Amino acids form prenucleation clusters: ESI-MS as a fast detection method in comparison to analytical ultracentrifugation

Matthias Kellermeier,[a] Rose Rosenberg,[a] Adrian Moise,[b] Ulrike Anders,[ab] Michael Przybylski[b] and Helmut Cölfen*[a]

Electrospray ionisation mass spectrometry (ESI MS) is a fast method which is able to provide molecular mass information with high precision. In this contribution, we show that prenucleation clusters   species recently found to play a pivotal role in crystallisation processes   are detected in addition to monomers by analytical ultracentrifugation (AUC) for the whole range of DL amino acids, while higher oligomers are simultaneously observed in ESI MS spectra. This suggests ESI MS is a fast method to identify systems, which form prenucleation clusters. The occurrence of these clusters as relevant precursors in non classical nucleation scenarios thus appears to be a more common phenomenon than so far assumed.

1   Introduction

Prenucleation clusters (PNCs) are a species currently receiving considerable interest because they are part of an alternative crystallisation pathway that opposes classical nucleation theory (CNT).[1] In contrast to the metastable clusters envisaged in CNT, PNCs exist in a reversible equilibrium with their dissolved components, which is characterised by an association constant and lies in a minimum of Gibbs energy.[2] Thus, they are stable and occur already in undersaturated solutions, where no ther modynamic driving force for crystallisation is expected. Increasing the concentration can lead to cluster aggregation and, eventually, precipitation of an amorphous phase, as shown for calcium carbonate[2,3] and phosphate.[4] PNCs are discussed to be relevant in biomineralisation processes like bone formation,[5,6] and also seem to be of impor tance for polymorph control[2,7] as well as the interaction of a precipitating system with crystallisation modifiers and scale inhibitors.[8] Recent theoretical work indicates that calcium carbonate PNCs are in fact highly dynamic and strongly hydrated poly meric species.[9] Experimental analyses of PNCs prove to be challenging, especially because the clusters are small and often present only in very low concentrations. Methods that were successfully applied to date for the detection of PNCs of inor ganic minerals are analytical ultracentrifugation (AUC),[2] cryo transmission electron microscopy (cryo TEM),[3 5] and electrospray ionisation mass spectrometry (ESI MS).[10] While AUC and cryo TEM are expensive and time consuming techniques, ESI MS was so far only reported for $CaCO_3$ mineral clusters occurring during a crys tallisation reaction.[10] Fast, robust and commonly available analytical methods for the characterisation of PNCs have thus not been established yet.

[a]University of Konstanz, Physical Chemistry, Universitatsstr. 10, D 78457 Konstanz. E mail: helmut.coelfen@uni konstanz.de; Fax: +49 7531 883091; Tel: +49 7531 884063
[b]University of Konstanz, Analytical Chemistry, Universitatsstr. 10, D 78457 Konstanz. E mail: michael.przybylski@uni konstanz.de; Fax: +49 7531 883097; Tel: +49 7531 882249

Interestingly, nanosized clusters have not only been detected in solutions of inorganic salts, but were also described for various organic molecules. For example, certain proteins tend to form dense, liquid like clusters as precursors to final crystals;[11] however, as opposed to PNCs, these clusters appear to be metastable species and represent a nucleated phase. On the other hand, evidence for the existence of non nucleated, solute clusters has been reported for supersaturated solutions of amino acids: for instance, Myerson and co workers inferred clustering of glycine from a decrease of the mean diffusion coefficient with increasing supersaturation[12] and ageing time.[13] In undersaturated systems, such effects were in turn not observed. In addition to diffusivity studies, cluster formation in supersaturated binary and ternary amino acid solutions (glycine and/or valine) was further supported by activity measurements in levitated droplets,[14] as well as by the development of a concentration gradient in long columns,[15] which is conceptually related to AUC experiments. However, the vast majority of the traced species was rather small, comprising mainly dimers and trimers,[16] whereas larger aggregates appeared to be rare and formed in significant amounts only when the point of nucleation was approached. Similar observations were made also for other small molecules such as sucrose,[17] urea,[18] or citric acid,[19] which in part showed drastically higher average numbers of monomers (up to 100) in the detected clusters.[16]

While these findings were at that time interpreted as evidence for the presence of metastable pre critical clusters in the framework of CNT, the observed species could indeed also have been stable PNCs. To address this question, it is necessary to study clustering in the undersaturated regime and at high dilution. Corresponding work has already been performed for amino acids by ESI MS.[20–25] Thereby, oligomers consisting of up to more than thirty molecules were detected for arginine,[20,21] and different amino acids were ranked with respect to their tendency to form higher clusters.[23] However, these studies were generally focussed on the characterisation of aggregates held together by non covalent interactions by means of ESI MS, that is amongst others, preventing that such weak interactions are ruptured during ionisation and hence successfully transferring respective clusters from solution into the gas phase, as realised previously for a number of biopolymer complexes in our laboratory.[26] Any relevance of the observed species as solution precursors for crystallisation was in turn hardly noted. However, when considering the size of the oligomers with molar masses of up to $\sim$4200 g mol^{-1} (as reported for arginine), a relation of these clusters to PNCs in solution seems likely, given that the molar mass expected for calcium carbonate PNCs is in a similar range.[2] If such a relation would exist, ESI MS could serve as fast and accurate method to identify systems in which PNC formation occurs. This would enable rapid screening of a large amount of crystallising systems, including organic as well as inorganic compounds, both of ionic or non ionic nature and in different solvents. Corresponding results would certainly contribute to improve our understanding of the role of prenucleation clusters in crystallisation processes. In this work, we have investigated all 20 DL amino acids by ESI MS concerning the occurrence of higher oligomers in the gas phase and, in parallel, characterised the species present in solution by AUC, in order to be able to assess possible correlations between amino acid oligomer detection in ESI MS and PNC formation in solution.

2 Materials and methods

2.1 Materials

The following amino acids were used as received: DL alanine (Acros, 99%), D arginine (Fluka, ≥99%), L arginine (Acros, ≥98%), DL asparagine monohydrate (Aldrich, 98%), DL aspartic acid (Acros, ≥99%), D cysteine (Aldrich, ≥99%), L cysteine (Fluka, ≥99%), D glutamine (Sigma, ≥98%), L glutamine (Sigma, ≥99%), DL glutamic acid monohydrate (Sigma, ≥98%), glycine (Aldrich, ≥99%),

DL histidine (Sigma, ≥99%), DL isoleucine (Alfa Aesar, 99%), DL leucine (Aldrich, ≥99%), DL lysine (Sigma, ≥98%), DL methionine (Sigma, 99.5%), DL phenylalanine (Aldrich, 99%), DL proline (Aldrich, 99%), DL serine (Aldrich, 99%), DL threonine (Sigma, ≥98.5%), DL tryptophan (Aldrich, ≥99%), DL tyrosine (Acros, 99%), and DL valine (Merck, 99%).

2.2 Sample preparation

First, 0.01 M stock solutions were prepared by dissolving appropriate amounts of amino acid in water of Milli Q quality. Due to its poor solubility, the stock concentration had to be decreased to 0.001 M in the case of tyrosine. DL Arg, DL Cys and DL Gln were obtained by mixing equimolar quantities of the respective L and D enantiomers. In a second step, the pH of the as prepared solutions was adjusted to 3.1 (being the native pH of the most acidic amino acid in our study, aspartic acid) by adding aliquots of acetic acid (HOAc, VWR Prolabo, AnalaR, 100%). These samples were then directly used for the AUC measurements. For the ESI MS screening experiments, solutions were in turn diluted 1 : 100 with water, giving typical final concentrations of around 10^{-4} M. Variations of the pH were done by doping neat 10^{-4} M solutions of DL Arg (native pH: 9.3) with small volumes of ammonium hydroxide (25% ammonia solution, Merck, p.a., and dilutions thereof) or 0.01 M HOAc to cover a pH range of 10 12 and 5 9, respectively, while concentrated acetic acid had to be used for pH 2 4. The sample at pH 1 was obtained by dissolving DL Arg directly in ∼60% HOAc. To study the concentration dependence of arginine oligomers detected by ESI MS, a stock solution at 0.1 M was prepared, set to pH 3.1 by addition of HOAc, and finally diluted with water to the desired concentration. Due to the high solubility of arginine in acidic media, the effect of concentration on the species observed by AUC was investigated by measuring different DL Arg solutions at their respective native pH, so as to gain data also from supersaturated systems. The solubility of the arginine used for these studies (DL arginine, Alfa Aesar, 98%) was determined to be about 1.3 M, and samples at concentrations in both the under (0.01, 0.1, 0.2, 0.5, 0.8 and 1 M) and supersaturated regime (1.56 and 1.82 M, corresponding to supersaturation levels of *ca.* 20 and 40%) were investigated. Supersaturated solutions were prepared by gently heating (∼30 °C) the samples for several hours until dissolution was completed.

2.3 Electrospray ionisation mass spectrometry (ESI-MS)

ESI MS measurements were carried out by direct infusion of the samples at a flow rate of 5 μL s^{-1} into an Esquire 3000+ ion trap mass spectrometer (Bruker Daltonik, Bremen, Germany) operated in positive ion mode. Mass spectra were recorded by scanning from 50 to 2500 m/z. The ion source parameters were as follows: 15 psi nebulising gas (nitrogen), 6 L min^{-1} of drying gas (nitrogen) at a temperature of 200 °C, capillary voltage 3000 V and capillary exit 80 V. For some amino acids (*e.g.* glutamic acid), higher sample concentrations were found to result in ion signal saturation due to insufficient ionisation and limited ion sampling, as well as low transmission efficiencies as a consequence of space charge effects. Furthermore, deposition of solid material was observed on the needle tip and end cap under these conditions, leading to needle clogging and contamination. Comparative studies showed that none of the amino acids caused such problems at a concentration of 10^{-4} M, which therefore was chosen for the screening experiments.

2.4 Analytical ultracentrifugation (AUC)

AUC is a fractionating, absolute first principle method for the detection of dissolved and dispersed species in solution. Its statistical basis is excellent since every molecule/particle is detected. It is therefore one of the few potential techniques which are able to fractionate and trace prenucleation clusters in solution.

Measurements were performed on an Optima XL I (Beckman Coulter, Palo Alto, CA, United States) using Rayleigh interference optics and 12 mm double sector tita nium centrepieces (Nanolytics, Potsdam, Germany). Samples were investigated at 20 °C and 60 000 rpm, corresponding to a centrifugal force as high as 280 000 g. Never theless, the effective force exerted on sedimenting clusters is very low and mainly results from the friction caused by Brownian motion, essentially because the clusters exhibit quite small sedimentation but very high diffusion coefficients. The true chal lenge is the evaluation of the AUC data, since the moving boundaries are extremely broad (due to diffusion) and certain scans need to be excluded from the analysis (*i.e.* those which do not describe a transport process anymore, but rather reflect sedimen tation diffusion equilibrium). Another challenge is the correct determination of the amount of clusters, which is very low as compared to the excess of monomeric units in solution. This has been tested and verified by numerous evaluations with simu lated data for various noise levels.

Sedimentation coefficient distributions with diffusion corrections were determined using the program SEDFIT by Schuck.[27,28] In addition, sedimentation and diffusion coefficients as well as concentrations of the species present in solution were derived by Lamm equation modelling,[29] using the model of non interacting species in SED FIT as an independent evaluation approach. All experiments were carried out in triplicate to ensure statistical relevance. Sizes for the distinct species were calculated on the basis of the sedimentation coefficients *via* the following equation:[30]

$$d_H = \sqrt{\frac{18\eta s}{\rho_P \quad \rho_S}} \qquad (1)$$

where η and ρ_S are the viscosity and density of the medium, and ρ_P is the density of the sedimenting species (which are assumed to be solid hard spheres in this approach). Values for the latter were obtained by inverting the partial specific volumes of the amino acids, which were calculated using the SEDNTERP soft ware.[31] Alternatively, hydrodynamic diameters were estimated with the simulta neously determined diffusion coefficient D *via* the Stokes Einstein relation:

$$d_H = \frac{kT}{3\pi\eta D} \qquad (2)$$

2.5 Molecular modelling

Calculations of the size and structure of the investigated amino acids were per formed using Materials Studio 5.5 (Accelrys Software Inc.). Thereby, the amino acid structure was simulated in vacuum with the aid of the discover module, leaving the minimiser setup at the default settings (*i.e.* smart minimiser, medium resolution, and a maximum of 5000 iterations using the compass force field). The longest measured molecule dimension was then taken as its predicted diameter. Even though this approach neglects the presence of a hydration layer around the charged amino acid molecules and will therefore underestimate their hydrodynamic radius (as deter mined by AUC), values calculated in this manner should at least be a rough approx imate for the true size in solution and hence allow for a comparison of measured and predicted diameters (in addition to theoretical values derived from partial specific volumes (*cf.* eqn (4) and 5), which enable a direct estimation of the expected hydro dynamic size).

3 Results and discussion

3.1 Cluster detection by ESI-MS

In a first series of experiments, the occurrence of amino acid oligomers in the gas phase was studied by ESI MS at a given set of conditions (10 ^{4} M aqueous solutions

of the DL form, pH 3.1, positive ion mode). The collected data confirm the presence of higher oligomers for all amino acids (note that, hereinafter, the term "oligomer" refers to aggregates of molecules that are not covalently linked, *i.e.* intermolecular complexes and not peptides). Exemplarily, the mass spectra recorded for arginine, phenylalanine, and proline are shown in Fig. 1, while the full dataset is reproduced in Section S3.1 of the Supplementary Information (SI).†

In the case of DL Arg, only singly charged species were observed, and the detected oligomers ranged from dimers to pentamers ($[(Arg)_n + H]^+$, n = 1 5). Thereby, isola tion and collision induced dissociation of an "n" mer resulted in the "n 1" mer by neutral loss of one Arg molecule. Similar observations were made also for the other amino acids. In addition to protonated species, sodium and potassium adducts ($[(AA)_n + M]^+$, M = Na or K) were formed in some cases (*cf.* Phe and Pro in Fig. 1; likely as a consequence of leaching of Na^+ (and K^+) from glass vessels and elution of K^+ from electrodes during pH adjustment). In general, the intensity of related signals differs between distinct amino acids, but was often found to exceed that of the corresponding protonated oligomer (*e.g.* $[(Phe)_n + M]^+$ and $[(Pro)_n + M]^+$, n = 2 3, M = Na or K).

Comparing the spectra of the different amino acids suggests that, although olig omer formation appears to be a general phenomenon, the degree of aggregation varies from system to system. At 10 4 M and pH 3.1, the highest oligomer observed was a hexamer for DL Ala, whereas merely dimers were detected for DL Tyr. In an attempt to quantify the tendency of the different amino acids to undergo oligomer isation, average cluster sizes (ACS) were calculated from the ESI MS data using the following equation, which was originally proposed by Nemes *et al.*:[23]

$$ACS = \frac{\sum\limits_{n} n \cdot I_n}{\sum\limits_{n} I_n} \tag{3}$$

wherein I_n represents the measured intensity of a cluster consisting of n monomer units. Values resulting for the average cluster size are listed in Table 1, together with the highest oligomer traced for a given amino acid and the respective most intense signal in the spectra.

The above findings evidence the presence of distinct amino acid oligomers in the gas phase. Essentially, this is well in line with previous ESI MS studies, in which higher oligomers were observed for all amino acids.[20 25] It has further been noted that glycine is less prone to form (or maintain) oligomers in the gas phase due to de stabilisation of its zwitterion under these conditions[21] which is in good agreement with our results, given that we could hardly resolve any corresponding signals in this case and therefore did not consider resulting ACS values as meaningful (*cf.* Table 1).

However, there are also certain discrepancies between the data obtained in the present work and those reported previously. For instance, Nemes *et al.* provided a ranking of oligomer formation ability (in terms of ACS) for the 20 natural L amino acids,[23] which differs in part from the order distinguished in our experiments (*cf.* Table 1 and Fig. S1 in the SI†). Although principal trends within distinct groups of amino acids are widely similar (*e.g.* Asp *vs.* Glu, Ser *vs.* Thr, or the rather weak clustering ability of basic amino acids), the position of individual amino acids in the sequence is subject to deviations (leucine, for example, showed the strongest propensity to form oligomers in the work of Nemes *et al.*,[23] whereas it exhibited only medium high ACS values in our experiments). In addition, average cluster sizes calculated from the present data are generally rather small (ACS $\leq$ 2.7) as compared to previous results (ACS of up to 7.5).[23] These differences may be caused by the fact that DL amino acids were used in this study; however, as described below, our anal yses indicate that stereochemical configuration does not have a strong influence on the observed oligomerisation behaviour, at least for arginine. On the other hand, the present measurements were performed at relatively low concentrations and different
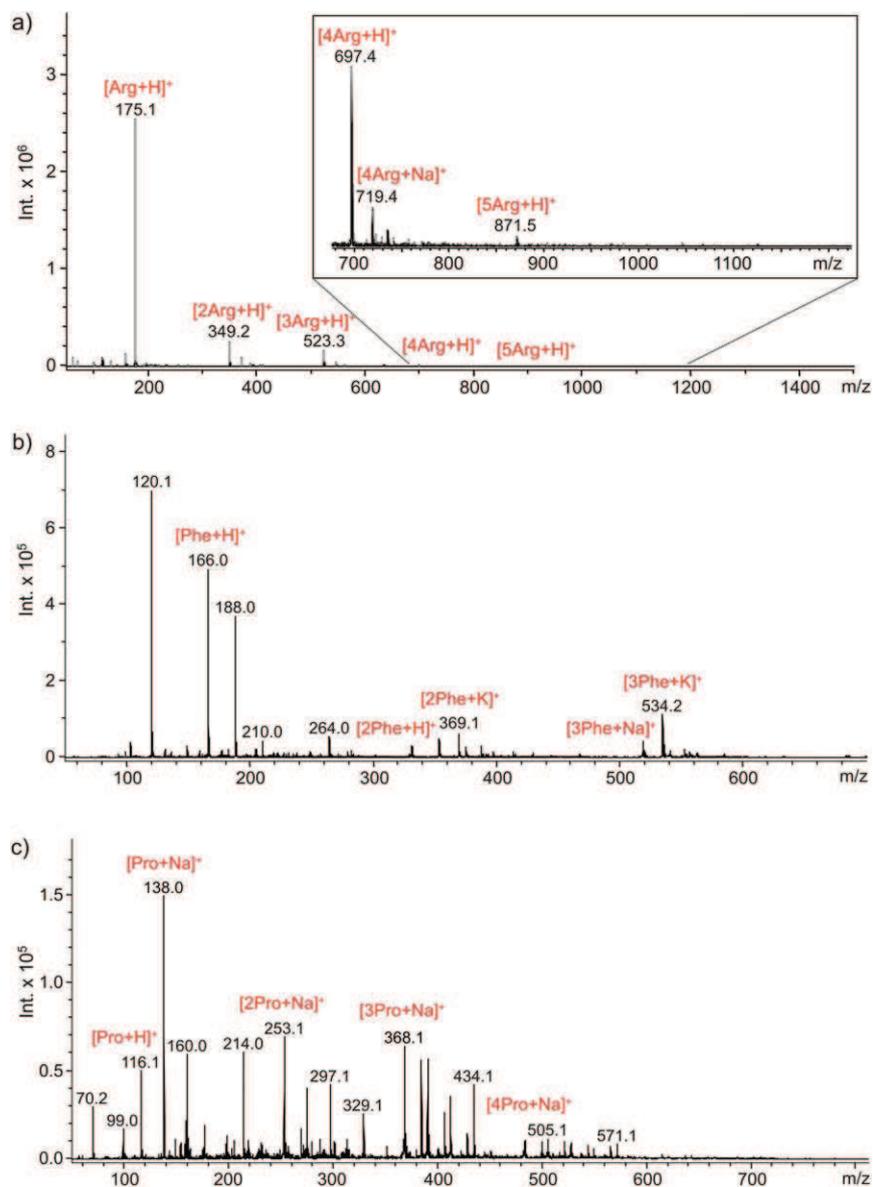
Fig. 1 ESI MS spectra recorded for 10⁻⁴ M aqueous solutions of a) DL arginine, b) DL phenyl alanine, and c) DL proline. The pH of the samples was adjusted to 3.1. Note the pronounced occurrence of sodium and potassium adducts for both Phe and Pro. The various unlabelled peaks in the spectrum of proline (and other amino acids, *cf.* Section S3.1†) originate, amongst others, from loss of ammonia (99, 214, and 329 *m/z* in case of the Pro mono , di , and trimer, respectively, or generally 17 *m/z* relative to the regular protonated signal) or addition of two sodium ions with concurrent loss of a proton (*e.g.* accounting for the Pro peak at 160 *m/z*). For the sake of clarity, these signals are not assigned in the spectra.

pH (10⁻⁴ M and pH 3.1 *vs.* 0.01 M and native pH in the work of Nemes *et al.*), which may likewise explain the distinct results. It is moreover interesting to note that most previous MS analyses of amino acid oligomers were carried out using water meth anol mixtures (usually 1 : 1 per volume) as solvent,[20,21,23 25] and Meng *et al.* as well as

Table 1 Overview of amino acid oligomers detected by ESI MS in the gas phase. Experiments were performed with 10⁻⁴ M solutions at pH 3.1. Values for the average cluster size (ACS) were calculated using eqn (3). Note that the amino acids are grouped according to their side chain functionality as non polar/hydrophobic, polar/neutral, acidic, or basic (from top to bottom). Corresponding mass spectra are shown in the Supplementary Information (Section S3.1)[†]

Amino acid	Largest oligomer detected	Oligomer with highest abundance	ACS
DL Ala	Hexamer	Trimer	2.67
DL Ile	Trimer	Monomer	1.41
DL Leu	Tetramer	Trimer	1.94
DL Met	Trimer	Monomer	1.72
DL Phe	Trimer	Monomer	2.49
DL Pro	Tetramer	Dimer	1.89
DL Trp	Trimer	Monomer	1.28
DL Val	Trimer	Trimer	2.21
DL Asn	Tetramer	Trimer	1.98
DL Cys	Tetramer	Trimer	2.47
DL Gln	Trimer	Monomer	1.57
Gly	Dimer	Monomer	[a]
DL Ser	Trimer	Trimer	2.66
DL Thr	Trimer	Trimer	1.98
DL Tyr	Dimer	Monomer	1.09
DL Asp	Tetramer	Monomer	1.81
DL Glu	Trimer	Monomer	1.64
DL Arg	Pentamer	Monomer	1.25
DL His	Trimer	Monomer	1.05
DL Lys	Trimer	Monomer	1.42

[a] The intensities of both the monomer and dimer signal were very low in the case of 10⁻⁴ M glycine and corresponding peaks could hardly be distinguished next to impurities and unidentified species. Therefore, reliable ACS values cannot be given for this amino acid.

Takats *et al.* reported that oligomerisation was depressed when purely aqueous solutions (as in this work) were employed.[20,22] This would also rationalise the relatively low ACS values obtained in our study, and we therefore investigated the effect of methanol on the resulting mass spectra of DL Arg. However, there were no significant changes discernible when water was replaced by 25 or 50% methanol, with respect to both the protonation level and the degree of aggregation (as reflected by ACS values). In this regard, the observed discrepancies are likely to originate either from distinct solution conditions (pH and concentration) or differences in the instrument settings, as it has previously been emphasised that the chosen ESI parameters and the type of spectrometer used can strongly affect the distribution of oligomers detected in the gas phase.[23] It is nevertheless clear that the size of the clusters does not correlate with the side chain functionality of the amino acid (*cf.* Fig. S1 in the SI†).

Moreover, our data do not provide unambiguous evidence for the existence of so called "magic number clusters", which show enhanced stability in the gas phase and would hence represent the most abundant species under various experimental conditions. This has been reported, amongst others, for tetramers of arginine[21] and especially the famous serine octamer, which aroused a great deal of interest in the context of homochirogenesis and chiral transmission during biomolecular evolution.[25,32 34] We cannot certainly explain why there was no preferential formation of Arg tetramers in our measurements and why the serine octamer could not be detected at all, but again speculate that distinct settings and possibly the rather low concentrations used in this study account for these circumstances.

In further experiments, different sample parameters were varied, including the pH as well as the concentration and configuration of the amino acid, using arginine as model system. The aim of these studies was to assess possible effects of solution conditions on the distribution of oligomeric species observed by ESI MS, and also to be able to directly correlate clustering in the gas phase with the ultracentrifugation results for the state in solution, given that the AUC measurements were performed at higher concentrations (usually 0.01 M) and in part at different pH (in case of the arginine concentration series, which was carried out at the respective native pH). Stepwise increases of the amount of DL Arg dissolved in the sampled solutions from 10^{-4} to 0.015 M did not induce the presence of oligomers larger than pentamers in the gas phase, with the monomer always being the most abundant species (see Section S3.2 for corresponding spectra†). However, the degree of aggregation as expressed by ACS values grew in a more or less linear fashion from initially 1.21 to about 1.8 in this concentration range (see Table S1 and Fig. S2 in the SI†). Starting from 0.025 M DL Arg, hexa and heptamers were detected in minor amounts, while the octamer was the largest observed oligomer between 0.045 and 0.075 M. At the highest concentration investigated (0.086 M), clusters consisting of up to 11 monomer units could be traced when fresh solutions were used, whereas only octamers remained upon ageing (this finding is difficult to explain at this stage and further experiments to clarify possible reasons are currently being carried out). Calculated average cluster sizes indicate a somewhat abrupt increase in the oligomerisation propensity between 0.025 and 0.045 M, where ACS values around 3.8 are reached (*cf.* Fig. S2 in the SI†); this corresponds to an estimated hydrodynamic diameter of 1.14 nm for the oligomers in average (see Section 3.2 for more details). Subsequently, ACS decreases gradually to about 3.0 at 0.086 M, suggesting that there is a maximum in the degree of aggregation at *ca.* 0.05 M. These observations agree well with previous results on the concentration dependence of oligomers detected by ESI MS.[20,21,23] Using 0.1 M arginine solutions, Zhang *et al.* found undecamers in the gas phase and obtained an oligomer distribution roughly similar to our data at 0.086 M.[21] They reported a maximum in the degree of oligomerisation at about 0.01 M, which is in line with the study by Nemes *et al.*[23] and fairly close to the value determined in the present work (0.045 M). Likewise, it was noted that smaller aggregates become favoured at higher concentrations and that sodiated oligomers are markedly suppressed under these conditions,[21,23] both being confirmed by our results (signals for Na and K adducts were completely absent in spectra recorded from solutions in the millimolar range). As already mentioned above, there are certain, though rather minor differences between the present data and those reported by Zhang *et al.*[21] in terms of the dominant oligomer in the spectra at various concentrations. Up to 0.025 M, the Arg monomer signal remained the most intense peak in our measurements, accompanied by roughly equal amounts of the dimer and trimer (with corresponding signals exhibiting intensities between 10 and 40% of the monomer peak, which is in good agreement with literature[21]). At concentrations of 0.035 M or higher, the trimer was generally found to be the most abundant species, followed by the tetramer and pentamer (see Fig. S3 in the SI†). Although the relative frequency of the tetramer (which has previously been described as magic number cluster with enhanced stability, thus giving the most intense signal at 0.1 M Arg)[21] is over a wide range close to that of the trimer and even slightly higher at 0.045 M, it does not represent a preferred state of aggregation in our experiments (which would rather apply for the trimer in the present case). In line with the calculated ACS values, peak intensities observed for higher oligomers decrease with respect to the monomer when increasing the concentration from 0.045 to 0.086 M, where the following order of abundance can eventually be discerned: trimer > tetramer > dimer > monomer > pentamer (*cf.* Fig. S3 in the SI†).

In turn, variations of the pH of 10^{-4} M arginine solutions (isoelectric point (IEP) at pH 9.675)[31] showed no systematic dependencies (see Table S2 and Fig. S4 in the SI,† and Section S3.3 for the raw spectra). Earlier work reported that

oligomerisation of Arg was depressed when samples were acidified,[20,22] which is to some extent supported by the fact that, in this study, the tetramer was the highest detected oligomer at pH 1, whereas pentamers could be observed at pH 2 and 3.1. However, average cluster sizes determined from the data indicate slightly stronger degrees of aggregation at lower pH values, and thus contradict the above notion. Above pH 3, the pentamer was also no longer visible, with tetramers and trimers being the largest species traced in the near neutral (pH 6 9) and alkaline (pH 10 12) range, respectively. Thereby, we could not distinguish any significant increase in the clustering propensity with pH, as observed by Zhang *et al.* for arginine solu tions at pH 10.[21] Indeed, the monomer was always the most abundant species and the intensities of the dimer, trimer and tetramer, as well as corresponding ACS values, did not exhibit any consistent trend with pH (*cf.* Fig. S4 in the SI†), thus sup porting the conclusions of Nemes *et al.* in that there is no clear correlation between the pH and the oligomers detected in the gas phase.[23]

Comparative studies using optically pure arginine samples (10 4 M D and L Arg) demonstrate that distinct chirality does not affect the resulting oligomer distribution to a noticeable extent. In fact, virtually identical mass spectra were obtained for both enantiomers and their racemic mixture (see Section S3.4 in the SI†), and the calcu lated average cluster size was about the same in all three systems (see Table S3 in the SI†). It is well known that certain amino acids can distinguish between enantiomers during oligomer formation, potentially displaying strong preference for homochiral aggregates as evidenced by isotope labelling for serine.[32 34] Similar, though less pronounced stereochemical selectivity was observed also for a series of other amino acids, including both homo (*e.g.* alanine) and heterochiral preferences (*e.g.* valine), partially depending on the particular cluster size.[23] Arginine, however, was not among the amino acids showing this behaviour, as verified by the present experi ments.

In essence, the results of the ESI MS measurements performed in this work agree reasonably well with previously reported data and confirm that higher oligomers are observed in the gas phase for all of the investigated amino acid samples. This provides a sound basis for characterising and comparing corresponding species in solution by means of analytical ultracentrifugation.

3.2 AUC analyses of cluster formation in solution

A general problem associated with mass spectrometric studies of molecular clus tering refers to the critical question of whether the detected oligomers truly reflect the original state of species in solution, or if they possibly formed only upon solvent removal and ionisation during transfer into the gas phase, as outlined explicitly in many of the previous MS studies.[20,22,23] Therefore, if the physical chemistry of presumed clusters in solution is to be elucidated, a second independent technique is required to ascertain that oligomers traced by ESI MS are relevant species also in the native solution state. Here, we have used analytical ultracentrifugation for this purpose.

Diffusion corrected sedimentation coefficient distributions c(s) calculated for samples of DL Arg demonstrate the existence of two populations in the solutions: individual amino acid monomers (s ≈ 0.05 0.25 S) as well as clusters with a sedi mentation coefficient of about 1 2 S (Fig. 2a).

Extensive AUC analyses of the whole range of DL amino acids prove that, indeed, species with s values typical for clusters can be detected in all cases (see Section S4 in the SI†). Monomers generally exhibit sedimentation coefficients in the range of 0.05 0.25 S (see Table 2), which is similar to values known for single ions in solution.[2]

Estimations of relative concentrations show that the monomers always represent the vast majority of species occurring in the system, usually in fractions higher than 99 wt%. In turn, the amount of clusters in the samples is lower than 0.5 wt% for most amino acids. Corresponding sedimentation coefficients vary between 0.8 and 2 S,
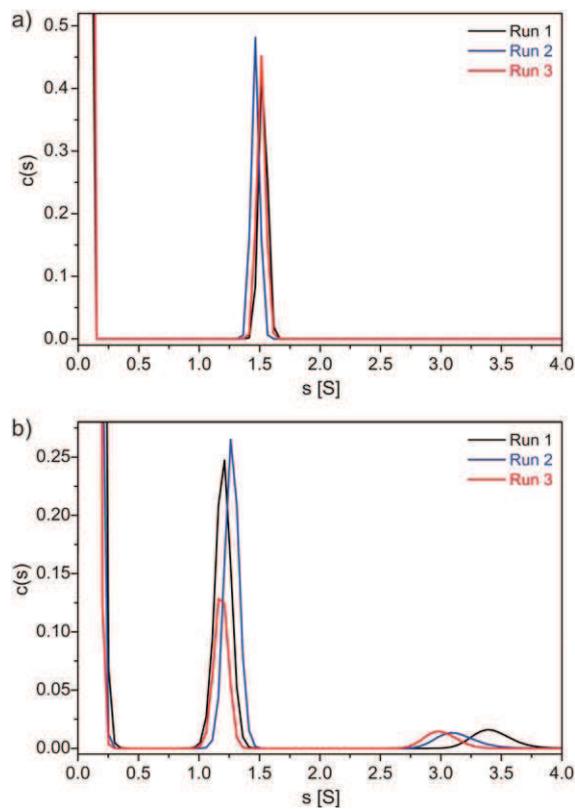
Fig. 2 Sedimentation coefficient distributions derived from AUC measurements using SED FIT for 0.01 M solutions of a) DL Arg and b) DL Asn, both adjusted to pH 3.1. Note that the dominant signal for amino acid monomers at s < 0.5 S has been cut off so as clearly illus trate peaks belonging to clusters ($1 \leq s \leq 1.5$) and their aggregates (s > 2.5 in case of Asn).

which is considerably higher than those of the monomers and moreover agrees well with s values reported for PNCs of calcium carbonate ($\sim$1.5 S).[2,3] This suggests that amino acids indeed form prenucleation clusters in undersaturated solutions. The low concentration of these clusters with respect to monomeric units indicates that their association constant is significantly smaller than in the $CaCO_3$ case (where typically 10 30% clusters occur in equilibrium with free ions at pH values between 9 and 10).[2] Nonetheless, they exhibit features that are characteristic of stable PNCs and distinct from what would be expected for metastable clusters in the framework of CNT. These are, on the one hand, the fact that measurable amounts of the clusters are present in solution far below the solubility limit and, on the other, the existence of a preferential cluster size that is clearly separated from monomers.[9] According to CNT, the probability of cluster formation decreases continuously with the number of monomer units in the cluster. Thus, one would expect a single peak in c(s) that is shifted slightly towards higher s values if a minor fraction of the amino acids was involved in the assembly (and subsequent disintegration) of metastable clusters (given that AUC is not capable of resolving lower oligomers next to monomers, see below).

For some amino acids (*i.e.* Ile, Leu, Met, Phe, Pro, Trp, Asn, Gln, Glu and His), even larger or respectively more dense species are detected as a third component, with sedimentation coefficients ranging from about 2 to 3 S (*cf.* Table 2 and Fig. 2b). Again, this complies well with previous results for calcium carbonate,

Table 2 Summary of the AUC results for 0.01 M solutions of the investigated amino acids at pH 3.1, whereby $s_{Monomer}$ is the sedimentation coefficient of single amino acid molecules, $s_{Cluster\ I}$ that of presumed single PNCs, and $s_{Cluster\ II}$ that of PNC aggregates, all with corresponding relative concentrations c_i. Data were obtained by Lamm equation modelling of non interacting species using the SEDFIT software

Amino acid	$s_{Monomer}$ [S]	$c_{Monomer}$ [wt%]	$s_{Cluster\ I}$ [S]	$c_{Cluster\ I}$ [wt%]	$s_{Cluster\ II}$ [S]	$c_{Cluster\ II}$ [wt%]
DL Ala	0.18 ± 0.04	99.6 ± 0.3	2.0 ± 0.3	0.4 ± 0.3		
DL Ile	0.08 ± 0.03	99.91 ± 0.04	0.89 ± 0.07	0.05 ± 0.03	2.6 ± 0.2	0.03 ± 0.02
DL Leu	0.10 ± 0.08	99.8 ± 0.2	0.78 ± 0.01	0.1 ± 0.2	1.4 ± 0.1	0.04 ± 0.05
DL Met	0.12 ± 0.02	99.94 ± 0.03	1.2 ± 0.2	0.05 ± 0.02	2.3 ± 0.4	0.02 ± 0.02
DL Phe	0.10 ± 0.01	99.98 ± 0.01	1.07 ± 0.03	0.01 ± 0.01	1.98 ± 0.03	0.01 ± 0.01
DL Pro	0.05 ± 0.01	99.99 ± 0.01	0.94 ± 0.03	0.006 ± 0.002	2.46 ± 0.04	0.004 ± 0.004
DL Trp	0.13 ± 0.01	99.91 ± 0.01	1.3 ± 0.2	0.009 ± 0.001	2.4 ± 0.2	0.002 ± 0.002
DL Val	0.17 ± 0.09	99.91 ± 0.06	1.3 ± 0.2	0.09 ± 0.06		
DL Asn	0.13 ± 0.03	99.93 ± 0.04	0.90 ± 0.07	0.05 ± 0.03	3.1 ± 0.4	0.02 ± 0.01
DL Cys	0.15 ± 0.05	99.89 ± 0.08	0.55 ± 0.06	0.08 ± 0.03	1.43 ± 0.07	0.02 ± 0.02
DL Gln	0.15 ± 0.03	99.8 ± 0.2	0.57 ± 0.08	0.2 ± 0.1	1.37 ± 0.05	0.07 ± 0.07
Gly	0.19 ± 0.09	99.86 ± 0.06	1.5 ± 0.2	0.14 ± 0.06		
DL Ser	0.11 ± 0.02	99.99 ± 0.01	1.4 ± 0.2	0.004 ± 0.002		
DL Thr	0.09 ± 0.03	99.98 ± 0.01	1.4 ± 0.1	0.02 ± 0.01		
DL Tyr	0.24 ± 0.09	98 ± 2	1.9 ± 0.1	2 ± 2		
DL Asp	0.22 ± 0.03	99.8 ± 0.1	2.1 ± 0.1	0.2 ± 0.1		
DL Glu	0.17 ± 0.03	99.8 ± 0.1	1.49 ± 0.03	0.07 ± 0.09	2.91 ± 0.07	0.17 ± 0.05
DL Arg	0.05 ± 0.01	99.99 ± 0.01	1.05 ± 0.04	0.002 ± 0.001		
DL His	0.11 ± 0.03	99.99 ± 0.01	0.8 ± 0.2	0.005 ± 0.001	1.8 ± 0.3	0.002 ± 0.001
DL Lys	0.08 ± 0.02	99.99 ± 0.01	1.6 ± 0.2	0.005 ± 0.004		

where species with similar s values were identified as aggregates of PNCs.[2] The concentration of such cluster assemblies in the present samples is even lower (<0.2 wt%) than that of individual PNCs (*cf.* Table 2).

If the amino acids are sorted according to their functionality (Fig. 3), it is obvious that a species with a sedimentation coefficient typical for PNCs is found regardless of the structure of the amino acid. This is fully in line with the ESI MS data, which show that at least dimers (Gly and Tyr) and in most cases oligomers of three or more molecules are detected in the gas phase, thus consistently giving ACS values higher than 1 (which implies significant oligomerisation, *cf.* Table 1). The larger aggregates observed for some systems by AUC (labelled Cluster II in Table 2) do also not evidently depend on the type of amino acid, although such species are particularly common in the group of non polar/hydrophobic amino acids (*cf.* Fig. 3). Further more, we could not trace any coherent trend of cluster size (and abundance) with the saturation of the solutions, which can thus likely be ruled out as a possible driving force for clustering (note that, due to the distinct solubility of the amino acids, the degree of (under)saturation at 0.01 M varies over two orders of magnitude; see Table S4 in the SI†).

Interestingly, the signals of both the amino acid monomers and the PNCs are very sharp in the sedimentation coefficient distributions, while those of the presumed
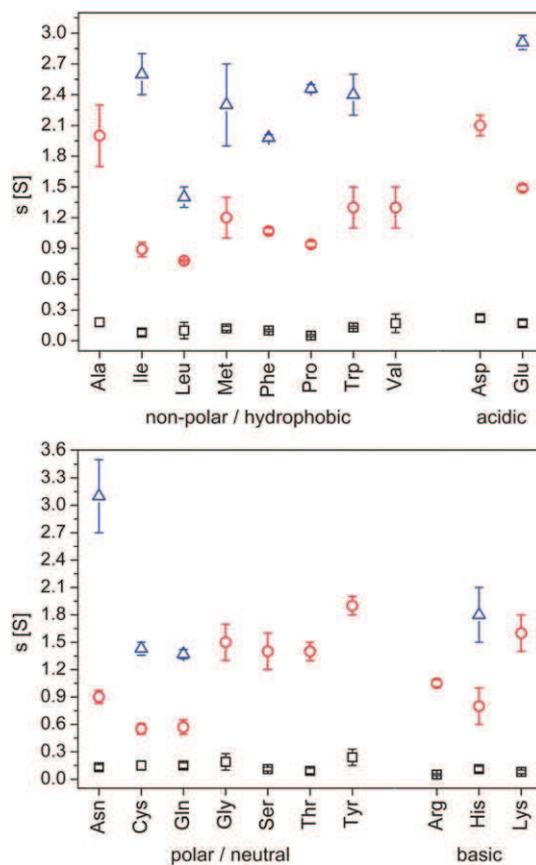

Fig. 3 Comparison of the sedimentation coefficients determined for the different amino acids. Fitting the experimental data required prescription of 2 3 distinct components (depending on the system), as indicated by different symbols (squares: Species 1/monomer, circles: Species 2/Cluster I, triangles: Species 3/Cluster II) and according to Table 2.

cluster aggregates are markedly broader (*cf.* Fig. 2b). This supports the notion that the larger species are assemblies of prenucleation clusters with varying aggregation numbers, whereas the amino acid molecules are by nature monodisperse. In light of recent modelling results, prenucleation clusters appear to be very dynamic, liquid like polymeric structures (called "dynamically ordered liquid like oxyanion poly mers" (DOLLOP) in the case of $CaCO_3$),[9] which steadily form and disintegrate on timescales much faster than that of an AUC experiment (which is hours). If such a rapid exchange takes place, only the time averaged signal will be detected by AUC, as is well known from the analysis of fast reversible interactions between biopolymers.[35] This could explain the sharp signal in the sedimentation coefficient distribution for the species denoted Cluster I (for a dynamic structure like DOLLOPs, fairly broad size distributions would in principle be expected).[9] Assuming that amino clusters form on the basis of a multiple binding equilibrium as observed for $CaCO_3$,[2] the detected signal would then correspond to the weighted average species between the monomer and the largest occurring PNC (*i.e.* polymer), as resulting from an evaluation of the datasets with a non interacting model.

Hydrodynamic diameters can be derived for the different species either from the sedimentation or diffusion coefficient using eqn (1) and 2, respectively. Further, it is possible to calculate d_H *via* the volume V of the amino acids and their oligomers, which can be estimated with the partial specific volume ν of hydrated amino acids and the molar mass M according to:[36]

$$V = \frac{\overline{\nu} \cdot M}{N_A} \tag{4}$$

The hydrodynamic diameter d_H is then readily calculated as follows:

$$d_H = 2 \cdot \left(\frac{3V}{4\pi} \right)^{1/3} \tag{5}$$

Values obtained in this manner for the smallest detected species are listed in Table 3 and compared to those calculated from s and D as well as diameters esti mated on the basis of molecular modelling of amino acids in vacuum (as described in Section 2.5).

It is evident that the sizes predicted *via* the partial specific volumes are generally similar to maximum extensions measured in vacuum (which, thus, appear to be reasonable estimates for molecular dimensions in solution even though hydration layers were neglected in the modelling approach). While corresponding values nearly coincide for certain amino acids (Pro, Val and Cys), ν based diameters are typically somewhat smaller than the simulated molecule length, although an opposite trend is observed in some cases (Ala, Ile, Gly, Ser and Thr). This indicates that the structure in solution is different from that in vacuum (likely due to hydration, *cf.* above). Nevertheless, the fairly good agreement between the results from calculations ac cording to eqn (4) and 5 and those obtained from modelling show that sizes derived from the partial specific volume can be considered realistic.

The experimental hydrodynamic diameters resulting from the sedimentation coef ficient and those determined *via* the diffusion coefficient differ slightly, with s based values generally exhibiting larger errors than those originating from D. This is reasonable since the sedimentation coefficients are close to the resolution limit and sedimentation is so slow under the given conditions that diffusion is the domi nating transport process. We therefore regard the hydrodynamic diameters calcu lated from the diffusion coefficients to be more reliable in this case. The fact that D based sizes comply well with values predicted with the aid of eqn (4) and 5 for single amino acid molecules suggest that all amino acids exist predominantly as

Table 3 Estimated hydrodynamic diameters of the smallest species observed by AUC. Experimental values were derived from both the sedimentation (s) and the diffusion coefficient (D) *via* eqn (1) and 2, respectively, using densities resulting from partial specific volumes ($\bar{v}$) calculated with SEDNTERP.[31] For comparison, predicted diameters obtained *via* the molar mass and the partial specific volumes (eqn (4) and 5) as well as maximum diameters determined by measuring the longest dimension in simulated amino acids molecules are also listed

Amino acid	$\bar{v}$ [cm³ g⁻¹]	$d_{H,Monomer}$ [nm]		*Predicted (specific volume)*	*Predicted (modelling)*
		Experimental (s)	*Experimental* (D)		
DL Ala	0.7379	1.0 ± 0.2	0.55 ± 0.05	0.59	0.48
DL Ile	0.8979	1.1 ± 0.3	0.55 ± 0.01	0.72	0.64
DL Leu	0.8979	1 ± 1	0.60 ± 0.04	0.72	0.79
DL Met	0.7479	0.8 ± 0.1	0.67 ± 0.01	0.71	0.80
DL Phe	0.7679	0.76 ± 0.07	0.66 ± 0.01	0.74	0.92
DL Pro	0.7579	0.5 ± 0.1	0.49 ± 0.01	0.65	0.65
DL Trp	0.7379	0.81 ± 0.02	0.74 ± 0.01	0.78	0.95
DL Val	0.8579	1.4 ± 0.8	0.49 ± 0.07	0.68	0.67
DL Asn	0.6179	0.6 ± 0.1	0.58 ± 0.02	0.64	0.74
DL Cys	0.6279	0.7 ± 0.2	0.61 ± 0.02	0.62	0.60
DL Gln	0.6679	0.7 ± 0.2	0.65 ± 0.03	0.68	0.86
Gly	0.6379	0.8 ± 0.4	0.41 ± 0.01	0.53	0.47
DL Ser	0.6279	0.6 ± 0.1	0.55 ± 0.01	0.56	0.46
DL Thr	0.6979	0.6 ± 0.2	0.58 ± 0.01	0.64	0.55
DL Tyr	0.7079	1.0 ± 0.4	0.53 ± 0.04	0.74	1.06
DL Asp	0.5979	0.8 ± 0.1	0.61 ± 0.07	0.63	0.67
DL Glu	0.6579	0.8 ± 0.1	0.69 ± 0.01	0.67	0.88
DL Arg	0.6979	0.45 ± 0.08	0.52 ± 0.01	0.73	1.07
DL His	0.6679	0.6 ± 0.2	0.52 ± 0.02	0.69	0.90
DL Lys	0.8179	0.8 ± 0.2	0.49 ± 0.01	0.72	1.00

monomers in solution (especially when considering that experimental d_H values derived from D are by trend even lower than predicted ones). Diameters obtained from the sedimentation coefficient for the smallest species are often larger than those expected for the respective monomer (in case of Ala, Ile, Leu, Val, Gly, Tyr, Asp, Glu). However, for many of the concerned amino acids, the determined size of the monomer is within the limits of accuracy (Leu, Val, Gly, Tyr), and the difference in case of the remaining amino acids (Ala, Ile, Asp, Glu) is not too significant with respect to the relatively large experimental error of such low sedimentation coefficients.

Thus, the diameters obtained from the diffusion coefficient likely represent a realistic estimate of the true hydrodynamic size, and resulting values strongly suggest that the major fraction of the amino acids are monomeric in solution, with a certain amount of larger oligomers coexisting. Using glycine as a model system, we have tested the capability of AUC to distinguish the presence of dimers next to a *ca.* 90% excess of monomers and traces of larger clusters. Datasets simulated for a corresponding mixture (89.87 wt% monomer, 9.99 wt% dimer, and 0.14 wt% PNCs) were found to be almost equally well described by fits according to a monomercluster and monomerdimercluster model, respectively. Therefore, we cannot unequivocally prove that all amino acids are present as monomers in our samples. However, since experimentally determined d_H values are usually lower than diameters expected for monomers, the fraction of dimers (if any) cannot be high.

This is an interesting finding, given that there has been much debate in the past years about whether glycine occurs preferentially as monomer or dimer in aqueous

solution. While early studies indicated considerable dimerisation (and in part also higher order clustering)[12] [16,37] and related observations were later taken as evidence to explain polymorph selection during Gly crystallisation,[38,39] more recent work argued that glycine indeed exists mainly in its monomeric form in water. The latter conclusion was supported by cryoscopy measurements,[40] diffusion coefficients determined via solution NMR spectroscopy,[40,41] dielectric relaxation studies,[42] as well as molecular dynamics (MD) simulations,[41,43] yielding approximate dimer fractions between 5 and 30%, depending on concentration and pH. Our results confirm this notion and, in addition, demonstrate that monomers are the dominant solution species for all of the investigated amino acids.

Apart from that, the MD simulations performed by Hamad et $al.$ also suggested the presence of a small fraction of glycine clusters (including species up to pentamers) next to a large excess of monomers in aqueous solutions.[43] Indeed, size distributions calculated in their work bear obvious resemblance to the sedimentation coefficient distributions obtained in our study. Typical radii of gyration (R_G) resulting for glycine oligomers from the simulations ranged from 0.2 (dimer) to more than 0.75 nm (pentamer). Assuming these clusters to be random coil polymeric structures in analogy to $CaCO_3$ DOLLOPs,[9] the reported R_G values would correspond to hydrodynamic diameters of about 1 nm in case of the pentamer ($R_H/R_G \approx 0.66$ for a random coil),[44] which is in fair agreement with cluster sizes estimated on the basis of the present AUC and ESI MS data (as described below, $cf.$ Table 4), and nearly coincides with predictions made for Gly pentamers by eqn (4) and 5 (giving $d_H = 0.91$ nm). Moreover, Hamad et $al.$ found that clustering of glycine relies on hydrogen bonds between monomers, which continuously break and reform on timescales of picoseconds.[43] This implies that the structure of the clusters is highly dynamic and hence, by nature, to some degree similar to the polymerlike DOLLOPs described for calcium carbonate.[9] While it remains to be proven that glycine clusters are indeed thermodynamically stable species (as their $CaCO_3$ counterparts), it seems clear that, if so, the equilibrium constant underlying their formation is much lower than in the $CaCO_3$ case (as judged from relative cluster fraction in equilibrium with monomers, $cf.$ above). Differences in the clustering behaviour appear plausible when considering that interactions between monomer units are not (only) based on electrostatics, but (additionally) involve hydrogen bonding in the case of glycine. In this context, it is finally worth noting that very recent modelling results corroborate clustering of amino acids in solution, as aspartate was shown to assemble into supramolecular polymers under conditions close to those used in the present AUC analyses.[45]

The sedimentation data also allow determination of the hydrodynamic diameters of the detected larger species. As discussed above and in light of the DOLLOP concept,[9] resulting values should be considered as an average of the supposedly polymeric PNCs. Under this assumption, the employed partial specific volumes should represent a good estimate for the cluster density and, thus, the sizes calculated from the sedimentation coefficient (Table 4) are expected to be more realistic in case of the amino acid clusters. Hydrodynamic diameters derived from the diffusion coefficient (see Table S5 in the SI†) are in turn judged to be less reliable and, accordingly, exhibit relatively high errors and deviate significantly from the values obtained via the sedimentation coefficient.

The data given in Table 4 show that the s based diameters of individual PNCs (Cluster I) vary between 1.4 and 3.7 nm, with Ile and Lys apparently forming the largest clusters. Mean sizes determined for the PNC aggregates (Cluster II) range from 2.1 nm for Cys over 4.7 nm for Leu up to more than 6 nm in the case of Ile, and thus fall into the same order of magnitude as reported for aggregates of $CaCO_3$ prenucleation clusters.[2] To directly correlate these results with the observations made by ESI MS, we have evaluated the dimensions of oligomers identified in the mass spectra by converting corresponding molar masses into approximate diameters via eqn (4) and 5, utilising the partial specific volumes listed in Table 3. This was done for both the highest detected oligomer and the average number of monomer

units in the gas phase clusters (as given by the ACS value), yielding sizes respectively denoted d_{Max} and d_{Av} in Table 4. It is evident that, in all of the investigated systems, (maximum) cluster diameters indicated by ESI MS are considerably smaller (around 1 nm) than what is obtained by AUC for single PNCs on average. Probably, this implies that larger clusters occurring in solution are less effectively transported into the gas phase than species of low molecular weight, or that they dissociate (*i.e.* break apart) upon ionisation. In either of these two scenarios, the oligomer distribution detected by ESI MS would not correctly represent the chemical specia tion in solution, but rather underestimate the mean size of the clusters (as observed).

Interestingly, much larger oligomers were traced in some of the previous MS studies on amino acids.[20,21,23] For example, Zhang *et al.* detected a 36 mer in their work on arginine, presumably as a consequence of more suitable experimental conditions regarding the ionisation of higher oligomers.[21] According to calculations based on the partial specific volume of hydrated amino acids, such a cluster should have a diameter of about 2.4 nm, which is indeed close to the average value deter mined by AUC in this case (2.1 nm). These considerations lead to the conclusion that although ESI MS can be used to rapidly screen distinct systems concerning the formation of clusters, it is not capable of delivering quantitative information on the size and relative frequency of such species, unless instrument settings (and probably also sample parameters like concentration) are fine tuned and optimised

Table 4 Comparison of sizes estimated for the different cluster species observed by AUC and the amino acid oligomers traced by ESI MS. Hydrodynamic diameters d_H were derived from the sedimentation coefficient *via* eqn (1) in the case of the AUC data, while the values quoted for the ESI MS experiments were obtained by eqn (4) and 5 using the partial specific volumes listed in Table 3 and the molar mass corresponding to the highest detected oligomer (d_{Max}, number of monomers indicated in brackets) and the average number of monomer units in the clusters (d_{Av}, as deduced from ACS values), respectively. Also included are the apparent percentages of clusters in the solutions (X_C, calculated according to eqn (6)) as resulting from the AUC experiments as well as, for direct comparison, the average cluster size obtained from ESI MS analyses (*cf.* Table 1)

Amino acid	AUC			ESI MS		
	$d_{H,Cluster\ I}$ [nm]	$d_{H,Cluster\ II}$ [nm]	X_C [%]	d_{Max} [nm]	d_{Av} [nm]	ACS
DL Ala	3.2 ± 0.4		2.3 ± 0.1	1.08 (6)	0.82	2.67
DL Ile	3.7 ± 0.3	6.4 ± 0.5	1.6 ± 0.3	1.04 (3)	0.81	1.41
DL Leu	3.49 ± 0.04	4.7 ± 0.4	1.3 ± 0.8	1.14 (4)	0.90	1.94
DL Met	2.6 ± 0.4	3.5 ± 0.7	0.3 ± 0.1	1.02 (3)	0.85	1.72
DL Phe	2.52 ± 0.08	3.44 ± 0.06	0.5 ± 0.1	1.06 (3)	1.00	2.49
DL Pro	2.30 ± 0.07	3.73 ± 0.05	0.5 ± 0.1	1.03 (4)	0.81	1.89
DL Trp	2.5 ± 0.5	3.5 ± 0.3	0.2 ± 0.1	1.13 (3)	0.85	1.28
DL Val	3.8 ± 0.7		0.9 ± 0.4	0.99 (3)	0.89	2.21
DL Asn	1.6 ± 0.1	3.0 ± 0.4	0.4 ± 0.2	1.01 (4)	0.80	1.98
DL Cys	1.3 ± 0.1	2.1 ± 0.1	1.0 ± 0.2	0.99 (4)	0.84	2.47
DL Gln	1.4 ± 0.2	2.18 ± 0.08	1.0 ± 0.1	0.98 (3)	0.79	1.57
Gly	2.2 ± 0.3		1.7 ± 0.1	0.77 (3)		
DL Ser	2.0 ± 0.2		0.3 ± 0.2	0.80 (3)	0.77	2.66
DL Thr	2.4 ± 0.2		0.3 ± 0.1	0.92 (3)	0.81	1.98
DL Tyr	2.9 ± 0.2		0.9 ± 0.5	0.93 (2)	0.76	1.09
DL Asp	2.4 ± 0.1		0.7 ± 0.2	1.00 (4)	0.77	1.81
DL Glu	2.28 ± 0.04	3.18 ± 0.07	1.2 ± 0.1	0.97 (3)	0.80	1.64
DL Arg	2.09 ± 0.09		0.1 ± 0.1	1.24 (5)	0.78	1.25
DL His	1.7 ± 0.5	2.5 ± 0.5	0.1 ± 0.1	1.00 (3)	0.70	1.05
DL Lys	3.6 ± 0.4		2.4 ± 0.2	1.04 (3)	0.81	1.42

for a given substance. In other words, while large clusters appear to generally exhibit low intensities in mass spectra due to insufficient transfer into the gas phase, they may well represent (one of) the most abundant species in solution (as it is possibly the case for the 36 mer of arginine).

In analogy to the ACS values used to quantify the degree of aggregation in the ESI MS experiments, apparent percentages of clusters (X_C) out of the total amount of amino acid molecules present were determined from the sedimentation coefficient distributions. This was achieved by relating the integrated area of cluster peaks in the c(s) diagrams to the sum of the integrals of both monomer and cluster signals, according to:

$$X_C = 100 \cdot \frac{A_{\text{Cluster peaks}}}{A_{\text{Cluster peaks}} + A_{\text{Monomer peak}}} \tag{6}$$

Corresponding results are listed in Table 4 and confronted with the average cluster size derived from ESI MS. Obviously, there is no coherent correlation between the two techniques in terms of the indicated clustering propensities. While fairly good agreement is observed for some amino acids (*e.g.* Ala or Tyr), ACS and X_C differ substantially for others (*e.g.* Ser, Phe, or Lys) and trends found with one method do not coincide with those obtained from the other (*cf.* Fig. S1 in the SI†). This suggests that the extent of clustering seen by ESI MS in the gas phase does not neces sarily reflect the state in solution as probed by AUC, again likely due to distinct abil ities (or probabilities) for the different amino acid oligomers to be ionised and transferred successfully into the gas phase. Nevertheless, these data confirm that ESI MS is able to trace systems which form clusters, while a detailed and quantita tive analysis of cluster sizes and distributions requires a more elaborate technique such as AUC.

In this regard, and given that clusters were observed for all studied amino acids, it would be of particular interest to find and investigate a compound that does not form PNCs but prefers to stay monomeric. Such a counter example would serve as a further proof of concept concerning the potential of the two methods as outlined above. Therefore, we performed AUC measurements with several other water soluble organic molecules and traced clustering in most cases   with the exception of guanidinium hydrochloride, a strong water structure breaker and denaturant of protein folding,[46] for which only monomeric species with s = 0.2 ± 0.2 S was de tected (at a concentration of 0.02 M). Indeed, ESI MS spectra recorded from a 10^{-4} M solution at pH 3.1 confirm the absence of higher order clusters (see Fig. S5 in the SI†), but a distinct peak of the dimer can nonetheless clearly be distin guished. Overall, the signals of both the monomer and the dimer are quite weak in this case (intensities being comparable to those of impurities, decomposition prod ucts and certain adducts), such that the clustering propensity appears to be roughly similar to what has been found for glycine (which, by contrast, forms clusters ac cording to AUC results). This implies that somewhat more appropriate criteria (or maybe simply other ESI parameters) must be defined in future work to render mass spectrometry a reliable screening method for the presence of PNCs, especially when it comes to systems in which no clustering takes place. On the other hand, it is important to stress that the absence of higher oligomers in the sedimentation data does not necessarily mean that PNCs do not exist. Under certain conditions (*i.e.* at very low species concentrations), this could also be an issue of the detection limit, which is in the picomolar range in case of ESI MS, while the Rayleigh interference optics used in AUC can resolve refractive index differences (Δn) on the order of *ca.* 10^{-6}.[47] For an amino acid with a molar mass of 100 g mol^{-1} and a refractive index increment (dn/dc) equal to that of a typical protein (*ca.* 0.19×10^{-3} mL mg^{-1}), this would correspond to a detectable concentration threshold of about 19 nM. Thus, AUC is capable of tracing PNCs down to quite low concentrations; however,
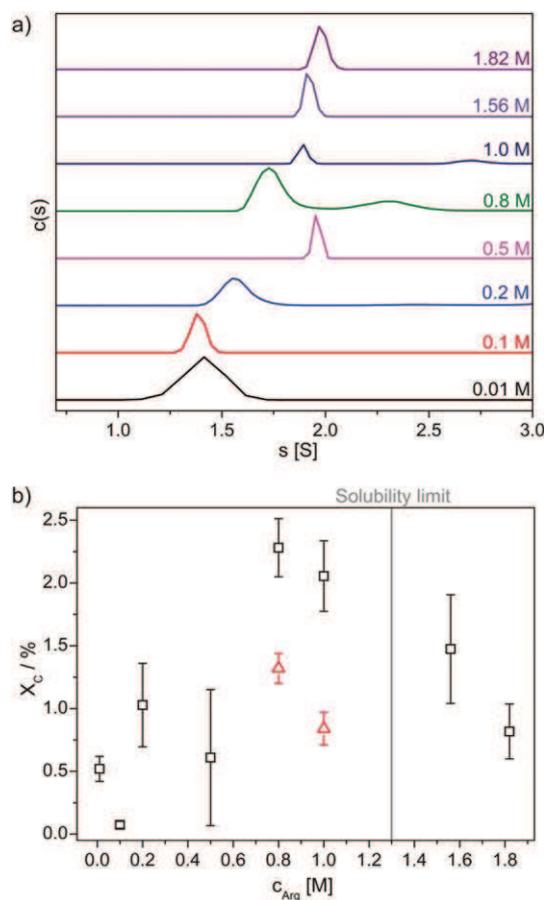
Fig. 4 AUC results for solutions of DL arginine at various concentrations (all at their respective native pH, ranging from about 10 to 12). a) Selected sedimentation coefficient distributions c(s) for both under (0.01 1 M) and supersaturated systems (1.56 and 1.82 M, the solubility being around 1.3 M). Note the shift of the cluster signal towards higher values between 0.2 and 0.5 M, as well as the occurrence of a broad peak at s > 2 S for 0.8 and 1.0 M, which is in turn absent in supersaturated solutions. b) Plot of the estimated total percentage of clusters at different concentrations (squares, calculated according to eqn (6), *i.e.* comprising both single clusters and cluster aggregates) and the fractional percentage of single clusters (triangles) at those concentrations where also cluster aggregates were detected (0.8 and 1.0 M).

if the equilibrium constant of cluster formation is very small, these species might indeed not be detectable anymore. ESI MS is a more sensitive and hence promising alternative in this respect, but obtained results may heavily depend on the chosen settings and state of the used samples, as evidenced by the present and previous findings. Therefore, the collected data do not allow us to ascertain unequivocally whether guanidinium hydrochloride forms clusters (which would be rare, though) or not.

Finally, we performed another set of AUC experiments in which solutions with various arginine contents (each left at its respective native pH) were investigated concerning the size and relative amount of occurring clusters, complementary to the concentration series carried out by ESI MS. Resulting sedimentation coefficient distributions (Fig. 4a) and cluster percentages calculated according to eqn (6) (Fig. 4b) both show clear trends with increasing concentration. At Arg contents of up to 0.1 M, the c(s) diagrams exhibit a single cluster species with s ≈ 1.4 S

next to a large excess of monomers (peak not shown in Fig. 4a). This is generally in line with the data obtained for a 0.01 M Arg solution at pH 3.1 (although s values of the cluster component differ to some degree, *cf.* Table 2 and Fig. 4), hence confirming the ESI MS measurements in that the pH does not strongly influence the clustering behaviour of arginine. As opposed to that, the clusters traced by AUC are again significantly larger (*ca.* 2.4 nm) than the highest oligomers detected in ESI MS for solutions at similar concentrations (11 mer with d $\approx$ 1.6 nm at 0.086 M), due to the reasons discussed above.

Raising the amount of dissolved Arg leads to a gradual shift of the maximum in c(s) to ~1.6 S at 0.2 M and ~2.0 S at 0.5 M, indicating that the mean (equilibrium) size of the clusters increases with the concentration. This is in line with considerations made for DOLLOPs of $CaCO_3$, where more frequent collisions between clusters and monomeric units (ions or ion pairs) at higher concentrations were proposed to account for a change in the size distribution of the clusters.[9] The estimated percentage of clusters varies inconsistently up to 0.5 M and appears to be constant within the limits of error (0 1%, *cf.* Fig. 4b). Starting from 0.8 M (*i.e.* when the solubility limit of ~1.3 M is slowly approached), X_C adopts considerably higher values (2 2.5% at 0.8 and 1 M). Simultaneously, a second population can be discerned in the s distributions, centred at about 2.3 and 2.7 S for 0.8 and 1 M, respectively. Corresponding peaks are markedly broadened as compared to the signal at lower s, suggesting that these species are cluster aggregates of variable size. Their enhanced occurrence at these concentrations rationalises the parallel increase in the cluster percentage, as new PNCs will be generated from Arg monomers upon aggregation of others provided that cluster formation is governed by equilibrium thermodynamics as in the case of calcium carbonate.[2]

This may also explain why the average size of the PNCs is smaller at 0.8 M than at 0.5 M (where pronounced aggregation does not yet take place). Further increasing the Arg concentration to 1 M leads to higher s values for both individual PNCs (~1.9 S) and their aggregates (~2.7 S), while the total percentage of clusters is more or less the same (note that X_C is not weighted by the number of monomers in the clusters). In the supersaturated regime, the relative amount of clusters drops significantly as aggregates are no longer detected and only the signal of single PNCs at ~1.9 S is left at both 1.56 and 1.82 M. This suggests that cluster aggregates were removed from the system by nucleation of a solid phase during (or even prior to) the AUC measurement, yielding particles that are too large to be observed in c(s) at the given rotor speed.

These concentration dependent data indicate that the traced clusters are relevant species for crystallisation: we found that aggregation of clusters occurs progressively when the saturation limit is approached and, therefore, speculate that nucleation proceeds *via* (or at the expense) of such aggregates both being directly analogous to $CaCO_3$.[1 3] Our results further evidence that at least the arginine clusters discussed in detail herein have characteristics similar to those of calcium carbonate,[2] since they appear to form on the basis of an equilibrium between monomeric units and PNCs, too. This is supported by the determined percentages of clusters as a function of concentration: as soon as significant amounts of cluster aggregates are formed in solution (*i.e.* at 0.8 and 1.0 M), X_C rises noticeably. Thereby, it is crucial to realise that cluster percentages calculated according to eqn (6) include contributions of both single clusters and aggregates, that is, $A_{Cluster\ peaks}$ is the sum of peak areas at s > 1 S. Thus, the observed increase in X_C is essentially due to the incorporation of clusters into aggregates which, in line with what is expected for an equilibrium reaction, causes new clusters to be formed from monomeric units (which then increase X_C). In fact, calculations of the actual percentage of single clusters (*i.e.* using only the area of the peak at 1 2 S in the numerator of eqn (6)) show that their fraction in equilibrium with monomers remains constant within the limits of error over the entire range of concentrations investigated (corresponding values are represented by triangles for 0.8 and 1.0 M in Fig. 4b). This behaviour agrees well with results

reported for calcium carbonate,[2] and suggests that amino acid clustering in solution relies on equilibrium thermodynamics.

We note that nucleation of glycine has previously been examined by means of small angle X ray scattering (SAXS) experiments using concentrated solutions that were gradually cooled in order to achieve increasing levels of supersaturation.[48,49] Analyses of the data indicated that a significant fraction of dimers was present at neutral pH from the beginning on (as judged from calculated radii of gyration) and that further oligomerisation took place with decreasing temperature (in line with early diffusivity and sedimentation studies),[12] [16,48] although this interpretation has later been debated.[43] Based on these findings and evaluations of the fractal behaviour of the samples, it was proposed that glycine nucleation is triggered by the formation of a liquid like cluster consisting of dimer building units and their subsequent reorganisation,[48] according to a typical two step nucleation model which assumes the intermediate clusters to be metastable.[50] The present results confront this notion in that amino acid PNCs were shown to exist already in undersaturated systems and therefore likely do not represent metastable crystallisation precursors. Our observations rather agree with preliminary neutron scattering (SANS) data on crystallising glycine solutions, which suggested the occurrence of larger aggregates (several nm and thus similar to the species denoted Cluster II in our analyses) before nucleation.[41] A general disadvantage of scattering techniques in this context is the fact that they probe the whole range of co existing species at once and often only yield average (size) information for the entire ensemble,[43] such that small fractions of possibly relevant crystallisation precursors may well be overlooked. This is by nature different in AUC, as it separates populations with sufficiently distinct sedimentation coefficient prior to detection. Consequently, even species with very low concentrations can be traced and a full distribution of sizes is obtained. On that basis, the measurements performed in this work have revealed the presence of nanosized clusters next to a huge excess of monomers for all amino acids in undersaturated solutions. Beyond that, we propose that the crystallisation of arginine relies on these clusters and essentially involves their aggregation as a crucial step towards nucleation; this, however, remains to be definitely proven, and related studies are required to shed light on the crystallisation pathways of the other amino acids as well.

4 Conclusions

Our findings demonstrate the existence of a qualitative correlation between amino acid oligomers observed in ESI MS and PNCs traced by AUC: when clusters are detected by AUC, higher oligomers occur in the mass spectra. Thus, it is likely that the species observed in the gas phase are in fact present in solution and not the result of artefacts during evaporation and ionisation. This implies that ESI MS may very well serve as rapid technique for the identification of systems in which prenucleation clusters occur. On the other hand, comparison of the oligomer distributions and clustering propensities indicated by the two methods reveals marked discrepancies, which we ascribe to distinct probabilities for the different amino acid oligomers to be transferred successfully into the gas phase. In particular, our data suggest that larger clusters are much less effectively transported and/or ionised, such that their contribution to size distributions is greatly underestimated by ESI MS. Moreover, experimental parameters and the type of instrument used seem to be further delicate factors influencing the final output. Therefore, even though mass spectrometry appears to be a promising alternative to search for prenucleation clusters, quantitative evaluation of cluster formation in solution by this technique remains challenging.

The AUC experiments carried out in this work have shown that measurable amounts of amino acid clusters occur even in dilute solutions. This together with size distributions estimated for these species (discrete peak separated from monomers), the traced concentration dependence of cluster formation (constant percentage of single clusters), and the aggregation behaviour observed at higher

Table 5 Sedimentation coefficients of ions and PNCs found for different salts. Measurement were carried out at a concentration of 0.02 M (and hence below the saturation limit) for ammonium acetate (NH$_4$OAc), MgSO$_4$, LiI, CsI, CsF, and Na$_2$CO$_3$, whereas saturated solutions were used in all other cases

Substance	s_{Ions} [S]	$s_{Cluster\ I}$ [S]	$s_{Cluster\ II}$ [S]	$s_{Cluster\ III}$ [S]
NH$_4$OAc	0.11 ± 0.07	0.4 ± 0.2	3 ± 1	
MgSO$_4$	0.23 ± 0.02	1.0 ± 0.2		
NaCl	0.04 ± 0.01	1.5 ± 0.2	2.75 ± 0.02	
LiI	0.14 ± 0.01	0.93 ± 0.04	3.27 ± 0.04	
CsI	0.81 ± 0.02	2.4 ± 0.3	5.1 ± 0.4	
CsF	0.4 ± 0.1	1.01 ± 0.03		
Na$_2$CO$_3$	0.19 ± 0.06	1.00 ± 0.34	3.1 ± 0.1	
MgCO$_3$	0.09 ± 0.01	1.0 ± 0.6	3.0 ± 0.1	
CaCO$_3$	0.11 ± 0.05	1.4 ± 0.1	5 ± 1	9 ± 2
SrCO$_3$	0.14 ± 0.04	0.9 ± 0.2		

concentrations serves as evidence that the detected clusters form on the basis of equilibrium thermodynamics and, hence, that they truly are PNCs in the sense of what has been described previously for calcium carbonate.[2] An obvious difference to the CaCO$_3$ case is the relatively low fraction of amino acid clusters existing in equilibrium with monomers (<1%, as opposed to several tens of percents for CaCO$_3$). However, due to the fairly high solubility of amino acids, the factual concentration of clusters (and aggregates) in solution amounts to the millimolar range when the point of nucleation is approached, as demonstrated for the example of arginine in the present study (for which the cluster concentration is on the order of 0.01 0.02 M close to the solubility limit). In this regard, it seems possible that these clusters (or respectively their aggregates) are precursors of initially nucleated particles and, thus, that they are relevant species for nucleation. Although the data obtained in this work support such a scenario, further insight to the nucleation process is required to draw definite conclusions.

Essentially, cluster formation was confirmed for all amino acids regardless of their functionality (anionic, cationic, hydrophilic or hydrophobic). This raises questions about the driving force of amino acid clustering, given that their solubility (and thus degree of saturation in this study) is quite distinct and, moreover, the hydration of the different amino acids should vary depending on their charge and hydrophilicity. Although we cannot answer this question on the basis of the present data, it seems evident that PNC formation is a more common phenomenon than hitherto believed. This notion is corroborated by additional AUC measurements of various inorganic salts other than the previously studied biominerals calcium carbonate, oxalate and phosphate.[2] The results shown in Table 5 evidence that PNCs were detected for all investigated salts which, remarkably, include both well and hardly soluble compounds and even salts of simple monovalent ions such as alkali halogenides.

We note that ion clustering has been reported for soluble salts in earlier work as well, but only concerning solutions at fairly high supersaturation.[19,51] First attempts to obtain good ESI MS spectra for ion pairs of inorganic salts and higher oligomers thereof were not yet successful in our lab, whereas meaningful results could be achieved in a previous study on calcium carbonate.[10] Most likely, our experimental parameters will have to be fine tuned to enable mass spectrometric determination of clusters also for inorganic systems. If a correlation similar to that presented here for the amino acid case can be established, ESI MS would be a fast method of choice to screen a large amount of samples for the presence of PNCs, which could then be precisely characterised for example by AUC. This combination of techniques would help to clarify how general prenucleation cluster formation is. The data collected in this work indicate that it is more widespread than previously assumed.

Acknowledgements

References

1 D. Gebauer and H. Colfen, *Nano Today*, 2011, **6**, 564.
2 D. Gebauer, A. Volkel and H. Colfen, *Science*, 2008, **322**, 1819.
3 E. M. Pouget, P. H. H. Bomans, J. A. C. M. Goos, P. M. Frederik, G. de With and N. A. J. M. Sommerdijk, *Science*, 2009, **323**, 1455.
4 A. Dey, P. H. H. Bomans, F. A. Muller, J. Will, P. M. Frederik, G. de With and N. A. J. M. Sommerdijk, *Nat. Mater.*, 2010, **9**, 1010.
5 F. Nudelman, K. Pieterse, A. George, P. H. H. Bomans, H. Friedrich, L. J. Brylka, P. A. J. Hilbers, G. de With and N. A. J. M. Sommerdijk, *Nat. Mater.*, 2010, **9**, 1004.
6 H. Colfen, *Nat. Mater.*, 2010, **9**, 960.
7 D. Gebauer, P. N. Gunawidjaja, J. Y, P. Ko, Z. Bacsik, B. Aziz, L. Liu, Y. Hu, L. Bergstrom, C. W. Tai, T. K. Sham, M. Eden and N. Hedin, *Angew. Chem., Int. Ed.*, 2010, **49**, 8889.
8 D. Gebauer, H. Colfen, A. Verch and M. Antonietti, *Adv. Mater.*, 2009, **21**, 435; A. Verch, D. Gebauer, M. Antonietti and H. Colfen, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16811.
9 R. Demichelis, P. Raiteri, J. D. Gale, D. Quigley and D. Gebauer, *Nat. Commun.*, 2011, **2**, 590.
10 S. E. Wolf, L. Muller, R. Barrea, C. J. Kampf, J. Leiterer, U. Panne, T. Hoffmann, F. Emmerling and W. Tremel, *Nanoscale*, 2011, **3**, 1158.
11 O. Gliko, W. Pan, P. Katsonis, N. Neumaier, O. Galkin, S. Weinkauf and P. G. Vekilov, *J. Phys. Chem. B*, 2007, **111**, 3106.
12 Y. C. Chang and A. S. Myerson, *AIChE J.*, 1986, **9**, 1567.
13 Y. C. Chang and A. S. Myerson, *AIChE J.*, 1987, **33**, 697; A. S. Myerson and P. Y. Lo, *J. Cryst. Growth*, 1990, **99**, 1048.
14 H. S. Na, S. Arnold and A. S. Myerson, *J. Cryst. Growth*, 1994, **139**, 104.
15 A. S. Myerson and P. Y. Lo, *J. Cryst. Growth*, 1991, **110**, 26.
16 R. M. Ginde and A. S. Myerson, *J. Cryst. Growth*, 1992, **116**, 41.
17 A. T. Allen, R. M. Wood and M. P. McDonald, *Sugar Tech. Rev.*, 1974, **2**, 165.
18 L. S. Sorell and A. S. Myerson, *AIChE J.*, 1982, **28**, 772.
19 M. A. Larson and J. Garside, *Chem. Eng. Sci.*, 1986, **41**, 1285.
20 C. K. Meng and J. B. Fenn, *Org. Mass Spectrom.*, 1991, **26**, 542.
21 D. Zhang, L. Wu, K. J. Koch and R. G. Cooks, *Eur. J. Mass Spectrom.*, 1999, **5**, 353.
22 Z. Takats, S. C. Nanita, R. G. Cooks, G. Schlosser and K. Vekey, *Anal. Chem.*, 2003, **75**, 1514.
23 P. Nemes, G. Schlosser and K. Vekey, *J. Mass Spectrom.*, 2005, **40**, 43.
24 N. Toyama, J. Kohno, F. Mafune and T. Kondow, *Chem. Phys. Lett.*, 2006, **419**, 369.
25 P. Yang, R. Xu, S. C. Nanita and R. G. Cooks, *J. Am. Chem. Soc.*, 2006, **128**, 17074.
26 M. Przybylski and M. O. Glocker, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 806; A. Marquardt, B. Bernevic and M. Przybylski, *J. Pept. Sci.*, 2007, **13**, 803; L. J. Deterding, J. Kast, M. Przybylski and K. B. Tomer, *Bioconjugate Chem.*, 2000, **11**, 335; H. Wendt, E. Durr, R. M. Thomas, M. Przybylski and H. R. Bosshar, *Protein Sci.*, 1995, **4**, 1563; M. Przybylski, M. O. Glocker, C. Maier, C. Borchers, E. Durr, W. Fiedler, J. Kast, H. Wendt and H. R. Bosshard, in *Peptides*, ed. H. L. S. Maia, Escom Science Publications, Leiden, 1994, pp. 42.
27 P. Schuck, *Biophys. J.*, 2000, **78**, 1606.
28 http://www.analyticalultracentrifugation.com/default.htm.
29 B. Demeler and H. Saber, *Biophys. J.*, 1998, **74**, 444.
30 H. Colfen and A. Volkel, *Prog. Colloid Polym. Sci.*, 2004, **127**, 31.
31 http://jphilo.mailway.com/download.htm.
32 K. J. Koch, F. C. Gozzo, D. Zhang, M. N. Eberlin and R. G. Cooks, *Chem. Commun.*, 2001, 1854; R. G. Cooks, D. Zhang, K. J. Koch, F. C. Gozzo and M. N. Eberlin, *Anal. Chem.*, 2001, **73**, 3646; K. J. Koch, F. C. Gozzo, S. C. Nanita, Z. Takats, M. N. Eberlin and R. G. Cooks, *Angew. Chem., Int. Ed.*, 2002, **41**, 1721.
33 A. E. Counterman and D. E. Clemmer, *J. Phys. Chem. B*, 2001, **105**, 8092; R. R. Julian, R. Hodyss, B. Kinnear, M. F. Jarrold and J. L. Beauchamp, *J. Phys. Chem. B*, 2002, **106**, 1219.
34 C. A. Schalley and P. Weis, *Int. J. Mass Spectrom.*, 2002, **221**, 9.
35 J. L. Cole, J. W. Lary, T. Moody and T. M. Laue, *Methods Cell Biol.*, 2008, **84**, 143.

36 H. P. Erickson, *Biol. Proced. Online*, 2009, **11**, 32.
37 W. C. M. Lewis, *Chem. Rev.*, 1931, **8**, 81; G. A. Anslow, M. L. Foster and C. Klingler, *J. Biol. Chem.*, 1933, **103**, 81.
38 D. Gidalevitz, R. Freidenhans'l, S. Matlis, D. F. Similgies, M. J. Christensen and L. Leiserowitz, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 955; I. Weissbuch, V. Y. Torbeev, L. Leiserowitz and M. Lahav, *Angew. Chem., Int. Ed.*, 2005, **44**, 3226; V. Y. Torbeev, E. Shavit, I. Weissbuch, L. Leiserowitz and M. Lahav, *Cryst. Growth Des.*, 2005, **5**, 2190.
39 C. S. Towler, R. J. Davey, R. W. Lancaster and C. J. Price, *J. Am. Chem. Soc.*, 2004, **126**, 13347.
40 J. Huang, T. C. Stringfellow and L. Yu, *J. Am. Chem. Soc.*, 2008, **130**, 13973.
41 C. E. Hughes, S. Hamad, K. D. M. Harris, C. R. A. Catlow and P. C. Griffiths, *Faraday Discuss.*, 2007, **136**, 71.
42 T. Sato, R. Buchner, S. Fernandez, A. Chiba and W. Kunz, *J. Mol. Liq.*, 2005, **117**, 93.
43 S. Hamad, C. E. Hughes, C. R. A. Catlow and K. D. M. Harris, *J. Phys. Chem. B*, 2008, **112**, 7280.
44 J. G. Kirkwood and J. Riseman, in *Rheology: theory and applications*, ed. F. Eirich, Academic Press, New York, 1956, pp. 495.
45 P. Raiteri, R. Demichelis, J. D. Gale, M. Kellermeier, D. Gebauer, D. Quigley, L. B. Wright and T. R. Walsh, *Faraday Discuss.*, 2012, DOI: 10.1039/C2FD20052J.
46 S. Lapanje, *Physicochemical aspects of protein denaturation*, Wiley, New York, 1978.
47 T. M. Laue, *Choosing which optical system of the OptimaTM XL I analytical ultracentrifuge to use*, Application Information A 1821A, Beckman Instruments, Fullerton (CA), 1996.
48 S. Chattopadhyay, D. Erdemir, J. M. B. Evans, J. Ilavsky, H. Amenitsch, C. U. Segre and A. S. Myerson, *Cryst. Growth Des.*, 2005, **5**, 523.
49 D. Erdemir, S. Chattopadhyay, L. Guo, J. Ilavsky, H. Amenitsch, C. U. Segre and A. S. Myerson, *Phys. Rev. Lett.*, 2007, **99**, 115702.
50 P. G. Vekilov, *Cryst. Growth Des.*, 2004, **4**, 671.
51 Y. C. Chang and A. S. Myerson, *AIChE J.*, 1985, **31**, 890; R. Mohan, O. Kaytancioglu and A. S. Myerson, *J. Cryst. Growth*, 2000, **217**, 393.