

Hydrophobicity Scale

 International Edition: DOI: 10.1002/anie.201813954
 German Edition: DOI: 10.1002/ange.201813954

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 Kokschi, Michael T. Bowers, Gert von Helden, and Kevin Pagel*

Abstract: 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 the 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 polarity of the side chain dictates the self-assembly. This allows for the derivation of a novel hydrophobicity scale, which is based solely 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 understanding the key aspects of biology and the application of noncanonical amino acids in the rational design of peptides and proteins. Many fundamental biological processes such as the 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 nonpolar 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 water/octanol partitioning,^[6] calculations of the accessible surface area,^[7] direct measurements of physical properties,^[8] and chromatographic techniques.^[9] Nevertheless, significant differences between 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 a protein rather than on its surface.^[7] Here, Cys is ranked as the most hydrophobic, because its thiol group can form disulfide bonds, which are 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 to be 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] in contrast, use amino acid derivatives or model peptides to define the hydrophobicity as a change in the retention time relative to a Gly-substituted analogue. In the case of the model peptide approach, a change in peptide sequence,^[9,11] peptide length,^[11] and substitution position^[12] strongly affects the hydrophobicity values.^[9,10,12] Additionally, the choice of the pore diameter, pH value and temperature of the aqueous buffer, as well as the bonding density of the alkyl chains in the stationary reverse phase 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 that is 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

[*] Dr. W. Hoffmann, Dr. S. Huhmann, Dr. J. Moschner, R. Chang, M. Accorsi, Prof. Dr. J. Rademann, Prof. Dr. B. Kokschi, 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
 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)
 Prof. Dr. M. T. Bowers
 University of California Santa Barbara
 Department of Chemistry & Biochemistry
 Santa Barbara, California 93106 (USA)
 Prof. Dr. J. Seo
 present address: University of Science and Technology (POSTECH)
 Department of Chemistry, 77 Cheongam-ro, Pohang 37673 (Korea)

 Supporting information and the ORCID identification numbers for the authors of this article can be found under:
<https://doi.org/10.1002/anie.201813954>

are particularly suitable for such investigations, since the underlying relative permittivity in a vacuum ($\epsilon_r = 1$) closely resembles that of the protein interior^[14] ($\epsilon_r = 6-7$). There are already promising studies in which the physicochemical properties of molecules are investigated in the gas phase, for example by differential mobility spectrometry.^[15] Our study utilizes the gas-phase technique of 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 interaction between the ion and the buffer gas and provides a measure of the unit volume of amino acids in clusters.^[17]

In this study, ${}^{\text{DT}}\text{CCS}_{\text{He}}$ (CCS measured in He buffer gas using a drift tube instrument,^[17] here denoted as Ω) is used to explore the relationship between the size of the amino acid cluster and the polarity of the side chain. Figure 1 a shows nano-electrospray ionization (n -ESI) mass spectra of Leu and Arg (5 mM) sprayed from an 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 stepwise manner and clusters up to a 24-mer are observed. Other amino acids assemble in a similar fashion (see the Supporting information and Ref. [18]).

Figure 1 b shows the CCSs as a function of the oligomer number n as measured by IM-MS 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] which represents the growth of an idealized spherical assembly. It is obtained from the equation $\Omega = \sigma_1 n^{2/3}$, 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 ($\Omega_{\text{Leu}} = 66 \text{ \AA}^2$ versus $\Omega_{\text{Arg}} = 72 \text{ \AA}^2$), which indicates that cluster formation is strongly influenced by the polarity of the side chains. A similar relationship between cluster growth and side-chain polarity was recently observed for selected amino acids^[18c] and is confirmed herein more systematically for all canonical amino acids (see the Supporting information). These data clearly show that hydrophobic amino acids generally form larger clusters than polar residues. Their nonpolar side chains likely orient themselves towards the low permittivity of the gas phase, which makes them “bulky” on the outside. Polar amino acids prefer to adopt more compact structures as their functional groups seek to maximize intermolecular interactions.

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, to give the equation $\Omega = \sigma_1 n^{2/3} \alpha^{2/3}$. This α value provides a measure

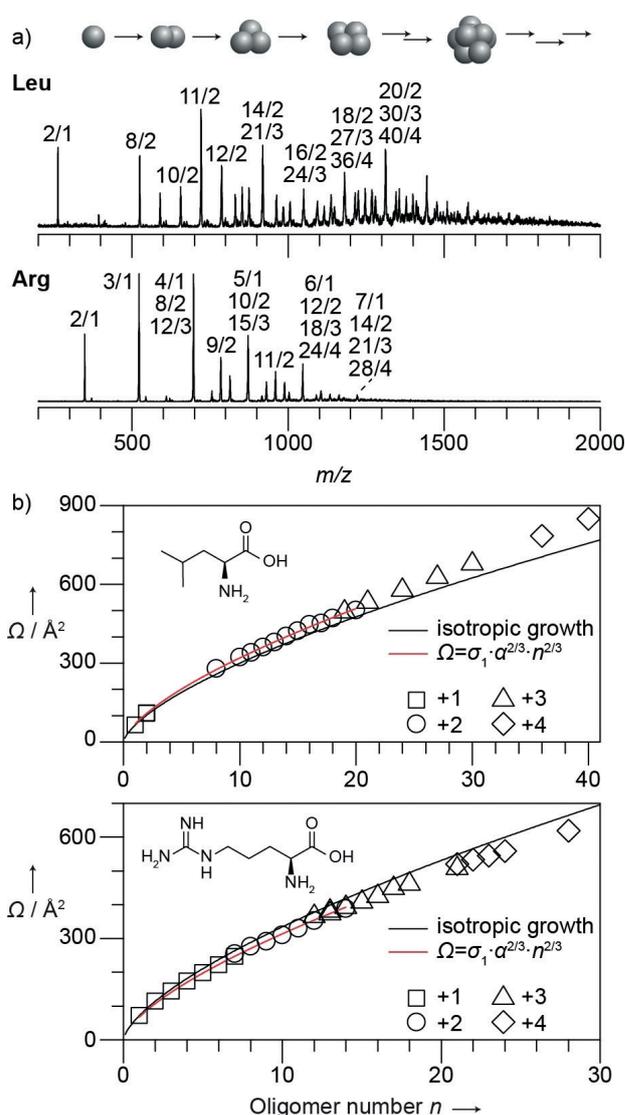


Figure 1. Mass spectra and collision cross-sections (Ω , ${}^{\text{DT}}\text{CCS}_{\text{He}}$) for Leu and Arg. a) n -ESI mass spectra obtained from concentrated (5 mM) aqueous amino acid solutions. The most abundant clusters are labeled with their n/z ratio, where n represents the number of amino acid units in the cluster and z the charge. b) Ω as a function of the oligomer number n . The black solid line represents a theoretical isotropic growth,^[19] that is, the 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 ${}^{\text{DT}}\text{CCS}_{\text{He}}$ is considerably smaller (typically $< 1\%$ for three independent replicates) than the size of the symbol.

of the packing efficiency in the cluster and directly correlates to the polarity of each side chain. Values of $\alpha > 1$ represent hydrophobic amino acids, whereas $\alpha < 1$ 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 propensity for cluster formation, 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 addi-

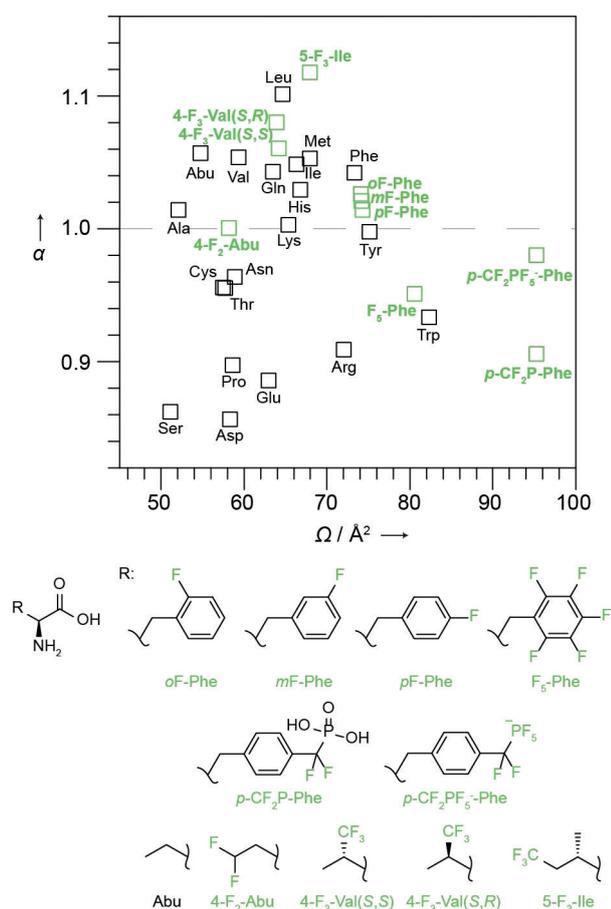


Figure 2. Relative hydrophobicity scale for amino acids. The hydrophobicity α as a function of the size of the amino acid (given as the monomer ${}^{\text{DTCCS}}\Omega$). Values for $\alpha > 1$ represent hydrophobic amino acids, whereas $\alpha < 1$ denotes hydrophilic side chains. Fluorinated variants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

tional influence on the cluster assembly. The strength of the 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 .

The resulting hydrophobicity scale ranks the natural amino acids $\text{Leu} > \text{Val} \approx \text{Met} > \text{Ile} > \text{Phe}$ 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 hydroxy 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 ($\alpha = 0.932$), which has a secondary alcohol.
- 3) Gln ($\alpha = 1.043$) is more hydrophobic than Asn ($\alpha = 0.964$) because of 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 carry either a guanidine group or a primary amine at the end of their long aliphatic chain. 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 scale presented here. 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 about 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 of the here-presented hydrophobicity scale and scales based on condensed-phase approaches, where absolute Pearson correla-

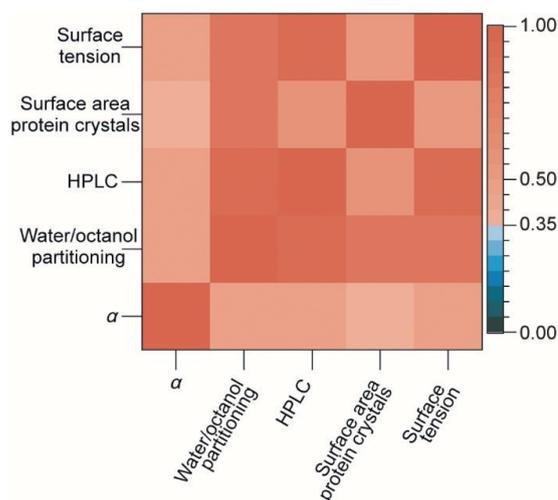


Figure 3. Heat map of the Pearson correlation coefficients $|R|$ between the here-presented hydrophobicity scale α of all the canonical amino acids and scales based on condensed-phase approaches such as water/octanol partitioning,^[6a] HPLC,^[9] calculation of the accessible surface area of a residue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

tion 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 versus condensed phase). Scales based on condensed-phase methods are influenced by solvent effects, the type of investigated species (e.g. different peptides), and parameters such as the pH value, chromatographic equipment, as well as solubility, which accounts for

the differences in 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 the other scales, which support the validity of the approach presented here.

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 as well as changes in the hydrophobicity and size, thus complicating the prediction of their properties.^[25] The hydrophobicity values for selected fluorinated amino acid analogues of Ile, Leu, and Phe as well as 2-aminobutyric acid (Abu) are shown in green in Figure 2.

In general, CF₃ fluorination of the aliphatic side chains increases the hydrophobicity compared to the unsubstituted analogues Ile and Val. The CF₃ substitution in (2*S*,3*S*)-5,5,5-trifluoroisoleucine (5-F₃-Ile), however, only marginally alters the overall size ($\Omega_{\text{Ile}} = 66 \text{ \AA}^2$ versus $\Omega_{5\text{-F}_3\text{-Ile}} = 68 \text{ \AA}^2$), whereas an increase of 8–12% in the CCS is observed for 4,4,4-trifluorovaline (4-F₃-Val; $\Omega_{\text{Val}} = 59 \text{ \AA}^2$ versus $\Omega_{4\text{-F}_3\text{-Val}} = 64 \text{ \AA}^2$). Interestingly, the fluorinated diastereomers of 4,4,4-trifluorovaline yield different hydrophobicity values: The (2*S*,3*S*)-4-F₃-Val isomer ($\alpha = 1.061$) is considerably more hydrophilic than 4-F₃-Val(*S*,*R*) ($\alpha = 1.080$), but both are more hydrophobic than Val ($\alpha = 1.053$). 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 CF₂ fluorination leads to a completely different behavior: 4,4-difluoroaminobutyric acid (4-F₂-Abu; $\alpha = 1.000$) exhibits a smaller α value than its non-fluorinated analogue (Abu; $\alpha = 1.057$). Thus, partial fluorination of aliphatic side chains can decrease the overall hydrophobicity of a given amino acid.^[25b] Such a prediction of the effects of fluorination is not trivial, but amino acids can be readily classified using the here-presented approach.

The incorporation of fluorine into phenyl rings leads to a special behavior: The H/F substitution reduces the hydrophobicity in the following order: Phe ($\alpha = 1.042$) > *o*F-Phe ($\alpha = 1.026$) > *m*F-Phe ($\alpha = 1.021$) > *p*F-Phe ($\alpha = 1.014$) > F₃-Phe ($\alpha = 0.951$). This rather unusual trend is likely a result of changes in the electronic structure of the ring. The change in the dipole moment upon fluorination leads to an increase in polarity, which results in more densely packed clusters. This effect is even more pronounced when a phosphonate group (R-CF₂-PO(OH)₂ for *p*-CF₂P-Phe; $\alpha = 0.906$) is attached to the phenyl ring (see Phe versus *p*-CF₂P-Phe). However, when the phosphonate group is perfluorinated to yield a hyper-valent R-CF₂-PF₅⁻ group that carries one permanent negative charge (see *p*-CF₂PF₅⁻-Phe),^[27] an increase in hydrophobicity ($\alpha = 0.980$) is observed compared to the neutral phosphonate group in *p*-CF₂P-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 nonpolar side chains are exposed to the gas-phase exterior, while polar residues form compact clusters to maximize electrostatically driven intermolecular interactions. 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 of 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 synthetic compounds with complex properties that make predictions of the hydrophobicity difficult or impossible. Thus, our method represents a valuable tool in the context of peptide, protein, and drug design.

Acknowledgements

We acknowledge D. A. Thomas for proof-reading. R.C., J.R., B.K., and K.P. thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for support under project number 387284271-SFB 1349. M.T.B. thanks the Alexander von Humboldt Foundation and the National Science Foundation (USA) for support under grant No. CHE-1565941.

Conflict of interest

The authors declare no conflict of interest.

Keywords: fluorinated amino acids · gas phase · hydrophilicity · ion mobility mass spectrometry · isotropic growth

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 8216–8220
Angew. Chem. **2019**, *131*, 8299–8303

- [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. Genet.* **2000**, *41*, 415–427.
- [2] a) K. Yutani, K. Ogasahara, T. Tsujita, Y. Sugino, *Proc. Natl. Acad. Sci. USA* **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 Struct. Funct. Genet.* **2001**, *42*, 243–255; b) K. M. Biswas, D. R. DeVido, J. G. Dorsey, *J. Chromatogr. 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. USA* **1980**, *77*, 1632–1636.
- [12] B. Tripet, D. Cepeniene, J. M. Kovacs, C. T. Mant, O. V. Krokhin, R. S. Hodges, *J. Chromatogr. A* **2007**, *1141*, 212–225.
- [13] C. Bleiholder, T. D. Do, C. Wu, N. J. Economou, S. S. Bernstein, S. K. Buratto, J.-E. Shea, M. T. Bowers, *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 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**, <https://doi.org/10.1002/mas.21585>.
- [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; *Angew. Chem.* **2006**, *118*, 568–583.
- [22] D. G. Isom, C. A. Castañeda, B. R. Cannon, E. B. García-Moreno, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5260–5265.
- [23] A. Vulpetti, C. Dalvit, *Drug Discovery 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. Kokschi, *Chem. Soc. Rev.* **2012**, *41*, 2135–2171; b) A. A. Berger, J. S. Voller, N. Budisa, B. Kokschi, *Acc. Chem. Res.* **2017**, *50*, 2093–2103.
- [26] J. R. Robalo, S. Huhmann, B. Kokschi, A. Vila Verde, *Chem* **2017**, *3*, 881–897.
- [27] S. Wagner, M. Accorsi, J. Rademann, *Chem. Eur. J.* **2017**, *23*, 15387–15395.

Manuscript received: December 7, 2018

Revised manuscript received: March 1, 2019

Accepted manuscript online: April 8, 2019

Version of record online: May 10, 2019