



A Monte Carlo approach for assessing the specificity of protein oligomers observed in nano-electrospray mass spectra

Laura A. Lane, Brandon T. Ruotolo, Carol V. Robinson, Giorgio Favrin*, Justin L.P. Benesch**

Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK

ARTICLE INFO

Article history:

Received 5 December 2008

Received in revised form 9 March 2009

Accepted 10 March 2009

Available online 24 March 2009

This contribution is dedicated to Mike Bowers on the occasion of his 70th birthday. His elegant development and applications of ion mobility mass spectrometry are a continuing source of inspiration to us and many other researchers.

Keywords:

Protein complex
Non-specific
Nano-electrospray
Amyloid
Monte Carlo

ABSTRACT

Nano-electrospray mass spectrometry is an emerging approach for studying the architecture and dynamics of complex oligomeric proteins. The analysis of such species can, however, be hindered by 'non-specific' protein–protein associations which arise as a result of the electrospray method. To understand the formation of specific *versus* non-specific protein oligomers detected by mass spectrometry we have developed a simple and rapid computational approach. This Monte Carlo algorithm characterizes the occupancy of protein species in the last offspring droplets created by the nano-electrospray process. As such it enables us to assess whether oligomers detected in mass spectra reflect solution populations, or instead are the result of associations during droplet fission and evaporation. We have trained and validated this method on three model protein complexes which are not known to form higher order oligomers, and one which has a tendency to self-associate in solution in a concentration dependent manner. We have then compared predictions of droplet occupancy to abundances obtained from mass spectra for the tetrameric amyloid-related protein transthyretin, which can cause cardiomyopathy and polyneuropathy in humans. Interestingly, when such comparisons were made for wild-type transthyretin we were able to observe a propensity for the protein to form specific oligomers larger than the tetramer. In contrast, the tendency for the leucine-55-proline variant to form such oligomers was considerably reduced. We contemplate the significance of these specific higher order wild-type oligomers, and absence of such species in the variant, on the pathway of amyloid formation in transthyretin. More generally this easy-to-implement computational approach promises to improve our ability to identify oligomers of biological significance within the mass spectra of heterogeneous protein complexes.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Mass spectrometry (MS) has become an extremely effective means for the investigation of architectural and dynamical aspects of proteins and the large biomolecular assemblies they form [1–5]. A number of amyloid-related proteins have been studied by MS including the aggregation of an SH3 domain [6], and β -2-microglobulin [7,8]. Additionally, MS has been used to investigate Parkinson's disease related protein α -synuclein [9,10], and the structural properties of the monomers, protofibrils and amyloid derived from the Alzheimer's peptide Amyloid- β [11–13]. The amyloid-related protein transthyretin (TTR) has previously been studied in our laboratory to characterize its dynamics [14], and ligand binding ability [15]. During these studies, in addition to the basic TTR tetramer [16], the presence of low abundances of octameric and dodecameric higher order oligomers was noted.

While one of the strengths of MS is its ability to interrogate poly-disperse systems [17], it can still remain a challenge to determine the specificity of interactions, and hence assess the biological significance of low intensity higher order oligomers such as these observed for TTR.

This difficulty arises because occasionally the presence of multiple oligomeric forms does not reflect the solution composition but rather is a result of the electrospray (ES) process. ES, particularly the miniaturised version nano-ES [18], is the method of choice for transferring fragile protein complexes into the gas phase [1], and is widely accepted to generate protein ions according to the charged residue model (CRM) [19,20]. The fundamental size difference between a protein complex and a nano-ES droplet, and the subsequent fission processes, can result in multiple copies of a protein species not only in the initial droplets formed, during nano-ES, but also in the progeny droplets. The presence of multiple protein species within the final offspring droplets, and the subsequent evaporation which fully desolvates the droplet-bound species, results in the production of so-called 'non-specific' assemblies [1]. This is an effect which therefore results from the ionization process, and is distinct from the tendency for some proteins to

* Corresponding author. Tel.: +44 0 1223 763848.

** Corresponding author. Tel.: +44 0 1223 331697; fax: +44 1223 336362.

E-mail addresses: g247@cam.ac.uk (G. Favrin), jlpb2@cam.ac.uk (J.L.P. Benesch).

aggregate in solution [21]. It is critical to be able to discriminate between these artefacts of the ES process, and those interactions which derive from the proteins having an affinity for each other in solution. Most recently these false positives have been studied using a non-binding reference protein [22] or reporter molecule [23], or by employing a hydrogen–deuterium exchange methodology [24].

In order to tackle the problem of identifying these ES-produced assemblies without labelling or addition of standards we have developed a computational method which enables the identification of non-specific associations. The Droplet Occupancy Predictor (DrOP)¹ algorithm predicts, purely from a probabilistic point of view, the distribution of proteins in the final nanospray droplets. Through calculating the number of protein complexes present in simulated droplets we are able to distinguish between two key scenarios: (i) are the higher order oligomers observed in nano-ES spectra reflective of solution populations, or (ii) are they a result of association as the droplets evaporate to produce ‘naked’ protein ions? Hereupon we will refer to these types of oligomers as being ‘specific’ or ‘non-specific’, respectively.

We trained and validated DrOP over a wide concentration range using four globular protein complexes: triosephosphate isomerase (TIM), pyruvate kinase (PK), alcohol dehydrogenase (ADH), and glutamate dehydrogenase (GDH). These protein complexes are of different sizes, and vary in their known propensities for self-association in solution. Thus validated, our *in silico* approach was then applied to TTR, the aggregation of which leads to a number of devastating diseases characterized by the extracellular deposition of insoluble amyloid fibrils in the tissues and organs [25,26]. The wild-type (WT) protein is implicated in senile systemic amyloidosis, a degenerative geriatric disease [27,28]. Additionally, many variant forms of TTR cause the lethal familial amyloidotic polyneuropathy (FAP) [29], the most virulent form of which results from the leucine-55-proline (L55P) point mutation [30]. Current models for *in vitro* TTR aggregation propose that the amyloid is formed as a result of tetramer dissociation to non-native but globular monomers which have undergone some misfolding; these species can then assemble into protofibrils and amyloid fibrils [31,32]. However, in both diseases, there remains uncertainty in the aetiology of the pathogenic species and the intermediates formed *en route* to amyloid deposition [33,34]. Recently, toxic non-fibrillar aggregates have been observed in FAP deposits and are thought to consist of species up to ~100 kDa in size [35,36]. The potential involvement of multiple oligomeric forms of TTR in the disease process encouraged us to investigate both the WT and the L55P variant form in order to assess the specificity of the octameric and dodecameric forms we had observed previously [14,15].

Through our use of DrOP we are able to differentiate between oligomers that exist in solution and those formed during nano-ES. Consequently we can identify specific interactions between WT tetramers in solution as contributing to higher order oligomers of WT TTR. In contrast, we find that the higher order oligomers of L55P TTR derive solely from the nano-ES process itself. By exploring the specificity of these higher oligomers in TTR we reflect on the extent of the involvement of higher order oligomers in stabilizing protein structure and in the mechanism of amyloidogenesis.

2. Experimental

2.1. Proteins

TIM and PK from *Oryctolagus cuniculus*, ADH from *Saccharomyces cerevisiae*, and GDH from *Bos taurus* were obtained from Sigma

¹ The DrOP algorithm is available from the authors, free of charge for academic use.

(Poole, Dorset, UK). TTR was expressed and purified in-house as described previously [37]. TIM, PK, ADH and GDH were buffer-exchanged into 100 mM ammonium acetate and TTR into 20 mM ammonium acetate (both pH 6.8) using micro bio-spin-6 columns (Bio-Rad Laboratories).

Protein concentration was obtained by measuring UV absorbance at 280 nm (Biomate, Thermo). An experimental extinction coefficient of 19400 M⁻¹ cm⁻¹ [38] was used to calculate TTR concentration, and theoretical extinction coefficients were used to calculate TIM, PK, ADH and GDH concentrations using equation (1),

$$\varepsilon_{280} = 5500n_{Trp} + 1490n_{Tyr} + 125n_{SS} \quad (1)$$

where ε_{280} is the extinction coefficient in M⁻¹ cm⁻¹, n_{Trp} is the number of tryptophan residues, n_{Tyr} is the number of tyrosine residues and n_{SS} the number of disulphide bonds [39].

Solutions of protein complexes for mass spectrometry were prepared from stock solutions to ensure accurate relative concentrations for successive dilutions. Concentrations (in μ M) of TIM dimer analysed were: 200, 150, 100, 75, 50, 25, 12.5, 5, and 2.5. Concentrations (in μ M) of PK tetramer, ADH tetramer, and GDH hexamer analysed were: 25, 6.25, and 1.56. Concentrations (in μ M) of TTR (WT and L55P) tetramer analysed were: 50, 25, 12.5, 6.25, 3.13, 1.56, and 0.78.

2.2. Mass spectrometry conditions

Nano-electrospray mass spectra were acquired on a Synapt HDMS mass spectrometer (Waters, Manchester, UK), using a previously described protocol optimized for the transmission of noncovalent protein complexes [40]. Variables such as needle tip orifice and needle position were maintained as carefully as possible and all instrument parameters and pressures were preserved for all concentrations of each protein.

The following instrument parameters were used for TIM, in positive ion mode: capillary voltage 1.7 kV; cone gas 80 L h⁻¹; sample cone 50 V; extraction cone 0 V; accelerating voltage into the trap cell (termed ‘trap collision energy’ as per the manufacturer’s terminology) 15 V; accelerating voltage into the transfer cell (termed ‘transfer collision energy’ as per the manufacturer’s terminology) 15 V; bias voltage 30 V; backing pressure 6.6 mbar, trap and transfer pressure 5.3×10^{-2} mbar (argon), T-wave drift cell pressure 5.1×10^{-1} mbar (nitrogen), and ToF pressure 2.3×10^{-6} mbar.

The following instrument parameters were used for PK, in positive ion mode: capillary voltage 1.7 kV; cone gas 80 L h⁻¹; sample cone 100 V; extraction cone 0 V; trap collision energy 10 V; transfer collision energy 10 V; bias voltage 20 V; backing pressure 4.5 mbar, trap and transfer pressure 3.5×10^{-2} mbar (argon), T-wave drift cell pressure 4.4×10^{-1} mbar (nitrogen), and ToF pressure 1.1×10^{-6} mbar.

The following instrument parameters were used for ADH, in positive ion mode: capillary voltage 2.0 kV; cone gas 80 L h⁻¹; sample cone 100 V; extraction cone 0 V; trap collision energy 10 V; transfer collision energy 10 V; bias voltage 20 V; backing pressure 4.6 mbar, trap and transfer pressure 5.2×10^{-2} mbar (argon), T-wave drift cell pressure 5.0×10^{-1} mbar (nitrogen), and ToF pressure 1.2×10^{-6} mbar.

The following instrument parameters were used for GDH, in positive ion mode: capillary voltage 1.8 kV; cone gas 80 L h⁻¹; sample cone 100 V; extraction cone 0 V; trap collision energy 10 V; transfer collision energy 10 V; bias voltage 20 V; backing pressure 4.7 mbar, trap and transfer pressure 3.4×10^{-2} mbar (argon), T-wave drift cell pressure 4.3×10^{-1} mbar (nitrogen), and ToF pressure 8.7×10^{-7} mbar.

The following instrument parameters were used for TTR, in positive ion mode: capillary voltage 1.5 kV; cone gas 80 L h⁻¹; sample

cone 100 V (for WT TTR) and 50 V (for L55P TTR); extraction cone 0 V; trap collision energy 10 V; transfer collision energy 10 V; bias voltage 20 V; backing pressure 4.3 mbar, trap and transfer pressure 5.4×10^{-2} mbar (argon), T-wave drift cell pressure 5.1×10^{-1} mbar (nitrogen), and ToF pressure 1.49×10^{-6} mbar. The reduced sample cone voltage for the L55P TTR compared to the WT form was used to ensure no dissociation of the complex and does not affect the relative intensities of the oligomers in the mass spectra.

2.3. Data analysis

Ion mobility MS (IM-MS) spectra were acquired and processed using MassLynx 4.1 with Driftscope (Waters, Manchester, UK). The m/z dimension of the spectra was calibrated externally using a 100 mg mL⁻¹ solution of caesium iodide. All spectra are shown with minimal smoothing and without background subtraction. A linear background was, however, subtracted prior to quantitative analysis to facilitate accurate measurement of the intensities of various oligomers. Relative intensities of oligomers were calculated from the mean peak height of the three most intense peaks in the charge state distribution for each oligomer. In all cases this accounts for >80% of the ion current and avoids the complications which arise from overlapping charge states. The deviation between triplicate experimental measurements is $\approx 18\%$ of the mean abundances shown.

2.4. Statistical fitting

DrOP predictions for different ξ values (for definition see Section 3.2) were compared quantitatively to the TIM training set by using root mean square deviation (RMSD). At each concentration, the ξ that fitted the data best was taken to be that with the minimum resulting RMSD. Furthermore, the difference in RMSD between the best and second-best fits represents a useful indicator of the quality of the overall correspondence between experiment and theory for that particular concentration. This is therefore used to give a weighted ξ which is subsequently applied to PK, ADH, and GDH for validation, and to investigate the self-association of TTR.

3. Theory

3.1. Non-specific association via the charged-residue model

The mechanism of charging of protein complexes during electrospray is generally accepted to follow the CRM [5,18–20]. After being expelled from the Taylor cone, a droplet, with charges localized at its surface, loses solvent through evaporation until it reaches the Rayleigh diameter D_R (equation (2)) [41]:

$$D_R = \sqrt[3]{\frac{n_e e}{\pi \gamma \epsilon_0}} \quad (2)$$

where n_e is the charge on the droplet surface, γ the solvent surface tension, ϵ_0 the electrical permeability of vacuum, and e the elementary charge.

At this point Coulombic repulsion overcomes the surface tension holding the droplet together, and an asymmetric fission occurs, whereby small offspring droplets taking $\sim 2\%$ of the volume and $\sim 15\%$ of the charge are released from the parent droplet [42,43]. These processes of evaporation and fission reoccur until the solvent is depleted and what remains is a residue composed of the components of the final droplet, with the same charge the droplet held an instant before [19]. It has been shown that the charge on globular proteins observed in electrospray mass spectra is at or below that predicted by the Rayleigh limit [20], a trend which appears to hold also for protein complexes [5], indicating that these species are likely formed *via* the CRM.

This leads to a simple model for the formation of non-specific oligomers, namely that they arise when two or more copies of a species are present in the final droplet, and therefore constitute the charged residue formed. This suggests that the number of non-specific oligomers might be deduced through predicting the occupancy of the droplets produced during electrospray [1].

3.2. Droplet occupancy predictor (DrOP)

To model the occupancy of nanospray droplets we incorporated the theory described in the previous section into a Monte Carlo simulation implemented in C++. The software relies on a number of parameters: the hard-sphere radius of the species being investigated (R_S); the concentration of this species in solution (C_S) and the radius of the final droplet (R_D).

R_S is estimated using the following relationship (equation (3)), based on an average protein density of $0.84 \text{ Da } \text{\AA}^{-3}$ and assuming a smooth sphere topology [1],

$$R_S = \sqrt[3]{\frac{m}{1120\pi}} \quad (3)$$

where m is the mass in Da of the protein complex under scrutiny. The implicit assumption that the density of gas-phase proteins and protein complexes can be approximated as constant is suggested by the literature [44]. As a starting point R_D is considered to be 9 nm, based on the literature suggesting the formation of initial nanospray droplets at $\sim 90 \text{ nm}$ radius [18,45], and their offspring having a radius $\sim 10\%$ of this value [42,43].

DrOP operates by filling a theoretical volume with a number of proteins, as defined from C_S , which are assigned random spatial coordinates. R_S is used to establish an excluded volume, such that different proteins cannot overlap with each other. This element is particularly important when C_S or R_S are very large, and overlap is likely. A sub-volume based on R_D is then specified, to represent the final nanospray droplet, and the number of proteins within each droplet are counted. This simulation is then repeated many ($\sim 50,000$) times in order to establish accurate statistics.

There are a number of assumptions made in this computational setup with the potential to affect the accuracy of the simulations. These are primarily the accuracy of the estimates of R_S , R_D and the accuracy of measurement of C_S . Of these the value for R_D is expected to be the most significant source of error, as it is derived from literature estimates based on extrapolation, rather than empirical measurement. Additionally, though there is likely to be a distribution of offspring droplet sizes we take a single value for R_D for our calculations. During our algorithm development we have found that this simplification makes a final difference in droplet occupancy of less than 3%. More significantly, however, R_D can be affected considerably by MS parameters such as the dimensions of the nanospray capillary, the nanospray voltage, the distance between capillary and sample cone, and the amount of backing pressure [18], and hence is likely to vary between different instruments and users. To address the inaccuracy and variability of R_D we have included a dimensionless correction factor into DrOP, which can act to scale simply R_D as follows (equation (4)),

$$x = xR_D \quad (4)$$

where x is the correction factor, and ξ is inputted into DrOP in place of R_D . It should be noted that this is equivalent to scaling C_S , as increasing the number of proteins counted within the sub-volume can not only be effected by increasing the size of the sub-volume (R_D) but also by increasing the particle density (C_S). Therefore scaling R_D not only addresses inaccuracy in our estimate of this parameter, but also any deviations in C_S which may result from solvent evaporation during nano-ES, or from imprecise concentration determination.

Additionally, it is known that there are MS instrumental factors which combine to affect the relationship between the actual abundances in solution, and those observed in the gas phase. These include detector efficiency, and other effects such as saturation and ion beam divergence [46]. In addition the relative ionization efficiency and signal suppression can further alter the observed distribution *versus* the actual [47]. Though these effects are likely to be relatively minor compared to the estimate of R_D , the replacement of R_D with ξ also has the benefit of reducing systematic deviations between computational prediction and MS experiment by manipulating the droplet occupancy and, as a consequence, the relative abundances of oligomers. It should be noted however that though ξ is derived from R_D and has units of nm, it should not be regarded as being the actual radius of the final droplets. Remarkably, we find that this simple approach is sufficient to ensure that the predicted distributions match well those observed within the accuracy of our MS measurements.

4. Results and discussion

4.1. The non-specific association of TIM

TIM is a monodisperse protein, physiologically active in higher eukaryotes as a 53.3 kDa dimer [48]. With no known propensity for TIM from *Oryctolagus cuniculus* to form higher order oligomers we conceived this as a reliable system to study non-specific interactions of protein complexes.

Fig. 1 shows nano-ES mass spectra of TIM over a 2.5–200 μM dimer concentration range. At 2.5 μM concentration there is one predominant charge state distribution, centred on the +14 charge state, which dominates the spectrum. This charge state series corresponds to a mass of $53,268 \pm 11$ Da, in good agree-

ment with the 53,251 Da mass calculated for the sequence of the TIM dimer. A very small amount of TIM tetramer is also noticeable at this concentration in the 4700–6000 m/z region. As the concentration is increased the number of charge state distributions per spectrum also increases. At a concentration of 200 μM species ranging from a dimer to a hexadecamer are observed for TIM. The different oligomeric states are labelled in Fig. 1 with the dimer occurring in the 3500–4300 m/z range, the tetramer in the 4700–6000 m/z range, hexamer in the 6000–7100 m/z range up to a hexadecamer form in the 10,000–10,500 m/z region.

These mass spectra clearly illustrate that with increasing protein concentration a wider distribution of oligomeric states is observed. Furthermore, we notice a decrease in the intensity of these non-specific higher oligomers which is distinctly different from the distributions observed for samples which exhibit polydispersity in solution [17]. These results therefore represent a ‘training set’ of values with which to constrain simulations predicting the occupancy of nano-ES droplets.

4.2. Development of the DrOP algorithm

To simulate the formation of non-specific clusters of protein complexes we developed a Monte Carlo based approach to assess the occupancy of protein species in the last offspring droplets created by the nano-ES process. Fig. 2 shows schematically the steps included in the programme to help us analyse the origins of interactions between protein complexes (for more detail see Section 3.2). Stage 1 calculates the number of proteins species in a cubic virtual volume of $0.125 \mu\text{m}^3$, based on the measured protein concentration. Stage 2 involves the random distribution of these protein species throughout the virtual volume. Following this the volume of a nano-ES progeny droplet is superimposed on the unit volume and the algorithm assesses if any proteins (stage 3), and how many (stage 4), are contained within this droplet. Stages 2–4 are repeated 50,000 times (stage 5) in order to produce statistically accurate data for occupancy of droplets. The final output gives details of the droplet occupancy distribution which we assume is directly related to the relative populations of the oligomeric species observed by MS.

The programme relies on three inputs: the radius of the species in question (R_S), the measured concentration (C_S), and a value ξ derived from the radius of the droplet and a correction factor (for more detail see Section 3.2). In the case of TIM R_S was approximated to 2.47 nm. Using this radius we simulated the droplet occupancy of a specified concentration of biological units (BUs) over a range of ξ values from 20 to 35 (Fig. 3). For the lowest value of ξ the simulation predicts the dimer (one BU) to be the most abundant species, with the intensity of higher oligomers decaying sharply with increasing numbers of BUs. As ξ increases the intensity of one BU drops and the intensity of the higher order oligomers increase. At $\xi = 26$ the intensity of higher oligomers continues to decrease with increasing oligomer size but much more gradually than for smaller ξ values. Additionally, for $\xi = 26$ the intensity of two BUs (a tetramer) is greater than that of one BU. This increasing trend in intensity from one to two BUs is also observed for larger values of ξ .

Qualitatively, the simulated intensities plotted for the 100 μM TIM dimer concentration (Fig. 3), and those calculated for other concentrations (data not shown), appear similar to the experimental MS oligomer intensities of TIM shown in Fig. 1. However, qualitative analysis alone is insufficient and it is important to be able to compare quantitatively the experimentally determined relative oligomeric abundances with those predicted computationally.

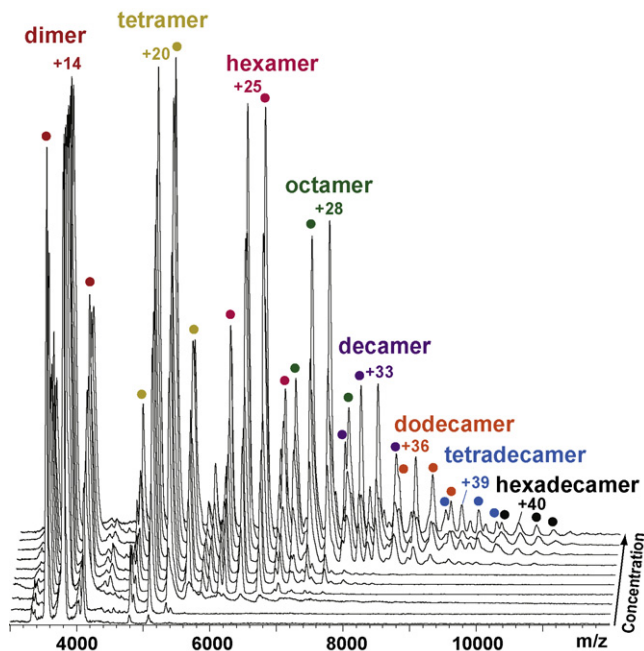


Fig. 1. Stack plot of nano-ES mass spectra of TIM over a range of concentrations. With increasing TIM concentration more non-specific assemblies larger than the dimeric biological unit are observed. At the lowest concentration (2.5 μM) mainly dimer (red) with a small amount of tetramer (yellow) are detected, whereas at the highest concentration (200 μM) there are many oligomers including: dimer (red); tetramer (yellow); hexamer (pink); octamer (green); decamer (purple); dodecamer (orange); tetradecamer (blue) and hexadecamer (black). From front to back dimer concentrations are: 2.5, 5, 12.5, 25, 50, 75, 100, 150, and 200 μM . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

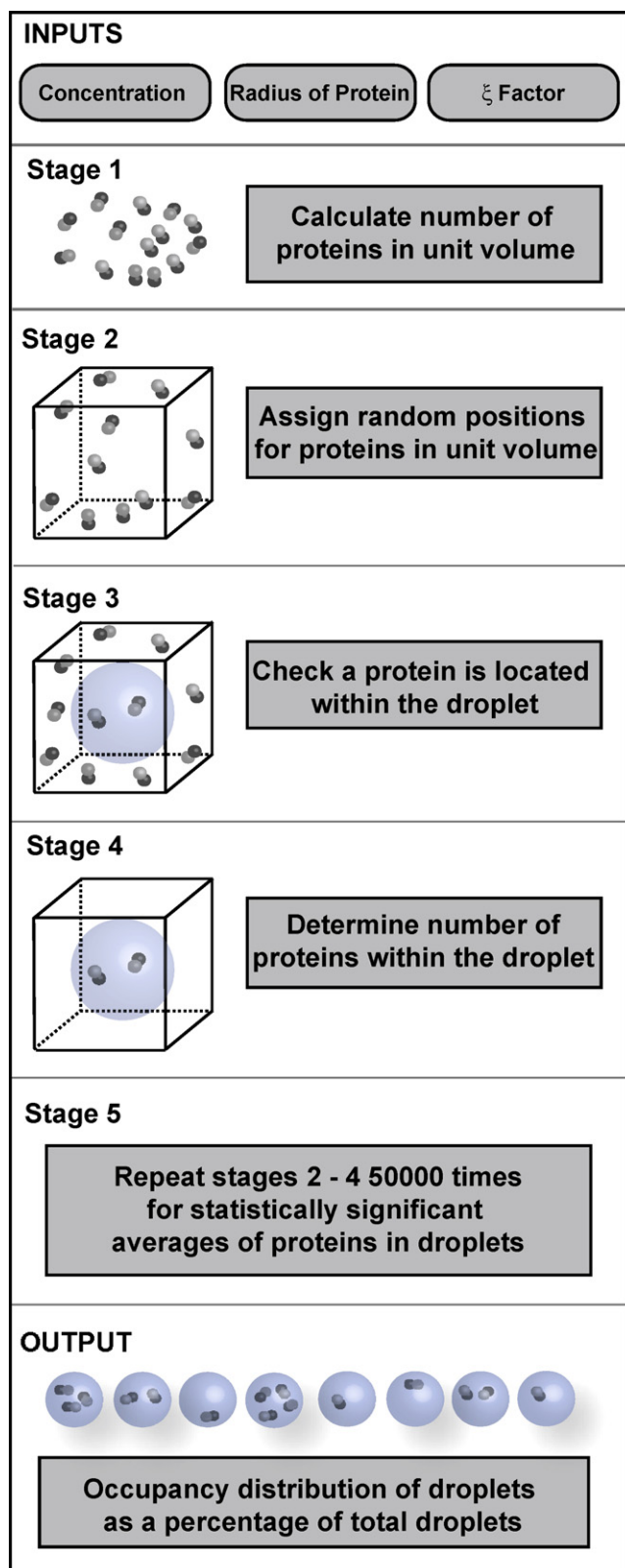


Fig. 2. Schematic of inputs, output and steps in the DrOP algorithm. Three inputs are required: sample concentration, protein radius, and ξ . Once the number of protein species in a unit volume has been calculated (Stage 1) the species are assigned random positions in a unit volume (Stage 2). Stage 3 locates a progeny nanospray droplet at the centre of the unit volume and checks that at least one protein species is located within the droplet. Stage 4 determines the number of protein species per droplet and stage 5 repeats stages 2–4 50,000 times to obtain statistically accurate occupancy values. The final step is the occupancy distribution output.

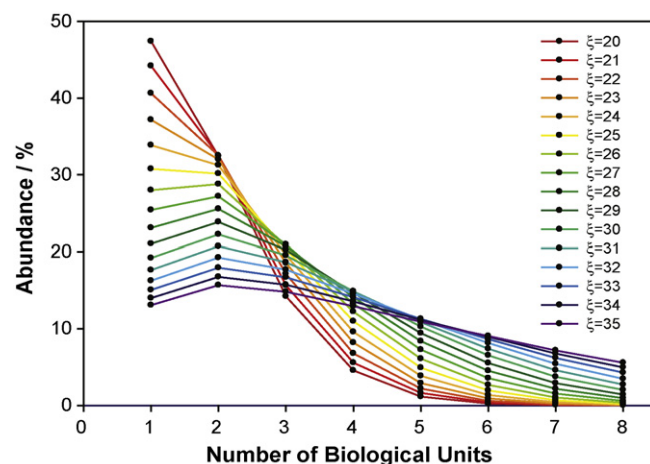


Fig. 3. Oligomer intensities simulated using DrOP for 100 μM TIM for a range of ξ values. For the lowest ξ (20) a sharp decline in higher order oligomer intensity is seen. As ξ increases the decay in intensity of higher order oligomers is more gradual, and at $\xi=26$ a small increase in intensity from one to two biological units, followed by a decrease in intensity for subsequent higher order species, is observed. This trend continues up to $\xi=35$ for which a slight increase in intensity is observed from one to two biological units followed by an almost linear decay in intensity for subsequent oligomers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

4.3. Finding the optimum ξ

Simulations of TIM droplet occupancy were computed for ξ values from 20 to 35 over a BU concentration range of 2.5–200 μM , and overlaid with the experimental oligomer abundances determined from the mass spectra (Fig. 4). At low sample concentration (5 μM) (Fig. 4A) the choice of ξ is relatively robust, with a wide range of ξ values appropriately fitting the experimental data, though the best fit is $\xi=22$. However, with increasing concentration, ξ must be more accurately defined in order to model the experimental findings. Panel B shows that at 75 μM dimer concentration there is a wider distribution of oligomeric species detected by MS and that the simulated intensities shown by the line graphs are more diverse. At this concentration $\xi=27$ is the closest fit to the experimental abundances. Finally panel C compares the simulated intensities to the experimental intensities of TIM with a dimer concentration of 200 μM . A ξ of 21 at this concentration most appropriately describes the experimental finding. The ξ value identified for each concentration as best describing the experimental intensities was chosen as the ξ value with the minimum RMSD between simulated and experimental intensities of each BU for a given concentration. These ξ values, and their corresponding RMSDs, were then used to calculate a weighted mean ξ of 24 ± 3 (error represents two standard deviations from the weighted mean), over the 2.5–200 μM concentration range. This average ξ value can now be used to examine the specificity of other protein complex associations.

It is notable that ξ is considerably greater than the literature value of R_D , demonstrating that the scaling factor is necessary. This is expected due to the assumptions of the model and the fact that the ξ value encompasses a range of systematic experimental errors, most significantly the accuracy of R_D (for more details see Section 3.2). While the ξ value addresses a number of complicated issues, the overall effect of all these influences on droplet occupancy are described adequately by this simple scaling. It should be noted that, due to ξ accommodating instrumental factors, the MS platform used will likely affect the optimal ξ . Moreover, while a single ξ value describes remarkably well the effects of concentration on oligomeric intensity for TIM, it is also important that we

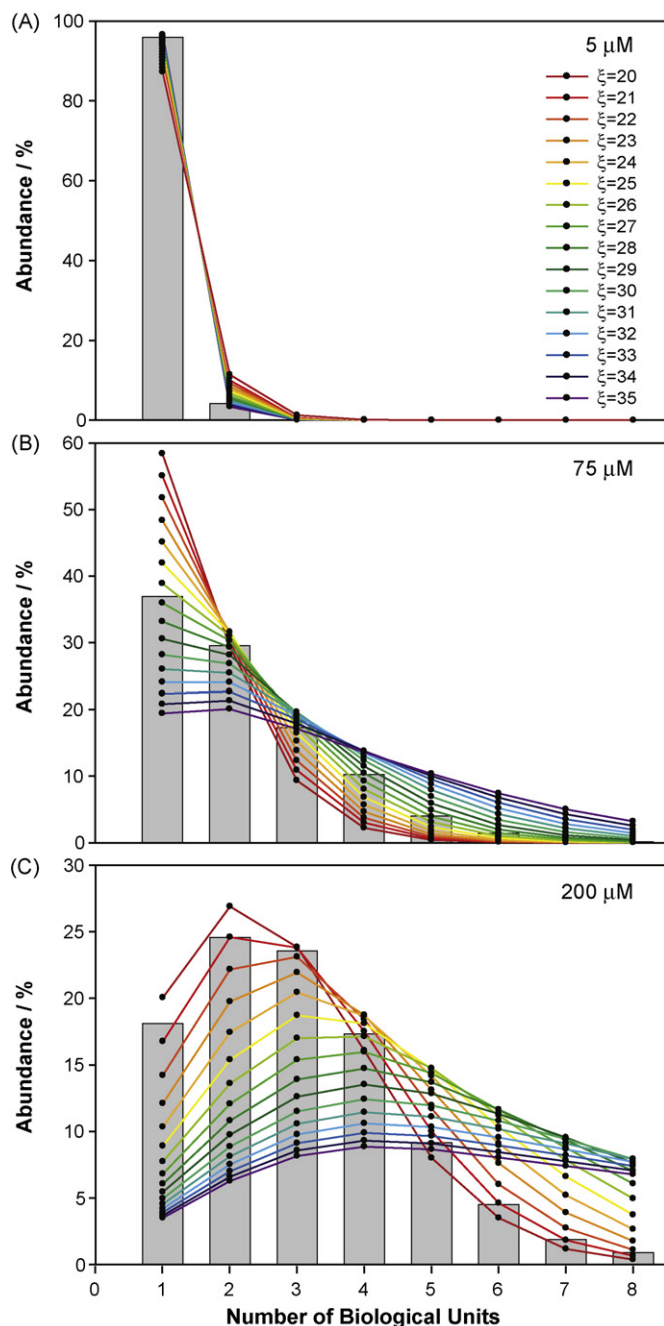


Fig. 4. Comparison of experimental MS intensities of TIM oligomers (bars) with simulated intensities of TIM oligomers (lines) for dimer concentrations of 5 μ M (A), 75 μ M (B) and 200 μ M (C). Panel A demonstrates that for a low concentration sample several simulated intensities, computed using a number of different ξ values, can each describe well the experimental intensities. However, with increasing sample concentration (panels B and C) the experimental oligomer distribution is most accurately described by a single simulation with a defined value for ξ . From these experiments we calculated an empirical value for ξ of 24 ± 3 . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

determine the general applicability of this method to other protein complexes.

4.4. Validation of DrOP method by application to PK, ADH and GDH

Having characterized the non-specific associations of the TIM dimer, in order to validate the DrOP method, we applied the

weighted average ξ value (24 ± 3) identified from TIM simulations to PK, ADH, and GDH to predict the level of non-specific oligomerization. We chose PK and ADH, which are homotetramers of 232 and 146 kDa, respectively, as they are not known to form higher order oligomers at ambient temperature [49,50]. This, in addition to them being considerably different in both size and oligomerization to TIM, means PK and ADH represent stringent positive controls. GDH, a 334 kDa hexamer, by contrast, is known to self-associate in solution in a concentration dependent manner [21], and therefore represents a negative control. As for TIM, as sample concentrations were increased larger assemblies containing multiple BUs were detected by MS. Fig. 5 compares the experimental oligomer abundances of PK, ADH and GDH for concentrations of 1.56, 6.25 and 25 μ M, to the simulated intensities of all the oligomers obtained using the range of ξ values (24 ± 3) determined from the TIM training set.

For PK and ADH the range of predicted non-specific associations obtained from DrOP matches well the observed oligomeric abundances extracted from mass spectra. This is remarkable given that ADH and PK are respectively approximately three and four times the size of TIM, and possess different structural and surface properties. Moreover, for our negative control, GDH, the predictions match the experimental abundances very poorly. Therefore it seems that, despite these significant differences between the proteins, we are able to account for the higher order non-specific oligomers by simply considering droplet occupancy.

The correspondence between experimental PK, ADH data and DrOP-derived intensities, and lack thereof for GDH, indicates the generic nature of this approach and its usefulness in identifying non-specific oligomers. The core advantage of our method is its simplicity: it requires only the concentration of protein sample in question, and an approximate radius for the biological unit. Nano-ES MS is an increasingly popular element of the structural biologist's tool kit when investigating molecular machines, but these fragile assemblies often have very particular buffer requirements and in order to study them by MS considerable time will often be invested in removing salts and contaminants. One of the strengths of the DrOP approach is that additional reference species, which may not be appropriate when investigating delicate and complex biological assemblies, are not required. Moreover, experiments need only be optimized for studying the complex of interest rather than to assess higher order oligomer specificity. Additionally, scarcity of sample may often restrict the running of extensive experiments solely to identify if larger oligomers are the result of non-specific associations or reflective of actual solution populations. The straightforwardness of the DrOP approach, which relies on computational analysis rather than additional experimentation, allows us to quickly discriminate between oligomers identified in MS experiments as being specific or non-specific.

4.5. The propensity of the amyloid-associated protein transthyretin to form oligomers

To explore the utility of our method for distinguishing specific versus non-specific oligomers we chose a protein complex known to aggregate both *in vitro* and *in vivo*. Such systems are among the most difficult to study by MS due to their propensity to self-associate in solution. An IM-MS spectrum of a polydisperse sample of WT TTR with a tetramer concentration of 63 μ M is shown in Fig. 6. By separating the oligomers by m/z and drift time in a nitrogen filled T-wave cell [51] we can distinguish oligomers for the basic biological unit which is a tetramer (in the 3800–4500 m/z range) up to an icosamer (in the 8000–8600 m/z range). An increase in drift time was observed for oligomers of increased mass from just over 6 ms, for the highest charge state of the WT TTR tetramer, to about 14.5 ms for the lowest charge state of the WT TTR icosamer. This increase

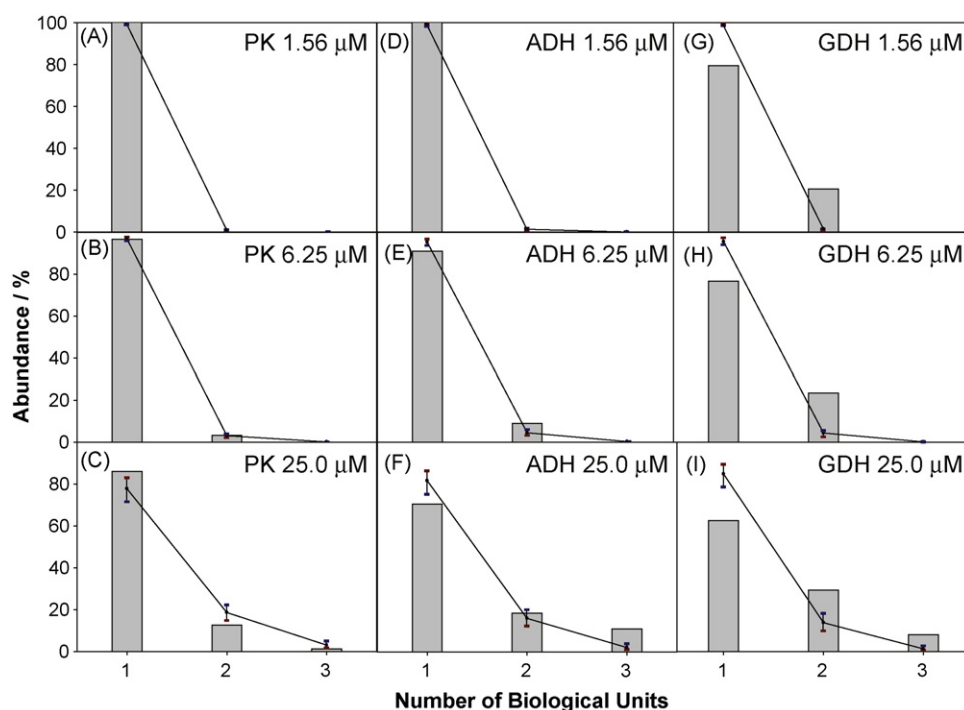


Fig. 5. The experimental abundances (bars) for PK (A–C), ADH (D–F), and GDH (G–I), are overlaid with those predicted by using DrOP for biological unit concentrations: 1.56 μ M (A, D, and G); 6.25 μ M (B, E, and H) and 25.0 μ M (C, F, and I). A value for ξ of 24 ± 3 , identified from TIM experiments, was used to simulate oligomeric intensities. $\xi = 24$ is plotted as a solid black line and the extremes of this range are shown in the figure as error bars; $\xi = 21$ is indicated by the red error bar and $\xi = 27$ is indicated by the blue error bar. Considering the differences between TIM, PK and ADH, both in size and structure, the range of ξ values identified from the TIM analysis describes well the experimental higher order oligomer intensities of PK and ADH. Additionally, GDH, which is known to self-associate in solution in a concentration-dependent manner, shows experimental oligomer intensities significantly above the level predicted by the simulation. These results provide a stringent validation of the range for ξ determined from the TIM measurements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

in drift time for the larger oligomers is related to an increase in the collision cross-section of the oligomer which can stem from both the size and shape of the protein complex, and allows improved deconvolution of the different oligomers [52]. As with other protein

complexes examined in this study, high sample concentrations give rise to the detection of multiple oligomeric forms.

The value for ξ of 24 (as identified from the TIM analysis and validated by PK, ADH and GDH results), along with an R_S for the TTR tetramer of 2.50 nm, was used to simulate the oligomers arising purely as a result of non-specific associations. In fact we had chosen TIM as the ideal 'calibrant' protein for TTR as they are very similar in size. Fig. 7A shows one raw mass spectrum obtained for a 3.13 μ M concentration of WT TTR. Three oligomeric states were detected (tetramer, octamer and dodecamer) and the heavily solvated peaks arise due to our desire to avoid excessive collisional activation [53], which may have a significant impact on the relative intensities of the different oligomers. Fig. 7B shows a histogram of all oligomer intensities, averaged from three experiments for WT TTR tetramer concentrations from 0.78 to 50 μ M. The plot shows a reduction in the intensity of the TTR tetramer with increasing concentration as the larger oligomers become more populated. Overlaid is a surface depicting the simulated oligomer intensities predicted for each concentration using a value of $\xi = 24$. The surface therefore represents the intensity of higher order oligomers which could be explained as a result of non-specific clustering of protein complexes in nano-ES progeny droplets. The fractions of the bars which protrude above the surface represent higher order oligomers formed as a result of specific solution phase affinities. In fact even if we apply the entire range of ξ values observed to fit any concentration of TIM, we find that DrOP predicts lower abundances for all higher order oligomers of WT TTR than obtained experimentally by MS (data not shown). This difference between the *in silico* and *in vacuo* abundances therefore represents the contribution by specific solution-phase interactions in the formation of these larger oligomers of WT TTR.

As with all proteins and protein complexes at high sample concentrations some contribution from non-specific association are

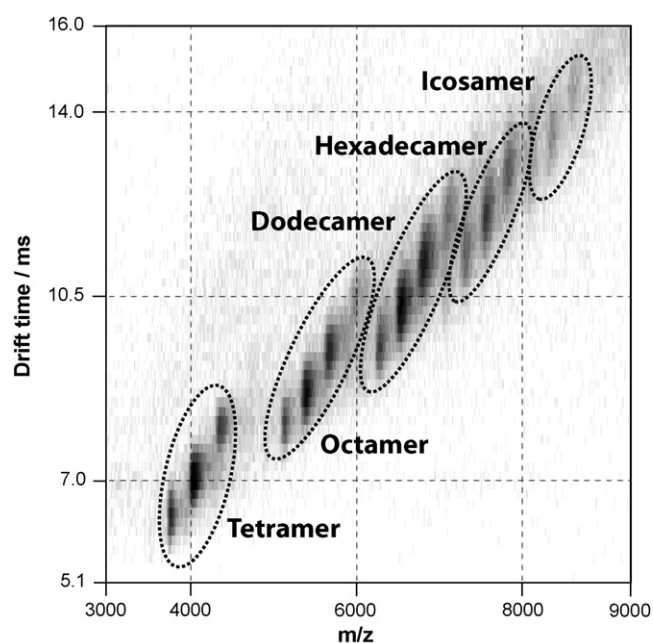


Fig. 6. Ion mobility mass spectrum of WT TTR. Five TTR oligomers: tetramer, octamer, dodecamer, hexadecamer, and icosamer, are separated by their ability to traverse a region of nitrogen gas under the influence of small DC pulses as well as by m/z . The drift-time dimension improves the deconvolution of each oligomer from the polydisperse ensemble.

likely to occur. In the case of WT TTR however it appears that higher order oligomers are formed from a composite effect of both non-specific associations arising in the nano-ES process and an inherent ability of the TTR tetramer to associate as a result of specific affinity between basic tetrameric units. Moreover, the WT octamer and dodecamer that we have observed previously [14,15] are to a large extent formed from a specific interaction between tetramers.

These findings reveal the ability of WT TTR to associate specifically and form large higher order oligomeric species in solution. Moreover, as the concentration of TTR in the body is $\approx 3.6\text{--}7.2\text{ }\mu\text{M}$ [54], and we observe such species at this concentration, it appears likely that this self-association also occurs *in vivo*. Furthermore, the WT tetramer compared to other more amyloidogenic forms of TTR is considered to be more stable [55]. This propensity for the WT TTR to form higher order structures in solution may contribute in some way to this stabilization.

4.6. Associations in a disease-associated variant of transthyretin

The L55P variant causes the most devastating form of early onset TTR diseases. A range of pre-fibrillar structures from single subunits up to multi-subunit systems with masses of 100 kDa are proposed to be the cytotoxic units for the amyloidogenic variant valine-30-methionine [36]. Given the importance of oligomers in TTR pathology we were motivated to consider the ability of L55P TTR tetramers to associate *in vitro*. Fig. 7C shows a mass spectrum

of a $3.13\text{ }\mu\text{M}$ concentration of L55P TTR. Similarly to Fig. 7A, and for the same reason, peaks are highly solvated. However, there is a striking lack of higher order oligomers for L55P, compared to the WT, with only one resolved charge state series which corresponds to an L55P TTR tetramer. The histogram in Fig. 7D displays all the oligomer abundances for a wide range of concentrations of L55P TTR determined by MS, and the red surface represents the oligomeric intensities calculated for $\xi = 24$ using DrOP. The experimental abundances deviate only marginally from the theoretical ones, indicating that the populations of higher order oligomers observed for L55P TTR can be attributed to the nano-ES process. Comparison with the equivalent results for WT TTR (Fig. 7B) show striking differences. The largest L55P oligomer formed at the highest protein concentration analysed ($50\text{ }\mu\text{M}$) was a dodecamer compared to the tetracosamer of the WT. Similarly, at the lowest concentration ($0.78\text{ }\mu\text{M}$) only a tetramer is observed for the L55P form whereas a small amount of octameric WT is observed at the equivalent concentration. Additionally, the intensity of the L55P tetramer dominates each spectrum at all concentrations, and a sharp decrease in the abundance of higher order oligomers is seen in all cases. Despite very similar MS conditions, the abundances of the higher order oligomers of each TTR form are considerably different, with the L55P form showing a greatly reduced population relative to the WT. This indicates, at least for the samples examined here, a reduced propensity of L55P TTR to oligomerize compared to WT TTR.

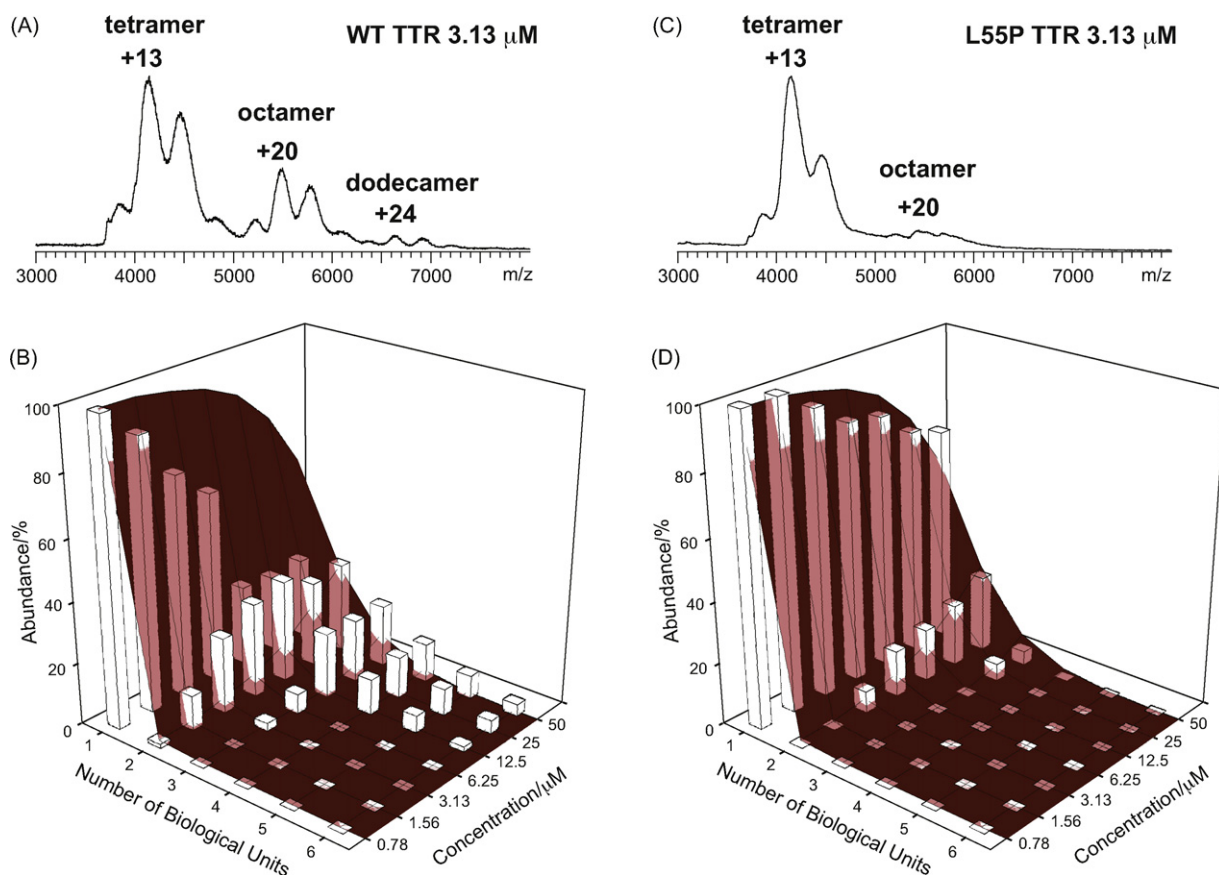


Fig. 7. Mass spectra, obtained under low activation MS conditions, are displayed for $3.13\text{ }\mu\text{M}$ WT (A) and L55P (C) TTR. Despite equal sample concentrations different levels of oligomerization for the WT and L55P are immediately noticeable. Tetramer, octamer and dodecamer forms of the WT TTR are easily identifiable at this concentration; whereas the L55P TTR displays high abundance of tetramer but only a very low intensity octamer. For each concentration of TTR three mass spectra were acquired and the relative intensities of each oligomer were averaged. The bar charts show these averaged TTR oligomer intensities, as determined by nano-ES MS, over a range of concentrations for WT (B) and L55P (D). The red surface on both graphs represents the simulated oligomer intensity with a ξ of 24. Graph B clearly shows specific oligomerization of WT TTR, with higher order oligomers observed at abundances larger than predicted for non-specific assemblies. Conversely, the mutant L55P oligomerizes to a much lesser extent (graph D). The disparity between the ability of WT and L55P to oligomerize may have implications on the mechanism by which TTR forms amyloid. Moreover, the higher order structures formed by the WT TTR may play an essential stabilizing role in protecting this species from forming toxic intermediate species and amyloid fibrils.

The attenuated level of oligomerization for L55P TTR compared to WT may indicate that these higher order oligomers in WT TTR play a crucial role in protection from pathogenicity and amyloidogenicity. Due to the lower toxicity of WT TTR compared to the L55P mutant form, and the ability of the WT to form larger oligomers in solution, it is a possibility that the association of WT tetramers in solution provides a form of protection against the formation of amyloid and toxic intermediates. This is in agreement with previous *in vitro* studies which suggest that TTR amyloid formation occurs via a mechanism which involves tetramer disassembly and partial unfolding of the resulting monomers which initializes the formation of pre-fibrillar species and amyloid [31,32].

5. Conclusions

This new method combines *in vacuo*, *in vitro*, and *in silico* approaches to aid our understanding and analysis of higher order protein oligomers, enabling us to reflect on the susceptibility of an amyloid protein to form higher order species. We have shown that, while there is some non-specific contribution to WT TTR oligomer populations, there is also a very significant contribution made by interactions which occur in solution. This specific association of the WT TTR tetramer *in vitro* perhaps reflects its ability to aggregate *in vivo*, revealing the complex nature of subunit interactions in solution and difficulties in defining the early stages of TTR amyloid fibril formation. Furthermore, we suggest, the propensity of WT TTR to form higher order oligomers, and the inability of the more virulent L55P to do so, may guard the WT TTR against the formation of amyloid and the smaller cytotoxic intermediates which are created *en route* to amyloid deposition. With the mechanism of TTR amyloid formation remaining the subject of much research [56], the biological role of these higher order oligomers, and how their abundances vary in different variants, remains a focus of research in our laboratory.

More generally we have presented here a new method which enables MS to become a more powerful approach for the analysis of soluble protein aggregates. This allows us to determine in a straightforward way the specificity of protein oligomers, facilitating accurate interpretation of oligomer populations in solution. Fundamentally, this coarse-grained approach of DrOP enables us to identify the origins of protein complex interactions as either: specific and present in solution, or as deriving from the nano-ES process through clustering in offspring droplets. The capability to distinguish between these two scenarios will improve our ability to interpret the complicated mass spectra which ensue for many polydisperse [17], and dynamic [57], protein ensembles. Such species are among the most challenging systems for structural biology. Aided by computational accessories such as DrOP, mass spectrometrists are well placed to provide unique insights into their architecture and associated dynamics.

Acknowledgments

We thank Sukjoon Hyung for generously purifying the TTR samples, and the entire Robinson group for helpful discussions. LAL is funded by an EPSRC/RSC Analytical Chemistry Trust Fund studentship; BTR is a Waters Research Fellow, CVR a Royal Society Professor, GF an EPSRC Life Science Interface Fellow, and JLPB a Royal Society University Research Fellow.

References

- [1] J.L.P. Benesch, B.T. Ruotolo, D.A. Simmons, C.V. Robinson, *Chem. Rev.* 107 (2007) 3544.

- [2] T. Wyttenbach, M.T. Bowers, *Annu. Rev. Phys. Chem.* 58 (2007) 511.
- [3] J.A. Loo, *Int. J. Mass Spectrom.* 200 (2000) 175.
- [4] A.E. Ashcroft, *Nat. Prod. Rep.* 22 (2005) 452.
- [5] A.J.R. Heck, R.H.H. van den Heuvel, *Mass Spectrom. Rev.* 23 (2004) 368.
- [6] N. Carulla, G.L. Caddy, D.R. Hall, J. Zurdo, M. Gairi, M. Feliz, E. Giral, C.V. Robinson, *C.M. Dobson, Nature* 436 (2005) 554.
- [7] A.M. Smith, T.R. Jahn, A.E. Ashcroft, S.E. Radford, *J. Mol. Biol.* 364 (2006) 9.
- [8] S.L. Myers, N.H. Thomson, S.E. Radford, A.E. Ashcroft, *Rapid Commun. Mass Spectrom.* 20 (2006) 1628.
- [9] M. Grabenauer, S.L. Bernstein, J.C. Lee, T. Wyttenbach, N.F. Dupuis, H.B. Gray, J.R. Winkler, M.T. Bowers, *J. Phys. Chem. B* 112 (2008) 11147.
- [10] Y.M. Xie, J. Zhang, S. Yin, J.A. Loo, *J. Am. Chem. Soc.* 128 (2006) 14432.
- [11] S.L. Bernstein, T. Wyttenbach, A. Baumketner, J.E. Shea, G. Bitan, D.B. Teplow, M.T. Bowers, *J. Am. Chem. Soc.* 127 (2005) 2075.
- [12] I. Kheterpal, M. Chen, K.D. Cook, R. Wetzel, *J. Mol. Biol.* 361 (2006) 785.
- [13] I. Kheterpal, K.D. Cook, R. Wetzel, *Methods Enzymol.* 413 (2006) 140.
- [14] C.A. Keetch, E.H. Bromley, M.G. McCammon, N. Wang, J. Christodoulou, C.V. Robinson, *J. Biol. Chem.* 280 (2005) 41667.
- [15] M.G. McCammon, D.J. Scott, C.A. Keetch, L.H. Greene, H.E. Purkey, H.M. Petrassi, J.W. Kelly, C.V. Robinson, *Structure* 10 (2002) 851.
- [16] C.C.F. Blake, M.J. Geisow, S.J. Oatley, B. Rérat, C. Rérat, *J. Mol. Biol.* 121 (1978) 339.
- [17] J.A. Aquilina, J.L.P. Benesch, O.A. Bateman, C. Slingsby, C.V. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 10611.
- [18] M.S. Wilm, M. Mann, *Int. J. Mass Spectrom. Ion Process.* 136 (1994) 167.
- [19] M. Dole, L.L. Mack, R.L. Hines, *J. Chem. Phys.* 49 (1968) 2240.
- [20] J.F. de la Mora, *Anal. Chim. Acta* 406 (2000) 93.
- [21] A.D. McCarthy, P. Johnson, K.F. Tipton, *Biochem. J.* 199 (1981) 235.
- [22] J. Sun, E.N. Kitova, W. Wang, J.S. Klassen, *Anal. Chem.* 78 (2006) 3010.
- [23] J. Sun, E.N. Kitova, N. Sun, J.S. Klassen, *Anal. Chem.* 79 (2007) 8301.
- [24] B.M. Hossain, D.A. Simmons, L. Konermann, *Can. J. Chem.* 83 (2005) 1953.
- [25] R.R.L. Smith, G.M. Hutchins, G.W. Moore, R.L. Humphrey, *Am. J. Med.* 66 (1979) 96.
- [26] K. Sletten, P. Westermark, J.B. Natvig, *Scand. J. Immunol.* 12 (1980) 503.
- [27] G.G. Cornwell, K. Sletten, B. Johansson, P. Westermark, *Biochem. Biophys. Res. Commun.* 154 (1988) 648.
- [28] P. Westermark, K. Sletten, B. Johansson, G.G. Cornwell 3rd, *Proc. Natl. Acad. Sci. U.S.A.* 87 (1990) 2843.
- [29] A. Coimbra, C. Andrade, *Brain* 94 (1971) 207.
- [30] D.R. Jacobson, D.E. McFarlin, I. Kane, J.N. Buxbaum, *Human Genet.* 89 (1992) 353.
- [31] A. Quintas, M.J. Saraiva, R.M. Brito, *J. Biol. Chem.* 274 (1999) 32943.
- [32] J.W. Kelly, *Curr. Opin. Struct. Biol.* 8 (1998) 101.
- [33] F. Chiti, C.M. Dobson, *Annu. Rev. Biochem.* 75 (2006) 333.
- [34] P.T. Lansbury Jr., *Proc. Natl. Acad. Sci. U.S.A.* 96 (1999) 3342.
- [35] M.M. Sousa, I. Cardoso, R. Fernandes, A. Guimaraes, M.J. Saraiva, *Am. J. Pathol.* 159 (2001) 1993.
- [36] N. Reixach, S. Deechongkit, X. Jiang, J.W. Kelly, J.N. Buxbaum, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 2817.
- [37] H.A. Lashuel, Z. Lai, J.W. Kelly, *Biochemistry* 37 (1998) 17851.
- [38] G.J. Miroy, Z. Lai, H.A. Lashuel, S.A. Peterson, C. Strang, J.W. Kelly, *Proc. Natl. Acad. Sci. U.S.A.* 93 (1996) 15051.
- [39] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, *Protein Sci.* 4 (1995) 2411.
- [40] H. Hernández, C.V. Robinson, *Nat. Protocols* 2 (2007) 715.
- [41] L. Rayleigh, *Lond. Philosophical Mag.* 5th Series 14 (1882) 184.
- [42] A. Gomez, K. Tang, *Phys. Fluids* 6 (1994) 404.
- [43] P. Kebarle, L. Tang, *Anal. Chem.* 65 (1993) 972.
- [44] J.L.P. Benesch, C.V. Robinson, *Curr. Opin. Struct. Biol.* 16 (2006) 245.
- [45] M. Wilm, M. Mann, *Anal. Chem.* 68 (1996) 1.
- [46] X. Chen, M.S. Westphall, L.M. Smith, *Anal. Chem.* 75 (2003) 5944.
- [47] J.L. Sterner, M.V. Johnston, G.R. Nicol, D.P. Ridge, *J. Mass Spectrom.* 35 (2000) 385.
- [48] R. Aparicio, S.T. Ferreira, I. Polikarpov, *J. Mol. Biol.* 334 (2003) 1023.
- [49] K.A. Markossian, N.V. Golub, H.A. Khanova, D.I. Levitsky, N.B. Poliansky, K.O. Muranov, B.I. Kurganov, *Biochim. Biophys. Acta* 1784 (2008) 1286.
- [50] J.O. Wooll, R.H. Friesen, M.A. White, S.J. Watowich, R.O. Fox, J.C. Lee, E.W. Czerwinski, *J. Mol. Biol.* 312 (2001) 525.
- [51] S.D. Pringle, K. Giles, J.L. Wildgoose, J.P. Williams, S.E. Slade, K. Thalassinou, R.H. Bateman, M.T. Bowers, J.H. Scrivens, *Int. J. Mass Spectrom.* 261 (2007) 1.
- [52] B.T. Ruotolo, J.L.P. Benesch, A.M. Sandercock, S.J. Hyung, C.V. Robinson, *Nat. Protocols* 3 (2008) 1139.
- [53] J.L.P. Benesch, *J. Am. Soc. Mass Spectrom.* 20 (2009) 341.
- [54] Y. Sekijima, J.W. Kelly, S. Ikeda, *Curr. Pharm. Des.* 14 (2008) 3219.
- [55] L.C. Serpell, G. Goldsteins, I. Dacklin, E. Lundgren, C.C.F. Blake, *Amyloid* 3 (1996) 75.
- [56] F. Chiti, C.M. Dobson, *Nature Chem. Biol.* 5 (2009) 15.
- [57] A.J. Painter, N. Jaya, E. Basha, E. Vierling, C.V. Robinson, J.L.P. Benesch, *Chem. Biol.* 15 (2008) 246.