2.1 Details of the fitting procedure

The quantities used as input for fitting Equation 1 in the main text can be found in SI Table 8. The parameters k_1 to k_7 were fit in stages. The value of k_1 is simply the Lennard-Jones component of the hydration free energy of methane per unit area, for the reasons indicated in the main text (also, see SI Table 9). The values of k_2 and k_3, corresponding to the contributions of the hydration free energy per water-carbonyl and water-amine hydrogen bonds, respectively, are from ref. 3. The remaining parameters were optimized by fitting Equation 1 in the main text to the hydration free energies of all the amino acids using the input data in SI Table 8.

5.2.2 Calculating P-values

The model captures the difference in hydration free energies between the partially fluorinated amino acids and their non-fluorinated counterparts less well than for the fully fluorinated ones (Figure 4 in the main text). The cause for this discrepancy could possibly be related to an inadequacy of a linear model to represent either or both the water–fluorine hydrogen bond and the electrostatic potential terms in Equation 1 in the main text. To test the validity of using Equation 1 (see main text) to calcu-

	ΔA, Å (k_1)	Δh_{CO} (k_2)	Δh_{NH} (k_3)	ΔΦ (k_4)	h_{CH_2F} (k_5)	h_{CHF_2} (k_6)	h_{CF_3} (k_7)
E1G	8.143	-0.118	-0.027	5.451	0.630	0.000	0.000
E2G	10.917	-0.178	0.001	3.206	0.000	0.703	0.000
E3G	12.987*	-0.224*	0.067*	0.007	0.000	0.000	0.492
P2G	5.611	-0.114	-0.013	2.451	0.000	0.656	0.000
V3S	10.993*	-0.164*	0.043*	0.299	0.000	0.000	0.493
V3R	10.582*	-0.195*	-0.080*	0.227	0.000	0.000	0.479
V6G	20.160*	-0.355*	-0.012*	0.146	0.000	0.000	0.723
I1G	6.676	-0.057	0.021	7.410	0.554	0.000	0.000
I3D	11.972*	-0.079*	0.161*	7.664	0.000	0.000	0.613
I3G	9.336*	-0.149*	0.092*	0.804	0.000	0.000	0.472
L1S	9.819	-0.062	0.045	3.211	0.790	0.000	0.000
L1R	8.182	-0.086	0.038	3.043	0.758	0.000	0.000
L4D	19.697	-0.191	0.058	2.894	0.000	1.271	0.000
L3S	14.142*	-0.163*	0.040*	0.222	0.000	0.000	0.665
L3R	11.698*	-0.140*	0.026*	1.247	0.000	0.000	0.607
L6D	24.188*	-0.265*	0.109*	0.535	0.000	0.000	0.875

Table 8 Input parameters associated with the coefficients k_{1-7} in Equation 1 in the main text. All quantities represent the difference between fluorinated amino acids and their non-fluorinated parent, in each case averaged over 25000 configurations of the corresponding system. * Retrieved from ref. 3. ΔA is change in the amino acid surface area (calculated with a spherical probe with a radius of 1.4 Å, corresponding to that of a water molecule), Δh is the change in the average number of the hydrogen bonds established between water and the functional group indicated as subscript, per amino acid; hydrogen bonds exist if the O···O distance <3.5 Å and the O-H···O angle is between 135° and 180°; $\Delta\Phi^+$ is the change in the number of water molecules, within 5.5 Å of the side chain carbon atoms, that experience a potential which is more positive than that of the waters at the 90^{th} percentile of the Φ probability distribution of the corresponding non-fluorinated amino acid. Related to Equation 1 in the main text.

	CH_4	CF_4
SASA (Å^2)	47.75	72.21

Table 9 Solvent accessible surface area (SASA) of methane (CH_4) and tetrafluoromethane (CF_4). Related to Equation 1 in the main text.

late the hydration free energy differences between fluorinated and non-fluorinated amino acids, we have calculated P-values for the $\Delta\Delta G_{Hyd}$ of each amino acid by excluding it from the fit to Equation 1 in the main text. These P-values are presented in SI Table 10 and show that, particularly for E1G, E2G, I3D, L1S, L4D and L3S, the hydration free energy differences calculated with this type of model are not representative of those calculated with FEP. Using all amino acids to perform the fit, the P-values show an improvement (SI Table 10), where only E2G, L1S, L4D and L3S are not representative at a confidence level of 99%. Apart from E1G, E2G, I3D, L1S, L4D and L3S, all $\Delta\Delta G_{Hyd}$ values are representative to a confidence level of 90%. The emphasis should then be placed at the qualitative description of the observed trends in $\Delta\Delta G_{Hyd}$, observed in Figure 4 of the main text, which is the central message of this work.

5.2.3 The interplay between cavity-formation costs and gains from water-fluorine hydrogen bonds

The correlation between the energy required to form an amino acid-sized cavity in water and the energy gain from the formation of water–fluorine hydrogen bonds is found in SI Figure 6. It is clear that both the penalty associated with the surface area and the gain from hydrogen bond formation decrease with the increasing number of fluorine atoms *per* fluoromethyl group.

	$\Delta\Delta G_{Hyd,Model}$	P-value$_{Model}$	C.L.$_{Model}$	$\Delta\Delta G_{Hyd,Excl}$	P-value$_{Excl}$	C.L.$_{Excl}$
E1G	-0.085	0.037	93	-0.233	0.000	100
E2G	0.311	0.000	100	0.125	0.000	100
E3G	0.906	0.181	64	0.896	0.161	68
P2G	-0.157	0.288	42	-0.183	0.251	50
V3S	0.690	0.268	46	0.703	0.251	50
V3R	1.156	0.067	87	1.129	0.053	89
V6G	1.831	0.169	66	1.789	0.121	76
I1G	-0.101	0.472	6	-0.089	0.456	9
I3D	0.835	0.009	98	1.594	0.000	100
I3G	0.478	0.468	6	0.477	0.468	6
L1S	-1.091	0.000	100	-0.851	0.000	100
L1R	-1.007	0.456	9	-1.019	0.433	14
L4D	-0.411	0.002	100	0.409	0.000	100
L3S	0.721	0.000	100	0.803	0.000	100
L3R	0.712	0.184	63	0.689	0.156	69
L6D	1.281	0.319	36	1.230	0.264	47

Table 10 Difference in hydration free energy ($\Delta\Delta G_{Hyd}$, kcal mol^{-1}), P-value and confidence level (C.L.) calculated either from Equation 1 in the main text (Model) or as the result of a multivariate linear fit, having the same parameters as Equation 1 in the main text, where each amino acid is excluded (Excl). P-values are obtained from a one-sided z-test comparing the $\Delta\Delta G_{Hyd}$ value calculated with the model to the average $\Delta\Delta G_{Hyd}$ from the FEP simulations; confidence levels are calculated from the P-values. Related to Equation 1 in the main text.

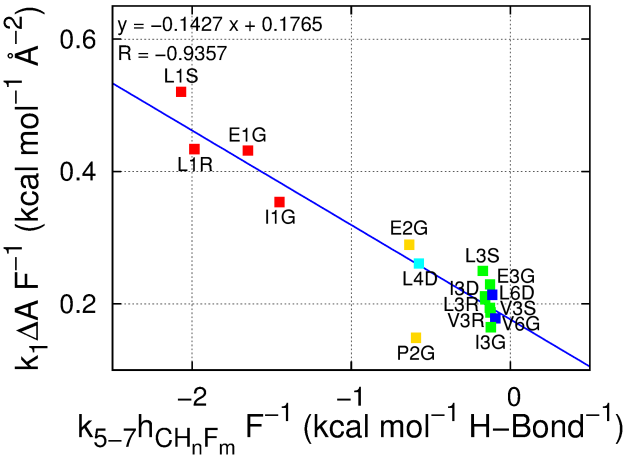

Fig. 6 Surface area contribution to the free energy of hydration *per* fluorine atom (k_1 ΔA F^{-1}) *versus* the water–fluorine hydrogen bonds contribution to the free energy of hydration *per* fluorine atom (k_{5-7} h$_{CH_nF_m}$ F^{-1}, where $n = 0,1,2$ and $m = 1,2,3$). Contributions are calculated using Equation 1 in the main text. The color code corresponds to the number of fluorine atoms in the side chain: red = one; yellow = two; green = three; cyan = four; blue = six. The blue line is a linear fit to the data points with the corresponding equation and regression coefficient at the top left. Related to Equation 1 in the main text.

5.2.4 Radial distribution functions and hydration free energy

It would be useful to interpret the hydration free energy changes we have reported in terms of methyl/fluoromethyl–water radial distribution functions (RDFs). These RDFs are shown in SI Figure 7. From the RDFs, an excess free energy required to form the hydration shell of each methyl/fluoromethyl group along the r reaction coordinate can be calculated as the product of the potential of mean force (PMF) required to bring a water molecule from bulk into the hydration shell and the number of water molecules in the hydration shell:

$$\Delta\Delta G_{Shell,PMF} = \sum_{r_0}^{r_{Shell}} -k_B T \ln\left[g(r)\right] g(r) V(r) \rho_{Bulk} \tag{7}$$

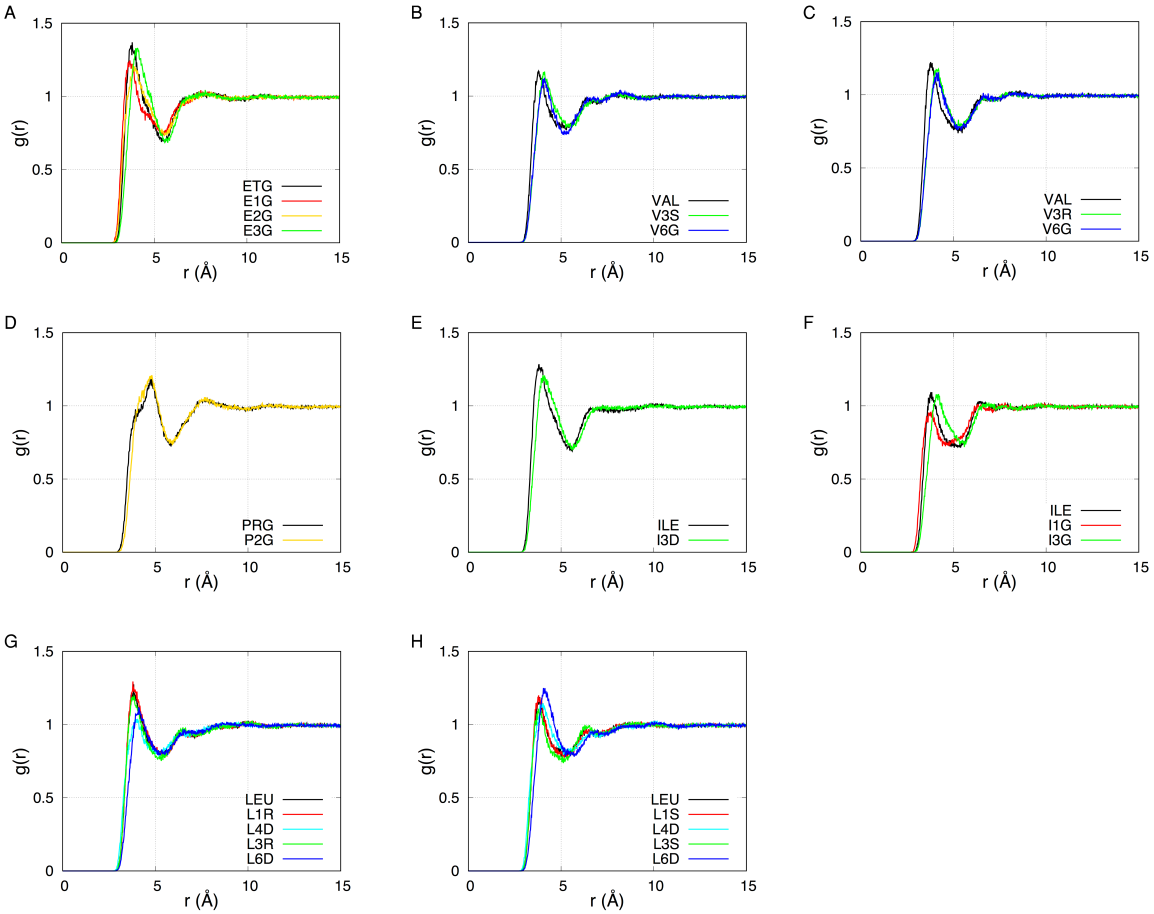

Fig. 7 A-H) radial distribution functions (g(r)) of water oxygen atoms relative to the carbon atom in the indicated fluoromethyl group, or its equivalent methyl group in the non-fluorinated amino acid, for A) the ethylglycine series, B,C) the valine series, D) the propylglycine series, E,F) the isoleucine series and G,H) the leucine series; the color code follows the total number of fluorine atoms in the side chain: black=0, red=1, yellow=2, green=3, cyan=4, blue=6.

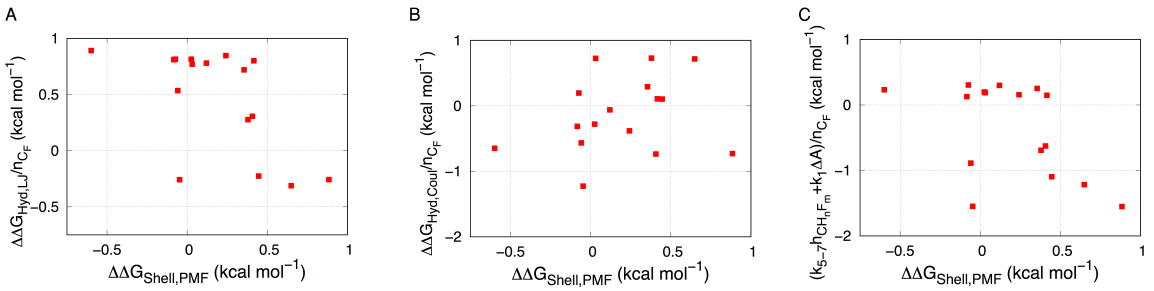

Fig. 8 The excess free energy, $\Delta\Delta G_{Shell,PMF}$, required to form the hydration shell of each methyl/fluoromethyl group, calculated from Equation 7, *vs.* A) the Lennard-Jones component per fluorinated carbon of the free energy of hydration calculated using FEP ($\Delta\Delta G_{Hyd,LJ}/n_{C_F}$; B) the Coulombic component per fluorinated carbon of the free energy of hydration calculated using FEP ($\Delta\Delta G_{Hyd,Coul}/n_{C_F}$); C) the sum of the energetic contributions of the surface area and the fluorine–water hydrogen bonds, per fluorinated carbon, to the free energy of hydration given by Equation 1 in the main text ($(k_{5-7}\, h_{CH_nF_m}+k_1\,\Delta A)/n_{C_F}$).

The term $-k_B T \ln[g(r)]$ corresponds to the PMF, with k_B being the Boltzmann constant, T the system temperature and $g(r)$ the radial distribution function. The term $g(r)V(r)\rho_{Bulk}$ is the particle density of water in a volume V(r) corresponding to the difference in volume of two spheres, one with radius r and the other with radius $r-0.02$ (the spacing between values of r at which the $g(r)$ is calculated); ρ_{Bulk} is the particle density of bulk TIP4P-Ew water. The summation in Equation 7 runs over values of r between r_0 (the C–O distance at which $g(r) > 0$) and r_{Shell} (the radius of the hydration shell of either methane, 5.531Å, or tetrafluoromethane, 5.690Å). Defining r_{shell} as the value for which each $g(r)$ reached its first minimum did not alter the trends described below.

We searched for correlations between the information contained in the RDFs, in the form of $\Delta\Delta G_{Shell,PMF}$, and both local and global measures of solvation. There is no correlation between $\Delta\Delta G_{Shell,PMF}$ and the Coulombic component of the hydration free energy per fluorinated carbon, $\Delta\Delta G_{Hyd,Coul}/n_{C_F}$, or the analogous Lennard-Jones component, $\Delta\Delta G_{Hyd,LJ}/n_{C_F}$ (Figure 8 A,B). $\Delta\Delta G_{Hyd,Coul}/n_{C_F}$ and $\Delta\Delta G_{Hyd,LJ}/n_{C_F}$ are global measures of hydration; these observables do not reflect only changes in local solvation around the fluorinated carbon. To isolate these local changes, we took advantage of the analytical model given by Equation 1 in the main text and summed over the contributions of the surface area and the fluorine-water hydrogen bonds. These results, shown in Figure 8 C, make clear that there is no correlation between this local measure of solvation and $\Delta\Delta G_{Shell,PMF}$.

6 Changes in side chain conformation induced by fluorination

To assess whether fluorination changes the occurrence of backbone-water hydrogen bonds by steric blockage, we calculated the distances between the carbon atom in a methyl or a fluoromethyl group (for example, C^{γ} in the ethylglycine series, C^{δ} in the leucine series) and the center of mass of either the carbonyl or amine group bound to the alpha carbon. The probability distributions for ETG (SI Figure 9) show that the side chain adopts a preferred conformation relative to the amine group where the methyl-amine distance is approximately 3 Å, while simultaneously adopting two equally probable conformations relative to the carbonyl group, corresponding to distances close to 3 and 4 Å. Fluorinating this methyl group progressively increases the probability of visiting a conformation that is both farther from the amine group and closer to the carbonyl, in agreement with the decrease in the number of carbonyl-water hydrogen bonds and the increase in amine-water hydrogen bonds observed for this series of amino acids (SI Table 8).

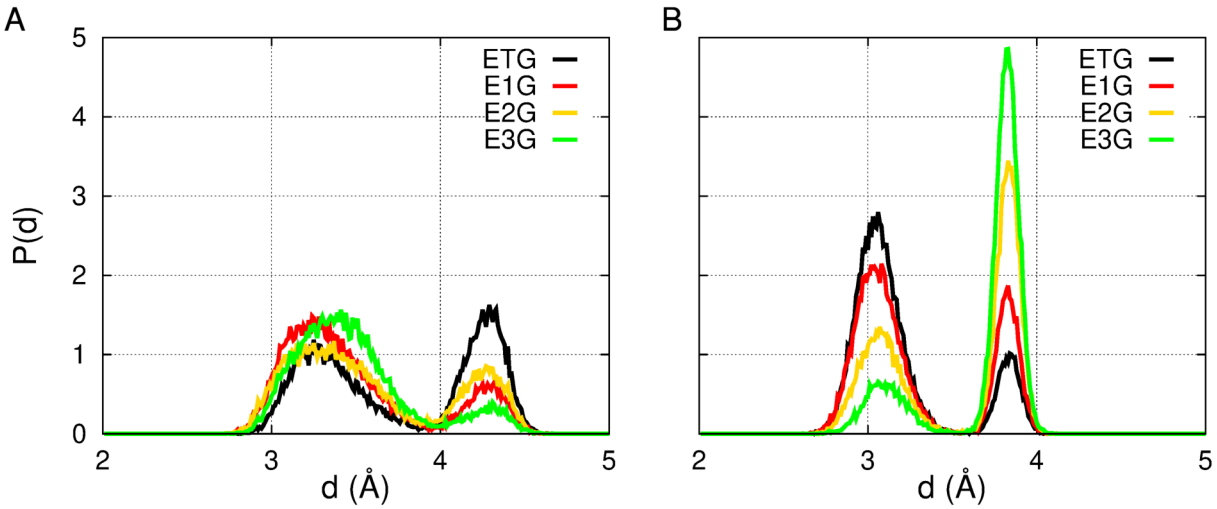

Fig. 9 Probability density (P(d)) associated with the distances (d) between methyl or fluoromethyl groups in the side chain of ETG, E1G, E2G and E3G and the A) carbonyl group or B) amine group bound to the alpha carbon in the backbone.

References

1 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew, *J. Chem. Phys.*, 2004, **120**, 9665–9678.

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

3 J. R. Robalo, S. Huhmann, B. Koksch and A. V. Verde, The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids, *Chem*, 2017, **3**, 881 – 897.

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

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

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

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

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

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

10 S. Páll, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, International Conference on Exascale Applications and Software, 2014, pp. 3–27.

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

12 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz,

I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision E.01*, 2014.

13 J. Wang, W. Wang, P. A. Kollman and D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.

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

15 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, LINCS: A linear constraint solver for Mol. Sim.s, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

16 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.

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

18 H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

19 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, 1989.

20 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 10037–10041.

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

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

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

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

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

26 M. Holst and F. Saied, Multigrid solution of the Poisson-Boltzmann equation, *J. Comput. Chem.*, 1993, **14**, 105–113.

27 M. J. Holst and F. Saied, Numerical solution of the nonlinear Poisson-Boltzmann equation: developing more robust and efficient methods, *J. Comput. Chem.*, 1995, **16**, 337–364.

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

29 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.

30 C. Dalvit and A. Vulpetti, Fluorine-Protein Interactions and F-19 NMR Isotropic Chemical Shifts: An Empirical Correlation with Implications for Drug Design, *ChemMedChem*, 2011, **6**, 104–114.