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

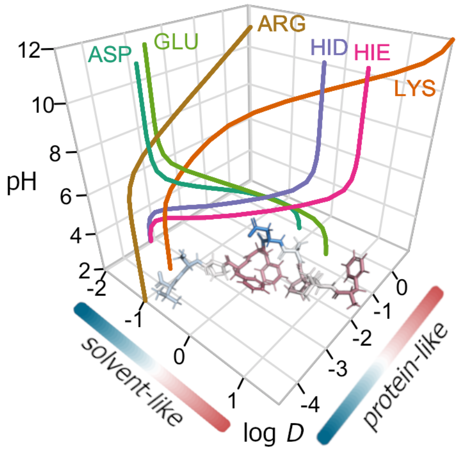
William J. Zamora, Josep Maria Campanera*, F. Javier Luque*

Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Science, Institute of Biomedicine (IBUB) and Institute of Theoretical and Computational Chemistry (IQTCUB), Campus Torribera, University of Barcelona, 08921 Santa Coloma de Gramenet, Spain

ABSTRACT

Lipophilicity is a fundamental property to characterize the structure and function of proteins, motivating the development of lipophilicity scales. Here we report a versatile strategy to derive a pH-adapted scale that relies on theoretical estimates of distribution coefficients from conformational ensembles of amino acids. This is accomplished by using an accurately parametrized version of the IEFPCM/MST continuum solvation model, as an effective way to describe the partitioning between *n*-octanol and water, in conjunction with a formalism that combines partition coefficients of neutral and ionic species of residues, and the corresponding pK_a of ionizable groups. Two weighting schemes are considered to derive *solvent-like* and *protein-like* scales, which have been calibrated by comparison with other experimental scales developed in different chemical/biological environments and pH conditions, as well as by examining properties such as the retention time of small peptides and the recognition of antigenic peptides. A straightforward extension to nonstandard residues is enabled by this efficient methodological strategy.

TOC GRAPHICS



Keywords: Lipophilicity scale, *n*-octanol/water partition, distribution coefficients, amino acids, pH-dependence, continuum solvation computations.

Lipophilicity is a cornerstone concept in chemistry and biology, as this property is crucial to understanding a variety of processes, such as the partitioning of molecules into inmiscible solvents, the formation of host-guest complexes, the folding of proteins, and the stability of supramolecular aggregates.[1,2] In proteins lipophilicity is mainly determined by the side chains of amino acids, and obtaining quantitative lipophilicity profiles of peptides and proteins is key to examine their structural and functional properties in biological environments. Accordingly, several strategies have been proposed to quantify the lipophilicity of amino acids, leading to lipophilicity scales that exploit the partitioning of small molecules between bulk solvents, the application of knowledge-based techniques to structural data, or experimental information derived from biological assays (for comprehensive reviews see refs. 3-5). Using these scales, lipophilicity profiles of peptides or proteins can be derived from the lipophilicity of single residues, generally assuming an additivity principle. Nevertheless, there are differences not only in the absolute magnitude of the residue lipophilicities, but also in the relative values, giving rise to a variable degree of correlation between scales that reflects the differences between the material systems, methods and experimental conditions that underlie the definition of each scale.

In this study our aim is to develop a lipophilicity scale from theoretical computations that takes into account the structural dependence of the conformational preferences of amino acids as well as the influence of pH in order to provide a consistent description of pH-adapted lipophilicity profiles in peptides and proteins. Here attention is focused on the set of natural amino acids, but the methodological strategy is intended to be easily adapted to nonstandard residues, such as nonproteinogenic residues, or to chemical modifications, such as phosphorylation, sulphonation and nitrosation, which regulate enzyme activity and signaling processes. To achieve this goal, each residue has been characterized by its distribution coefficient (D_{pH}) using as model system the corresponding *N*-acetyl-*L*-amino acid amides, taking into account the potential contribution of ionizable species at a given pH as noted in Eq. 1, which has recently been shown to reproduce the pH-dependent lipophilicity profiles of amino acid analogues.[6]

$$\log D_{pH} = \log\left(P_{\mathrm{N}} + P_{\mathrm{I}} \cdot 10^{\delta}\right) - \log(1 + 10^{\delta}) \qquad (1)$$

where P_{N} and P_{I} denote the partition coefficient of neutral and ionized species of an ionizable amino acid, and δ is the difference between the pK_a of the ionizable group and the pH of the environment.

Let us note that choice of *N*-acetyl-*L*-amino acid amides in this study enables a direct comparison with the experimental results reported by Fauchère and Pliska,[7] as their experimental lipophilicity scaled was determined using these model systems in their study. The partition coefficients P_{N} and P_{I} were determined from theoretical computations using the B3LYP/6-31G(d) version of the quantum mechanical IEFPCM-MST continuum solvation method,[8] which relies on the Integral Equation formalism (IEF) of the Polarizable Continuum Model (PCM).[9,10] Following our previous study of the hydration free energy of the natural amino acids,[11] the backbone-dependent conformational library compiled by Drunback and coworkers[12-14] (http://dunbrack.fccc.edu) was used to extract the conformational preferences of residues, which defined the ensemble of structures used to estimate the log D_{pH} values from IEFPCM-MST calculations in *n*-octanol and water (see SI for a detailed description of the computational methods).

Two schemes were explored for weighting the contribution of each conformational state to the differential solvation in the two solvents. In one case, P_{N} and P_{I} were determined using a

Boltzmann`s weighting scheme to the relative stabilities of the conformational species of a given residue in the two solvents, leading to the *solvent-like* scale (SolvL). In the second scheme, named *protein-like* scale (ProtL), the contribution of each conformation was directly taken from the population distribution reported in the backbone-dependent conformational library. Therefore, these weighting schemes are expected to yield scales better suited for reflecting the lipophilic balance of amino acids well exposed to bulk solvent or in a protein-like environment, respectively. Finally, the effect of pH on the logD_{pH} values was introduced from the experimental pK_as of ionizable residues in peptide models in aqueous solution[15,16] and in folded proteins[17,18] for the SolvL and ProtL scales, respectively.

The values of these lipophilicity scales for the amino acids at physiological pH are shown in Table 1 (ProtL data are averages of the log $D_{7.4}$ values determined separately for α-helix and β-sheet structures, which are reported in SI Table S1). Taken Gly as reference, the ProtL scale comprises log $D_{7.4}$ values ranging from -3.91 (Arg) to 3.99 (Phe), reflecting the extreme values of hydrophilic residues (Arg, Asp, Glu and Lys), and hydrophobic ones (Trp, Phe) (see also SI Figure S1). These trends are also found in the SolvL scale, although the distribution of log $D_{7.4}$ values varies from -1.35 (Glu) to 2.62 (Phe). This trait is also found in other scales, as knowledge-based methods generally give rise to a narrower range of lipophilicites compared to other experimental scales.[19] In our case, this arises from the distinct weighting factors used in ProtL and SolvL scales, leading to larger differences in the log $D_{7.4}$ values of polar and ionizable amino acids, which show a preference for extended conformations (SI Figure S2), likely reflecting the formation of stabilizing interactions (*e.g* salt bridges) or the solvent exposure to bulk water in proteins.[20,21]

The sensitivity of the lipophilicity of ionizable residues to pH changes is shown in Figure 1, which compares the log D_{pH} values at pH 2.1, 7.4 and 9.0, chosen as representative values of the pH changes along the gastrointestinal tract. The hydrophilicity of acid/basic amino acids is enhanced at basic/acidic pHs, as expected from the predominance of the ionic species. In the SolvL scale, it is worth noting the hydrophilic nature of protonated His at acidic pH, and the slight hydrophobicity of protonated Glu. In contrast, the ProtL scale exhibits a higher sensitivity to pH, as noted in the large changes in the log D_{pH} values of Asp and Glu, which are decreased 2-3 log D_{pH} units upon deprotonation, the reduced hydrophilicity of Lys at basic pH, and the change from hydrophobic (at acid and physiological pH) to hydrophilic (at basic pH) of Cys. This reflects the ability of these scales to present the pH influence on the lipophilicity of ionizable residues, which may be affected by the local environment in proteins.[22,23]

To calibrate the suitability of these scales, comparison was made with the log $D_{7.4}$ values reported by Fauchère and Pliska,[7] which were experimentally determined from the partitioning of *N*-acetyl-*L*-amino acid amides between *n*-octanol and water at physiological pH (Figure 2). Comparison with the SolvL values gives satisfactory results, as noted in a correlation coefficient (*r*) of 0.96 and a mean unsigned error (mue) of 0.33 log $D_{7.4}$ units for a set of experimental values ranging from -3.36 to 0.61. The correlation coefficient is slightly worse (r = 0.92) and the mue increases to 1.68 for the ProtL scale. For the sake of comparison, the same analysis was performed by using log $D_{7.4}$ values obtained from computations with the SMD solvation model,[24] in conjunction with the two weighting schemes, and the results also revealed a better performance for the solvent-adapted scheme (r = 0.85, mue = 0.83; SI Figure S3). On the other hand, the SolvL scale also performed better than the empirical estimates of log $D_{7.4}$ obtained from ACD/ILab[25] (r = 0.88, mue=0.60) and ChemAxon[26] (r = 0.92, mue=0.65) when compared with the experimental values reported by Fauchère and Pliska (SI Figure S4).

Table 2 shows the comparison of the SolvL and ProtL lipophilicities with experimental scales, including four bulk solvent-based scales (Fauchère-Pliska,[7] Eisenberg-McLachlan,[27] Hopp-Woods,[28] Wimley et al.[29]), two biological-derived (Moon-Fleming,[30] Hessa et al.[31]) and two knowledge-based (Koehler et al,[19] Janin et al.[32]) scales, and a consensus (Kyte-Doolittle[33]) one. The bulk solvent-based scales rely on experimental measurements of the transfer between *n*-octanol and water (Fauchère-Pliska, Eisenberg-McLachlan) at physiological pH or at basic conditions (pH = 9.0; Wimley et al.), and between ethanol and the vapor phase (Hopp-Woods). Excellent correlations are found with Fauchère-Pliska, Eisenberg-McLachlan, and Hopp-Woods scales ($0.89 < r < 0.92$). A worse correlation ($r \approx 0.60$) is found in the comparison with Wimley et al. scale, but at large extent this can be attributed to the formation of salt bridges between Arg/Lys residues with the terminal carboxyl group in *n*-octanol for the AcWL-X-LL pentapeptides used as model systems, as noted by ^{13}C-NMR studies.[34] Exclusion of Arg and Lys enhances the correlation coefficient to 0.87. On the other hand, the bulk solvent-based lipophilicities are consistently closer to the values collected in the SolvL scale (mue of 0.36-0.92 log *P/D* units) than to the ProtL ones (mue of 0.84-1.24 log *P/D* units).

The correlation coefficients obtained with biological-, knowledge-based and consensus scales are satisfactory ($0.74 < r < 0.94$; Table 2), but tend to be lower than the values obtained with the bulk solvent-based transfer scales. This is not unexpected keeping in mind that the lipophilicites are derived from statistical analysis of topological distributions of residues in proteins (Koehler et al, Janin et al.), or from complex biochemically-adapted assays, such as the transfer of amino acids from water to a phospholipid bilayer (Moon-Fleming), the recognition of artificial helices by the Sec61 translocon (Hessa et al.), or the combination of water-vapor transfer free energies with the interior-exterior distribution of amino acids in the consensus (Kyle-Doolittle) scale. Keeping in mind the notable differences in the material systems and protocols used to derive these experimental scales, the correlation coefficients obtained from the comparison with the SolvL scale are still remarkable.

The sensitivity of the results to the pH was examined by extending the comparison to the lipophilicities determined for the SolvL and ProtL scales at pH values of 3.8, 7.4, and 9.0 (note that the acidic and basic pH values were chosen in the studies reported by Moon and Fleming and Wimley et al., respectively). In general, there is little difference between the correlation coefficients obtained at pH 7.4 and 9.0 (Figure 3). However, a larger effect is found in the comparison of the log $D_{3.8}$, as there is a general decrease in the correlation coefficient, which is remarkable for the bulk solvent-based transfer scales, especially in the case of Hoop -Woods and Wimley et al. The only exception is found in the comparison with the Moon-Fleming scale, as the highest correlation coefficient is found for the ProtL values corrected at pH 3.8. These findings support the suitability of the SolvL/ProtL scales to account for the pH influence on the lipophilicity of amino acids.

The reliability of the SolvL/ProtL scales has been calibrated by comparing the cumulative lipophilicity with the (RP-HPLC) retention time determined for different sets of peptides.[35,36] Given the small size of the peptides (≤ 13 residues) and the lack of well defined secondary structures, non-additivity effects can be expected to play a minor role.[37] Accordingly, the cumulative lipophilicity was determined assuming an additive scheme (Eq. S3 in SI Computational Methods).

The first test comprises eight 10-mer peptides with equal charge that differ in the content of hydrophobic residues (SI Table S2).[38] The SolvL cumulative lipophilicity yields a correlation coefficient of 0.96 (Figure 4A), which compares with the value estimated from the hydrophobic surfaces of peptides derived from molecular dynamics simulations (r = 0.97),[38] whereas a slightly lower correlation was found for the ProtL scale (r = 0.91; SI Table S3). For this simple set of homogeneous peptides, most of the experimental lipophilicity scales generally yielded correlations higher than 0.9 (SI Table S3).

A more challenging test is the set of 248 peptides with equal length, but different net charge at the experimental acidic conditions (pH = 2.1),[39,40] comprising 36 peptides with two charged amino acids (Arg combined with His or Lys), 105 peptides with a single charged residue (Arg, Lys, or His), and finally 17 neutral peptides. The SolvL cumulative lipophilicity correlates satisfactorily with the retention time determined for the whole set of peptides (r = 0.85; Figure 4B). Among bulk solvent-based scales, Fauchère-Pliska, Eisenberg-McLachlan and Hopp-Woods also provided reasonable correlations coefficients ($0.74 < r < 0.85$; SI Table S2 and Figure S6), but a worse correlation was found for Wimley et al., although this may be attributed to the different pH used in this latter scale (pH = 9.0) and the experimental assay conditions (pH = 2.1). The performance of biological-, knowledge-based and consensus scales was also worse ($0.55 < r < 0.64$; SI Table S3 and Figure S5), but for Moon-Fleming (r = 0.78), likely reflecting the acidic pH conditions considered in the derivation of this lipophilicity scale.

Finally, given the relevance of partition (log P_N)/distribution (log $D_{7.4}$) coefficients for ADME properties of peptides,[41] the suitability of the SolvL scale was further checked for reproducing the differences in log P_N /log $D_{7.4}$ of a set of random peptides.[42] The SolvL-based additive scheme yielded promising results, as noted in r values of 0.93 and 0.83 in reflecting the experimental range of log P_N and log $D_{7.4}$ for sets of 118 and 116 peptides, respectively (Figure 4C,D). Compared to experimental scales, a similar predictive power was attained for Fauchère-Pliska and Eisenberg-McLachlan scales ($r \approx 0.90$) for the set of 118 log P_{N} data, and for Hopp-Woods ($r \approx 0.88$) for the set of 116 log$D_{7.4}$ values, but with a larger mue (around 2.3 versus 0.7 for the SolvL scale; SI Tables S4 and S5).

In these test cases, the ProtL scale performed worse ($0.60 < r < 0.91$; SI Figure S6) than the SolvL one, suggesting that the Boltzmann-weighting scheme is better suited for describing the lipophilicity of residues in structureless peptides. However, one might expect an improved performance of the ProtL scale in the analysis of the lipophilic complementarity in peptide-protein and protein-protein complexes. To this end, we have examined the relationship between the ProtL cumulative lipophilicity and the experimental binding free energies of 19 peptides to MHC (HLA-A*02:01 allele) proteins (SI Table S6). These peptides were chosen subject to the availability of (i) a precise structural information of the peptide-protein complex in the Protein Data Bank,[43] and (ii) an estimate of the binding affinity in the Immune Epitope Database and Analysis Resource[44] (SI Table S6). The cumulative lipophilicity was determined taking into account the fraction of solvent-exposed area of the peptide residues in the MHC complex, supplemented with two correction parameters that account for the contribution due to the involvement of the backbone in hydrogen bonds,[45] and to the burial of apolar residues from water to hydrophobic environments[30] (Eq. S4 in SI Computational Methods).

The results show that the ProtL scale works better than the SolvL scale (correlation coefficients of 0.58 and 0.42, respectively; Figure 5) when the whole set of 19 peptides is considered, yielding correlation coefficients that are comparable with Moon-Fleming and Eisenberg-

McLachlan scales (*r* of 0.61 and 0.51, respectively; SI Table S7). This correlation is remarkable keeping in mind the heterogeneity of the peptides, and the uncertainty arising from the combination of data taken from different studies and determined using distinct experimental approaches. Further, a significant improvement is observed upon exclusion of the two Cys-containing peptides (PDB codes 3MRG and 2PYE), perhaps reflecting a quenching effect of cysteine in fluorescence assays.[46,47] Thus, upon exclusion the correlation coefficient of ProtL and SolvL scales increases up to 0.80 and 0.73, respectively, leading to regression equations with increased statistical significance (*p*-values of 2×10^{-4} and 2×10^{-3}, respectively). Finally, let us note that this improvement outperforms the results obtained with the experimental scales ($r < 0.67$; SI Table S7).

Overall, the results point out the versatility of the SolvL/ProtL scales to examine the relationships between lipophilicity and physicochemical properties of peptides under different pH conditions. From a methodological point of view, the strategy relies on the combination of accurately parametrized version of continuum solvation models with an elaborate formalism to derived distribution coefficients from the partition of neutral and ionic species, in conjunction with the pK_a of ionizable groups. The simplicity of the computational strategy and the low cost of required calculations permit an straigthforward extension to non-standard residues, such as effect of chemical modifications on lipophilicity maps of proteins, thus providing information valuable to explore biomolecular recognition, and to modulate the properties of engineered polymeric materials.

ASSOCIATED CONTENT

Supporting Information.

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Detailed description of the computational strategy, Tables and figures showing complementary information about the SolvL and ProtL scales, and their application to several test systems.

AUTHOR INFORMATION

Corresponding Authors

E-mail: campaxic@gmail.com

E-mail: fjluque@ub.edu

ORCID

William J. Zamora: 0000-0003-4029-4528

Josep M. Campanera: 0000-0002-6698-874X

F. Javier Luque: 0000-0002-8049-3567

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministerio de Economia y Competitividad (SAF2017-88107-R), the Spanish María de Maeztu program (MDM-2017-0767), the Generalitat de Catalunya (2017SGR1746), and the Consorci de Serveis Universitaris de Catalunya (Molecular Recognition project). WJZ is fellow from the Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT), Consejo Nacional para Investigaciones Científicas y Tecnológicas (CONICIT; Costa Rica) and University of Costa Rica (UCR).

REFERENCES

(1) Tanford, C. The Hydrophobic Effect and the Organization of Living Matter. *Science* **1978**, *200*, 1012–1018.

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

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

(4) Peters, C.; Elofsson, A. Why is the Biological Hydrophobicity Scale More Accurate than Earlier Experimental Hydrophobicity Scales? *Proteins* **2014**, *82*, 2190–2198.

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

(6) 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.

(7) Fauchere, J. L.; Pliska, V. Hydrophobic Parameters Pi of Amino Acid Side Chains from the Partitioning of N-Acetyl-Amino Acid Amides. *Eur. J. Med. Chem.* **1983**, *18*, 369–375.

(8) 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.

(9) Cancès, 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.

(10) Mennucci, B. Polarizable Continuum Model. *WIRES Comput. Mol. Sci*. **2012**, *2*, 386–404.

(11) 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.

(12) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins: Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230*, 543–574.

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

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

(15) 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.

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

(17) 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.

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

(19) Koehler, J.; Woetzel, N.; Staritzbichler, R.; Sanders, C. R.; Meiler, J. A Unified Hydrophobicity Scale for Multispan Membrane Proteins. *Proteins* **2009**, *76*, 13–29.

(20) Musafia, B.; Buchner, V.; Arad, D. Complex Salt Bridges in Proteins: Statistical Analysis of Structure and Function. *J. Mol. Biol.* **1995**, *254*, 761–770.

(21) Tomlinson, J. H.; Ullah, S.; Hansen, P. E.; Williamson, M. P. Characterization of Salt Bridges to Lysines in the Protein G B1 Domain. *J. Am. Chem. Soc.* **2009**, *131*, 4674–4684.

(22) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno, E. B. Large Shifts in pKa Values of Lysine Residues Buried Inside a Protein. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5260–5265.

(23) André, I.; Linse, S.; Mulder, F. A. A. Residue-Specific pKa Determination of Lysine and Arginine Side Chains by Indirect ^{15}N and ^{13}C NMR Spectroscopy: Application to *apo* Calmodulin. *J. Am. Chem. Soc*. **2007**, *129*, 15805–15813.

(24) 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.

(25) ACD/I-Lab. Advanced Chemistry Development, Inc.: Toronto, ON, Canada; http://www.acdlabs.com.

(26) ChemAxon, Budapest, Hungary; http://www.chemaxon.com.

(27) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199–203.

(28) Hopp, T. P.; Woods, K. R. Prediction of Protein Antigenic Determinants from Amino Acid Sequences. *Proc. Natl. Acad. Sci.* **1981**, *78*, 3824–3828.

(29) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host-Guest Pentapeptides. *Biochemistry* **1996**, *35*, 5109–5124.

(30) 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.

(31) Hessa, T.; Kim, H.; Bihlamaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S.; Von, G. Recognition of Transmembrane Helices by the Endoplasmic Reticulum Translocon. *Nature* **2005**, *433*, 377–381.

(32) Janin, J. Surface and inside Volumne in Globular Proteins. *Nature* **1979**, *277*, 491–492.

(33) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

(34) Wimley, W. C.; Gawrisch, K.; Creamer, T. P.; White, S. H. Direct Measurement of Salt-Bridge Solvation Energies Using a Peptide Model System: Implications for Protein Stability. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2985-2990.

(35) Wilce, M. C. J.; Aguilar, M. I.; Hearn, M. T. W. Physicochemical Basis of Amino Acid Hydrophobicity Scales: Evaluation of Four New Scales of Amino Acid Hydrophobicity Coefficients Derived from RP-HPLC of Peptides. *Anal. Chem.* **1995**, *67*, 1210–1219.

(36) Biswas, K. M.; DeVido, D. R.; Dorsey, J. G. Evaluation of Methods for Measuring Amino Acid Hydrophobicities and Interactions. *J. Chromatogr. A* **2003**, *1000*, 637–655.

(37) König, G.; Bruckner, S.; Boresch, S. Absolute Hydration Free Energies of Blocked Amino Acids: Implications for Protein Solvation and Stability. *Biophys. J.* **2013**, *104*, 453–462.

(38) Amrhein, S.; Oelmeier, S. A.; Dismer, F.; Hubbuch, J. Molecular Dynamics Simulations Approach for the Characterization of Peptides with Respect to Hydrophobicity. *J. Phys. Chem. B* **2014**, *118*, 1707–1714.

(39) Houghten, R. A.; Degraw, S. T.; Met, M.; Phe, F.; Pro, P.; Ser, S.; Thr, T. Effect of Positional Environmental Domains on the Variation of High-Performance Liquid Chromatographic Peptide Retention Coefficients. *J. Chromatogr.* **1987**, *386*, 223–228.

(40) Reimer, J.; Spicer, V.; Krokhin, O. V. Application of Modern Reversed-Phase Peptide Retention Prediction Algorithms to the Houghten and DeGraw Dataset: Peptide Helicity and Its Effect on Prediction Accuracy. *J. Chromatogr. A* **2012**, *1256*, 160–168.

(41) Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discov. Today*. **2015**, *20*, 122–128.

(42) Buchwald, P.; Bodor, N. Octanol-Water Partition of Nonzwitterionic Peptides: Predictive Power of a Molecular Size-Based Model. *Proteins* **1998**, *30*, 86–99.

(43) Rose, P. W.; Prlic, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Di Costanzo, L.; Duarte, J. M.; Dutta, S.; Feng, Z; et al. The TCSB Protein data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nuc. Acids Res*. **2017**, *45*, D271–D281.

(44) Vita, R.; Mahajan, S.; Overton, J. A.; Dhanda, S. K.; Martini, S.; Cantrell, J. R.; Wheeler, D. K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 Update. *Nuc. Acids Res*. 2018, in press. DOI: 10.1093/nar/gky1006.

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

(46) Chen, Y.; Barkley, M. D. Toward Understanding Tryptophan Fluorescence in Proteins. *Biochemistry* **1998**, *37*, 9976–9982.

(47) D'Auria, S.; Staiano, M.; Kuznetsova, I.; Turoverov, K. K. The Combined Use of Fluorescence Spectroscopy and X-Ray Crystallography Greatly Contributes to Elucidating Structure and Dynamics of Proteins. *Reviews in Fluorescence 2005*; Geddes, C. D.; Lakowicz, J. R., Eds.; Springer: Boston, MA. 2005, 25–61.

Table 1. Solvent-like (SolvL) and Protein-like (ProtL) Lipophilicity Scales Based on the logD_{pH} Values Determined for *N*-Acetyl-*L*-Amino Acid Amides at Physiological pH. The experimental pK_a of Side Chain Ionizable Groups, and Calculated Partition Coefficients of Neutral (logP_N) and Ionized (logP_{I}) Residues Are Also Given.

Residue	Exp. pK_a		log P_N		log P_{I}		log $D_{7.4}$ [a]	
	SolvL	**ProtL**	**SolvL**	**ProtL**	**SolvL**	**ProtL**	**SolvL**	**ProtL**
Ala	-	-	-1.16	-2.47	-	-	-1.16 (0.85)	-2.47 (0.66)
Arg	**12.5**[b]	**12.5**[b]	**-2.86**	**-3.66**	**-2.99**	**-7.38**	**-2.99 (-0.98)**	**-7.04 (-3.91)**
Asn	-	-	-2.98	-3.97	-	-	-2.98 (-0.97)	-3.97 (-0.84)
Asp	**3.90**[c]	**3.50**[d]	**-2.26**	**-3.18**	**-2.80**	**-8.54**	**-2.80 (-0.79)**	**-5.87 (-2.74)**
Cys	**9.83**[e]	**6.80**[d]	**-0.16**	**-1.47**	**-4.19**	**-5.78**	**-0.16 (1.85)**	**-2.17 (0.96)**
Gln	-	-	-2.22	-4.00	-	-	-2.22 (-0.21)	-4.00 (-0.87)
Glu	**4.20**[c]	**4.20**[d]	**-1.49**	**-3.79**	**-3.38**	**-6.20**	**-3.36 (-1.35)**	**-5.96 (-2.83)**
Gly	-	-	-2.01	-3.13	-	-	-2.01 (0.00)	-3.13 (0.00)
His (δ)	**7.00**[c]	**6.60**[d]	**-1.20**	**-4.67**	**-4.06**	**-5.97**	**-1.35 (0.66)**	**-4.56 (-1.43)**
His (ε)	**7.00**[c]	**6.60**[d]	**-0.72**	**-4.98**	**-4.06**	**-5.97**	**-0.87 (1.14)**	**-4.97 (-1.84)**
Ile	-	-	-0.50	-0.38	-	-	-0.50 (1.51)	-0.38 (2.75)
Leu	-	-	0.05	-1.36	-	-	0.05 (2.06)	-1.36 (1.77)
Lys	**11.1**[c]	**10.5**[d]	**-0.40**	**-2.19**	**-3.24**	**-6.81**	**-3.18 (-1.17)**	**-5.08 (-1.95)**
Met	-	-	-0.51	-1.83	-	-	-0.51 (1.50)	-1.83 (1.30)
Phe	-	-	0.61	0.86	-	-	0.61 (2.62)	0.86 (3.99)
Pro	-	-	-0.77	-1.44	-	-	-0.77 (1.24)	-1.44 (1.69)
Ser	-	-	-2.04	-4.12	-	-	-2.04 (-0.03)	-4.12 (-0.99)
Thr	-	-	-1.22	-3.01	-	-	-1.22 (0.79)	-3.01 (0.12)
Trp	-	-	0.33	0.16	-	-	0.33 (2.34)	0.16 (3.29)
Tyr	**10.3**[c]	**10.3**[d]	**-0.49**	**-1.80**	**-4.21**	**-9.59**	**-0.49 (1.52)**	**-1.80 (1.33)**
Val	-	-	-0.93	-1.68	-	-	-0.93 (1.08)	-1.68 (1.45)

[a] Values for ionizable residues are shown in bold. Log $D_{7.4}$ values relative to glycine are given in parenthesis.
[b] Ref 14. [c] Ref 15. [d] Ref 16. [e] Ref 17.

Table 2. Statistical Parameters of the Comparison of the SolvL and ProtL Scales with Other Lipophilicity Scales. Comparison Was Made Using the Values Adapted to the Specific pH of Each Scale and Relative to Gly.

Scale[a]	**SolvL**				**ProtL**			
	mse[b]	***mue***	***rsmd***	***r*** ***p*-value**	***mse***	***mue***	***rsmd***	***r*** ***p*-value**
Bulk-Solvent Adapted Scale								
Fauchère - Pliska	-0.20	0.36	0.46	0.94 2×10^{-10}	0.36	0.98	1.28	0.92 6×10^{-9}
Eisenberg - McLachlan	-0.20	0.44	0.57	0.90 3×10^{-8}	0.36	1.08	1.35	0.91 2×10^{-8}
Hopp - Woods	-0.49	0.60	0.74	0.91 2×10^{-8}	0.07	0.84	1.08	0.89 9×10^{-8}
Wimley et al.	-0.60	1.02	1.16	0.59 0.006	0.04	1.24	1.64	0.61 4×10^{-3}
	-0.87[c]	0.92	1.03	0.87 2×10^{-6}	-0.30	1.03	1.25	0.87 2×10^{-6}
Biological-Based Scale								
Moon - Fleming	-0.12	0.57	0.67	0.94 4×10^{-10}	0.24	0.72	0.93	0.91 7×10^{-9}
Hessa et al.	-0.92	0.93	1.18	0.79 3×10^{-5}	-0.36	1.08	1.46	0.82 6×10^{-6}
Knowledge-Based Scale								
Koehler et al.	-0.91	1.10	1.33	0.78 4×10^{-5}	-0.35	1.55	1.87	0.80 2×10^{-5}
Janin et al.	-1.06	1.11	1.32	0.78 3×10^{-5}	-0.51	1.36	1.71	0.74 2×10^{-4}
Consensus Scale								
Kyte-Doolittle	-0.81	1.43	1.71	0.72 3×10^{-4}	-0.25	1.13	1.41	0.78 3×10^{-5}

[a] A physiological pH was considered in all cases, but for Wimley at al. and Moon-Fleming, since the corresponding pH was fixed at 9.0 and 3.8 following the specific experimental conditions.

[b] 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.

[c] Values in this row were obtainined upon exclusion of Arg and Lys. Since this scale was built up using model pentapeptides (AcWL-X-LL) at pH 9.0, Arg and Lys formed a salt bridge with the terminal carboxyl group in *n*-octanol as noted by ^{13}C-NMR studies.[34]

Figure 1. Representation of the pH Dependence of the SolvL (left) and ProtL (rigth) Lipophilicity Scales for Ionizable Amino Acids (Values Relative to Gly). Values Determined at pH of 2.1, 7.4 and 9.0 are Shown in Orange, Green and Blue, Respectively, and the Values of the Neutral Species (log P_N) are Shown in Black.

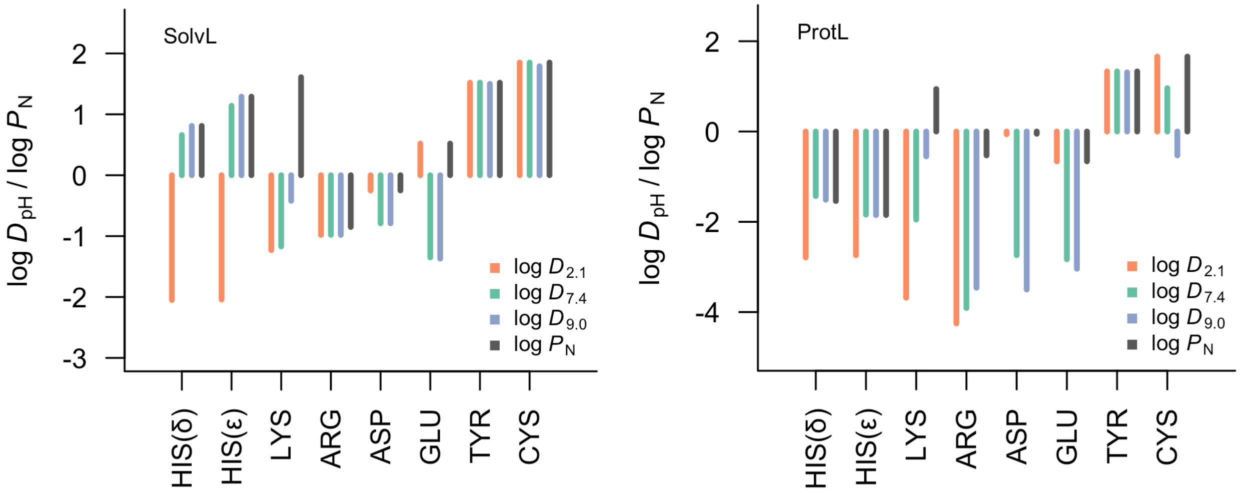


Figure 2. Comparison Between (left) SolvL and (right) ProtL Lipophilicity Scales Derived From the IEF/MST Solvation Model (Expressed as log $D_{7.4}$) and Fauchère-Pliska Experimental Values for the Twenty *N*-Acetyl-*L*-Amino Acid Amides (*r*: Pearson correlation coefficient; mse: Mean signed error; mue: Mean Unsigned Error; rmsd: Root-Mean Square Deviation). Regression Equations Shown in Table S8.

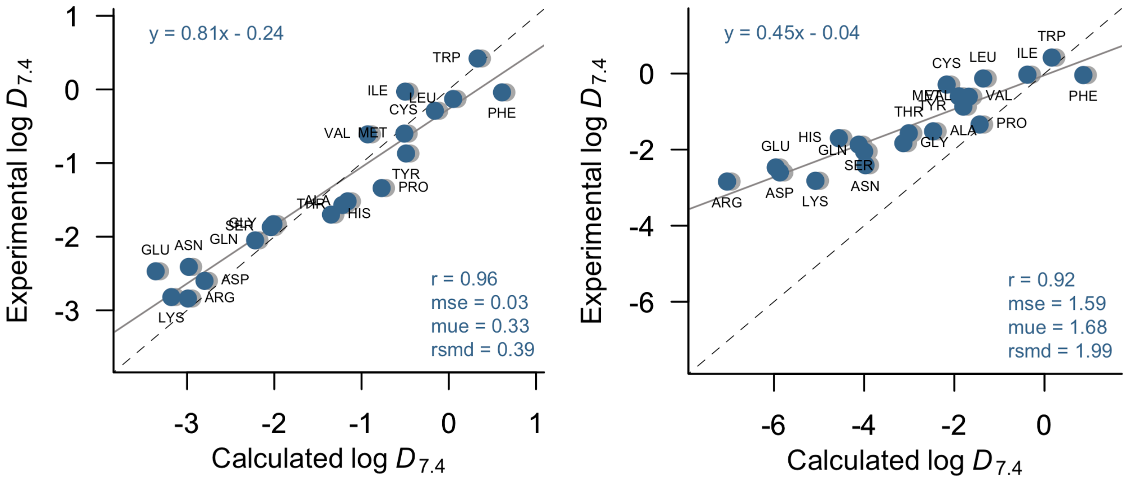


Figure 3. Representation of the Pearson Correlation Coefficient in the Comparison of the SolvL scale with Bulk Solvent-Based Scales, and ProtL Scale with Biological-Based, Knowledge-Based and Consensus Lipophilicity Scales at pH 3.8, 7.4 and 9.0 (Shown as Green, Red and Blue Lines, Respectively).

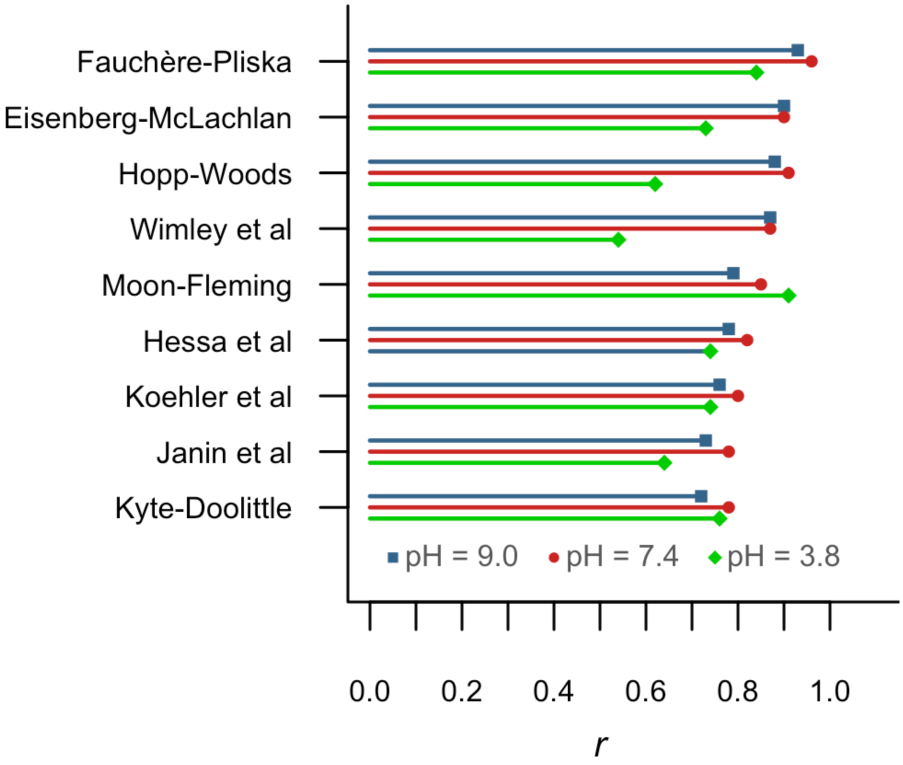


Figure 4. Relationship Between the Cumulative Lipophilicities Determined from the SolvL Scale Versus (A) the Retention Time for Eight 10-mer Peptides (pH 7.4; Ref. 38), (B) 248 Unique 13-mer Peptides (pH 2.1; Ref. 39,40), (C) log P_N for 118 Random Peptides (Ref. 42), and (D) log $D_{7.4}$ for 116 Random Peptides (Ref. 42). Regression Equations Shown in Table S8.

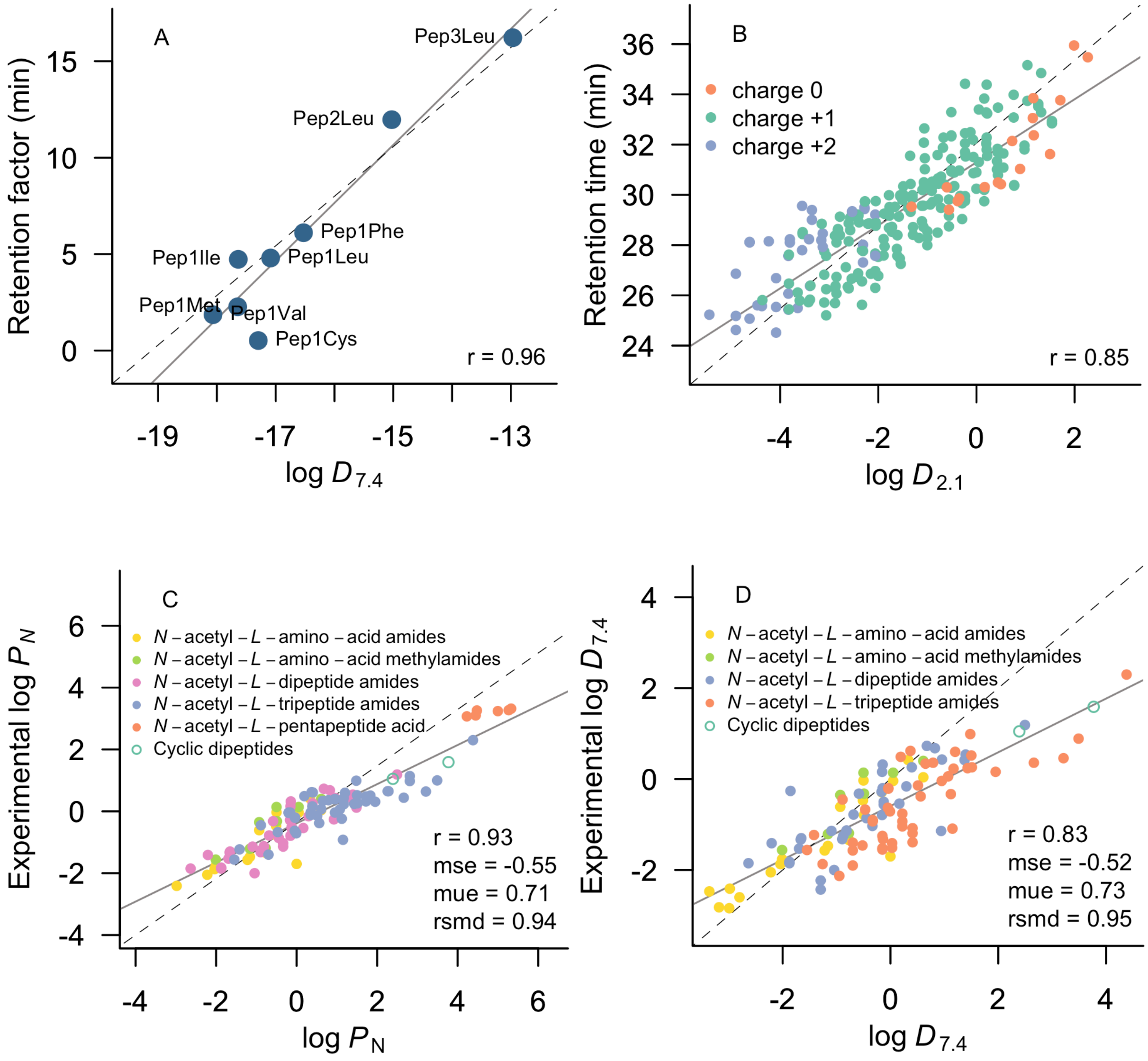
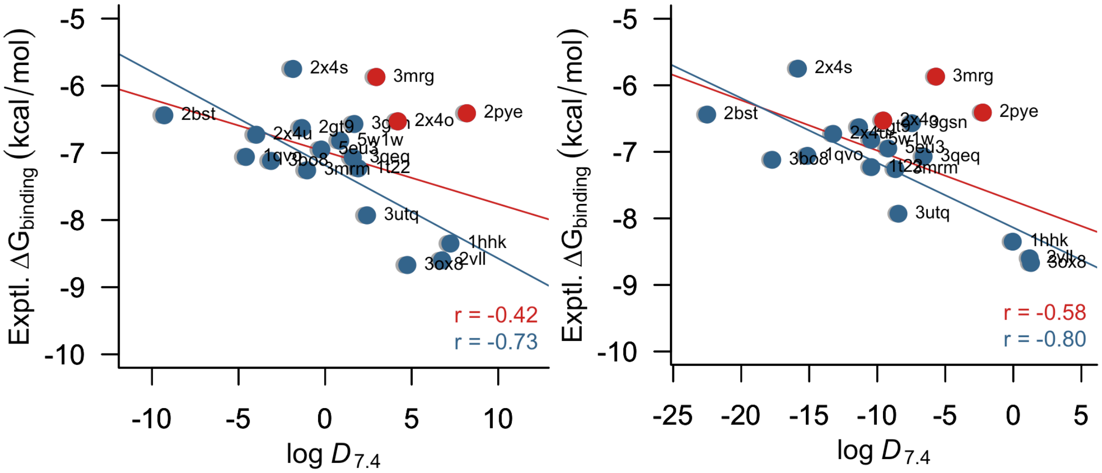


Figure 5. Relationship Between the Cumulative Lipophilicities Determined from (left) SolvL and (right) ProtL Scales Versus Experimental Binding Affinities of MHC-Bound Peptides. Cys-Containing Peptides Are Indicated as Red Dots. Regression Equations Shown in Table S8.


Exptl. ΔG_binding (kcal/mol)
2x4s
3mrg
2bst
2pye
2x4u
2gt9
3g n
2x4o
5w1w
5eu3
1qvo
3bo8
3qeq
3mrm
1t22
3utq
1hhk
3ox8
2vll
r = -0.42
r = -0.73
log D 7.4
Exptl. ΔG_binding (kcal/mol)
2x4s
3mrg
2bst
2pye
2x4o
3gsn
2x4u
2gt9
5w1w
5eu3
3qeq
3bo8
1qvo
1t22
3mrm
3utq
1hhk
2vll
3ox8
r = -0.58
r = -0.80
log D 7.4