A new algorithm to characterise the degree of concaveness of a molecular surface relevant in ion mobility spectrometry

Thomas Wyttenbach\textsuperscript{a}, Christian Bleiholder\textsuperscript{b}, Stanley E. Anderson\textsuperscript{a} and Michael T. Bowers\textsuperscript{a,*}

\textsuperscript{a}Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA, USA; \textsuperscript{b}Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL, USA

(Received 21 November 2014; accepted 14 April 2015)

A new algorithm to describe the convex or concave nature of the surface of a given molecular structure was developed and incorporated into the projected superposition approximation (PSA). The new method searches for a reference structure characterised by a minimal surface area fully enclosing the molecule. The ratio of the molecular surface area to the surface area of this reference structure was found to be a good descriptor of the molecular concaveness required for the accurate calculation of theoretical cross sections using PSA. These computations are an essential step in the interpretation of ion mobility data in terms of ion molecular structure.

Keywords: ion mobility; molecular structure; cross section; projected superposition approximation (PSA)

Introduction

Ion mobility spectrometry coupled to mass spectrometry has evolved into a powerful method to obtain structural information of polyatomic ions \cite{1,2,3}. However, accurate interpretation of the ion mobility drift time in terms of molecular structure requires a detailed understanding of how the ion structure affects the ion drift process which is governed by the collisions of the ion with the buffer gas. Empirically, it has been found nearly 100 years ago that charged macroscopic particles, such as oil droplets, travel in an electric field a factor of 1.36 slower than theoretically expected based on the particle size and its spherical shape \cite{4}. This increased resisting force has been attributed to a partially diffuse reflection process occurring in collisions with the buffer gas \cite{4}.

A deeper understanding of diffuse reflections in the oil droplet experiment, the origin of the increased resisting force, and the empirical factor 1.36 is relevant in establishing a rigorous connection between ion mobility data and ion structure. This is particularly important for large polyatomic ions such as protein complexes \cite{5,6} in the size range between small ions governed by Langevin collision dynamics \cite{7} and macroscopic particles (>1 μm). A recent theoretical study indicates that a polyatomic ion with concave surface elements (or surface ‘roughness’) drifts with an increased resisting force compared to a particle corresponding to the fully convex envelope surrounding the ion in question \cite{8}. The magnitude of this effect depends on the ion shape but as a general trend it is observed that for globular proteins the effect increases on average with ion size from the kilodalton to the megadalton range and it might converge to a factor near 1.36 for macroscopic particles (much larger than a megadalton) \cite{8}.

Hence, not only the size of the average projection cross section determines the resisting force of a drifting ion, but also the ion shape, the makeup of the ion surface. A smooth, convex surface causes less resistance than a rough, concave surface. The reason is that the resisting force is coupled to momentum transfer occurring in collisions \cite{9–11} and that buffer gas collisions with a rough surface result on average in larger scattering angles and therefore a larger momentum transfer per collision \cite{8,12}. Therefore it is important for any theoretical model attempting to quantify the resisting force of a drifting polyatomic ion with a given projection cross section to incorporate an accurate and appropriate description of the concaveness of the ion surface.
In earlier theoretical studies [10–14] the resisting force of a drifting ion with a given molecular geometry was calculated by propagating trajectories of ion-buffer gas collisions, evaluating deflection angles, and thereby directly computing the momentum transfer integral numerically. Whereas this is a rigorous approach converging for a large number of trajectories to a cross section value which is naturally scaled by the molecular concavity, it is in practice a very time-consuming calculation for large molecules and for an accurate ion-buffer gas interaction potential.

Therefore a computationally much less demanding different approach, the projected superposition approximation (PSA), was recently introduced [15]. In this model an estimate of the shape effects or surface effects on the resisting force of a drifting ion is an integral part described by the quantity \( \rho \), the PSA shape factor. By definition, \( \rho \) accounts for the difference between the momentum transfer cross section of the real molecule (which always includes regions of concaveness) and that of a hypothetical, ideally convex reference structure of identical size (for which the projectionapproximationisaccurate[12]). Therefore a reason able approach for the computation of \( \rho \) implemented in the original PSA version is defining \( \rho \) as the ratio of the molecular surface, \( A_{\text{mol}} \), to the surface of the convex envelope, \( A_{\text{ce}} \), enclosing the molecular structure [15],

\[
\rho = \frac{A_{\text{mol}}}{A_{\text{ce}}}. \tag{1}
\]

This approach has been successfully applied to many molecular systems including carbon clusters, supramolecular coordination complexes, peptides, and proteins [15–18]. However, while this concept is perfectly reasonable for the systems included in the initial studies [15–18] and for most other commonly encountered molecular structures, we discovered recently that it fails for certain less common geometries. In particular we find that Equation (1) holds well for compact geometries with or without cavities such as globular proteins, but fails for extended structures occurring, for example, for high charge states of proteins under denaturing conditions in electrospray mass spectrometry [19]. In this paper we address the shortcomings of the initial approach and propose a more rigorous concept to evaluate the PSA shape factor.

Methods

The sample structures used in this work are selected for illustration purposes. They include fullerene \( \text{C}_{60} \), a number of additional (hypothetical) carbon clusters assembled and annealed employing the HyperChem [20] software tools, an annealed gas-phase structure of the 76-residue protein ubiquitin [19] in charge state +13, and the 125-residue haemoglobin structure 2KSC (cyanobacterial 2/2 haemoglobin from Synechococcus sp PCC 7002) [21] downloaded from the Protein Data Bank (PDB) database [22]. Drawings of the structures are done with the VMD software version 1.9.1 [23]. Ion mobility cross sections are calculated by the PSA algorithm which has previously been described in detail [15]. PSA parameters for carbon clusters drifting in helium buffer gas and for proteins drifting in helium or nitrogen have previously been tested and validated [15–18,24]. Explicit knowledge of the charge state or charge distribution of the ions is not required for PSA; the typical average effect of charge is rolled into the parameterisation for the type of molecule used. All PSA calculations carried out here employ the PSA program version 2.4 and are based on helium buffer gas at 298 K. For biomolecules this software version is presently publically available via a web interface [25].

In this work, we are mostly interested in the PSA shape factor \( \rho \) which is a measure of the concaveness of the molecular surface. The algorithm employed to evaluate \( \rho \) for a given molecular shape in the present PSA version 2.4 is the subject of this paper and is discussed below in the ‘Results and discussion’ section. For comparison, the results based on Equation (1), the initial PSA approach to compute \( \rho \) (versions 2.3 and older), are also given for selected examples.

Results and discussion

Figure 1 shows a range of carbon fullerene structures. The first structure is fullerene \( \text{C}_{60} \), a nearly perfect convex structure when using appropriate atomic radii relevant in an ion mobility experiment in helium. The PSA result compiled in Table 1 reflects the convex property of the geometry with a computed shape factor \( \rho \) very close to unity. The structure in Figure 1(c) is a concave structure, a cup-shaped carbon cluster (approximately half of a fullerene \( \text{C}_{240} \) cage). The shape factor of \( \rho = 1.09 \) captures the concaveness and indicates a 9% increased resisting force for this ion compared to a fully convex molecule of the same projection size. Whereas the nine fullerene \( \text{C}_{20} \) cages in the bottom structure (Figure 1(d)) are positioned...
in a linear arrangement ($\rho \approx 1.0$), they form an L-shaped geometry in the fourth structure (Figure 1(b)). And even though the concaveness is similar for the two ($C_{20})_9$ structures, the PSA shape factors calculated according to Equation (1) are quite different. In fact, the $\rho$ value of 0.823 for the L-shaped structure is substantially smaller than unity. Since $\rho$ is supposed to reflect the increase in the resisting force of a drifting ion compared to an ideal fully convex object with a smooth surface of identical projection cross section (see Introduction), a value $\rho < 1$ is not physically meaningful. Hence, Equation (1) does not apply to this type of molecule and we have to re-evaluate the rationale behind this equation used in the PSA calculations.

The PSA approach to calculate a theoretical ion mobility cross section $\sigma_{PSA}^{(1,1)}$ for a given model structure is a two-step process [15],

$$
\sigma_{PSA}^{(1,1)} = \rho \times \langle \Omega_{PSA} \rangle.
$$

First, an appropriate orientation-averaged projection cross section $\langle \Omega_{PSA} \rangle$ is evaluated which includes superposition effects caused by the simultaneous interaction of the buffer gas particle with multiple atoms in the polyatomic ion. Second, the surface of the polyatomic ion is analysed for its concave nature. This analysis results in a shape factor $\rho$ which is used for scaling up the projection cross section value obtained in the first step.

Figure 2(a) shows an object with concave surface elements, aspheric with dents in its surface similar to a golf ball. Panel 2(b) displays the same object but with the concave

---

**Figure 1.** Carbon cluster geometries for (a) fullerene $C_{60}$ and several hypothetical structures, (b) L-shaped ($C_{20})_9$, (c) cup-shaped $C_{124}$, and (d) quasi linear ($C_{20})_9$. Also shown in panel (a) is the convex nature of the $C_{60}$ surface sampled in ion mobility experiments in helium (surface based on PSA calculations).
parts covered by planar lids. This results in a perfectly convex body. Shining a light source on both of these objects side by side will cast two shadows onto a plane that look identical. Since the concave parts of the golf ball will never show in the shadow, you will never be able to determine whether the dents are actually present in the real object or not on the basis of the shape of the shadow. Hence, the projection cross section is the same for the two objects in Figure 2, but their resistive force and therefore their momentum transfer cross section are different due to the different surface makeup.

Thus, application of the PSA method to molecules resembling the shape of the two objects in Figure 2 would yield identical projection cross sections for the two molecules but different shape factors. Since \( \rho = 1 \) holds Table 1, PSA shape factor calculated for a range of molecular shapes.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Molecular shape</th>
<th>Equation (1)</th>
<th>Equation (3)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(_{60})</td>
<td>Spherical</td>
<td>0.992</td>
<td>0.993</td>
<td>0.08%</td>
</tr>
<tr>
<td>C(_{124})</td>
<td>Cup-shaped</td>
<td>1.088</td>
<td>1.090</td>
<td>0.2%</td>
</tr>
<tr>
<td>(C(_{20}))(_9)</td>
<td>Quasi-linear</td>
<td>0.993</td>
<td>1.020</td>
<td>3%</td>
</tr>
<tr>
<td>(C(_{20}))(_9)</td>
<td>L-shaped</td>
<td>0.823</td>
<td>1.025</td>
<td>25%</td>
</tr>
<tr>
<td>Haemoglobin(^b)</td>
<td>Globular</td>
<td>1.151</td>
<td>1.199</td>
<td>4%</td>
</tr>
<tr>
<td>Ubiquitin(^c)</td>
<td>Unfolded/partially helical</td>
<td>0.727</td>
<td>1.110</td>
<td>53%</td>
</tr>
</tbody>
</table>

\(^a\)Shape factor \( \rho \) calculated according to Equation (1) (PSA versions 2.3 and older) or Equation (3) where the reference corresponds to the structure with minimal surface but completely surrounding the molecular shape (PSA version 2.4). \(^b\)NMR structure 2KSC of cyanobacterial haemoglobin [21]. \(^c\)Gas-phase structure of ubiquitin charge state +13 [19].

Figure 2. (a) Sphere with several dents in its surface. (b) Same object as in (a) but with the dents filled in; the resulting object is fully convex. Cuts through the centre of the respective objects are given in the bottom panels.

is essentially fully convex. Only on the inside of the bend, in the elbow, is a small surface area with concave curvature. In this case, the object’s convex envelope shown in panel 3(b) does not really capture the minimal concaveness: Applying Equation (1) to this system actually yields a shape factor \( \rho < 1 \), since the convex envelope has such a large surface area. Since \( \rho < 1 \) is physically meaningless, Equation (1) has to be rewritten in a more general form,

\[
\rho = \frac{A_{\text{mol}}}{A_{\text{ref}}} \quad (3)
\]

where \( A_{\text{ref}} \) is the surface area of an appropriate reference structure, not necessarily the convex envelope. In some cases (see example in Figure 2(a)), it might be the convex envelope, in other cases (see example in Figure 3(a)) we have to find a better reference structure.

What is a good reference structure for an L-shaped molecule? By definition, \( \rho \) reflects the increase of the momentum transfer cross section due to shape (Equation (2))
Figure 3. (a) Picture of an L-shaped rod and (b) its convex envelope.

compared to an ideally convex reference with the same projection cross section as the molecule itself. Therefore, the overall size of the reference structure should resemble that of the molecule. Clearly, the convex envelope shown in Figure 3(b) is not a suitable choice as a reference for the object in Figure 3(a) because its projection cross section is much larger than that of the object itself. Hence, a better choice in general would be a structure where all cup-like concaveness is smoothed out, a locally but not necessarily a globally convex structure. This ensures that the projection cross sections of molecule and reference structure are alike. Thus, the question appears to be analogous to this: What is the minimal amount of wrapping paper needed to completely wrap an object? Very often, this will be the convex envelope (Figure 2). However, for the L-shaped rod it is more economical (in terms of paper use) to wrap the paper tightly around the rod than to place it into a convex envelope.

Whereas we are not able to deliver a rigorous proof that the structure with minimal surface (but completely surrounding the object) is the relevant reference in Equation (3), common sense suggests that this structure is certainly an improvement over simply choosing the convex envelope. Maximum momentum transfer in a collision occurs when the deflection angle is 180° [8]. Hence, an object with a surface makeup resembling that of a reflector, an everyday article reflecting light back where it came from independent of incident angle, will have a large shape factor. This indicates that it is indeed cup or cone-like cavities (like in a reflector) that account for an increased resisting force of a drifting object and that saddle-shaped concave curvatures like bends in a rod have much less of an effect.

Molecular Physics

How do we find the envelope with minimal surface area butfullyenclosingthemolecule? The PSA method employs established algorithms described in the literature to characterise the surface of an object [26–29]. In essence, the method is analogous to rolling a ball along the molecular surface. A probing ball with a very large radius will not be able to enter any concave parts of the surface thereby mapping out the convex envelope. A very small probe radius, on the other hand, will probe the true molecular surface. Hence, finding the minimum enclosing surface is equivalent to finding the probe radius which delivers a minimum surface area, the approach taken in the newest version of PSA (version 2.4).

Figure 4 shows how the probed surface area changes as a function of probe radius for three representative molecular shapes: sphere, cup, and ‘L’ (Figure 1(a)–(c)). In all cases, the surface area approaches the molecular surface for a probe radius of zero and it reaches the surface of the convex envelope for very large radii. The asymptotic limit of the convex hull is obtained by the QuickHull algorithm [30]. For quasi-spherical fullerene C60 the surface does not depend on the probe radius and the molecular surface is the same as the convex envelope. For the cup shape,
the surface area decreases monotonically from the value of the molecular surface to that of the convex envelope with increasing probe radius. For the L-shaped molecule, the surface decreases initially with increasing probe radius, reaches a minimum, and subsequently increases and converges to the convex envelope which is larger than the molecular surface area in this case. Using the structure corresponding to the minimum of the curve in Figure 4(c) as a reference to calculate \( \rho \) following Equation (3), yields a physically meaningful value of \( \rho = 1.025 \). This value is just slightly larger than the \( \rho \) value of the structurally closely related linear geometry (Figure 1(d) vs. 1(b)) as expected (Table 1). In these examples and in general we find that the computational (CPU) time required to evaluate \( \rho \) by the new algorithm roughly doubles compared to simply finding the convex envelope (from 300 to 700 s for the (C_{20}) examples).

Since the convex envelope value agrees very closely with the minimum of the curves in Figure 4(a) and 4(b) (sphere and cup), the PSA shape factor based on version 2.4 1sthesameasthathadeterminedbyEquation(1)1sthesecases (Table 1) within the numerical precision of the methods involved (rolling ball [26–29] vs. QuickHull [30]). As a result, close agreement between the two methods is also expected for any other mostly convex body such as a small peptide or for any shape with small or large, deep or shallow cup-like cavities such as a globular protein. In practice, preliminary work confirms this expectation and indicates that the \( \rho \) values for small peptides stay essentially the same and that the \( \rho \) values for globular proteins are usually within a few per cent for the two algorithms. Therefore, the conclusions drawn from our original PSA validation study [16] based predominantly on a large number of protein and protein complex structures downloaded from the PDB database [22] remain unchanged. In fact, structural data for extended denatured proteins which would potentially lead to an underestimation of the shape factor by Equation (1) are very hard to find not only in the PDB database but from any source. In addition, since the \( \rho \) values presented in the previous study [16] are centred near a rather large value of 1.3, a substantial underestimation of the shape factor by Equation (1) in that study can be excluded.

Cyanobacterial haemoglobin is an example for a globular protein. For its solution structure (2KSC [21]) we compute similar \( \rho \) values of 1.151 and 1.199 using the old and new algorithms, respectively (Table 1). Inspection of this structure indicates that several polar and ionic amino acid side chains extend out into the solvent. These extrusions are responsible for an increased \( A_{cp} \) over \( A_{ref} \) and therefore for the slight underestimation of \( \rho \) using Equation (1). After desolvation occurring in the electrospray process [31] preceding the ion mobility experiment, those side chains will fold back onto the surface of the molecular framework [32,33] and the \( \rho \) values evaluated by the two algorithms for the resulting annealed structure – relevant in an ion mobility study – will then necessarily agree even closer with each other.

For extended structures, on the other hand, the result depends very strongly on the method used to calculate \( \rho \). For instance, Equation (1) yields a completely meaningless value of 0.727 for an extended partially helical structure of the protein ubiquitin [19] (fully annealed in charge state \( z = 13 \)), whereas application of the new algorithm results in \( \rho = 1.110 \), a shape factor of reasonable magnitude in comparison with previous theoretical shape analyses of ubiquitin [8,19] and with experimental ion mobility data [33–35]. Previously reported ubiquitin shape factors are 1.06 [19], a lower limit estimate for the \( z = 13 \) geometry, and 1.23 for the native state [8]. The previously measured ion mobility cross section is 2020 A\(^2\) [33] for \( z = 13 \), while \( \rho = 1.110 \) yields a comparable PSA value of 1960 A\(^2\) for the corresponding \( z = 13 \) structure [19]. A more comprehensive study of ubiquitin PSA cross sections for various charge states both in helium and nitrogen buffer gas and as a function of temperature in comparison with experimental data is in progress and will be presented in a future publication [24]. Preliminary results are promising and show broad agreement between PSA and experimental cross sections.

**Conclusions**

Connecting ion mobility data with the ion molecular structure is not trivial. In order to make this connection the PSA method employs a shape factor \( \rho \), a measure for the concaveness of the molecular surface. The original approach to calculate \( \rho \) is based on a comparison of the molecular surface area with the convex envelope enclosing the molecule. This approach yields satisfactory results for globular structures but fails for extended geometries. Finding the structure with a minimal surface area but still fully enclosing the molecule is a better strategy to describe concaveness in the general case. In some cases, this reference structure is identical with the convex envelope, in other cases, a structure with a smaller surface can be found. For extremely extended structures, such as a bent skinny rod or a skinny donut, the surface of the convex envelope is very large, larger than the molecular surface itself. The new method of comparing the molecular surface with the minimal surface structure yields a physically
meaningful description of the concaveness of a molecular surface for all types of molecular shapes and the resulting quantity \( \rho \) appears to be useful as a shape factor in PSA calculations.

**Acknowledgements**

We acknowledge support from the Air Force Office of Scientific Research under grant FA9550-11-1-0113 and partial personnel support from Bruker Daltonics.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

Air Force Office of Scientific Research [grant number FA955011-1-0113].

**References**


T. Wyttenbach et al.