International Journal of Mass Spectrometry 378 (2015) 360–363

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

A divide-and-conquer approach to compute collision cross sections in the projection approximation method

Béla Paizs*

School of Chemistry, Bangor University, Bangor LL57 2UW, UK

ARTICLE INFO

Article history:
Received 1 June 2014
Received in revised form 12 October 2014
Accepted 13 October 2014
Available online 22 October 2014

Keywords:
Ion mobility spectrometry
Collision cross section
Projection approximation
Modeling

ABSTRACT

The prevalent method to compute collision cross sections of large molecules is the projection approximation (PA) method that involves Monte Carlo (MC) integration of molecular projections on randomly chosen planes. Here we propose a new strategy to compute these projections based on a divide-and-conquer (DC) strategy. It is demonstrated that the DC method is faster and results in more accurate molecular projections than MC integration for large biomolecules using similar integration criteria. A new software tool (CCS) is presented for efficient computation of collision cross sections in the PA method.

1. Introduction

Recent developments in ion mobility spectrometry [1,2] coupled to mass spectrometry (IMS–MS) led to such instrumentation available to a growing number of laboratories. These instruments allow structural studies on molecular systems ranging from small molecules [3] to large biomolecules and their complexes [4]. Typical workflow to probe gas-phase ion structures in IMS–MS [5] involves measuring experimental collision cross sections (CCSs) and theoretical studies to compute potential ion structures and corresponding theoretical CCSs for comparison to experimental values.

A handful of computational strategies are available to calculate theoretical CCSs for ion structures. Bowers and co-workers developed the projection approximation (PA) method [5,6] that projects 3D molecular structures onto randomly chosen planes where atoms are represented by circles with pre-defined collision radius. The projection area ('molecular projection') is then integrated by using Monte Carlo (MC) strategies [7] where randomly chosen points ('projectiles') are considered a hit if located on any circle representing an atom ('atomic collision areas'). Such molecular projections are determined for a number of orientations until a statistically meaningful average is obtained. The PA method is computationally efficient but it becomes inaccurate for large biomolecules, this issue is often treated by simple scaling of PA CCSs [8,9].

More accurate CCSs can be calculated by strategies explicitly considering ion neutral collisions. The exact hard-sphere scattering (EHSS) model [10] is based on a hard sphere two-body interaction potential while the trajectory method (TM) considers Lennard–Jones and dipole terms [11]. The latter is the 'gold standard' of CCS calculation, the TM method is clearly capable to accurately describe the physics of low-energy ion-molecule collisions. Unfortunately, this efficiency comes at a price, the currently available TM software [11] is inefficient for large biological ions. To overcome this issue Bowers and coworkers have recently redesigned the PA method [12] to account for size and shape effects on CCSs.

Theoretical CCSs can be calculated by using a few software tools [6,10–12]. Unfortunately, all of these are subjected to some limitations in terms of efficiency, availability or robustness. Here we introduce a new program (CCS) to calculate PA collision cross section implementing the traditional Monte Carlo and a new divide-and-conquer strategy to integrate molecular projections on random planes.

2. Results and discussion

Calculation of collision cross sections in the PA framework involves generation and integration of molecular projections on randomly selected planes. The first task can be achieved by rotation of the whole molecule around its center of mass using uniform random rotation matrices built from Euler angles or quaternions [13]. Once the ion is rotated one simply takes the *x*-*y* plane and integrates molecular shapes generated by placing circles around

* Tel.: +44 1248 382734.
E-mail address: B.Paizs@bangor.ac.uk

http://dx.doi.org/10.1016/j.ijms.2014.10.005
1387-3806/

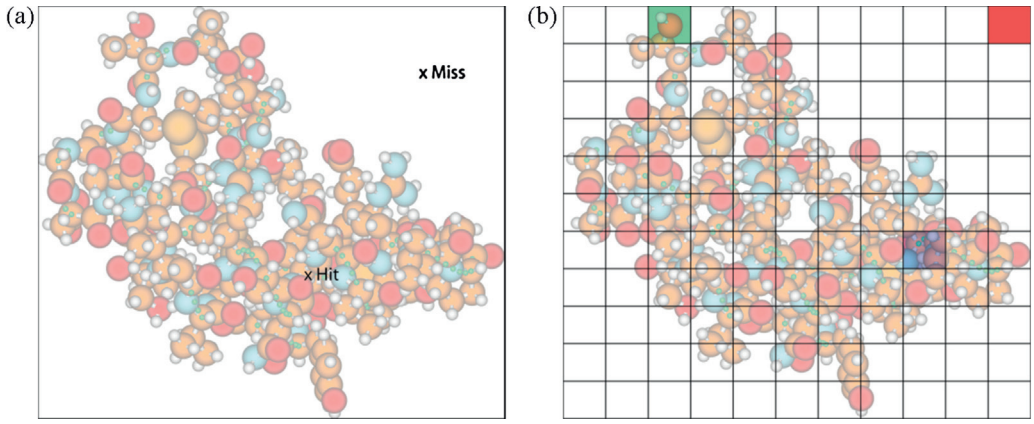

Fig. 1. (a) A biomolecular ion projected onto a plane and placed into a closely bounding box. Random 'projectiles' shot at the box are classified as hit (hitting an atomic collision area represented by a circle) or miss in MC integration. (b) The same molecular projection splitted into a rectangular grid for DC integration.

atoms with pre-defined collision radius. This is illustrated in Fig. 1a where 'atomic collision areas' are represented by color-coded circles and the molecular projection is placed into a closely bounding box. Due to the complexity of 3D molecular structures no closed form for the integration of molecular projections exists. Rather, a Monte Carlo strategy [5,7] is applied to compute this by randomly picking points in the box of Fig. 1a and checking these whether a hit on the molecular shape was chosen. The MC molecular projection is then calculated as

$$P_{MC} = A \times \frac{n_{hit}}{n} \quad (1)$$

where A is the box area, n_{hit} and n are the number of hits on the molecular area and the number of tries, respectively. When a statistically meaningful set of 'projectiles' is generated the error of the MC integration can be approximated [6,7] as

$$Err_{MC} = \sqrt{\frac{(n \times n_{hit} - n_{hit} \times n_{hit})}{n^3}} \quad (2)$$

that is Err_{MC} is proportional to $n^{-1/2}$. Err_{MC} can be reduced and the convergence of the MC method be concurrently improved by minimizing the size of the integration box defined around the molecular projection.

The question whether a 'projectile' is a 'hit' or a 'miss' (Fig. 1a) is answered by executing a loop in MC programs that checks projectiles whether they are in the 'atomic collision area' of any atom. If for any atom the answer is positive, the loop is left and both n_{hit} and n are increased by one. If the 'projectile' does not hit any 'atomic collision area' only n is increased by one. This means that for such 'missed projectiles' the MC program needs to actually assess all atomic centers and these areas can be substantial as shown in Fig. 1a. In other words, substantial computational resources are used to explore projection areas where no hits are expected and this is realized only by performing full loops over all atoms in the MC method. The efficiency of the standard MC scheme can also be low for areas where hits are expected. For example, assume that atoms are ordered for the ion in Fig. 1b so that all atoms in the green shaded area are close to the end of the list of atoms. In such cases 'projectiles' hitting 'atomic collision areas' in this segment are confirmed to be hits after assessing many other atoms in remote parts of the molecular ion.

This short discussion indicates that while MC strategies to integrate molecular areas are simple and relatively easy to implement, their performance for large molecular systems can be poor. First, as the size of the molecular system increases, more and more 'projectiles' are needed to appropriately explore molecular projections to reach the same accuracy. Furthermore, a substantial part of computational efforts is used to explore either 'empty' areas or atoms that are remote to the actual 'projectile'. These limitations can be overcome by implementing an integration strategy based on a divide and conquer (DC [14]) approach. DC algorithms first split a 'large problem' into 'smaller problems' and solve these either accurately (analitically) or more accurately than is feasible for the 'large problem'. Once achieved, solutions of the 'smaller problems' are combined into solution of the 'large problem'.

For example, the molecular projection in Fig. 1a can be split into a number of smaller areas (called cells in the following) defined by a rectangular grid as shown in Fig. 1b. The individual cells substantially differ from another in terms of the number of atoms they include. For example, the cell colored by red is empty. The blue cell is densely populated while the green one contains only a few atoms. Atoms in the molecule can be unambiguously assigned to cells and by generation of such information one can easily distinguish empty and densely populated areas of molecular projections.

Furthermore, by appropriately choosing the lengths of cell sides one can accelerate the integration of molecular projections and generate more accurate integrals than is computed in the MC scheme. For example, if the cell sides are chosen to be slightly longer than the largest atomic radius (r_{max}) for the investigated ion augmented by the atomic or molecular radius of the colliding neutral ($r_{max} + r_{He}$ for He as neutral for example), then only atoms belonging to the present and adjacent cells need to be considered when individual 'projectiles' are assessed for being a hit. For example, the red cell in Fig. 1b is empty as are all its adjacent cells, so without firing a single projectile at this area one can state that this cell will not contribute to the overall molecular projection. The blue cell is densely populated as are nearly all of the adjacent cells indicating that the whole cell area will contribute to the overall molecular projection. The green cell means an intermediate case between the red and blue cells: both the cell itself and its neighbors are partly populated so this cell's contribution to the overall projection will be larger than zero but smaller than the cell area. It is worth noting here that once a cell is assigned either as 'red' or 'blue' its contribution to the overall molecular projection is defined as zero or the cell area without any further computation. The contributions from the green cells need to be evaluated using the MC method. However, the atomic lists to be considered here are substantially shorter than the number of atoms in the whole system since only atoms in the green and adjacent cells need to be considered, all other atoms are positioned farther than the maximum augmented atomic radius ($r_{max} + r_{He}$ for He for example).