Angewandte Chemie

Akzeptierter Artikel

Autoren: Waldemar Hoffmann, Jennifer Langenhan, Susanne Huhmann, Johann Moschner, Rayoon Chang, Matteo Accorsi, Jongcheol Seo, Rademann Jörg, Gerard Meijer, Beate Koksch, Michael T. Bowers, Gert von Helden, and Kevin Pagel

WILEY-VCH

WILEY-VCH

COMMUNICATION

An Intrinsic Hydrophobicity Scale for Amino Acids and Its Application to Fluorinated Compounds

Waldemar Hoffmann*, Jennifer Langenhan, Susanne Huhmann, Johann Moschner, Rayoon Chang, Matteo Accorsi, Jongcheol Seo, Jörg Rademann, Gerard Meijer, Beate Koksch, Michael T. Bowers, Gert von Helden and Kevin Pagel*

Abstract: The classification of amino acids according to their intrinsic properties, such as the physico-chemical properties and structure, yields fundamental insights into their role in interactions in biological processes. More than 100 hydrophobicity scales have been introduced, with each being based on a distinct condensed-phase approach. However, a comparison of the hydrophobicity values gained from different techniques, and their relative ranking is not straightforward, as the interactions between the environment and amino acid are unique to each method. Here, we overcome this limitation by studying the properties of amino acids in the clean-room environment of the gas phase. In the gas phase, entropic contributions from the hydrophobic effect are by default absent and only the side-chain´s polarity dictates self-assembly. This allows for the derivation of a novel hydrophobicity scale, which is solely based on the interaction between individual amino acid units within the cluster and thus more accurately reflects the intrinsic nature of a side-chain. This principle can be further applied to classify non-natural derivatives, as shown here for fluorinated amino acid variants.

The accurate determination of the intrinsic hydrophobicity of amino acids is crucial for the understanding of key aspects in biology and application of non-canonical amino acids in the rational design of peptides and proteins. Many fundamental biological processes such as folding,[1] stability[2] and oligomerization[3] of proteins as well as protein-ligand interactions[4] are strongly influenced by the hydrophobic effect in solution where the entropically unfavored solvent shell around non-polar residues is released to the bulk water. To date, more than 100 hydrophobicity scales[5] have been established, with most of them being derived from condensed-phase methods such as partitioning,[6] accessible surface area calculations,[7] direct measurements of physical properties[8] and chromatographic techniques.[9] Nevertheless, significant differences among these scales exist as they utilize markedly different principles or vary in the type of species investigated.[7, 10]

A more detailed assessment of hydrophobicity measurements reveals the limitations of current approaches. Scales based on partitioning use organic solvents such as octanol to mimic the protein interior and rank Trp as the most hydrophobic amino acid.[6] Organic solvents, however, often dissolve in water to a certain extent, thus altering the characteristics of both phases. This mixing makes it difficult to obtain an unbiased hydrophobicity scale. In contrast, surface area calculations utilize a database of protein crystal structures and define the hydrophobicity as the tendency of a residue to be found inside of a protein rather than on its surface.[7] Here, Cys is ranked as the most hydrophobic, because its thiol group can form disulfide bonds, frequently located inside a globular structure. The most popular scale based on physical properties was developed by measuring the surface tension of amino acid solutions in reference to a Gly solution.[8] Here, Leu is reported as the most hydrophobic, because it yields the largest decrease in surface tension. Pro, Arg, and Lys, however, exist in a different ionic state at their isoelectric points compared to the reference Gly, which introduces discrepancies compared to other hydrophobicity scales. Chromatographic techniques,[9] on the contrary, use amino acid derivatives or model peptides to define the hydrophobicity as a change in retention time relative to a Gly-substituted analog. In case of the model peptide approach, a change in peptide sequence,[9, 11] peptide length[11] and substitution position[12] strongly affects hydrophobicity values.[9-10, 12] Additionally, the choice of pore diameter, aqueous buffer pH and temperature, or bonding density of the stationary reverse phase alkyl chains also influence the hydrophobicity scale.[5b]

Most common hydrophobicity scales generally do not allow a universal comparison and classification of amino acids because they are often biased by the employed methodology. Here, we suggest an alternative hydrophobicity ranking obtained by studying the interaction of amino acids in the clean-room environment of the gas phase. Although it may appear counterintuitive at first glance, gas-phase conditions are particularly suitable for such investigations, since the underlying relative permittivity in vacuum (ε_r = 1) closely resembles that of

[*]   Dr. W. Hoffmann, Dr. S. Huhmann, Dr. J. Moschner, R. Chang, M. Accorsi, Prof. Dr. J. Rademann, Prof. Dr. B. Koksch, Prof. Dr. K. Pagel
        Freie Universität Berlin
        Department of Biology, Chemistry and Pharmacy
        Takustraße 3 / Königin-Luise-Straße 2+4, 14195 Berlin (Germany)
        E-Mail: w.hoffmann@fu-berlin.de
        E-mail: kevin.pagel@fu-berlin.de

        Dr. W. Hoffmann, J. Langenhan, R. Chang, Prof. Dr. J. Seo≠, Prof. Dr. G. Meijer, Prof. Dr. G. von Helden, Prof. Dr. K. Pagel
        Fritz-Haber-Institut der Max-Planck-Gesellschaft
        Department of Molecular Physics
        Faradayweg 4-6, 14195 Berlin (Germany)
        ≠ present address: University of Science and Technology (POSTECH), Department of Chemistry, 77 Cheongam-ro, Pohang 37673 (Korea)

        Prof. Dr. M. T. Bowers
        University of California Santa Barbara
        Department of Chemistry & Biochemistry
        Santa Barbara, California 93106 (USA)

        Supporting information for this article is given via a link at the end of the document.

COMMUNICATION
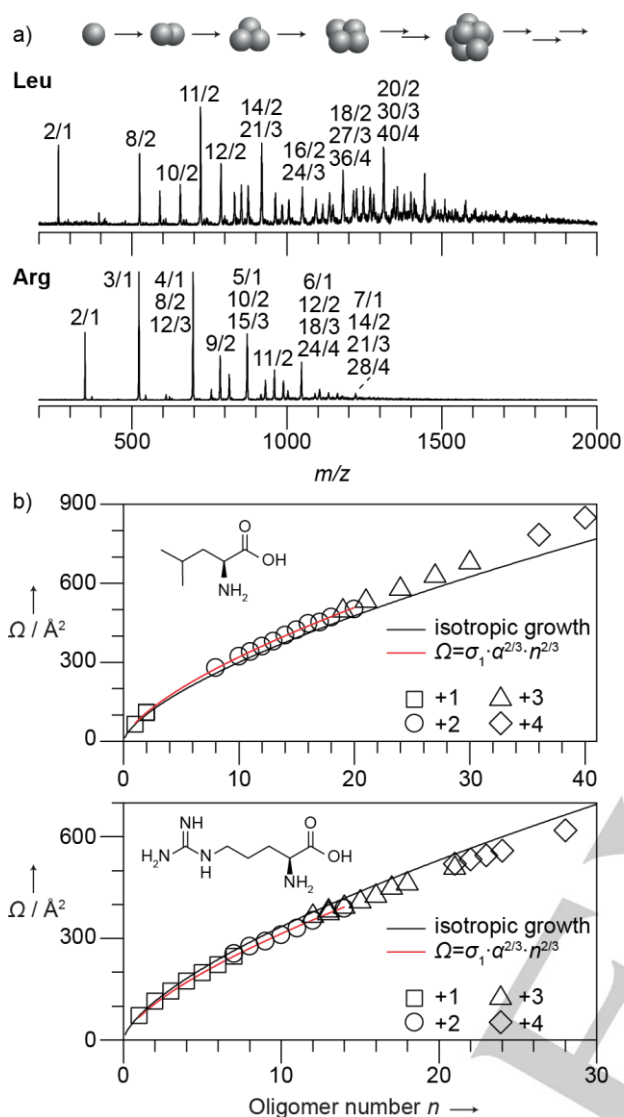
Figure 1. Mass spectra and collision cross-sections (Ω, $^{DT}CCS_{He}$) for Leu and Arg. a) *n*-ESI mass spectra obtained from concentrated (5 mM) aqueous amino acid solutions. The most abundant clusters are labelled with their *n/z* ratio, where *n* represents the number of amino acid units in the cluster and *z* the charge. b) Ω as a function of the oligomer number *n*. The solid black line represents a theoretical isotropic growth,[13] *i.e.* growth of an idealized spherical assembly, whereas the red line shows the fit to derive the respective hydrophobicity value *α*. The error of the measured $^{DT}CCS_{He}$ is considerably smaller (typically < 1% for three independent replicas) than the size of the symbol.

the protein interior[14] (ε_r = 6–7). There are already promising studies in which the physico-chemical properties of molecules are investigated in the gas phase, as for example *via* differential mobility spectrometry.[15] Our study utilizes the gas-phase technique ion mobility-mass spectrometry (IM-MS), which separates ions according to their mass-to-charge ratio (*m/z*) as well as their size and shape.[16] It provides a rotationally-averaged collision cross-section of an ion (Ω, CCS)–a molecular property that is specific to the ion-buffer gas interaction and provides a measure for the unit volume of amino acids in clusters.[17]

In this work, $^{DT}CCS_{He}$ (CCS measured in He buffer gas using a drift tube instrument,[17] here denoted as Ω) is used to explore the relationship between amino acid cluster size and side-chain polarity. Figure 1 a shows *nano*-electrospray ionization (*n*-ESI) mass spectra of Leu and Arg (5 mM) sprayed from a pure aqueous solution. Leu assembles into a dimer *n/z* = 2/1 along with larger clusters starting from an octamer up to a 36-mer with *n/z* = 36/4, where *n* stands for the number of Leu units in the cluster and *z* for the charge. The more polar Arg, which carries a guanidine moiety, behaves differently: It aggregates in a more consecutive manner and clusters up to a 24-mer are observed. Other amino acids assemble in a similar fashion (see Supporting information and literature[18]).

The CCSs as a function of the oligomer number *n* as measured by IM-MS are shown in Figure 1 b for Leu (top) and Arg (bottom). The uncertainties in the measured CCSs are considerably smaller than the actual size of the corresponding symbol. The black solid line corresponds to the theoretical isotropic growth,[19] representing the growth of an idealized spherical assembly. It is obtained from the equation $\Omega = \sigma_1 \cdot n^{2/3}$, where σ_1 is the CCS of the monomer and *n* the number of amino acid units in the cluster. From a visual inspection, it appears that Leu forms more extended clusters than predicted by theoretical isotropic growth, whereas the polar Arg assembles into more compact oligomers. The resulting packing efficiency does not depend on the overall size of the monomeric units (Ω_{Leu} = 66 Å^2 *vs.* Ω_{Arg} = 72 Å^2), which indicates that cluster formation is strongly influenced by the polarity of the side-chains. A similar dependence between cluster growth and side-chain polarity was recently observed for selected amino acids[18c] and is confirmed more systematically for all canonical amino acids within this work (see Supporting information). This data clearly shows that hydrophobic amino acids generally form larger clusters than polar residues. Their non-polar side-chains likely orient themselves towards the low permittivity of the gas phase, which makes them "bulky" on the outside. Polar amino acids rather adopt more compact structures as their functional groups seek to maximize intermolecular interactions.

In order to systematically evaluate the aforementioned trend in cluster growth, a correction factor *α* was derived to account for the deviation from the theoretical isotropic growth, according to $\Omega = \sigma_1 \cdot n^{2/3} \cdot \alpha^{2/3}$. This *α* value provides a measure for the packing efficiency in the cluster and directly correlates to the polarity of each side-chain. Values of *α* > 1 represent hydrophobic amino acids, whereas *α* < 1 indicates hydrophilic side-chains. The typical error of *α* is lower than 1%. As such, *α* represents the ideal basis for a novel, unbiased hydrophobicity scale for amino acids.
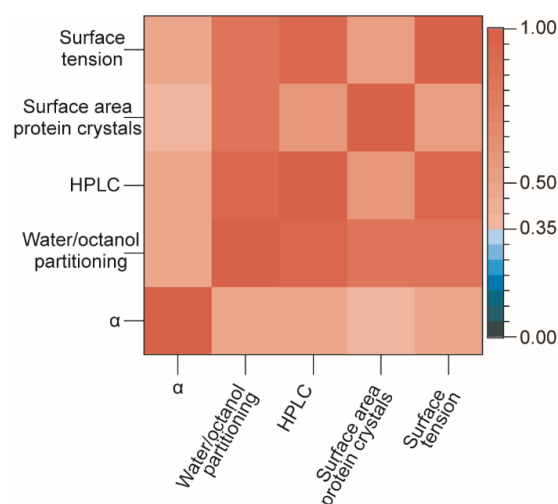
A summary of the *α* values as a function of the size (Ω) is given in Figure 2. The investigated amino acids differ in their cluster formation propensity, but all of them form clusters up to charge state 2+. In addition, amino acids with side-chains that carry an additional charge would have an additional influence on the cluster assembly. The strength of resulting interactions would depend on the nature of the side-chains.[20] Thus, to ensure comparable datasets and to circumvent the influence of possible electrostatic interactions such as ion-dipole/ -induced dipole or Coulomb interactions on the packing efficiency in higher charge states, only charge states 1+ and 2+ were used to derive *α*.

COMMUNICATION
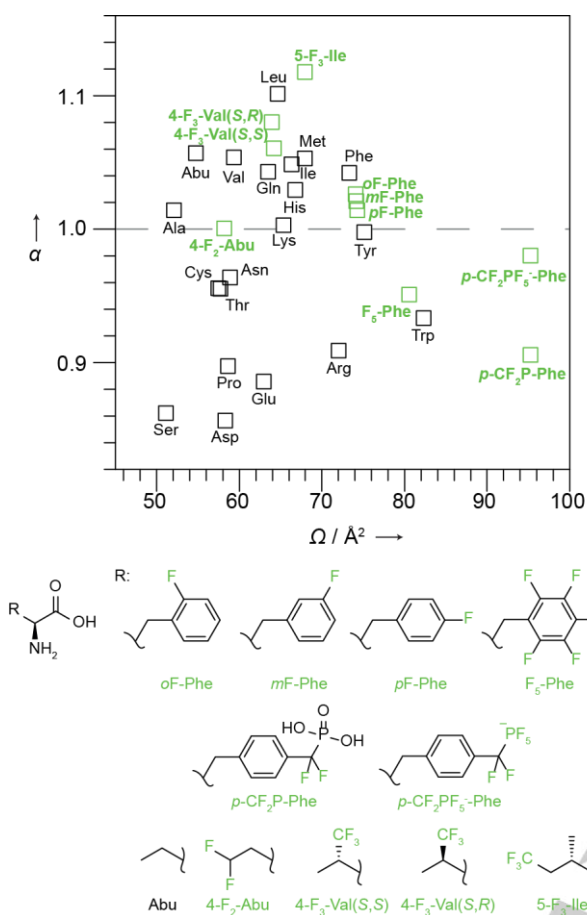
Figure 2. Relative hydrophobicity scale for amino acids. The hydrophobicity α as a function of amino-acid size (given as the monomer $^{DT}CCS_{He}$ Ω). Values for $\alpha > 1$ represent hydrophobic amino acids, whereas $\alpha < 1$ stands for hydrophilic side-chains. Fluorinated variants are shown in green.

Figure 3. Heat map of the Pearson correlation coefficients $|R|$ between the here-presented hydrophobicity scale α of all canonical amino acids and scales based on condensed-phase approaches such as water/octanol partitioning,[6a] HPLC,[9] calculation of the accessible surface area of a residue within a protein crystal[7] and measurement of the surface tension of an amino acid solution.[8] A positive correlation is shown in red. No linear correlation is indicated in blue.

The resulting hydrophobicity scale ranks the natural amino acids Leu > Val ≈ Met > Ile > Phe as most hydrophobic, which is in good qualitative agreement with previous scales.[6a, 9] In addition, the new scale indicates a plausible relative ranking of amino acids from a chemical point of view: 1) Phe ($\alpha = 1.042$) is more hydrophobic than Tyr ($\alpha = 0.998$), which carries one additional hydroxyl group at the phenyl ring. 2) Ser is one of the most hydrophilic amino acids ($\alpha = 0.862$), which is in good agreement with previous studies.[18c, 21] The primary alcohol makes Ser more hydrophilic than Thr ($\alpha = 0.932$) with a secondary alcohol. 3) Gln ($\alpha = 1.043$) is more hydrophobic than Asn ($\alpha = 0.964$) due to the longer aliphatic chain, whereas both Gln and Asn are less hydrophilic than their corresponding carboxylic acid analogues (Glu; $\alpha = 0.886$ and Asp; $\alpha = 0.856$). 4) Lys and Arg share their long aliphatic chain with either a guanidine group or a primary amine at its end. However, the guanidine group is more polar, and consequently Arg ($\alpha = 0.909$) has a lower α value than Lys ($\alpha = 1.003$). Interestingly, Lys shows neither a very polar nor a hydrophobic character within the here-presented scale. This result contradicts condensed-phase scales[6a, 7, 9] which rank Lys as one of the most polar amino acids. In solution, the Lys side-chain is predominantly protonated, whereas we examined the

intrinsic hydrophobicity of an on average neutral side-chain in the gas phase. Thus, the long aliphatic chain outweighs the hydrophilic character of the neutral amine and yields an α value of ~ 1. We believe that this relative ranking for Lys more accurately depicts the underlying nature of its Janus-headed side-chain, for which protonation can vary drastically when buried within a protein environment.[22]

Figure 3 depicts a quantitative comparison between the here-presented hydrophobicity scale and scales based on condensed-phase approaches, where absolute Pearson correlation coefficients $|R|$ are displayed as a heat map. A value of $|R| = 1$ (red) indicates a perfect correlation where all data points lie on a line, whereas an $|R|$ value of 0 (blue) implies no correlation between the two scales. A very high correlation ($|R| > 0.6$) between α and other hydrophobicity scales is generally not observed, as they are based on vastly different approaches (gas phase *vs.* condensed phase). Condensed-phase methods are influenced by solvent effects, the type of investigated species (e.g. different peptides), and parameters such as pH, chromatographic equipment, as well as solubility, thus contributing to the pattern of the correlation matrix. However, the data indicate a positive relationship ($|R| > 0.35$) between the relative ranking of hydrophobicity values based on α and all other scales, which support the validity of the here-presented approach.

Moreover, the robustness of the new approach to classify non-natural derivatives was tested for a particularly challenging class: fluorinated amino acids. Fluorine substitution is a common strategy to modulate the properties of pharmaceuticals[23] and peptides/proteins.[24] Its impact on folding is determined by a complex interplay of the interaction partner, changes in hydrophobicity and size, complicating the prediction of their properties.[25] The hydrophobicity values for selected fluorinated amino acid analogous of Ile, Leu and Phe as well as 2-aminobutyric acid (Abu) are shown in green in Figure 2.

COMMUNICATION

In general, CF$_3$-fluorination in aliphatic side-chains increases the hydrophobicity compared to the unsubstituted analogues Ile and Val. The CF$_3$-substitution in (2S,3S)-5,5,5-trifluoroisoleucine (5-F$_3$-Ile), however, only marginally alters the overall size (Ω_{Ile} = 66 Å^2 *vs.* Ω_{5-F_3-Ile} = 68 Å^2), whereas for 4,4,4-trifluorovaline (4-F$_3$-Val), an increase of ~ 8–12% in CCS is observed (Ω_{Val} = 59 Å^2 *vs.* Ω_{4-F_3-Val} = 64 Å^2). Interestingly, the fluorinated diastereomers of 4,4,4-trifluorovaline yield different hydrophobicity values: The (2S,3S)-4-F$_3$ Val isomer (α = 1.061) is considerably more hydrophilic than 4-F$_3$-Val(S,R) (α = 1.080), but both are more hydrophobic than Val (α = 1.053). This observation is in good agreement with HPLC results[3b] as well as theory,[26] and indicates that the here-presented approach is sensitive to small variations within a given structure.

Moreover, a CF$_2$-fluorination leads to a completely different behavior: 4,4-difluoroaminobutyric acid (4-F$_2$-Abu; α = 1.000) exhibits a smaller α value than its non-fluorinated analogue (Abu; α = 1.057). Thus, partial fluorination in aliphatic side-chains can decrease the overall hydrophobicity of a given amino acid.[25b] Such prediction upon fluorination is not trivial, but amino acids can be readily classified using the here-presented approach.

The incorporation of fluorine into phenyl rings exhibits a different behavior: The H/F substitution reduces hydrophobicity in the following order: Phe (α = 1.042) > oF-Phe (α = 1.026) > mF-Phe (α = 1.021) > pF-Phe (α = 1.014) > F$_5$-Phe (α = 0.951). This rather unusual trend is likely a result of changes in the electronic structure of the ring. The change of dipole moment upon fluorination leads to an increase in polarity resulting in more densely packed clusters. This effect is even more pronounced when a phosphonate group (R-CF$_2$-PO(OH)$_2$ for p-CF$_2$P-Phe; α = 0.906) is attached to the phenyl ring (see Phe *vs.* p-CF$_2$P-Phe). Yet when the phosphonate group is perfluorinated to yield a hypervalent R-CF$_2$-PF$_5^-$-group that carries one permanent negative charge (see p-CF$_2$PF$_5^-$-Phe),[27] an increase in hydrophobicity (α = 0.980) is observed compared to the neutral phosphonate group in p-CF$_2$P-Phe. This confirms that subtle changes in the fluorination pattern of amino acids can indeed lead to vast changes in their hydrophobicity and aggregation behavior.

In summary, we present a novel and unbiased hydrophobicity scale based on the clustering of amino acids in the gas phase. Under these clean-room conditions the entropic contribution from solvation, which leads to the hydrophobic effect, is explicitly absent. As a result, this low-permittivity environment resembles that of a densely packed protein interior. Typically, hydrophobic residues form extended clusters where their non-polar side-chains are exposed to the gas-phase exterior, while polar residues form compact clusters to maximize electrostatically driven intermolecular interactions. In order to perform a quantitative assessment and classify natural as well as several non-natural fluorinated amino acids, a correction factor α was employed, which provides a measure for the deviation from isotropic cluster growth. The here-presented method represents a general approach that allows the precise determination of the intrinsic, unbiased hydrophobicity of amino acids. This approach not only includes natural building blocks, but also artificial compounds with complex conditions that make predictions of this property difficult or impossible. Thus, our method represents a valuable tool in the context of peptide, protein and drug design.

[1] a) G. D. Rose, R. Wolfenden, *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381-415; b) V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 415-427.

[2] a) K. Yutani, K. Ogasahara, T. Tsujita, Y. Sugino, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4441-4444; b) B. Tripet, K. Wagschal, P. Lavigne, C. T. Mant, R. S. Hodges, *J. Mol. Biol.* **2000**, *300*, 377-402.

[3] a) F. Chiti, N. Taddei, F. Baroni, C. Capanni, M. Stefani, G. Ramponi, C. M. Dobson, *Nat. Struct. Mol. Biol.* **2002**, *9*, 137; b) U. I. M. Gerling, M. Salwiczek, C. D. Cadicamo, H. Erdbrink, C. Czekelius, et al., *Chem. Sci.* **2014**, *5*, 819-830.

[4] a) L. Young, R. L. Jernigan, D. G. Covell, *Protein Sci.* **1994**, *3*, 717-729; b) R. G. Efremov, A. O. Chugunov, T. V. Pyrkov, J. P. Priestle, A. S. Arseniev, E. Jacoby, *Curr. Med. Chem.* **2007**, *14*, 393-415.

[5] a) C. C. Palliser, D. A. Parry, *Proteins* **2001**, *42*, 243-255; b) K. M. Biswas, D. R. DeVido, J. G. Dorsey, *J. Chrom. A* **2003**, *1000*, 637-655.

[6] a) J. -L. Fauchère, V. Pliska, *Eur. J. Med. Chem.* **1983**, *18*, 369-375; b) Y. Nozaki, C. Tanford, *J. Biol. Chem.* **1971**, *246*, 2211-2217.

[7] G. Rose, A. Geselowitz, G. Lesser, R. Lee, M. Zehfus, *Science* **1985**, *229*, 834-838.

[8] H. B. Bull, K. Breese, *Arch. Biochem. Biophys.* **1974**, *161*, 665-670.

[9] J. M. Kovacs, C. T. Mant, R. S. Hodges, *Biopolymers* **2006**, *84*, 283-297.

[10] C. T. Mant, J. M. Kovacs, H. M. Kim, D. D. Pollock, R. S. Hodges, *Biopolymers* **2009**, *92*, 573-595.

[11] J. L. Meek, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1632-1636.

[12] B. Tripet, D. Cepeniene, J. M. Kovacs, C. T. Mant, O. V. Krokhin, R. S. Hodges, *J. Chrom. A* **2007**, *1141*, 212-225.

[13] C. Bleiholder, T. D. Do, C. Wu, N. J. Economou, S. S. Bernstein, et al., *J. Am. Chem. Soc.* **2013**, *135*, 16926-16937.

[14] L. Li, C. Li, Z. Zhang, E. Alexov, *J. Chem. Theory Comput.* **2013**, *9*, 2126-2136.

[15] C. Liu, J. C. Y. Le Blanc, B. B. Schneider, J. Shields, J. J. Federico, et al., *ACS Cent. Sci.* **2017**, *3*, 101-109.

[16] a) G. von Helden, M. T. Hsu, N. Gotts, M. T. Bowers, *J. Phys. Chem.* **1993**, *97*, 8182-8192; b) D. E. Clemmer, M. F. Jarrold, *J. Mass Spectrom.* **1997**, *32*, 577-592.

[17] V. Gabelica, A. A. Shvartsburg, C. Afonso, P. Barran, J. L. P. Benesch, et al., *Mass Spectrom. Rev.* **2019**, *0*.

[18] a) C. K. Meng, J. B. Fenn, *Org. Mass Spectrom.* **1991**, *26*, 542-549; b) P. Nemes, G. Schlosser, K. Vékey, *J. Mass Spectrom.* **2005**, *40*, 43-49; c) T. D. Do, N. E. C. de Almeida, N. E. LaPointe, A. Chamas, S. C. Feinstein, M. T. Bowers, *Anal. Chem.* **2016**, *88*, 868-876.

[19] C. Bleiholder, N. F. Dupuis, T. Wyttenbach, M. T. Bowers, *Nat. Chem.* **2011**, *3*, 172-177.

[20] J. Seo, W. Hoffmann, S. Malerz, S. Warnke, M. T. Bowers, K. Pagel, G. von Helden, *Int. J. Mass Spectrom.* **2018**, *429*, 115-120.

[21] S. C. Nanita, R. G. Cooks, *Angew. Chem. Int. Ed.* **2006**, *45*, 554-569.

[22] D. G. Isom, C. A. Castañeda, B. R. Cannon, B. García-Moreno E., *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5260-5265.

[23] A. Vulpetti, C. Dalvit, *Drug Discov. Today* **2012**, *17*, 890-897.

[24] L. M. Gottler, R. de la Salud Bea, C. E. Shelburne, A. Ramamoorthy, E. N. G. Marsh, *Biochemistry* **2008**, *47*, 9243-9250.

[25] a) M. Salwiczek, E. K. Nyakatura, U. I. Gerling, S. Ye, B. Koksch, *Chem. Soc. Rev.* **2012**, *41*, 2135-2171; b) A. A. Berger, J. S. Voller, N. Budisa, B. Koksch, *Acc. Chem. Res.* **2017**, *50*, 2093-2103.

[26] J. R. Robalo, S. Huhmann, B. Koksch, A. Vila Verde, *Chem.* **2017**, *3*, 881-897.

[27] S. Wagner, M. Accorsi, J. Rademann, *Chem. Eur. J.* **2017**, *23*, 15387-15395.

WILEY-VCH

COMMUNICATION

COMMUNICATION
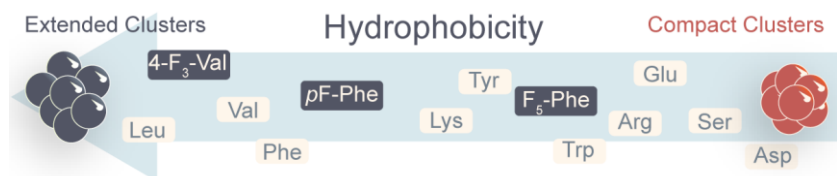
The size and shape of gas-phase clusters were used to classify natural and fluorinated amino acids according to their intrinsic hydrophobicity.

W. Hoffmann*, J. Langenhan, S. Huhmann, J. Moschner, R. Chang, M. Accorsi, J. Seo, J. Rademann, B. Koksch, G. Meijer, M. T. Bowers, G. von Helden and K. Pagel*

Page No. – Page No.

An Intrinsic Hydrophobicity Scale for Amino Acids and Its Application to Fluorinated Residues