Supporting Information

Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations

William J. Zamora, Josep Maria Campanera*, F. Javier Luque*

Index

Computational methods S2

Tables

S1 S9
S2 S11
S3 S11
S4 S12
S5 S13
S6 S14
S7 S15
S8 S16

Figures

S1 S17
S2 S18
S3 S19
S4 S19
S5 S20
S6 S22
COMPUTATIONAL METHODS

SolvL and ProtL lipophilicity scales.

Following a previous study on the hydration free energy of the natural amino acids, the N-acetyl-L-amino acid amides (CH₃-CO-NH-CHR-CO NH₂) were chosen as molecular models. Using the backbone-dependent conformational library reported by Dunbrack and coworkers, a total of 572 rotamers (i.e., conformers with a probability contribution higher than 5% to the total conformational space of each residue) were compiled. These structures were then used to compute the n-octanol/water transfer free energies, which were performed with the B3LYP/6-31G(d) MST version of the IEF-PCM model.

Computation of the distribution coefficients at a given pH (log $D_{pH}$) was performed by combining the partition coefficient of neutral and ionic species (for ionizable residues) using Eq. S1.

$$\log D = \log \left( P_N + P_I \times 10^\delta \right) \log(1 + 10^{\delta})$$

(S1)

where $P_N$ and $P_I$ denote the partition coefficient of the neutral and ionized species of the amino acid, and $\delta$ is the difference between the $pK_a$ of the ionizable group and the pH of the environment.

Let us note that Eq. S1 represents one of the formalisms considered to estimate the pH-dependent lipophilicity profile of small (bio)organic compounds, and was found to reproduce satisfactorily the change in pH-dependent distribution coefficients for amino acid analogues.

The contribution of the conformational species in water and n-octanol was accounted for considering two weighting schemes, giving rise to the Solvent-like (SolvL) and Protein-like (ProtL) lipophilicities scales, respectively.
(i) In the SolvL scale, the contribution of each conformational state to the partition coefficient of the neutral/ionized species was determined using a Boltzmann weighting scheme, where the effective free energy was estimated by combining the internal energy of the conformer and its solvation free energy in water and \(n\)-octanol. To this end, the geometry of all rotamers was optimized at the B3LYP/6-31G(d) level of theory while keeping the backbone dihedrals fixed to the torsional values of the Dunbrack’s library, and subsequently single-point calculations in the gas phase and in solution. The \(\log D_{\text{pH}}\) was then computed using Eq. 1, adopting the \(pK_a\) values reported for ionizable residues from experimental peptide models in aqueous solutions.\(^{S8,S9}\)

(ii) In the ProtL scale, the contribution of each conformation to the partition between the two solvents was determined by using the weights reported in the Dunbrack’s library, which reflect the rotameric distribution in a protein environment. The \(pK_a\)s of ionizable residues were taken from values in folded proteins.\(^{S10,S11}\)

For the sake of comparison, we also computed both approaches with the SMD model using the B3LYP/6-31G(d) level of theory.\(^{S12}\) All calculations were performed using a locally modified version of Gaussian 09.\(^{S13}\)

**Comparison with experimental hydrophobicity scales.**

Due to the diversity of experimental lipophilicity scales of amino acids, generally expressed in terms of transfer free energies, comparison was made by converting them to partition/distribution coefficients, which were subsequently normalized to Gly following Eq. S2.

\[
\log P_N / D_{\text{pH}} = \left(\frac{G_{\text{trans,AA}} - G_{\text{trans,Gly}}}{RT\ln10}\right)
\]

(S2)
where \( G_{\text{trans,AA}} \) is the transfer free energy of a given amino acid from the aqueous phase to the organic/biological environment, and \( G_{\text{trans,Gly}} \) is the transfer free energy of Gly.

**Determination of the cumulative lipophilicity.**

Most of the experimental scales present in the literature compute the lipophilicity of a given peptide as the sum of individual lipophilicities of the constituent amino acids relative to a reference residue, usually Gly or Ala. Since the MST solvation model gives atomic contributions to the transfer free energy, \( S^{14-S^{16}} \) we can separate the global lipophilicity in contributions corresponding to the backbone (\( bb \)), side-chain (\( sc \)), and the capping groups (\( cg \)). Combination of the \( bb \) and \( sc \) contributions yields the amino acid lipophilicity (reported in Table 1 in the manuscript), whereas the contribution of the capping groups has been estimated to be (N-terminus) \( \text{CH}_3\text{CO}^- \) (\( \log P_{N}=0.20 \)), \( \text{NH}_3^+ \) (\( \log D_{7.4}=-2.99 \)), and (C-terminus) \( \text{NH}_2^- \) (\( \log P_{N}=-1.08 \)), \( \text{NMe}^- \) (\( \log P_{N}=0.35 \)), \( \text{COO}^- \) (\( \log D_{7.4}=-4.89 \)).

The cumulative lipophilicity of a peptide with \( N_{\text{res}} \) residues may be estimated by using Eq. S3.

\[
\log(P_{N} / D_{\text{pH}})^{\text{peptide}} = \sum_{i=1}^{N_{\text{res}}} \log(P_{N}^{i} / D_{\text{pH}}^{i})^{bb+sc} + \sum_{i=1}^{N_{cg}} \log(P_{N}^{i} / D_{\text{pH}}^{i})^{cg}
\]

where \( P_{N}^{i} / D_{\text{pH}}^{i} \) stands for the fragment (\( bb+sc \) or \( cg \)) partition/distribution coefficient, \( N_{\text{res}} \) and \( N_{cg} \) being the total number of residues and capping groups in the peptide.

For practical applications, this simple expression is convenient when there is no explicit knowledge about the 3D structure of peptides, as may occur in structureless peptides. For our purposes here, this is the expression adopted to evaluate the lipophilicity of small, flexible peptides in solution.
On the other hand, if the 3D structure of the peptide is known from experimental (X-ray, NMR) or computational (Molecular Dynamics) approaches, then the cumulative lipohilicity may be estimated taking into account the specific structural features of peptides/proteins, as noted in Eq. S4.

\[
\log (P_N / D_{pH})^{\text{peptide}} = \sum_{i=1}^{N_{\text{res}}} \left( i \log (P_N / D_{pH})^{bb+sc} + i \log (P_N / D_{pH})^{cg} + i \right) \tag{S4}
\]

In Eq. S4, \( i \) stands for the fraction of solvent-exposed surface area (SASA) of the amino acid (\( bb+sc \)) or capping group (\( cg \)) according to the local structural environment of in a peptide/protein. For our purposes, the SASA was determined using NACCESS.\(^{S17}\)

In addition, two correction factors were also introduced. The parameter \( i \) introduces a correction to the hydrophobic contribution when the backbone participates in a hydrogen bond (HB). This contribution can be estimated to amount, on average, to 0.73 (log P units) per HB.\(^{S18}\) The occurrence of this kind of HBs in a given 3D structural model was determined with the DSSP program.\(^{S19}\) Finally, the \( i \) factor accounts for a correction due to the burial of the side chain of hydrophobic residues (Ala, Leu, Ile, Val, Pro, Phe, Trp, Met and Tyr) from water to a lipophilic environment. This contribution has been estimated to be 0.023 kcal mol\(^{-1}\) Å\(^{-2}\) according to the studies reported by Moon and Fleming for the transfer of nonpolar side chains from water into a lipid bilayer.\(^{S20}\) Therefore, the \( i \) term has been estimated from the fraction of the buried side chain with respect to the fully buried side chain, as noted in Eq. S5.

\[
i = H^{i}_{\text{res}} x (1 - i)^{sc}
\]

\( i \)
where $H_{\text{res}}^i$ stands for the hydrophobic contribution (in logP units) of a specific apolar residue, which was estimated as noted in Eq. S6.

$$H_{\text{res}}^i = \frac{0.023 \times \text{SASA}_{\text{res}}^{\text{sc}}}{2.303 R T}$$ (S6)

where $\text{SASA}_{\text{res}}^{\text{sc}}$ is the average SASA of a given residue type, R is the gas constant, and T is temperature.

The $H_{\text{res}}^i$ values for nonpolar residues are given in Table S0.

Table S0. Average solvent accessible surface area for the side-chain of the hydrophobic residues and the hydrophobic effect contribution value when the side chain is fully buried.

| Residue | Average SASA (Å²) | $H_{\text{res}}^i$ (log P units) |
|---------|------------------|----------------------------------|
| Ala     | 69               | 1.2                              |
| Val     | 130              | 2.2                              |
| Leu     | 158              | 2.7                              |
| Ile     | 157              | 2.6                              |
| Met     | 166              | 2.8                              |
| Pro     | 115              | 1.9                              |
| Phe     | 188              | 3.2                              |
| Trp     | 232              | 3.9                              |
| Tyr     | 201              | 3.4                              |

References
(S1) Campanera, J. M.; Barril, X.; Luque, F. J. On the Transferability of Fractional Contributions to the Hydration Free Energy of Amino Acids. *Theor. Chem. Acc.* 2013, 132, 1–14.
(S2) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins:
Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230*, 543–574.

(S3) Dunbrack, R. L.; Karplus, M. Conformational Analysis of the Backbone-Dependent Rotamer Preferences of Protein Sidechains. *Nat. Struct. Biol.* **1994**, *1*, 334–340.

(S4) Shapovalov, M. V.; Dunbrack, R. L. A Smoothened Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844–858.

(S5) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Javier Luque, F. Extension of the MST Model to the IEF Formalism: HF and B3LYP Parametrizations. *J. Mol. Struct. THEOCHEM* **2005**, *727*, 29–40.

(S6) Cances, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(S7) Zamora, W. J.; Curutchet, C.; Campanera, J. M.; Luque, F. J. Prediction of pH-Dependent Hydrophobic Profiles of Small Molecules from Miertus–Scrocco–Tomasi Continuum Solvation Calculations. *J. Phys. Chem. B* **2017**, *121*, 9868–9880.

(S8) Arnold, M. R.; Kremer, W.; Lüdemann, H. D.; Kalbitzer, H. R. 1H-NMR Parameters of Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapeptides Gly-Gly-X-Ala at Pressures between 0.1 and 200 MPa. *Biophys. Chem.* **2002**, *96*, 129–140.

(S9) Kortemme, T.; Creighton, T. E. Ionisation of Cysteine Residues at the Termini of Model α-Helical Peptides. Relevance to Unusual Thiol pK\(_a\) Values in Proteins of the Thioredoxin Family. *J. Mol. Biol.* **1995**, *253*, 799–812.

(S10) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A Summary of the Measured pK\(_a\) Values of the Ionizable Groups in Folded Proteins. *Protein Sci.* **2009**, *18*, 247–251.

(S11) Harms, M. J.; Schlessman, J. L.; Sue, G. R.; Garcia-Moreno E., B. Arginine Residues at Internal Positions in a Protein Are Always Charged. *Proc. Natl. Acad. Sci.* **2011**, *108*, 18954–18959.

(S12) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(S13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et Al. Gaussian 09, Revision D.01; Gaussian, Inc.: Wallingford CT, 2009.
(S14) Luque, F. J.; Curutchet, C.; Muñoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I.; Morreale, A.; Gelpí, J. L.; Orozco, M. Continuum Solvation Models: Dissecting the Free Energy of Solvation. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3827–3836.

(S15) Ginex, T.; Muñoz-Muriedas, J.; Herrero, E.; Gibert, E.; Cozzini, P.; Luque, F. J. Development and Validation of Hydrophobic Molecular Fields Derived from the Quantum Mechanical IEF/PCM-MST Solvation Model in 3D-QSAR. *J. Comput. Chem.* **2016**, *37*, 1147-1162.

(S16) Vázquez, J.; Deplano, A.; Herrero, A.; Ginex, T.; Gibert, E.; Rabal, O.; Oyarzabal, J.; Herrero, E.; Luque, F. J. Development and Validation of Molecular Overlays Derived from Three-Dimensional Hydrophobic Similarity with PharmScreen. *J. Chem. Inf Model.* **2018**, *58*, 1596–1609.

(S17) Hubbard S.; Thornton, J. M. *NACCESS V.2.1.1* 1993; http://wolf.bms.umist.ac.uk/naccess.

(S18) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; et al. Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.* **2014**, *23*, 652–661.

(S19) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.

(S20) Moon, C. P.; Fleming, K. G. Side-Chain Hydrophobicity Scale Derived from Transmembrane Protein Folding into Lipid Bilayers. *Proc. Natl. Acad. Sci.* **2011**, *108*, 10174–10177.
Table S1. Protein-like (ProtL) Lipophilicity Scale Based on the log\(D_{pH}\) Values Determined for \(N\)-Acetyl-L-Amino Acid Amides at Physiological pH. The Lipophilicity Obtained for Conformational Distributions in \(\alpha\)-Helix and \(\beta\)-Sheet Structures, the Experimental \(pK_a\) of Side Chain Ionizable Groups, and the Calculated Partition Coefficients of Neutral (log \(P_N\)) and Ionized (log \(P_I\)) Residues Are Also Given.

| Residues | Exp. \(pK_a\) | log \(P_N\) | log \(P_I\) | log \(D_{pH}\) |
|----------|---------------|---------------|---------------|---------------|
| ALA      | -             | -2.47         | -             | -2.47         |
| \(\alpha\)-helix | -         | -2.87         | -             | -2.87         |
| \(\beta\)-sheet | -        | -2.03         | -             | -2.03         |
| ARG      | 12.51         | -3.66         | -7.38         | -7.04         |
| \(\alpha\)-helix | -       | -3.75         | -8.09         | -7.59         |
| \(\beta\)-sheet | -    | -3.49         | -5.98         | -5.98         |
| ASN      | -             | -3.97         | -             | -3.97         |
| \(\alpha\)-helix | -       | -4.09         | -             | -4.09         |
| \(\beta\)-sheet | -    | -3.39         | -             | -3.39         |
| ASP      | 3.50          | -3.18         | -8.54         | -5.87         |
| \(\alpha\)-helix | -       | -3.26         | -7.37         | -5.63         |
| \(\beta\)-sheet | -    | -3.07         | -10.07        | -6.19         |
| CYS      | 6.80          | -1.47         | -5.78         | -2.17         |
| \(\alpha\)-helix | -       | -2.06         | -5.75         | -2.76         |
| \(\beta\)-sheet | -    | -1.09         | -5.81         | -1.78         |
| GLN      | -             | -4.00         | -             | -4.00         |
| \(\alpha\)-helix | -       | -5.00         | -             | -5.00         |
| \(\beta\)-sheet | -    | -1.64         | -             | -1.64         |
| GLU      | 4.20          | -3.79         | -6.20         | -5.96         |
| \(\alpha\)-helix | -       | -3.67         | -6.42         | -0.14         |
| \(\beta\)-sheet | -    | -4.03         | -5.76         | -5.58         |
| GLY      | -             | -3.13         | -             | -3.13         |
| HID      | -             | -4.67         | -5.97         | -4.56         |
| \(\alpha\)-helix | 6.60 | -5.12         | -6.16         | -5.00         |
| \(\beta\)-sheet | -    | -4.26         | -5.79         | -4.15         |
| HIE      | 6.60          | -4.98         | -5.97         | -4.97         |
| \(\alpha\)-helix | -       | -5.49         | -6.16         | -5.46         |
| \(\beta\)-sheet | -    | -4.49         | -5.79         | -4.52         |
| ILE      | -             | -0.38         | -             | -0.38         |
| \(\alpha\)-helix | -       | -0.55         | -             | -0.55         |
| \(\beta\)-sheet | -    | -0.24         | -             | -0.24         |
| LEU      | -             | -1.36         | -             | -1.36         |
| \(\alpha\)-helix | -       | -1.59         | -             | -1.59         |
| \(\beta\)-sheet | -    | -1.09         | -             | -1.09         |
| LYS      | 10.53         | -2.19         | -6.81         | -5.08         |
| \(\alpha\)-helix | -       | -2.32         | -7.18         | -5.29         |
| \(\beta\)-sheet | -    | -1.98         | -6.16         | -4.73         |
| MET      | -             | -1.83         | -             | -1.83         |
| \(\alpha\)-helix | -       | -2.06         | -             | -2.06         |
| \(\beta\)-sheet | -    | -1.56         | -             | -1.56         |
| PHE      | -             | 0.86          | -             | 0.86          |
| \(\alpha\)-helix | -       | 2.23          | -             | 2.23          |
| \(\beta\)-sheet | -    | -0.18         | -             | -0.18         |
| PRO      | -             | -1.44         | -             | -1.44         |
| \(\alpha\)-helix | -       | -1.42         | -             | -1.42         |
| \(\beta\)-sheet | -    | -1.45         | -             | -1.45         |
| SER      | -             | -4.12         | -             | -4.12         |
| \(\alpha\)-helix | -       | -3.21         | -             | -3.21         |
|   | β-sheet |   |   |   |
|---|---------|---|---|---|
| THR | -       | -4.92 | - | -4.92 |
| α-helix | - | -3.33 | - | -3.33 |
| β-sheet | - | -2.80 | - | -2.80 |
| TRP | - | 0.16 | - | 0.16 |
| α-helix | - | 0.51 | - | 0.51 |
| β-sheet | - | -0.10 | - | -0.10 |
| TYR | 10.33 | -1.80 | -9.59 | -1.80 |
| α-helix | -1.96 | -9.65 | -1.96 |
| β-sheet | -1.69 | -9.55 | -1.69 |
| VAL | - | -1.68 | - | -1.68 |
| α-helix | -2.19 | - | -2.19 |
| β-sheet | -1.38 | - | -1.38 |
Table S2. Experimental RP-HPLC Retention Time for Eight Model Decapeptides and Cumulative Hydrophobicity Determined with the SolvL and ProtL Lipophilicity Scales.

| Peptide | Sequence   | Retention factor k<sup>a</sup> (min) | log D<sub>7.4</sub> | SolvL | ProtL |
|---------|------------|-------------------------------------|---------------------|-------|-------|
| Pep1Leu | DKDKGGGGLG  | 4.80                                | -17.09              | -34.04|       |
| Pep2Leu | DKDKGGGGLG  | 11.97                               | -15.03              | -32.27|       |
| Pep3Leu | DKDKGGGLLG  | 16.22                               | -12.97              | -30.50|       |
| Pep1Cys | DKDKGGGGCG  | 0.52                                | -17.30              | -34.85|       |
| Pep1Ile | DKDKGGGGIG  | 4.73                                | -17.64              | -33.06|       |
| Pep1Met | DKDKGGGCGG  | 2.27                                | -17.65              | -34.51|       |
| Pep1Phe | DKDKGGGGFG  | 6.11                                | -16.53              | -31.82|       |
| Pep1Val | DKDKGGGLVG  | 1.86                                | -18.07              | -34.36|       |

<sup>a</sup> Ref. 38.

Table S3. Correlation of Retention Time for Eight Model Decapeptides with the Same Charge<sup>a</sup> and for 218 Peptides<sup>39,40</sup> with Three Different Charge States Using the Cumulative Hydrophobicity with Our Adaptive Hydrophobicity Scale and with Others Experimental Scales.

| Scale       | <i>r</i>  | <i>p</i>-value<sup>a</sup> |
|-------------|----------|-----------------------------|
|             | Ref. 38 (pH = 7.4) | Refs. 39,40 (pH = 2.1)     |
| Faucheré-Pliska | 0.96    | 0.85 < 1 × 10^-16         |
|              | 2 × 10^-4 |                        |
| Eisenberg-McLachlan | 0.95    | 0.79 < 1 × 10^-16        |
|              | 3 × 10^-4 |                        |
| Hopp-Woods   | 0.99    | 0.74 < 1 × 10^-16        |
|              | 7 × 10^-6 |                        |
| Wimley et al.| 0.99    | 0.36 < 1 × 10^-9         |
|              | 4 × 10^-7 |                        |
| Moon-Fleming | 0.99    | 0.78 < 1 × 10^-16        |
|              | 3 × 10^-6 |                        |
| Hessa et al. | 0.96    | 0.61 < 1 × 10^-16        |
|              | 2 × 10^-4 |                        |
| Koehler et al.| 0.76    | 0.64 < 1 × 10^-16        |
|              | 0.03     |                            |
| Janin et al. | 0.39    | 0.55 < 1 × 10^-16        |
|              | 0.3      |                            |
| Kyte-Doolittle | 0.93   | 0.60 < 1 × 10^-16        |
|              | 8 × 10^-4 |                        |
| SolvL        | 0.96    | 0.85 < 1 × 10^-16        |
|              | 2 × 10^-4 |                        |
| ProtL        | 0.91    | 0.80 < 1 × 10^-16        |
|              | 0.002   |                            |

<sup>a</sup> <i>r</i>: Pearson correlation coefficient, <i>p</i>: statistical <i>p</i>-value.
Table S4. Statistical Parameters of the Comparison of the SolvL and ProtL Scale with Others Hydrophobicity Scales Against log \( P_N \) Values for 118 Random Peptides.

| Scale           | \( r \)  |
|-----------------|---------|
| Fauchère-Pliska | 0.90    |
|                 | < 1 \times 10^{-16} |
|                 | -2.53   |
|                 | 2.53    |
|                 | 2.64    |
| Eisenberg-McLachlan | 0.89   |
|                  | < 1 \times 10^{-16} |
|                  | -2.29   |
|                  | 2.29    |
|                  | 2.38    |
| Hopp-Woods      | 0.74    |
|                 | < 1 \times 10^{-16} |
|                 | -2.07   |
|                 | 2.11    |
|                 | 2.31    |
| Wimley et al.   | 0.70    |
|                 | < 1 \times 10^{-16} |
|                 | -1.54   |
|                 | 1.67    |
|                 | 1.81    |
| Moon-Fleming    | 0.69    |
|                 | < 1 \times 10^{-16} |
|                 | -0.80   |
|                 | 1.12    |
|                 | 1.34    |
| Hessa et al.    | 0.22    |
|                 | 0.02    |
|                 | 0.29    |
|                 | 0.98    |
|                 | 1.29    |
| Koehler et al.  | 0.45    |
|                 | 3 \times 10^{-7} |
|                 | -0.35   |
|                 | 0.87    |
|                 | 1.12    |
| Janin et al.    | 0.38    |
|                 | 2 \times 10^{-5} |
|                 | -0.65   |
|                 | 1.08    |
|                 | 1.28    |
| Kyte-Doolittle  | 0.50    |
|                 | 6 \times 10^{-9} |
|                 | -2.85   |
|                 | 3.00    |
|                 | 3.60    |
| ProtL           | 0.60    |
|                 | 5 \times 10^{-13} |
|                 | 1.35    |
|                 | 1.68    |
|                 | 2.00    |
| SolvL           | 0.93    |
|                 | < 1 \times 10^{-16} |
|                 | -0.55   |
|                 | 0.71    |
|                 | 0.94    |

*a* mse: mean signed error, mue: mean unsigned error, rmsd: root-mean square deviation, \( r \): Pearson correlation coefficient, \( p \): statistical p-value. mse, mue and rmsd are given in log \( P_N/D \) units.
Table S5. Statistical Parameters of the Comparison\(^a\) of the SolvL and ProtL Scale with Others Hydrophobicity Scales Against log \(D_{7.4}\) Values for 116 Random Peptides.

| Scale                   | \(r\)          | \(p\)-value | mse    | mue    | rmsd |
|-------------------------|-----------------|--------------|--------|--------|------|
| Fauchère-Pliska         | 0.76            | \(< 1 \times 10^{-16}\) | -2.76  | 2.76   | 2.88 |
| Eisenberg-McLachlan     | 0.75            | \(< 1 \times 10^{-16}\) | -2.58  | 2.58   | 2.69 |
| Hopp-Woods              | 0.88            | \(< 1 \times 10^{-16}\) | -2.32  | 2.33   | 2.43 |
| Wimley et al.           | 0.52            | 2 \times 10^{-9} | -1.94  | 1.94   | 2.23 |
| Moon-Fleming            | 0.79            | \(< 1 \times 10^{-16}\) | -1.16  | 1.24   | 1.48 |
| Hessa et al.            | 0.72            | \(< 1 \times 10^{-16}\) | -0.22  | 0.60   | 0.73 |
| Koehler et al.          | 0.76            | \(< 1 \times 10^{-16}\) | -0.90  | 1.01   | 1.19 |
| Janin et al.            | 0.61            | 4 \times 10^{-13} | -1.12  | 1.21   | 1.38 |
| Kyte-Doolittle          | 0.52            | 2 \times 10^{-9} | 3.04   | 3.17   | 3.76 |
| ProtL                   | 0.79            | \(< 1 \times 10^{-16}\) | 1.46   | 1.82   | 2.11 |
| SolvL                   | 0.83            | \(< 1 \times 10^{-16}\) | -0.52  | 0.73   | 0.95 |

\(^a\)mse: mean signed error, mue: mean unsigned error, rmsd: root-mean square deviation, \(r\): Pearson correlation coefficient, \(p\): statistical p-value. mse, mue and rmsd are given in log \(P_N/D\) units.
Table S6. Length (L), Net Charge (Q) and Cumulative Lipophilicity Determined Using ProtL, SolvL and Experimental Lipophilicity Scales of Peptides and Experimental Binding Affinities (BA; kcal/mol) Toward MHC (HLA-A*02:01 allele) from the Immune Epitope Database and Analysis Resource (Ref. 44).

| PDB | Sequence     | L  | Q    | BA *           | SolvL | ProtL | Fauchère-Pliska | Eisenberg-McLachlan | Hopp-Woods | Winley et al. | Moon-Fleming | Hessa et al. | Koehler et al. | Janin et al. | Kyle-Doolittle |
|-----|--------------|----|------|----------------|-------|-------|----------------|----------------------|------------|---------------|--------------|--------------|---------------|--------------|----------------|
| 2BST| SRYWAIRTR    | 9  | 3    | -6.4           | -9.29 | -22.50| 2.51           | 0.75                 | -0.67      | 7.75          | -8.55        | -6.99        | -0.72         | -3.60        | -7.95         |
| 3BO8| EADPTGHSY    | 9  | -1   | -7.1±0.3       | -3.12 | -17.72| 0.93           | 1.97                 | -1.91      | -4.12         | -8.81        | -9.75        | -1.35         | -1.80        | -9.63         |
| 1QVO| QVPLRPMTYK   | 10 | 2    | -7.1±0.3       | -4.58 | -15.13| 4.59           | 5.43                 | 0.79       | 5.18          | -4.68        | -9.02        | -1.47         | -3.60        | -5.25         |
| 2X4S| AMSNTELEL    | 9  | -2   | -5.8           | -1.85 | -15.83| 3.15           | 3.54                 | -0.53      | -2.59         | -5.33        | -6.21        | -0.72         | -0.40        | -0.50         |
| 2X4U| ILKEPVHGV    | 9  | 0    | -6.7±0.6       | -3.98 | -13.24| 5.16           | 5.34                 | 0.78       | 1.92          | -5.15        | -6.31        | -0.80         | -0.20        | 2.99          |
| 2GT9| EAAGIGILTV   | 10 | -1   | -6.6           | -1.34 | -11.34| 6.76           | 6.07                 | 3.84       | 1.24          | -0.85        | -2.07        | 1.00          | 2.80         | 11.38         |
| 1T22| SLYNTVATL    | 9  | 0    | -7.2±0.1       | 1.92  | -10.44| 5.77           | 5.86                 | 5.96       | 2.89          | -2.46        | -2.42        | 0.45          | 0.50         | 4.82          |
| 5W1W| VMAPRTLVL    | 9  | 1    | -6.8           | 0.87  | -10.44| 7.35           | 6.93                 | 4.21       | 5.40          | 1.43         | 2.64         | 0.32          | 1.00         | 9.42          |
| 2X4O| KLTPLCVTLE   | 9  | 1    | -6.5           | 4.20  | -9.56 | 8.11           | 6.76                 | 4.14       | 5.24          | -1.25        | -2.85        | -0.17         | 0.50         | 8.16          |
| 5EU3| YLEPGVPTA    | 9  | -1   | -7.0           | -0.22 | -9.19 | 5.25           | 5.73                 | 2.54       | -0.12         | 1.15         | -6.09        | -0.46         | -0.20        | 0.51          |
| 3MRM| KLVALGINAV   | 10 | 1    | -7.3           | -1.04 | -8.66 | 6.67           | 6.47                 | 4.49       | 4.96          | -2.80        | -2.48        | 0.45          | 1.50         | 11.89         |
| 3UTQ| ALWGPDPAAA   | 10 | -1   | -7.9           | 2.42  | -8.44 | 5.86           | 6.13                 | 3.04       | 0.28          | 0.41         | -6.49        | -0.38         | 1.10         | 2.17          |
| 3GSN| NLVPMVATV    | 9  | 0    | -6.6±0.1       | 1.69  | -7.45 | 7.28           | 7.72                 | 6.03       | 2.93          | 0.86         | -2.43        | 0.32          | 2.00         | 10.45         |
| 3EQ  | AAGIGILTV    | 9  | 0    | -7.01          | 1.60  | -6.59 | 7.40           | 6.62                 | 6.03       | 3.72          | 0.35         | -0.11        | 1.51          | 3.50         | 13.93         |
| 3MRG | CINGVCWTC    | 9  | 0    | -5.9±0.1       | 2.96  | -5.68 | 9.23           | 5.97                 | 7.57       | 3.81          | -3.25        | -1.56        | 0.44          | 3.60         | 9.05          |
| 2PYE | SLMMWITQC    | 9  | 0    | -6.4           | 8.19  | -2.24 | 10.22          | 8.33                 | 8.01       | 5.05          | -0.59        | -1.53        | 0.51          | 2.30         | 7.73          |
| 1HKK | LLGYPVYV     | 9  | 0    | -8.41          | 7.24  | -0.04 | 10.27          | 9.86                 | 9.98       | 5.76          | 6.84         | -1.68        | 1.11          | 1.90         | 10.36         |
| 2VLL | GILGFVTLE    | 9  | 0    | -8.6±0.1       | 6.76  | 1.21  | 10.26          | 9.00                 | 8.95       | 6.46          | 3.75         | 0.47         | 1.89          | 3.70         | 14.88         |
| 30X8 | FLPSDFPPSV    | 10 | -1   | -8.7±0.1       | 4.75  | 1.30  | 8.88           | 8.42                 | 5.23       | 3.13          | 4.11         | -5.71        | 0.09          | 1.20         | 5.91          |

* Estimated generally using cellular MHC/competitive/fluorescence half maximal inhibitory concentration (IC50), and exceptionally from radioactive assays. When several data were available, the binding affinity is given as the mean value together with the standard deviation.
Table S7. Correlation Coefficient of Cumulative Hydrophobicity Determined Using Different Lipophilicity Scales of MHC(HLA-A*02:01 allele)-Bound Peptides with Experimental Estimates of Binding Affinities.

| Scale                    | $r$          | $p$-value*     |
|--------------------------|--------------|----------------|
|                          | Entire Set, n=19 | No Cys set, n=17 |
| Fauchère-Pliska          | 0.34         | 0.67           |
|                          | 0.007        | 0.005          |
| Eisenberg-McLachlan      | 0.51         | 0.66           |
|                          | 0.008        | 0.006          |
| Hopp-Woods               | 0.36         | 0.62           |
|                          | 0.012        | 0.010          |
| Wimley et al.            | 0.18         | 0.31           |
|                          | 0.24         | 0.25           |
| Moon-Fleming             | 0.61         | 0.65           |
|                          | 0.008        | 0.006          |
| Hessa et al.             | 0.07         | 0.25           |
|                          | 0.36         | 0.35           |
| Koehler et al.           | 0.32         | 0.41           |
|                          | 0.11         | 0.12           |
| Janin et al.             | 0.18         | 0.39           |
|                          | 0.12         | 0.14           |
| Kyte-Doolittle           | 0.21         | 0.34           |
|                          | 0.21         | 0.20           |
| ProtL                    | 0.58         | 0.80           |
|                          | 0.009        | $2 \times 10^{-4}$ |
| SolvL                    | 0.42         | 0.73           |
|                          | 0.075        | 0.002          |

* $r$: Pearson correlation coefficient, $p$: statistical p-value.
Table S8. Equations of the Linear Regression Models Shown in Figures 2, 4 and 5.\(^a\)

| Figure  | Regression Equation     | \(r\)  | \(p\)-value       |
|---------|-------------------------|--------|-------------------|
| 2, left | 0.81 x - 0.24           | 0.96   | \(6 \times 10^{-11}\) |
| 2, right| 0.45 x - 0.04           | 0.92   | \(7 \times 10^{-9}\) |
| 4A      | 3.01 x + 55.76          | 0.96   | \(2 \times 10^{-4}\) |
| 4B      | 1.26 x + 31.30          | 0.85   | \(< 1.0 \times 10^{-16}\) |
| 4C      | 0.63 x - 0.39           | 0.93   | \(< 1.0 \times 10^{-16}\) |
| 4D      | 0.59 x - 0.59           | 0.83   | \(< 1.0 \times 10^{-16}\) |
| 5, left | -0.08 x - 6.98          | 0.42   | \(0.075\)         |
|         | -0.14 x - 7.18\(^b\)    | 0.73   | \(0.002\)         |
| 5, right| -0.08 x - 7.74          | 0.58   | \(0.009\)         |
|         | -0.10 x - 8.14\(^b\)    | 0.80   | \(3 \times 10^{-4}\) |

\(^a\) \(r\): Pearson correlation coefficient, \(p\): statistical \(p\)-value.

\(^b\) Obtained upon exclusion of two Cys-containing peptides.
Figure S1. Representation of SolvL (blue) and ProtL (yellow) Lipophilicity Scales (Values Relative to Gly) at Physiological pH.
Figure S2. Distribution of the Accessible Surface Area (Backbone + Side Chain) for Conformational Species of Arg (top, left), Lys (top, right), Asp (middle, left), Glu (middle, right), Asn (bottom, left) and Gln (bottom, right) Using the Dunbrack’s Backbone-Dependent Conformational Library. The Weight of Each Rotameric Species in This Conformational Library Is Also Shown (blue line).
Figure S3. Comparison Between Fauchère-Pliska Experimental log $D_{7.4}$ Values and Theoretical Estimates Obtained by Using the SMD Model with (left) Solvent-Adapted and (right) Protein-Adapted Weighting Factors for the Twenty N-Acetyl-L-Amino Acid Amides ($r$: Pearson Correlation Coefficient; mse: Mean Signed Error; mue: Mean Unsigned Error; rmsd: Root-Mean Square Deviation).

Figure S4. Comparison Between Fauchère-Pliska Experimental log $D_{7.4}$ Values and Theoretical Estimates Obtained by Using (left) ACD/I-Lab and (right) ChemAxon for the Twenty N-Acetyl-L-Amino Acid Amides ($r$: Pearson Correlation Coefficient; mse: Mean signed error; mue: Mean Unsigned Error; rmsd: Root-Mean Square Deviation).
Figure S5. Representation of the RP-HPLC Retention Time of 248 13-mer Peptides (Refs. 39,40) Versus The Lipophilicity determined from Experimental Scales in Table 2.
Figure S6. Representation of the Cumulative Lipophilicities Determined from the ProtL Scale Versus (A) the Retention Time for Eight 10-mer Peptides (pH 7.4; Ref. 38), (B) 248 Unique 13-mer Peptides (pH 2.1; Ref. 39,40), (C) log $P_N$ for 118 Random Peptides (Ref. 42), and (D) log $D_{7.4}$ for 116 Random Peptides (Ref. 42).