# A Hydrodynamic Comparison of Solution and Gas Phase Proteins and Their Complexes

**ARTICLE** *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2014

Impact Factor: 3.3 · DOI: 10.1021/jp501950d · Source: PubMed



**5 AUTHORS**, INCLUDING:

Erik Marklund
University of Oxford
**17** PUBLICATIONS **387** CITATIONS
SEE PROFILE

David Jan Scott
University of Nottingham
**93** PUBLICATIONS **1,585** CITATIONS
SEE PROFILE

Carol V Robinson
University of Oxford
**448** PUBLICATIONS **22,859** CITATIONS
SEE PROFILE

Antoni Borysik
King's College London
**13** PUBLICATIONS **226** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

# A Hydrodynamic Comparison of Solution and Gas Phase Proteins and Their Complexes

Dominic Hewitt,[†] Erik Marklund,[‡] David J. Scott,[§] Carol V. Robinson,*[,†] and Antoni J. Borysik*[,†]

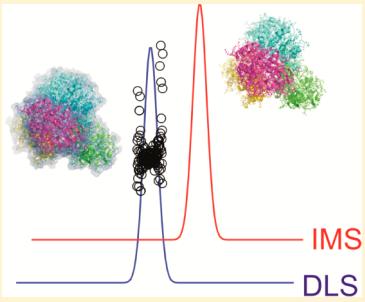
[†]Chemistry Research Laboratory, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3TA, United Kingdom

[‡]Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3QZ, United Kingdom

[§]National Centre for Macromolecular Hydrodynamics, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, United Kingdom

Supporting Information

**ABSTRACT:** The extent to which protein structures are preserved on transfer from solution to gas phase is a central question for native mass spectrometry. Here we compare the collision cross sections ($\Omega$) of a wide range of different proteins and protein complexes (15−500 kDa) with their corresponding Stokes radii ($R_S$). Using these methods, we find that $\Omega$ and $R_S$ are well correlated, implying overall preservation of protein structure in the gas phase. Accounting for protein hydration, a scaling term is required to bring $\Omega$ and $R_S$ into parity. Interestingly, the magnitude of this scaling term agrees almost entirely with the drag factor proposed by Millikan. $R_S$ were then compared with various different predicted values of $\Omega$ taken from their atomic coordinates. We find that many of the approaches used to obtained $\Omega$ from atomic coordinates miscalculate the physical sizes of the proteins in solution by as much as 20%. Rescaling of $\Omega$ estimated from atomic coordinates may therefore seem appropriate as a general method to bring theoretical values in line with those observed in solution.




## INTRODUCTION

Gas phase sizing of proteins and protein complexes by ion mobility mass spectrometry (IMS) has been gaining in popularity particularly since the recent commercial availability of travelleling wave (TWIMS) instrumentation.[1−3] Compared with other low resolution methods such as small angle X-ray scattering (SAXS), analytical ultracentrifugation (AUC), or dynamic light scattering (DLS), IMS has particular advantages in rapid speed of analysis and low sample requirements. Moreover, the coupling of IMS to electrospray ionization mass spectrometry (ESIMS) with further separation of proteins based on their mass-to-charge ratios ($m/z$) has given IMS an unparalleled ability to investigate the shapes of individual components within highly complex interacting systems. Despite these advantages, the obvious questions relating to IMS concern the effects of desolvation and ionization on protein structure. Though many desolvated proteins and protein complexes remain stable in the gas phase, for time scales significantly longer than the standard analysis time frame, the apparent gas phase stability of protein ions is largely charge state and system dependent.[4,5] A more complete picture of how gas phase protein structure corresponds to protein structure in solution is required.

IMS permits the determination of ion collision cross sections ($\Omega$), which are obtained by measuring their drift times in a neutral gas under the influence of a weak electric field, where ion velocity ($v$) and field strength ($E$) are related through the mobility constant $K$ (eq 1).

$$v = KE \tag{1}$$

$K$ is proportional to ion charge ($z$) and inversely related to temperature ($T$), pressure ($N$), and $\Omega$ through the Mason−Schamp equation, wherein $\mu$ is the reduced ion/neutral mass (eq 2).

$$K = \frac{3z}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{1}{\Omega} \tag{2}$$

More compact ions have smaller $\Omega$ than their more elongated counterparts. Conventional approaches to compare protein structures in the gas phase and solution rely on the availability of high resolution structures or models.[6] These structures can then be used to predict the $\Omega$ of proteins allowing the native solution conformations to be compared with experiment. However, the $\Omega$ of an ion also depends on its degree of interaction with the buffer gas. The magnitude of this drag factor ($\xi$) increases in line with the number of inelastic buffer-ion collisions to approach a limiting value of ∼1.39. This is close to the value of $\xi$ = 1.36 that was obtained empirically by Millikan. According to Epstein, $\xi$ = 1.36 relates to the fraction of neutrals (0.91) that interact inelastically with the ion.[7−9] Alternatively, Millikan's term could relate to the distribution of different collision angles somewhere between the specular and

**Received:** February 25, 2014
**Revised:** June 19, 2014
**Published:** June 19, 2014

   dx.doi.org/10.1021/jp501950d | J. Phys. Chem. B 2014, 118, 8489−8495

diffuse models that result in $\xi = 1.36$, as argued by Larriba and co-workers.[10,11]

Different approaches to predict the magnitude of ion/neutral interactions have led to a plethora of methods to compute Ω. The most straightforward and readily implemented approach does not consider ion/neutral interactions and simply determines Ω from the rotationally averaged projected area of a structure. This projection approximation ($\Omega_{PA}$) method generally underestimates Ω but due to its speed and ease of application is popular when Ω are required for protein complexes; an empirically derived scaling factor (PA × 1.14) is typically used to bring experiment and theory into parity.[12,13] More sophisticated methods, such as the exact hard sphere scattering ($\Omega_{EHSS}$) method and the trajectory method ($\Omega_{TM}$), consider ion neutral interactions in different ways. The TM adjusts ion size based on a complex set of ion/neutral interaction potentials.[14] While considered the gold standard for theoretical predictions of Ω, the high computational demands of the TM mean that it is not readily applicable for use with larger protein structures. The EHSS can circumvent the computational demands of the TM, by using a hard sphere approximation of ion/neutral interactions; however, $\Omega_{EHSS}$ generally overestimates the experimental Ω of proteins by ∼15%.[15] In a recent approach, Ω is obtained by approximating the ion/neutral potential and modifying this term by a shape factor that broadly encapsulates the effects of molecular surface on momentum transfer. The broad aim of this projection superapproximation ($\Omega_{PSA}$) method is to permit the generation of TM-based Ω without the time and computational limitations imposed by these more rigorous calculations.[16] There is also a range of various other less common but equally useful approaches to compute Ω such as the DHSS.[10] How to understand the relationship between the gas phase and solution structures of proteins is compounded by the abundance of different methods to calculate Ω and the large (∼30%) variation in Ω depending on the method used. With these caveats in mind, there is a surprising lack of direct biophysical comparisons of protein size in the gas phase and in solution.

Here we perform a direct comparison between the gas phase and solution sizes of a broad range of proteins and their complexes with masses ranging from 15−500 kDa. Protein Ω obtained from atomic structures using a range of different methods are also compared. Experimental protein Ω were obtained from literature values or measured directly with an in-house RF-confining linear field IMS in helium.[17] The solution sizes of proteins were obtained from their Stokes radii ($R_S$) using DLS coupled to size exclusion chromatography (SEC). DLS permitted the determination of the translational diffusion coefficients ($D^\circ$) of the proteins in solution. From this, their $R_S$ were obtained according to the Stokes−Einstein relationship, where $k_B$ is the Boltzmann constant, $T$ is equal to 293.25 K, and $\eta$ is the viscosity of water at 20 °C (eq 3).

$$R_S = \frac{k_B T}{6\pi\eta D^\circ_{20,w}} \tag{3}$$

$R_S$ describes the radius of a sphere with the same translational properties as the protein under investigation and is one of the most commonly reported direct measures of protein size and shape in solution. However, for molecules >5 kDa, stick boundary conditions are operative and the $R_S$ also contains a hydration contribution ($\delta$) from solvation layers. Consideration of this hydration layer is important given the lack of bulk solvent in the gas phase. While there are some uncertainties regarding the magnitude of protein hydration, here we use the most widely accepted value of $\delta$ = 0.3 g of $H_2O$/g of protein.[18,19] Different shape responses between IMS and DLS were also considered. This was explored by comparing the translational Perrin friction factors of ellipsoids with well-known size/surface area relationships of convex bodies.[20,21]

## EXPERIMENTAL METHODS

**Solution Sizing of Proteins and Protein Complexes Using SEC/DLS.** Dry powders of the various proteins were dissolved in 200 mM ammonium acetate (2−10 mg/mL) and syringe filtered using 4 mm 0.22 μm Millex filters (Millipore, U.K.). To reduce self-association processes and remove any unfiltered protein aggregates, all DLS measurements were performed online using a Dawn Heleos II modified with an inboard quasielastic light scatterer (Wyatt Tech., CA, USA). Each protein was eluted in turn down a SuperDex 75 10/300 size exclusion column pre-equilibrated in 200 mM ammonium acetate at 4 °C at a flow rate of 0.3 mL/min. The UV absorbance, diffractive index, and diffusion coefficient were obtained as a function of time. Protein diffusion coefficients were corrected for buffer viscosity and temperature by making use of a calibration curve. The mean viscosity of 200 mM ammonium acetate at 5, 15, and 30 °C was obtained. The viscosity of 200 mM ammonium acetate could then be obtained at any temperature using eq 4, where $\eta_{AA,Tm}$ is the viscosity of the 200 mM ammonium acetate at temperature $T$, $\eta_{AA,Tc}$ is the measured viscosity of 200 mM ammonium acetate, $\eta_{water,Tm}$ is the viscosity of water at temperature $T$, and $\eta_{water,Tc}$ is the viscosity of water at the calibration temperature. The viscosity calibration curve for 200 mM ammonium acetate is shown in the Supporting Information (Figure S1).

$$\eta_{AA,Tm} = \frac{\eta_{water,Tm}}{\eta_{water,Tc}} \times \eta_{AA,Tc} \tag{4}$$

**Acquisition of Experimental Ω.** Aside from lysozyme, chymotrypsin, ovalbumin, and aldolase all Ω reported are taken from literature sources and are reported and the mean value across all charge states.[17] The remaining Ω were obtained directly using an in-house RF-confining IMS in helium as reported previously.[17] Briefly, proteins in dry powder were reconstituted in 200 mM ammonium acetate to a concentration of <10 μM. Samples were then filtered using 4 mm 0.22 μm Millex filters (Millipore, U.K.) and desalted using $P\bar{6}$ Micro Bio-Spin columns (Bio-Rad, CA, USA) pre-equilibrated with 200 mM ammonium acetate. Ω were obtained in positive ion mode in helium. Six different drift time ($t_d$) measurements were obtained between 75 and 200 V at 2.5 Torr. The Ω of the ions were obtained from the gradient of $t_d$ against $1/V$ according to eq 5, wherein $L$ is the flight tube length and $t_0$ is the flight independent $t_d$. All reported Ω are the mean values reported across all charge states.

$$t_d = \frac{L}{KE} + t_0 \tag{5}$$

**Theoretical Calculations of Ω.** Theoretical Ω were calculated from protein crystal structures using the following coordinates: cytochrome *c* (1HRC), lysozyme (1HEL), myoglobin (1WLA), chymotrypsin (1EX3), ovalbumin (1OVA), transthyretin (1F41), bovine serum albumin (3V03), alcohol dehydrogenase (2HCY), aldolase (1ADO),


dx.doi.org/10.1021/jp501950d | J. Phys. Chem. B 2014, 118, 8489−8495

pyruvate kinase (1F3W), glutamate dehydrogenase (1NR7), and galactosidase (1DP0). Missing atoms were added using the WHATIF server http://swift.cmbi.ru.nl/servers/html/index.html. Missing residues were built onto the structures using the ModLoop.[22] The correct oligomeric states of each protein were made using UCSF Chimera.[23] $\Omega_{PA}$, $\Omega_{EHSS}$, and $\Omega_{TM}$ calculations were performed on protonated structures using Mobcal.[15] $\Omega_{PSA}$ calculations were performed on the PSA web server at 300 K in helium.[16] Reported errors for the $\Omega_{PA}$, $\Omega_{EHSS}$, and $\Omega_{TM}$ calculations are of ±1 SD of the mean of 16 calculations. The error in $\Omega_{PSA}$ was obtained directly from the PSA report summary.

## RESULTS AND DISCUSSION

**Hydrodynamic Sizing of Proteins and Protein Complexes.** The diffusion coefficients ($D°$) of each protein were obtained by SEC-DLS in 200 mM ammonium acetate (Experimental Methods). $D°$ was determined from the autocorrelation function of the proteins ($g^2(\tau)$) which for a single ridged species follows first order decay as given by eq 6, wherein $D°$ is the translational diffusion coefficient and $q$ is the Bragg wave vector in which $n$ is the refractive index of the medium, $\theta$ is the scattering angle, and $\lambda$ is the wavelength of the incident light (eq 7).[24]

$$g^2(\tau) - 1 = e^{(-D°q^2\tau)} \tag{6}$$

$$q = (4\pi n/\lambda)\sin(\theta/2) \tag{7}$$

Individual $g^2(\tau)$ values were obtained at 2 s intervals to yield >100 independent values for $D°$ during the elution of a single protein or protein complex. An initial filter was then applied manually to remove the peripheral 10% of values obtained due to their poor signal-to-noise ratios. To the remaining distribution of values, outliers were then identified and removed with $D° > \pm 3$ standard deviations from the mean value. No upward curvature was observed in the remaining data, suggesting the absence of any nonideal behavior in the proteins during the measurement of the translational diffusion coefficients. The fit to the $g^2(\tau)$ of each protein was extremely good with randomly distributed low residual values (Figure 1a). An example of protein elution, $D°$ acquisition, and filtering is shown in Figure 1b.

The remaining $D°$ values were then corrected to standard conditions of pure water at 20 °C using eq 8, wherein $D°_{20,w}$ is the corrected diffusion coefficient, $T$ is the experimental temperature, $\eta_{soln,T}$ is the solution viscosity (200 mM ammonium acetate) at the experimental temperature, and $\eta_{20,w}$ is the viscosity of pure water at 20 °C (Experimental Methods and Figure S1, Supporting Information). This procedure allowed the experimental $D°$ of the proteins obtained in ammonium acetate to be compared directly with literature values for other proteins in different buffers for which the comparison was found to be highly favorable (data not shown).

$$D°_{20,w} = D°\left(\frac{293.15}{T}\right)\left(\frac{\eta_{soln,T}}{\eta_{20,w}}\right) \tag{8}$$

The protein Stokes radii ($R_S$) were obtained using the Stokes–Einstein relationship (eq 3). Each experiment was repeated at least two times. The hydrodynamic properties of the various proteins and protein complexes are given in Table 1.

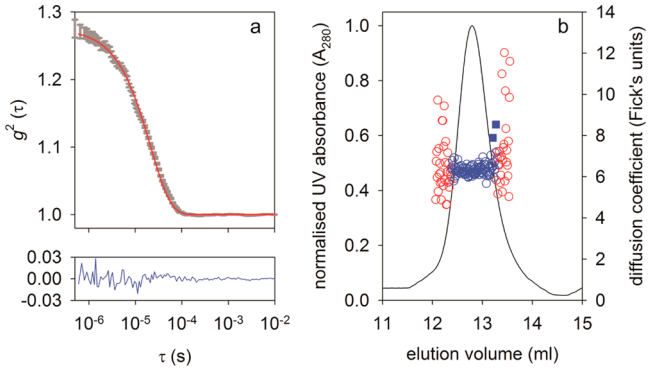



**Figure 1.** SEC-DLS of alcohol dehydrogenase (ADH) tetramer obtained in 200 mM ammonium acetate on a Dawn Heleos II fitted with an online DLS. (a) Autocorrelation data $g^2(\tau)$ (upper panel) is represented as gray error bars to the fit shown in red. The residuals are also shown in blue (lower panel). (b) Peak showing the elution of ADH as measured by $A_{280}$. The scattered data represent the measured diffusion coefficients during elution of the protein. The peripheral 10% of values on either side of the center of the eluting peak were initially removed (red circles). Of the remaining data, outliers were identified with $D° > \pm 3$ standard deviations from the mean value and removed (blue filled squares). $D°$ was obtained from the average value of the remaining data points (blue circles).

**Protein Stokes Radii Comply with Millikan Drag.** The protein $R_S$ were then compared with their equivalent Ω. Experimental Ω were obtained from literature values under native conditions in helium or in house on a linear field RF-confining IMS as described previously (Experimental Methods).[17] The Ω of the proteins correlated well with their equivalent Stokes radii, suggesting that protein structure was broadly preserved on transition from solution to gas phase. We then used this data to gauge the value of any gas phase drag term needed to bring Ω and $R_S$ into parity with each other. As a first approximation, the Ω of a protein can be determined by eq 9, wherein $R_i$ and $R_{He}$ are the respective ion and neutral radii and $\xi$ is a drag enhancement term used to modify the hard sphere Ω to account for inelastic ion/neutral collisions.

$$\Omega = \pi(R_i + R_{He})^2 \times \xi \tag{9}$$

The $R_S$ of a hard sphere is related to its hydrated volume by eq 10, wherein $\delta$, $\bar{v}$, and $\rho$ are the hydration factor (0.3 g/g), partial volume, and solvent density (0.9982 g/mL), $\bar{v}$ were obtained using SEDNTERP,[25] and $V_{hy}$ and $V_{anhy}$ relate to the hydrated and anhydrous protein volumes. It is instructive to clarify that hydrodynamic hydration relates to solvation layers that stick to the surface of a protein and diffuse with the protein, resulting in an apparent swollen volume. These solvation layers must be removed, as gaseous ions do not possess surface solvation layers.

$$\delta = \bar{v}\rho\left(\frac{V_{hy}}{V_{anhy}} - 1\right) \tag{10}$$

Rearrangement of eq 10 provides the relationship between the anhydrous protein radius ($R_{anhy}$) and $R_S$ (eq 11).

$$R_{anhy} = \frac{R_S\bar{v}\rho}{\sqrt[3]{\bar{v}^2\rho^2(\bar{v}\rho + \delta)}} \tag{11}$$


dx.doi.org/10.1021/jp501950d | J. Phys. Chem. B 2014, 118, 8489–8495

**Table 1. Hydrodynamic Properties of the Different Proteins Studied**[a]

| protein | mass (kDa) | $n$ | $D^{\circ}_{20,w}$ | $R_S$ (nm) | Ω (nm$^2$) |
|---|---|---|---|---|---|
| cytochrome *c* | 11.8 | 1 | 13.0 ± 1.6 | 1.62 ± 0.2 | 12.6 ± 0.3 |
| lysozyme | 14.3 | 1 | 11.6 ± 0.4 | 1.85 ± 0.1 | 14.7 ± 0.9 |
| myoglobin | 17.7 | 1 | 11.7 ± 1.1 | 1.83 ± 0.2 | 16.7 ± 0.1 |
| chymotrypsin | 25.7 | 1 | 9.6 ± 0.3 | 2.23 ± 0.1 | 18.7 ± 0.4 |
| ovalbumin | 44.3 | 1 | 7.0 ± 0.3 | 3.06 ± 0.1 | 31.8 ± 0.1 |
| transthyretin | 55.1 | 4 | 6.3 ± 0.1 | 3.38 ± 0.1 | 34.0 ± 0.2 |
| bovine serum albumin | 69.3 | 1 | 6.1 ± 0.6 | 3.50 ± 0.3 | 40.7 ± 0.3 |
| alcohol dehydrogenase | 147.4 | 4 | 4.8 ± 0.2 | 4.48 ± 0.2 | 68.6 ± 1.0 |
| aldolase | 157.5 | 4 | 4.9 ± 0.2 | 4.34 ± 0.3 | 80.1 ± 0.5 |
| pyruvate kinase | 232.6 | 4 | 4.1 ± 0.2 | 5.17 ± 0.2 | 102.2 ± 1.2 |
| glutamate dehydrogenase | 329.6 | 6 | 3.6 ± 0.1 | 6.02 ± 0.2 | 128.0 ± 0.0 |
| galactosidase | 475.6 | 4 | 3.4 ± 0.1 | 6.35 ± 0.1 | 155.2 ± 0.5 |

[a]The protein mass and stoichiometry ($n$) were obtained from previously reported values. $D^{\circ}_{20,w}$ were determined experimentally in 200 mM ammonium acetate (Experimental Methods) and are reported in Fick's units. Protein $R_S$ were obtained from $D^{\circ}_{20,w}$. Ω is the mean experimental collision cross section across all charge states. The reported error values are ±1 SD of the mean.

Thus, assuming $R_i \equiv R_{anhy}$, substitution of eq 11 into eq 9 yields eq 12, which describes the relationship between Ω and $R_S$ for globular folded proteins.

$$\Omega = \pi\left[\frac{R_S\bar{v}\rho}{\sqrt[3]{\bar{v}^2\rho^2(\bar{v}\rho + \delta)}} + R_{He}\right]^2 \times \xi \quad (12)$$

Fitting the experimental relationship between $R_S$ and Ω to eq 12 yields $\xi = 1.40 \pm 0.03$ ($R^2 = 0.99$) (Figure 2). As expected,

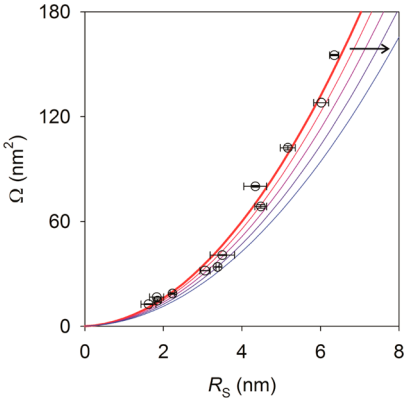



**Figure 2.** Comparison of $R_S$ with Ω for a range of proteins and their complexes. $R_S$ were obtained in 200 mM ammonium by SEC-DLS. Protein Ω were obtained either from literature values or in house with an RF-confining linear field IMS (Experimental Methods and Table 1).[17] Protein Ω are the average values across all charge states. Fitting the relationship between $R_S$ and Ω to eq 12 yields $\xi = 1.40 \pm 0.03$ (red line, $R^2 = 0.99$). Projected relationships between $R_S$ and Ω are also shown for reference with various different values of $\xi$ at −0.1 increments (as shown by the arrow, red to blue).

therefore, these results show that $R_S$ is well correlated with Ω for a range of different proteins and protein complexes with molecular weights in the range of ∼15−500 kDa. Interestingly, after accounting for protein hydration, the $R_S$ of the proteins require additional scaling by $\xi = 1.40$, which is in close agreement with the drag term defined by Millikan for nanoscale droplets.

**Direct Comparison of Protein Size with Theoretical Predictors of Ω.** We then compared the solution sizes of the proteins and protein complexes with various different theoretical predictors of Ω. Protein Ω were obtained from their $R_S$ according to eq 12 assuming $\xi = 1.36$. Theoretical Ω were obtained from atomic coordinates in helium (Experimental Methods). As expected, the solution sizes of the various protein and protein complexes are on average underestimated by $\Omega_{PA}$. More interestingly, protein Ω obtained directly from their $R_S$ are significantly smaller (∼20%) than $\Omega_{TM}$ and $\Omega_{EHSS}$, with a marginal (∼10%) overestimation by the $\Omega_{PSA}$. These large overestimations cannot reflect the use of an inadequate drag factor to reconstitute Ω from protein $R_S$, since $\xi = 1.36$ is in close proximity to the macroscopic limit of these processes where $\xi = 1.39$. Although protein hydration values can vary for proteins ($\delta = 0.2-0.6$ g/g), a value of $\delta = 0.3$ g/g is typical for most proteins and protein complexes. Furthermore, while reducing $\delta$ would increase the Ω of proteins obtained from their $R_S$, the magnitude of this reduction would need approach values less than $\delta = 0.05$ g/g to bring the solution sizes of the proteins into parity with $\Omega_{TM}$ and $\Omega_{EHSS}$. This large reduction in protein hydration is not feasible for large particles (>5 kDa), such as proteins where stick boundary conditions are in operation. These comparisons suggest therefore that many of the theoretical predictors of Ω overestimate the sizes of proteins in solution. Certainly, $\Omega_{TM}$ and $\Omega_{EHSS}$ apparently overestimate solution protein size by ∼20%.

**Protein Shape Effects on Ω and $R_S$.** A plausible explanation for some of the discrepancies seen above could be differences in protein shape dependencies between IMS and DLS. Significantly different shape dependencies between these methods would result in an apparent size difference between the techniques for the same protein. If DLS showed a reduced dependency to protein shape than IMS, this could reconcile the comparative reduction in protein size derived by this method. This is feasible, as hydrodynamic methods based on molecule translation, such as DLS, are notoriously poor measures of protein asymmetry. To make these comparisons, we invoked the common hydrodynamic practice of performing calculations based of two-component ellipsoids.[26] The shape dependence of ellipsoids is known analytically for a range of methods allowing direct comparison. For DLS, size and shape relationships are described analytically by the Perrin translational friction factors, $f(P)$, for two-component ellipsoids.[20] The term $f(P)$ is a radial shape factor ($R_{sf}$) defining the relationship between apparent radius and ellipsoid axial ratio. This relationship is given in eqs 13 and 14 for respective prolate and oblate shapes, wherein $q$ is


dx.doi.org/10.1021/jp501950d | J. Phys. Chem. B 2014, 118, 8489−8495

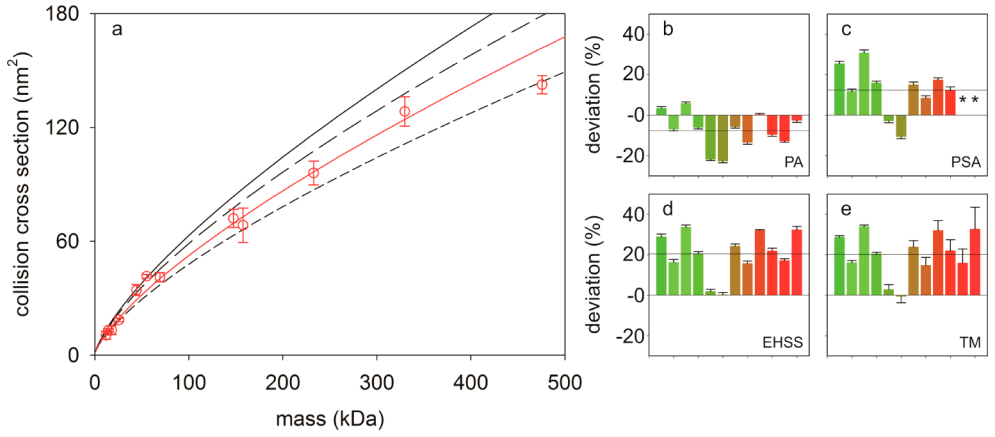



**Figure 3.** (a) Direct comparison of protein size in solution and theoretical predictors of Ω from atomic coordinates. $R_S$ derived Ω were obtained using eq 12 and assuming $\xi = 1.36$ (red circles). Average trends for the PA (short dashed line), PSA (medium dashed line), and EHSS and TM (solid line) are shown ±1 SD. Average trends were obtained from fitting power relationships to the theoretical Ω (Experimental Methods, Figure S2, Supporting Information). Note the EHSS and TM trends superimpose. (b–e) Percent model-by-model deviations in Ω of the PA (b), PSA (c), EHSS (d), and TM (e) from values obtained from $R_S$ are shown. Values for proteins are shown increasing in mass from green to red as in Table 1. The mean deviation across all proteins is also indicated (solid line). Note PSA-derived Ω for glutamate dehydrogenase and galactosidase could not be obtained and are marked as asterisks.

the axial ratio of the semimajor and semiminor ellipsoid axes given by $a/b$ (prolate) and $b/a$ (oblate).

$$f(P) = \frac{\sqrt{1-q^2}}{2^{2/3}\ln\frac{1+\sqrt{1-q^2}}{q}} \quad (13)$$

$$f(P) = \frac{\sqrt{q^2-1}}{q^{2/3}\arctan\sqrt{q^2-1}} \quad (14)$$

Equivalent Ω shape factors ($\Omega_{sf}$) for elliptical size and shape relationships were estimated by well-known relationships between the projected area (PA) and surface area (SA) of convex bodies, as given by eq 15.[21]

$$SA = 4PA \quad (15)$$

Equation 15 permits the derivation of $\Omega_{sf}$ for both prolate and oblate ellipsoids, eqs 16 and 17, respectively.

$$\Omega_{sf} = \frac{2\pi\left(a^2 + \frac{ab\alpha}{\sin(\alpha)}\right)}{4\pi\left(\sqrt[3]{3V/4\pi}\right)^2} \quad (16)$$

$$\Omega_{sf} = \frac{2\pi\left[a^2 + \frac{b^2}{\sin(\alpha)}\ln\left(\frac{1+\sin(\alpha)}{\cos(\alpha)}\right)\right]}{4\pi\left(\sqrt[3]{3V/4\pi}\right)^2} \quad (17)$$

where $a$ and $b$ are the dimensions of the respective semiminor axes for prolate and oblate shapes, $V$ the ellipsoid volume, and $\alpha$ the angular eccentricity; prolate = $\arccos(a/b)$ and oblate = $\arccos(b/a)$. $\Omega_{sf}$ was then rescaled $[(\Omega_{sf})^{1/2}]$ to bring it in line with $R_{sf}$. Comparison of the shape dependencies of both methods shows moderate increases in apparent size with changes in geometry. For comparison, the ellipsoid shape dependencies of SAXS are also included which differ greatly from those of IMS and DLS due to the different underlying principles of this technique. Importantly, the shape dependencies of both IMS and DLS are comparable with each other. Thus, the apparent differences between solution and gas phase protein sizes cannot be reconciled by different shape dependencies between the methods (Figure 4).

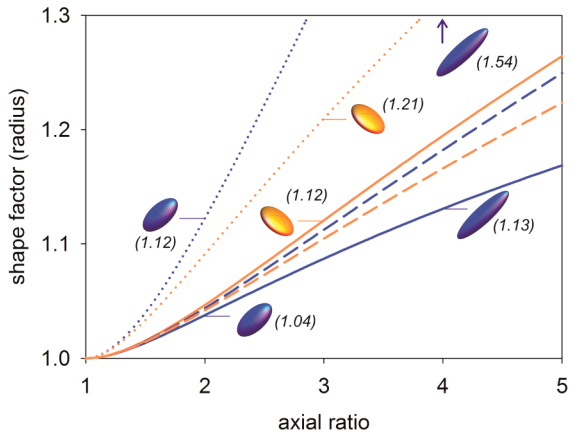



**Figure 4.** Shape dependencies of various low resolution methods for two-component ellipsoids. The shape factors for prolate (blue) and oblate (orange) ellipsoids obtained by IMS (unbroken lines), DLS (dashed lines), and SAXS (dotted lines) are shown. Various different ellipsoids are shown for reference along with their corresponding shape in parentheses for the various different techniques. The theoretical SAXS dependencies are shown for reference.

## CONCLUSION

We have compared protein size in solution and gas phase over a wide mass range. This comparison has been performed using DLS to obtain protein $R_S$ which were then rescaled to account for surface hydration and ion–neutral scattering. To our knowledge, this is one of the first direct comparisons of protein size in solution and in the gas phase. Previous studies that compared the gas phase structures of proteins with their $R_S$ concluded that proteins shrank by ~5–20% in the gas phase.[27] However, these studies did not consider protein hydration or ion/neutral interactions. In contrast, our comparison presented here reveals a broad similarity between experimental protein Ω and their equivalent $R_S$. These data can therefore be added to a


dx.doi.org/10.1021/jp501950d | J. Phys. Chem. B 2014, 118, 8489–8495

growing catalogue of evidence supporting the idea that the structures of globular proteins are largely preserved in the gas phase.

The TM and EHSS methods predict protein cross sections that are on average 20% larger than those derived from $R_S$ dropping to 10% when the PSA is utilized. Deviations in these overall trends are observed for cytochrome *c* and myoglobin which could reflect the larger errors associated with the $R_S$ obtained for these proteins (Table 1). Thus, the $R_S$ of a broad range of different proteins and protein complexes are significantly smaller than predicted by many commonly used theoretical predictors of Ω. Protein $R_S$ were reduced to take protein hydration and surface solvation layers into account that would not survive in the gas phase. The degree of reduction of $R_S$ is consistent with well accepted hydration values of $\delta = 0.3$ g/g.

Typical values for hydrodynamic hydration are within the range 0.2−0.6 g/g with glycosylated protein occupying the upper levels of this range.[28] Hydration estimates based on amino acid composition using the method of Kuntz typically result in values significantly greater than $\delta = 0.3$ g/g.[29] In our experiments, parity between calculated Ω and $R_S$ would require an unfeasibly large reduction in the contribution of protein hydration to $R_S$ of $\delta < 0.05$ g/g. We also adjusted the $R_S$ to consider gas phase scattering effects. This was performed using a value of 1.36 as outlined by Millikan, which is close to the macroscopic limit of these processes. Any possible differences in the shape dependencies of DLS and IMS appear to cancel each other out, particularly for globular proteins with roughly spherical shapes. Thus, in spite of these reasonable and necessary modifications to $R_S$, the sizes of a range of different proteins and protein complexes in solution appear too large when judged by many different theoretical predictors of Ω.

The tendency of the EHSS and TM to overestimate the experimental Ω of proteins has been well documented.[30] Our data suggests that these size discrepancies also occur in solvent when protein sizes are obtained from their $R_S$, which is one of the most commonly used and robust measures of protein size in solution. These size overestimates are significant and are on average 20% over the range of proteins we investigated. Vacuum MD relaxation of atomic coordinates may represent one way to increase the parity between $\Omega_{TM}$, $\Omega_{EHSS}$, and $\Omega_{PSA}$ and experimentally obtained measures of Ω for proteins and protein complexes. However, this would not account for the apparent size differences in solution, as judged by $R_S$. Moreover, possible dependencies on choice of force field, protein charge state, and charge orientation would bring further ambiguities into play, possibly making this approach imprecise for general preprocessing of atomic structures. Thus, broad rescaling approaches may be more suitable for bridging the gap between theoretical and experimental measures of protein size in gas phase and solution. While empirically derived scaling terms lack a theoretical basis, and this needs to be addressed, a single scaling term permits highly correlated theoretical and experimental parity of Ω for globular proteins and protein complexes over a large mass range. Rescaling of $\Omega_{PA}$ would seem most appropriate therefore, given its ease of implementation and greater parity with protein structures in solution as measured by their $R_S$.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Figures showing a viscosity calibration curve for 200 mM ammonium acetate and average trends for the PA, PSA, EHSS, and TM for the proteins used in this study and equations for determining the viscosity of water. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Authors

*E-mail: carol.robinson@chem.ox.ac.uk. Phone: 44 (0) 1865 275473.

*E-mail: antoni.borysik@chem.ox.ac.uk. Phone: 44 (0) 1865 275975.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank Dr Argyris Politis for useful discussions and Dr Stephen Carr for assistance with DLS.

## ■ REFERENCES

(1) Parent, D. .; Bowers, M. T. Temperature-Dependence of Ion Mobilities - Experiment and Theory. *Chem. Phys.* **1981**, *60* (2), 257−275.

(2) Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F. High-Resolution Ion Mobility Measurements. *Rev. Sci. Instrum.* **1997**, *68* (2), 1122−1129.

(3) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun. Mass Spectrom.* **2004**, *18* (20), 2401−2414.

(4) Benesch, J. L.; Robinson, C. V. Biological Chemistry: Dehydrated but Unharmed. *Nature* **2009**, *462* (7273), 576−577.

(5) Schennach, M.; Breuker, K. Proteins With Highly Similar Native Folds Can Show Vastly Dissimilar Folding Behavior When Desolvated. *Angew. Chem., Int. Ed. Engl.* **2014**, *53* (1), 164−168.

(6) Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS One* **2010**, *5* (8), e12080.

(7) Millikan, R. A. The General Law of Fall of a Small Spherical Body Through a Gas, and its Bearing Upon the Nature of Molecular Reflection From Surfaces. *Phys. Rev.* **1923**, *22* (1), 1−23.

(8) Epstein, P. S. On the Resistance Experienced by Spheres in their Motion Through Gases. *Phys. Rev.* **1924**, *23*, 710−733.

(9) Hogan, C. J., Jr.; de la Mora, J. F. Ion Mobility Measurements of Nondenatured 12−150 kDa Proteins and Protein Multimers by Tandem Differential Mobility Analysis-Mass Spectrometry (DMA-MS). *J. Am. Soc. Mass Spectrom.* **2011**, *22* (1), 158−172.

(10) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* **2013**, *251*, 344−363.

(11) Larriba, C.; Hogan, C. J., Jr. Ion Mobilities in Diatomic Gases: Measurement Versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **2013**, *117* (19), 3887−3901.

(12) Vonhelden, G.; Hsu, M. T.; Gotts, N.; Bowers, M. T. Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *J. Phys. Chem.* **1993**, *97* (31), 8182−8192.

(13) Benesch, J. L.; Ruotolo, B. T. Mass Spectrometry: Come of Age for Structural and Dynamical Biology. *Curr. Opin. Struct. Biol.* **2011**, *21* (5), 641−649.

(14) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100* (40), 16082−16086.

(15) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261* (1−2), 86−91.
(16) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (I). Method. *Int. J. Mass Spectrom.* **2011**, *308* (1), 1−10.
(17) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: a Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557−9565.
(18) Squire, P. G.; Himmel, M. E. Hydrodynamics and Protein Hydration. *Arch. Biochem. Biophys.* **1979**, *196* (1), 165−177.
(19) Dam, J.; Schuck, P. Calculating Sedimentation Coefficient Distributions by Direct Modeling of Sedimentation Velocity Concentration Profiles. *Methods Enzymol.* **2004**, *384*, 185−212.
(20) Perrin, F. Mouvement Brownien d'un Ellipsoide (II). Rotation Libre et Dépolarisation des Fluorescences. Translation et Diffusion de Molécules Ellipsoidales. *J. Phys. Radium* **1936**, *7*, 1−11.
(21) Vouk, V. Projected Area of Convex Bodies. *Nature* **1948**, *162* (4113), 330−331.
(22) Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9* (9), 1753−1773.
(23) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612.
(24) Harding, S. E.; Jumel, K. Light Scattering. *Curr. Protoc. Protein Sci.* **2001**, Chapter 7, Unit 7.8.
(25) Hayes, D.; Laue, T.; Philo, J. *Program Sednterp: Sedimentation Interpretation Program*; University of New Hampshire: 1995.
(26) Carrasco, B.; Garcia de la Torre, J. Hydrodynamic Properties of Rigid Particles: Comparison of Different Modeling and Computational Procedures. *Biophys. J.* **1999**, *76* (6), 3044−3057.
(27) Kapellios, E. A.; Karamanou, S.; Sardis, M. F.; Aivaliotis, M.; Economou, A.; Pergantis, S. A. Using Nanoelectrospray Ion Mobility Spectrometry (GEMMA) to Determine the Size and Relative Molecular Mass of Proteins and Protein Assemblies: a Comparison with MALLS and QELS. *Anal. Bioanal. Chem.* **2011**, *399* (7), 2421−2433.
(28) Pessen, H.; Kumosinski, T. F. *Methods in Enzymology*, Vol. 117; Academic Press: New York, 1985.
(29) Kuntz, I. D., Jr Hydration of Macromolecules. III. Hydration of Polypeptides. *J. Am. Chem. Soc.* **1971**, 514−516.
(30) Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. Travelling Wave Ion Mobility Mass Spectrometry Studies of Protein Structure: Biological Significance and Comparison with X-ray Crystallography and Nuclear Magnetic Resonance Spectroscopy Measurements. *Rapid Commun. Mass Sp.* **2008**, *22* (20), 3297−3304.