Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behaviour of small molecules and their protonation site isomers

Jasper Boschmans[1], Sam Jacobs[2], Jonathan P. Williams[3], Martin Palmer[3], Keith Richardson[3], Kevin Giles[3], Cris Lapthorn[4], Wouter A. Herrebout[2], Filip Lemière[1], Frank Sobott[1,5,6]

[1]*Biomolecular & Analytical Mass Spectrometry group, Department of Chemistry, University of Antwerp, Antwerp, Belgium,* [2]*Molecular Spectroscopy group, Department of Chemistry, University of Antwerp, Antwerp, Belgium,* [3]*Waters Corporation, Stamford Avenue, Wilmslow, SK9 4AX, UK,* [4]*Faculty of Engineering & Science, University of Greenwich, Medway Campus, Chatham, ME4 4TB, UK,* [5]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK,* [6]*School of Molecular and Cellular Biology, University of Leeds, LS2 9JT, UK*

1. Abstract

Electrospray ion mobility-mass spectrometry (IM-MS) data show that for some small molecules, two (or even more) ions with identical sum formula and mass, but distinct drift times are observed. In spite of showing their own unique and characteristic fragmentation spectra in MS/MS, no configurational or constitutional isomers are found to be present in solution. Instead the observation and separation of such ions appears to be inherent to their gas-phase behaviour during ion mobility experiments. The origin of multiple drift times is thought to be the result of protonation site isomers ('protomers'). Although some important properties of protomers have been highlighted by other studies, correlating the experimental collision cross-sections (CCS) with calculated values has proven to be a major difficulty. As a model, this study uses the pharmaceutical compound melphalan and a number of related molecules with alternative (gas-phase) protonation sites. Our study combines density functional theory (DFT) calculations with modified MobCal methods (e.g. nitrogen-based Trajectory Method algorithm) for the calculation of theoretical CCS

34 values. Calculated structures can be linked to experimentally observed signals, and a

35 strong correlation is found between the difference of the calculated dipole moments of the

36 protomer pairs and their experimental CCS separation.

37

38

39

40 **2. Introduction**

41

42 Ion mobility-mass spectrometry (IM-MS) is a separation and characterization technique

43 that has proven to be applicable in many research fields since it started to gain popularity

44 more than a decade ago with the introduction of the commercial Synapt system[1]. The drift

45 time in ion mobility is determined by the collision cross-section (CCS) of an ion, which is a

46 parameter related to its size, shape and charge. Originally used in structural studies

47 investigating protein folding[2-9] and protein complexes[10-15], more and more researchers

48 are starting to use commercial IM-MS instrumentation to investigate the separation,

49 identification and gas-phase behaviour of small molecules. A possible application is the

50 rapid separation of all types of isomers, based on their mobilities (drift times)[16-24].

51 Recently, a number of ion mobility studies have reported on the observation of protomers

52 for aniline and the antibacterial agent norfloxacin[25-27]. These isobaric ions are gas-phase

53 protonation site isomers, where the protons are located on different atoms. Although the

54 position of a single H atom and the positive charge appear to have a very subtle effect on

55 these small molecules, they can cause clear differences in drift times. Such differences

56 can be relatively large for small molecules, and one would therefore typically expect that

57 they are due to the presence of isomers or conformers (i.e. size and shape differences).

58 In a recent study, Warnke *et al.* used IM-MS in combination with infrared multiple photon

59 dissociation (IRMPD) spectroscopy to study the origin of benzocaine protomers[28].

60 Differences in N-H and O-H stretch vibrations showed that two alternative sites are

61 protonated: the amine and, unexpectedly, also the carbonyl group. These data confirm

62 that the large difference between the observed CCS values for this compound is caused

63 by the different charge sites, rather than e.g. the consequence of a subsequent

64 gas-phase rearrangement reaction. The appearance of alternative gas-phase protonation

65 sites highlights the possibility of intra-molecular charge transfer during the electrospray

66 process[28]. Anionic species show similar phenomena, as was recently reported by

67 Galaverna *et al.* for benzoic acid de-protomers[29]. It also questions the localization of

68 charges in multiply protonated peptide and protein ions, important for computational

69 structure and fragmentation prediction, which are frequently assumed to remain on basic,

70 surface-exposed residues such as lysine and arginine during the ESI process.

71

72 Computational methods, such as molecular dynamics and quantum mechanical

73 calculations, can support IM-MS observations (see Figure 1). These methods have

74 become important tools for understanding and interpreting the experimental data, and

75 they can potentially also be used to predict the separation of hypothetical charge isomers

76 in ion mobility[30,31]. Interpretation of ion mobility data typically requires a conformational

77 analysis of each protomer, after which all structures are optimized using DFT. This yields

78 a set of geometries and associated partial atomic charges that can be used to compute

79 the corresponding CCS values.

80 In this study we utilised the MobCal software[32], which provides three different algorithms;

81 the Projection Approximation (PA), Exact Hard Sphere Scattering (EHSS) and the

82 Trajectory Method (TM). Of these, the most widely used are the PA and TM. In both the

83 PA and EHSS methods the molecule is represented as a collection of overlapping hard

84 spheres. The CCS calculated by the PA is simply the rotational average of the projected

85 area of this collection. While fast, the PA fails to model momentum transfer between the

86 gas and analyte molecules as well as concave analyte structure and long range ion

87 molecule interactions. In the EHSS method, a full hard sphere trajectory is calculated for

88 each analyte-gas collision. This is a significantly more sophisticated model, and it has

89 found utility as a fast method for calculating CCS of large molecular structures[33].

90 Long-range interactions, which are often significant for drug-like molecules, are

91 necessarily ignored. The TM is the most sophisticated and computationally intensive of

92 the commonly used methods. It involves a simulation of the trajectory of gas atoms

93 through a superposition of Lennard-Jones potentials corresponding to the atoms in the

94 analyte molecule. Polarisation of the gas molecules by the charge on the analyte

95 molecule is also taken into account, but modifications to the original algorithm are

96 required to adequately model interactions with polyatomic gases. More recently, the

97 Projected Superposition Approximation (PSA) algorithm was introduced by Bleiholder *et.*

98 *al.*[34-38] In this approach, which is used mainly for larger molecules, the PA cross section is

99 modified to take into account the detailed three-dimensional structure of the analyte.

100 Although previous studies of protomers made use of methods originally available in

101 MobCal, they were not always able to accurately reproduce the experimental CCS

102 values[25,27]. It is believed that IM-MS separations using polyatomic drift gases (such as N_2

103 or CO_2) require a better representation of long-range interactions. The passage of a

104 charged particle induces higher-order multipoles in the gas molecules, leading to

105 additional (retarding) forces on the ion, and more collision geometries must be

106 considered. Kim *et al.* proposed a modification to the existing trajectory method CCS

107 calculation algorithms for N_2, which takes ion-quadrupole interactions and the orientation

108 of non-spherical gas molecules into account[40,41]. This modification leads to significantly

109 higher calculated CCS values, which better conform to the experimentally determined

110 data. Only a few studies have used this new approach so far to correctly reproduce

111 experimental CCS values[28,30,31,40-42]. Apart from protomer-related studies, Lavanant *et al.*

112 used the modified algorithm to calculate CCS values for phosphoric acid clusters, which

113 can be used for negative ion mode IM calibrations[43].

114

115 The present study investigates 1. the experimental separation of hypothetical protomers

116 for 7 related compounds which share an amino and carbonyl function (in aniline, a

117 benzene ring) as alternative protonation sites; 2. the effect of using different levels of

118 theory for optimization of molecular geometry and charge distribution, 3. the ability to

119 obtain theoretical CCS that closely match experimental values; 4. the importance of the

120 analyte charge distribution itself (and the resulting dipole moment) in contrast to possible

121 charge-driven changes in molecular structure, and 5. the ability to predict protomer

122 separation in ion mobility experiments based on the difference in the calculated molecular

123 dipole moments for hypothetical protomer pairs. This study tests the hypothesis that

124 experimentally found protomers can be predicted reasonably well by differences in the

125 calculated dipole moments. The results reported here inform choices of computational

126 approaches for the prediction of protomer separation in ion mobility so that spectral

127 interpretation software (e.g. in metabolomics) could be trained to detect such

128 phenomena.

129

130

131

132 **3. Results and discussion**

133

134 3.1 IM-MS separation of protonation site isomers

135 Two distinct peaks are observed in the arrival time distribution (ATD) of melphalan (*m/z*

136 305) using nitrogen as drift gas: **I'** and **I''** (see Figure 2), which are centred around 169.9

137 $Å^2$ and 179.1 $Å^2$. However, for two other, closely related compounds,

138 dimethoxymelphalan ($DOCH_3$; **II**) and dihydroxymelphalan (DOH; **III**), we observe only

139 single and unique peaks (Figure 2), at 172.2 $Å^2$ for $DOCH_3$ and 165.3 $Å^2$ for DOH. We

140 also used CO_2 to perform ion mobility separations of $DOCH_3$ and DOH, but again only

141 single peaks were observed (data not shown).

142 Similar observations to those with melphalan were made for the local anaesthetic

143 *para*-benzocaine, an ethyl ester derivative of *para*-aminobenzoic acid (Figure 3). Two

144 peaks were found (**IV'** and **IV''**) at 131.7 $Å^2$ and 147.5 $Å^2$. For comparison, positional

145 isomers of benzocaine were also studied (Figure 3): ethyl 2-aminobenzoate

146 ("*ortho*-benzocaine") and ethyl 3-aminobenzoate ("*meta*-benzocaine"). The selected-ion

147 ATD of *ortho*-benzocaine shows only one peak at 135.2 $Å^2$ (**V**). For *meta*-benzocaine, two

148 peaks are observed (**VI'** and **VI''**) which correspond to CCS values of 133.6 $Å^2$ and 146.4

149 $Å^2$. For aniline, which we included here as a reference compound, we find two peaks as

150 reported previously[25] (**VII'** and **VII''**; see Figure 3), with CCS values of 112.9 $Å^2$ and 118.9

151 $Å^2$. Table 1 summarizes all experimental CCS values.

152

153 ### 3.2 Calculating CCS values of melphalan protomers

154 Three possible protomers of melphalan were taken into account here: protonation at the

155 nitrogen atom of the NH_2 group (N_{AA}), the carboxyl group (O_{CO}) on the amino acid side,

156 and the nitrogen atom adjacent to the phenyl ring on the chloroethyl side (N_{Cl}). Based on

157 the solution basicity (i.e. pK_a) of the various functional groups in melphalan[44], O_{CO} and N_{Cl}

158 protonation seem less plausible (see Figure 2). Nonetheless, other protomer studies

159 have reported on oxygen-rich functional groups or even aromatic rings[25-28] as preferred

160 protonation sites. After performing a conformational analysis of melphalan and

161 subsequent DFT optimisation at the B3LYP/6-31G(d,p) level (hereafter referred to as

162 "standard" level), the CCS values for the optimised structures were calculated using a

163 modified version of MobCal where the TM code optimized for use with nitrogen (see

164 Figure 1). Table 2 gives an overview of the top 5 lowest-energy conformers of each

165 melphalan protomer, together with energies, overall Boltzmann weights, dipole moments

166 and calculated CCS values. Figure 4 visualizes the conformation and molecular

167 electrostatic potential (MEP) of each lowest-energy melphalan protomer.

168 From the three protomers considered here, the N_{AA} and N_{Cl} forms best match the

169 experimentally determined CCS_{N2}. This would indicate that the O_{CO} protomer is not

170 observed during the ion mobility experiments. The ΔCCS_{N2} between the calculated N_{AA}

171 and N_{Cl} protomers is 9.0 $Å^2$, which is a good match with the experimentally determined

172 value of 9.2 $Å^2$.

173

174 <u>3.3 Melphalan-related compounds: dihydroxymelphalan and dimethoxymelphalan</u>

175 The study of melphalan derivatives, which unlike melphalan itself show only one

176 observed drift time, allows us to investigate the factors that govern formation and

177 separation of melphalan protomers more closely. A conformational analysis was

178 performed for possible protomers of these compounds, and the resulting structures were

179 optimised at "standard" level. Although DOH and $DOCH_3$ are chemically less complex

180 structures than melphalan itself (i.e. no halogen atoms), the additional rotational flexibility

181 yields more conformers and thus entails an added computational cost. For each

182 lowest-energy protomer, the values are reported in Table 3 and structures are given in

183 Figure 2. Calculated CCS values were also compared to the experimentally derived ones

184 (172.2 $Å^2$ for $DOCH_3$ and 165.3 $Å^2$ for DOH). This allows us to evaluate the

185 nitrogen-modified MobCal code, but could also show whether significantly different CCS

186 values are calculated for protomers in cases where they are not experimentally resolved.

187 The N_{AA} protomer for DOH has a considerably smaller calculated CCS (156.1 $Å^2$) than the

188 experimentally observed value. The CCS of the N_{OH} protomer on the other hand (164.6

189 $Å^2$) is a close match with the experiment. For $DOCH_3$, the calculated CCS values of both

190 hypothetical protomers (181.3 $Å^2$ and 182.3 $Å^2$) over-estimate the experimental CCS of

191 172.2 $Å^2$.

192

193 <u>3.4 Other related small molecules: benzocaine isomers and aniline</u>

194 Our calculation strategy was further evaluated against experimentally determined CCS

195 values of benzocaine isomers and aniline (Table 4). For *para*-benzocaine, protonation of

196 the carbonyl group gives a CCS_{N2} of 132.0 $Å^2$ for the lowest-energy structure after

197 standard-level optimization, while the equivalent procedure for the amine-protonated

198 species results in a CCS_{N2} value of 140.9 $Å^2$, which is significantly smaller than the

199 experimental value (147.5 $Å^2$). While the theory correctly predicts the separation of the

200 two protomers, it remains unclear why the calculated value for the amine-protonated

201 *para*-benzocaine deviates so much from the experiment.

202 After standard-level optimization, O_{CO} and N_{NH2} protomers of *ortho*-benzocaine have

203 computed CCS values of 131.0 and 131.5 $Å^2$, respectively. Based on these values,

204 assignment of the single experimentally observed peak to either protomer is difficult, as

205   they are expected to be almost indistinguishable. It is therefore impossible to say whether

206   they both occur in the gas phase, with their peaks overlapping, or if only one of them is

207   present. In this context it is worth noting that a recent report proposed the two alternative

208   forms of deprotonated *ortho*-hydroxybenzoic acid to be connected by a relatively low

209   isomerization barrier[29]. It might therefore be considered likely that the protomers of

210   *ortho*-benzocaine could also easily convert due to intramolecular H-bonding, leading to

211   only one mobility signal.

212   Two distinct CCS values are calculated after standard-level optimization for

213   *meta*-benzocaine: 133.9 $Å^2$ for the O_{CO} isomer and 140.8 $Å^2$ for the N_{NH_2} protomer. While

214   such calculations predict reasonably well if the postulated protomers will be separated by

215   ion mobility (one or two peaks expected), the absolute CCS values do not always match

216   well with the measured ones, e.g. for the N_{NH_2} form of *meta*-benzocaine.

217   The calculated CCS value of the N protomer of aniline (at "standard" level) is also not well

218   matched with either of the experimental values (112.9 $Å^2$ and 118.0 $Å^2$). The

219   ring-protonated species should be assigned to the first peak in the ATD, based on data

220   reported in the literature[25].

221

222   <u>3.5 Evaluating different levels of DFT calculations</u>

223   Although the majority of ion mobility studies employ the commonly used B3LYP functional

224   with 6-31G or 6-311G basis sets, a wide variety of other functionals and basis sets are

225   available. Because the CCS values computed so far, using the "standard" level combined

226   with the N_2-modified MobCal code, still show some discrepancies with the experiment, we

227   also used B3PW91/def2-TVZP (hereafter called "high" level) in order to assess how

228   sensitive the computed values are to the DFT parameters. Tables 1, 3 and 4 and Figures

229   S.2 to S.9 contrast "standard" with high-level calculated CCS values.

230   For the benzocaine isomers and aniline, we found that the high-level calculations yield

231   generally better matches against the experimental values (Table 1).

232   Both postulated protomers of DOH are found to have similar CCS at "high" level, which

233   also closely match the single observed peak centred at 165.3 $Å^2$. The "standard" level

234   result, where the CCS value for the N_{AA} protomer was underestimated (156.1 $Å^2$), falsely

235   suggested that two distinct DOH protomers should exist, with a ΔCCS_{N2} of 8.5 $Å^2$ (Table

236   3).

237   For the $DOCH_3$ form of melphalan on the other hand, the CCS values calculated at both

238   levels of structural optimization (around 180 $Å^2$) are significantly higher than the

239   measured one (172.2 Å^2). It is not apparent though why CCS calculations for this

240   compound deviate so much from the experiment.

241   For melphalan itself, the expected improvement in the theoretical values is also less

242   pronounced when using the B3PW91 functional and larger basis set. Notably the CCS of

243   the N_{Cl} protomer is now overestimated: 170.6 Å^2 ("standard" level) vs. 174.1 A^2 ("high"

244   level). Since the experimentally derived value for this protomer is 169.9 Å^2, the "standard"

245   level result is in better agreement in this particular case. While the reason for this anomaly

246   is not entirely clear, melphalan stands out as a compound with the highest conformational

247   "flexibility" (see below) in the group studied here.

248   We show here that for a number of structurally related compounds, the calculation of

249   "best" molecular geometries and charge distributions using two different levels of DFT

250   calculations leads to mixed results, with respect to how well the derived CCS match with

251   experimentally observed CCS (see Figure 5). Contrary to what we might have expected,

252   the high-level calculations do not always agree better with experiment. A more thorough

253   investigation of different basis sets and functionals is needed, as well as a

254   re-parametrization of CCS calculation methods, which currently rely on modifications to

255   the existing MobCal code. Such efforts are now underway in different research groups.

256

257   3.6 Effect of charge distribution on CCS calculations

258   Although this study assumes that different protonation site isomers can be formed in the

259   electrospray process and observed via their characteristic mobility peaks, the question

260   still remains to what extent different factors contribute to protomer separation in ion

261   mobility: is it the charge site itself, or rather the conformational change driven by

262   alternative charge sites? As an example of the latter, the rotation of the chloroethyl

263   groups in the mustard moiety of melphalan could lead to the observation of multiple

264   conformers. Interaction of these chloroethyl groups with the phenylalanine moiety of the

265   molecule could possibly also result in more compact ions.

266   Protonation at different sites, which results in different charge distributions (after

267   optimization) across the molecule, can potentially affect its geometry (i.e. the atomic

268   positions) in a way that contributes to a change in CCS. In order to assess the magnitude

269   of this effect separately from that of the charge distribution itself, we simply recalculated

270   the CCS, while removing the atomic partial charges. These re-calculated CCS values are

271   reported in Table 5 ("no charge distribution"). Removing the charges drastically lowers the

272   computed CCS values, as expected in N_2 drift gas particularly for the smaller analytes

273 (benzocaine isomers and aniline), and we do not expect them to match the experimental

274 values anymore. More importantly, what this exercise can show is if the calculated CCS

275 difference between two postulated protomers is maintained even in the absence of any

276 charge, i.e. whether it is largely caused by a conformational change of the molecule. This

277 is the case only for the melphalan protomers. All other molecules studied here show

278 virtually identical CCS (within the error margin of the experiment) for the "uncharged"

279 protomer pairs. This signifies that the potentially different molecular geometries of the

280 protomers, optimised in the presence of charge, would not account for any possible CCS

281 difference. Rather the position of the proton and the resulting relatively large differences

282 in charge distributions and dipole moments are held responsible for the observed

283 protomer separation in ion mobility. We can speculate that of the molecules studied here,

284 only melphalan is "flexible" enough to undergo a charge-site driven conformational

285 change which is sufficiently large to contribute to the separation of its protomer peaks.

286 These calculations show that different protonation sites can yield significantly different ion

287 mobilities in nitrogen, indicating that the long-range electrostatic contribution of the

288 charge to the overall CCS is substantial.

289

290 <u>3.7 Protomers and dipole moments</u>

291 The analysis of the effect of molecular geometry on CCS independent of charge (see

292 section 3.6) showed that for some of the small molecules studied here, charge

293 distributions are the determining factor for their separation in ion mobility when using

294 polarizable gases. A close look at the structures of the protomer pairs shows that,

295 although their mobilities can differ considerably, their geometries may indeed be relatively

296 similar. Since the atomic coordinates of these structures only vary slightly (apart from

297 melphalan), the CCS difference is thought to be predominantly the result of the charge

298 distribution. The differences between the molecular dipole moment of various protomers

299 could therefore be used as a possible predictor for the separation of these protomers in

300 ion mobility[30].

301 As an example, the charge distributions of the three melphalan protomers considered

302 here are visualized as molecular electrostatic potentials (MEPs) in Figure 4. Comparison

303 of the 5 lowest-energy structures per protonation site (see Table 2) shows that they share

304 similar dipole moments. Furthermore, the structures with the smallest dipole moment (ca.

305 6 Debye) also correspond to the ion with the smallest CCS value, i.e. the N_{Cl} protomer.

306 The structures that have a dipole moment of around 11 Debye correspond to the ion with

307 the largest observed CCS value (N_{AA} protomer). We plotted all 4 experimentally observed

308 protomer pairs with their ΔCCS values against the corresponding Delta dipole values,

309 calculated at the best-fitting DFT level (apart from melphalan, all "high" level; see Figure

310 6). Although the correlation is not very strong, the trend is clear: the larger the calculated

311 Delta dipole values, the higher the measured Delta CCS.

312 To test this hypothesis further, we also plotted predicted ΔCCS values for all possible

313 protomer pairs, calculated at both "standard" and "high" levels using MobCal, against

314 their corresponding Delta dipole values. These data highlight that the correlation between

315 CCS and dipole moment is quite poor with standard-level calculations (red squares in

316 Figure 6). The high-level structure calculations on the other hand (blue diamonds) yield a

317 reasonably good correlation (linear fit: R^2=0.8784) between differences in dipole moment

318 of protomer pairs and their separation in ion mobility experiments where polarisable drift

319 gases such as nitrogen are used[30,31]. Aniline shows a ΔCCS larger than expected based

320 on the calculated Δ dipole value, which may be due to the fact that the smaller protomer is

321 a ring-protonated (charge-delocalized) form.

322 Taken together, these data suggest that rather than geometry or net charge alone, the

323 charge distribution – characterized by the dipole moment and, as recently proposed[29], the

324 polarizability of the analyte – plays a major role for the observed CCS values as well,

325 particularly for relatively rigid molecules and their specific interaction with a polarizable

326 drift gas (i.e. N_2). Experimentally observed protomer separation is found to be explained

327 reasonably well by differences between the calculated dipole moments of alternatively

328 protonated forms of the analyte.

329

330

331

332 **4. Experimental**

333

334 Caution: melphalan and degradation products are carcinogenic and should be handled

335 with care.

336

337 4.1. Chemicals and sample preparation

338 Chemicals purchased from Sigma-Aldrich (Bornem, Belgium): acetaminophen (> 99.0

339 %), alprenolol (Eur. Pharmacop. Ref.), aniline (99.8 %), colchicine (> 95 %),

340 dexamethasone (> 97 %), ethyl 4-aminobenzoate (98 %), ethyl 3-aminobenzoate (97 %),

341 ethyl 2-aminonenzoate (> 99 %), N-ethylaniline (98 %), melphalan (min. 95 %),

342 ondansetron (> 98 %), poly-DL-alanine, sodium formate (HPLC, > 99.0 %) and verapamil

343 (> 99 %). Acetonitrile (ACN; HPLC grade), methanol (MeOH; HPLC grade) and formic

344 acid (FA; 99+ %) were obtained from Acros (Geel, Belgium). Reversed osmosis (RO)

345 water was prepared using a Silex water filtering system from Eurowater (Nazareth-Eke,

346 Belgium). Ammonium hydroxide (solution of 25 % v/v) was purchased from Merck

347 (Overijse, Belgium). Dimethoxymelphalan was synthesized in-house, and

348 dihydroxymelphalan formed during synthesis as an additional reaction product. Stock

349 solutions (10^{-2} M) of all analytes and calibrants were prepared in MeOH.

350

351 <u>4.2 Optimisation of molecular structures and charge distributions</u>

352 A conformational analysis was performed to find the best structure of melphalan ions in

353 the gas phase. The conformational space of the different protonated species

354 ('protomers') was explored using TINKER (version 6)[45] with the Merck molecular force

355 field (MMFF94). The resulting structures were further optimised with Gaussian 09[46] at the

356 B3LYP/6-31G(d,p) ("standard") and B3PW91/def2-TZVP ("high") levels. For each

357 calculation, the optimised structure was verified to be a local minimum by performing a

358 vibrational analysis. Atomic charges were computed using the Merz-Singh-Kollman

359 scheme with the constraint to reproduce the molecular dipole ('pop=mk,dipole'). The

360 uncharged structures were generated by simply removing the atomic partial charges. As

361 the dipole moment for charged species depends on the origin chosen, the center of

362 charge was used as a reference point instead of the center of mass for all calculations.

363 Three-dimensional structures were visualized using Avogadro (version 1.1.1)[47] and

364 molecular electrostatic potentials (MEPs) using VMD (version 1.9.2)[48,49].

365

366 <u>4.3 Calculation of CCS values</u>

367 MobCal was used to calculate CCS values[32,33]. MobCal is available as freeware[50].

368 MobCalPARSER, also available as freeware[51], allowed the direct use of Gaussian output

369 (.log) files.

370 The modified version of MobCal[41] used in this study calculates CCS values for

371 experiments in nitrogen drift gas and takes into account ion-quadrupole interactions and

372 the orientation of non-spherical gases during collisions (TM algorithm only).

373 Lennard-Jones potentials were re-tuned by scaling universal force field (UFF) parameters

374 such as the atomic energy and van-der-Waals distance, in order to represent the ion

375 motion through N_2 drift gas better. The code was also expanded with other types of

376 atoms.

377

378 ### 4.4 Instrumentation

379 Travelling-Wave Ion Mobility Spectrometry (TWIMS) experiments were performed on a

380 Synapt G2 HDMS instrument, and data acquisition and processing were carried out using

381 MassLynx (V 4.1).

382 The instrument (Waters Corporation, Wilmslow, UK) was equipped with a

383 nano-electrospray source and used gold-coated glass capillary needles, which were

384 fabricated in-house. In order to obtain clean spectra, the ions under investigation were

385 *m/z* selected in all IM-MS experiments. Prior to the experiments, the instrument was

386 calibrated in the *m/z* 50-600 mass range using sodium formate solution (positive ion

387 mode). All prepared solutions were checked for the presence of impurities, which might

388 overlap with the signal of the analyte.

389 Typical instrument parameters in time-of-flight mode are: sample and extraction cone

390 voltage: 10 V and 3 V, trap bias: 2 V, trap cell gas flow: 2 mL/min, trap and transfer CE: 4

391 V and 0 V. Experimental CCS are determined after a single calibration of the TWIMS cell

392 using both poly-DL-alanine (0.5 mg/mL in 1:1 H_2O:ACN) and a set of drug-like

393 compounds (5 μM each in 1:1 H_2O:ACN) as reported before[12,39]. In ion mobility mode,

394 drift times were determined for different IM wave velocities in order to eliminate

395 energy-dependent phenomena, which could affect the ion mobility separation. Some

396 parameters differ in ion mobility mode: trap bias: 40 V, IMS wave velocity: 600 m/s or, for

397 melphalan and related compounds: 1000 m/s, IMS wave height: 40 V, He and IMS

398 (nitrogen) gas flow: 180 mL/min and 90 mL/min.

399

400
401

402 ## 5. Conclusions

403

404 Most small compounds show only one, unique drift time in ion mobility experiments, which

405 allows the use of such data as additional identifiers for mass-spectrometry based

406 molecular characterization, e.g. in metabolomics. There is now considerable interest in

407 assembling databases which contain ion mobilities of compounds under standardized

408 conditions, not dissimilar to retention times in chromatography.

409 Occasionally, small molecules exhibit multiple ion mobility drift times, due to the ability to
410 form different protomers. Protomers are constitutional isomers, or more precisely,
411 isomeric catiomers. They are molecular species which originate from the same chemical
412 entity in solution, but where partial, intramolecular proton transfer during electrospray
413 ionization causes the formation of charge isomers in the gas phase. This is often
414 encountered for, but not limited to, small molecules containing an amino and a carbonyl
415 or carboxyl moiety. Protonation at different sites may not only distort the molecular
416 geometry, affecting the hard sphere cross-sections, but also lead to significantly different
417 charge distributions. The latter can be represented by the dipole moment, which has a
418 large effect on the ion's mobility when polarisable drift gases such as nitrogen are used.

420 In the most systematic analysis to date, we utilized a panel of 7 closely related small
421 molecules, 4 of which are found to show two drift time signals, to better understand what
422 determines protomer separation in ion mobility and evaluate computational approaches
423 for their characterization. IM-MS of the chemotherapeutic agent melphalan revealed the
424 presence of two mobility peaks, whereas molecules closely related to this compound (i.e.
425 dihydroxy- and dimethoxymelphalan) only showed one protonated form. For comparison,
426 aniline and three isomers of benzocaine were also included in this study.
427 By comparing experimental with calculated CCS values from molecular modelling, we
428 could assign the protonation site and structure of the observed ions. More flexible
429 structures with bulky side chains (e.g. $DOCH_3$) however appear to have their CCS
430 overestimated with both types of calculations used. While results of what we call
431 "high-level" calculations match experimental data much better for most molecules studied
432 here, a more systematic investigation of functionals and basis sets is required to
433 determine the most appropriate computational strategy for the optimization of structure
434 and charge of protomers. With different protonation sites in these compounds available
435 under electrospray conditions, a number of alternative charge distributions and molecular
436 geometries have to be evaluated for how well they match the corresponding collision
437 cross sections in the experiment. More straightforward and efficient calculation methods
438 would make this step much faster and more accurate, and enable "high-throughput"
439 approaches for ion mobility data processing such as would benefit, e,g., compound
440 identification in complex samples.
441 The use of polarisable drift gases (e.g. N_2), which has become common due to the
442 widespread use of travelling wave IM-MS, leads to a more frequent observation of

protomer phenomena, and puts the issue of their structual assignment into the spotlight. We found a good agreement between experimental and theoretical CCS data in this study when using a modified version of the trajectory method, optimised for use with nitrogen as drift gas. Our data show that the molecular dipole moment, rather than the hard sphere collision cross section, is a useful determinant for the ion mobility separation of protomers. Furthermore, a good correlation appears to exist between the different calculated dipole moments, and both experimental and theoretical CCS differences, in protomer pairs investigated here. As calculated dipole moments are readily available, they may be useful "predictors" of protomer separation in experiments which target rapid small molecule isomer separation and identification using ion mobility.

Acknowledgements

References

1.   K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose and R. H. Bateman, *Rapid Comm. Mass Spectrom.*, 2004, **18**, 2401-2414.

2.   K. Thalassinos, S. E. Slade, K. R. Jennings, J. H. Scrivens, K. Giles, J. Wildgoose, J. Hoyes, R. H. Bateman and M. T. Bowers, *Int. J. Mass Spectrom.*, 2004, **236**, 55-63.

3.   H. A. Sawyer, J. T. Marini, E. G. Stone, B. T. Ruotolo, K. J. Gillig and D. H. Russell, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 893-905.

4.   S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinos, R. H. Bateman, M. T. Bowers and J. H. Scrivens, *Int. J. Mass Spectrom.*, 2007, **261**, 1-12.

5.   C. A. Scarff, K. Thalassinos, G. R. Hilton and J. H. Scrivens, *Rapid Commun. Mass Spectrom.*, 2008, **22**, 3297-3304 .

478 6. D. P. Smith, T. W. Knapman, I. Campuzano, R. W. Malham, J. T. Berryman, S. E. Radford
479   and A. E. Ashcroft, *Eur. J. Mass Spectrom.*, 2009, **15**, 113-130.

480 7. C. W. N. Damen, Chen, A. B. Chakraborty, M. van Oosterhout, J. R. Mazzeo, J. C. Gebler,
481   J. H. M. Schellens, H. Rosing and J. H. Beijnen, *J. Am. Soc. Mass Spectrom.*, 2009, **20**,
482   2021-2033.

483 8. C. Atmanene, S. Petiot-Bécard, D. Zeyer, A. Van Dorsselaer, V. Vivat Hannah and S.
484   Sanglier-Cianférani, *Anal. Chem.*, 2012, **84**, 4703-4710.

485 9. A. Konijnenberg, A. Butterer and F. Sobott, *BBA - Proteins and Proteomics*, 2013, **1834**,
486   1239-1256.

487 10. B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung and C. V. Robinson, *Nat.*
488   *Protoc.*, 2008, **3**, 1139-1152.

489 11. C. Uetrecht, R. J. Rose, E. Van Duijn, K. Lorenzen, and Albert J. R. Heck, *Chem. Soc.*
490   *Rev.*, 2010, **39**, 1633-1655.

491 12. M. F. Bush, I. D. G. Campuzano, and C. V. Robinson, *Anal. Chem.*, 2012, **84**, 7124-7130.

492 13. J. L. Benesch and B. T. Ruotolo, *Curr. Opin. Struct. Biol.*, 2011, **21**, 641-649.

493 14. J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson and A. J. R. Heck, *Angew. Chem. Int. Ed.*,
494   2013, **52**, 4020-4023.

495 15. L. Han and B. T. Ruotolo, *Int. J. Ion Mobil. Spectrom.*, 2013, **16**, 41-50.

496 16. M. Benassi, Y. E. Corilo, D. Uria, R. Augusti and M. N. Eberlin, *J. Am. Soc. Mass*
497   *Spectrom.*, 2009, **20**, 269-277.

498 17. J. P. Williams, T. Bugarcic, A. Habtemariam, K. Giles, I. Campuzano, P. M. Rodger and P.
499   J. Sadler, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1119-1122.

500 18. M. Zhu, B. Bendiak, B. Clowers and H. H. Hill, *Anal. Bioanal. Chem.*, 2009, **394**,
501   1853-1867.

502 19. G. J. Dear, J. Munoz-Muriedas, C. Beaumont, A. Roberts, J. Kirk, J. P. Williams and I.
503   Campuzano, *Rapid Commun. Mass Spectrom.*, 2010, **24**, 3157-3162.

504 20. M. Fasciotti, A. F. Gomes, F. C. Gozzo, B. A. Iglesias, G. F. de Sá, R. J. Daroda, M.
505   Toganoh, H. Furuta, K. Araki and M. N. Eberlin, *Org. Biomol. Chem.*, 2012, **10**, 8396-8402.

506 21. L. Ahonen, M. Fasciotti, G. B. A. Gennäs, T. Kotiaho, R. J. Daroda, R. M. Eberlin and R.
507   Kostiainen, *J. Chromatogr. A*, 2013, **1310**, 133–137.

508 22. V. E. Wright, F. Castro-Gómez, E. Jurneczko, J. C. Reynolds, A. Poulton, S. D. R. Christie,
509   P. Barran, C. Bo and C. S. Creaser, *Int. J. Ion Mobil. Spec.*, 2013, **16**, 61–67.

510 23. J. P. Williams, J. M. Brown, I. Campuzano and P. J. Sadler, *Chem. Commun.*, 2010, **46**,
511   5458-5460.

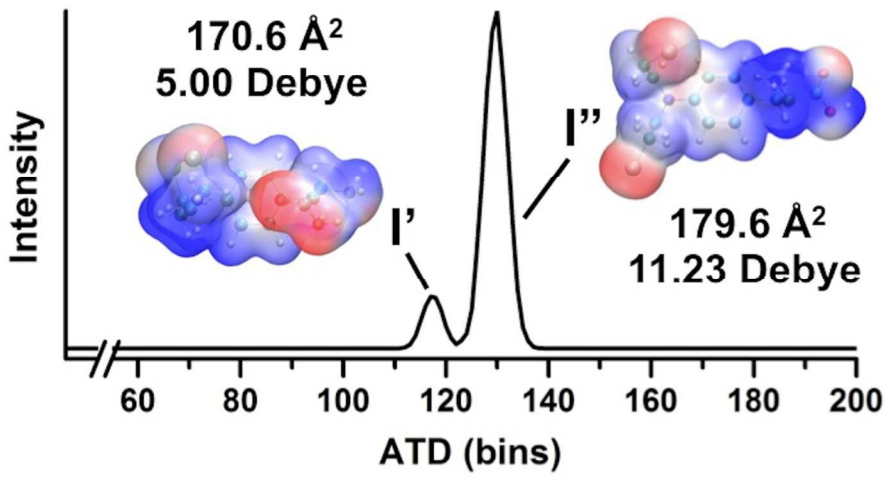512 24. J. Hofmann, H. S. Hahm, P. H. Seeberger and K. Pagel, *Nature*, 2015, **526**, 241-244.

513　25.　P. M. Lalli, B. A. Iglesias, H. E. Toma, G. F. Sa, R. J. Daroda, J. C. Silva Filho, J. E.

514　　　　Szulejko, K. Araki and M. N. Eberlin, *J. Mass Spectrom.*, 2012, **47**, 712-719.

515　26.　A. Kaufmann, P. Butcher, K. Maden, M. Widmer, K. Giles and D. Uria, *Rapid Commun.*

516　　　　*Mass Spectrom.*, 2009, **23**, 985-998.

517　27.　C. Lapthorn, T. J. Dines, B. Z. Chowdhry, G. L. Perkins and F. S. Pullen, *Rapid Commun.*

518　　　　*Mass Spectrom.*, 2013, **27**, 2399-2410.

519　28.　S. Warnke, J. Seo, J. Boschmans, F. Sobott, J. H. Scrivens, C. Bleiholder, M. T. Bowers, S.

520　　　　Gewinner, W. Schöllkopf, K. Pagel and G. von Helden, *J. Am. Chem. Soc.*, 2015, **137**,

521　　　　4236-4242.

522　29.　R. S. Galaverna, G. A. Bataglion, G. Heerdt, G. F. de Sa, R. Daroda, V. S. Cunha, N. H.

523　　　　Morgon, M. N. Eberlin, Eur. *J. Org. Chem.*, 2015, **2015**, 2189-2196.

524　30.　P. M. Lalli, Y. E. Corilo, M. Fasciotti, M. F. Riccio, G. F. de Sá, R. J. Daroda, G. H. M. F.

525　　　　Souza, M. McCullagh, M. D. Bartberger, M. N. Eberlin and I. D. G. Campuzano, *J. Mass*

526　　　　*Spectrom.*, 2013, **48**, 989-997.

527　31.　T. G. Flick, I. D. G. Campuzano and M. D. Bartberger, *Anal. Chem.*, 2015, **87**, 3300-3307.

528　32.　M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, *J. Phys.*

529　　　　*Chem.*, 1996, **100**, 16082-16086.

530　33.　A. A. Shvartsburg and M. F. Jarrold, *Chem. Phys. Lett.*, 1996, **261**, 86-91.

531　34.　T. W. Knapman, J. T. Berryman, I. Campuzano, S. A. Harris and A. E. Ashcroft, *Int. J. Mass*

532　　　　*Spectrom.*, 2010, **298**, 17-23.

533　35.　C. Bleiholder, T. Wyttenbach and M. T. Bowers, *Int. J. Mass Spectrom.*, 2011, **308**, 1-10.

534　36.　S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang and M. T. Bowers, *Int. J. Mass*

535　　　　*Spectrom.*, 2012, **330-332**, 78-84.

536　37.　C. Bleiholder, S. Contreras, T. D. Do and M. T. Bowers, *Int. J. Mass Spectrom.*, 2013,

537　　　　**345-347**, 89-96.

538　38.　C. Bleiholder, S. Contreras and M. T. Bowers, *Int. J. Mass Spectrom.*, 2013, **354-355**,

539　　　　275-280.

540　39.　E. G. Marklund, M. T. Degiacomi, C. V. Robinson, A. J. Baldwin and Justin L.P. Benesch,

541　　　　Collision Cross Sections for Structural Proteomics, *Structure*, http://impact.chem.ox.ac.uk,

542　　　　(accessed May 2015).

543　40.　D. Guillarme, H. Kim, J. Ruta, H. I. Kim, S. Rudaz, P. V. Johnson, J.-L. Veuthey, L. W.

544　　　　Beegle, J. L. Beauchamp, W. A. Goddard and I. Kanik, *Anal. Chem.*, 2008, **80**, 1928-1936.

545　41.　I. Campuzano, M. F. Bush, C. V. Robinson, C. Beaumont, K. Richardson, H. Kim, and H. I.

546　　　　Kim, *Anal. Chem.*, 2012, **84**, 1026–1033.

547　42.　C. Lapthorn, F. S. Pullen, B. Z. Chowdhry, P. Wright, G. L. Perkins, and Y. Heredia,

548   *Analyst*, 2015, **140**, 6814-6823.

549   43.   H. Lavanant, V. Tognetti and C. Afonso, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 572-580.

550   44.   S. A. Stout and C. M. Riley, *Int. J. Pharm.*, 1985, **24**, 193-208.

551   45.   TINKER – Software Tools for Molecular Design, http://dasher.wustl.edu/tinker, (accessed

552   January 2015).

553   46.   M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman,

554   G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.

555   P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K.

556   Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T.

557   Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E.

558   Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A.

559   Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene,

560   J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann,

561   O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma,

562   V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö.

563   Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision D.01,

564   Gaussian, Inc., Wallingford CT, 2009.

565   47.   Avogadro – An Open-source Molecular Builder and Visualization Tool,

566   http://avogadro.cc/wiki/Main_Page, (accessed May 2015).

567   48.   VMD – Visual Molecular Dynamics, http://www.ks.uiuc.edu/Research/vmd, (accessed

568   January 2015).

569   49.   W. Humphrey, A. Dalke and K. Schulten, VMD - Visual Molecular Dynamics, *J. Molec.*

570   *Graphics*, 1996, **14**, 33-38.

571   50.   MobCal – A Program to Calculate Mobilities, http://www.indiana.edu/~nano/software.html,

572   (accessed May 2015).

573   51.   MobCalPARSER, http://sourceforge.net/projects/mobcalparser, (accessed May 2015).

574   52.   Collision Cross Section Database, http://depts.washington.edu/bushlab/ccsdatabase,

575   (accessed February 2015).

576

577

578   **<u>Figures</u>**

579

580   Graphical abstract

581

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
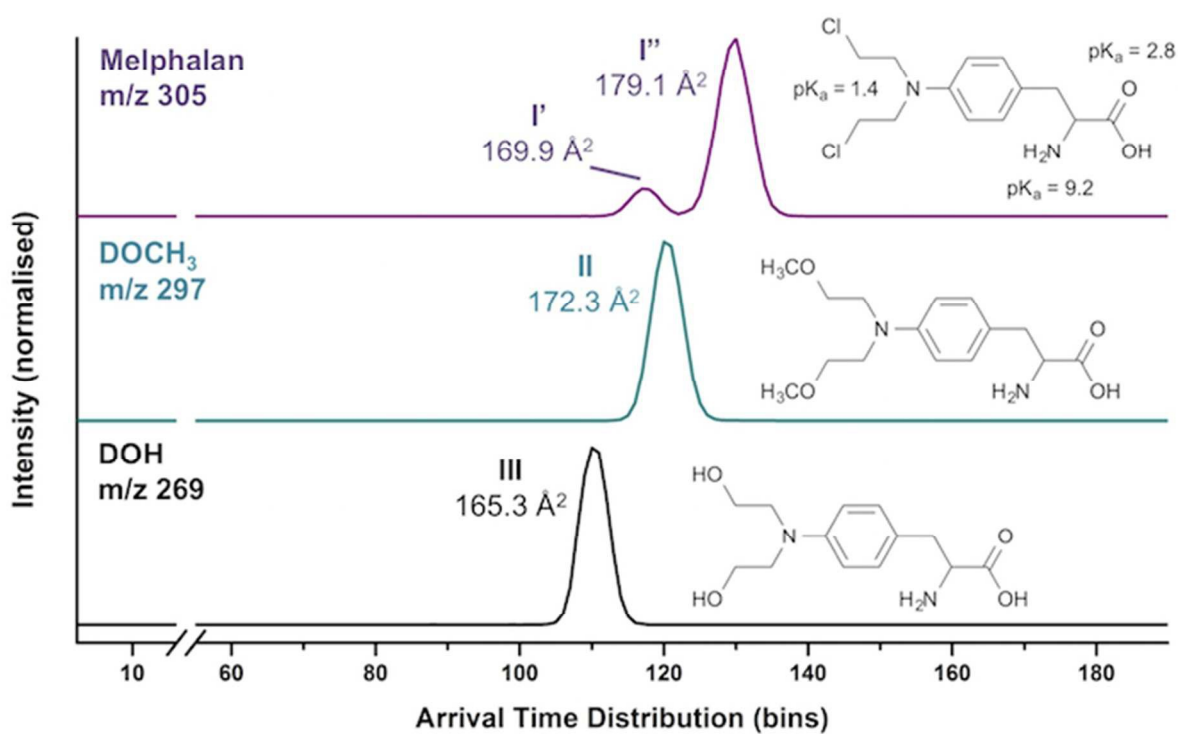
582   Figure 1

583   Overview of the sequence and output of the various experiments and calculations.

Ion mobility experiments
(N₂ drift gas)
→ separation of protomers
→ determine experimental CCS

Conformational analysis
and DFT optimisation
→ explore protomer geometry
→ determine dipole moment

Calculate theoretical CCS
(MobCal - trajectory method)
→ shape/size (i.e. CCS)
→ charge distribution (i.e. dipole)

584

1
2
3
4   585   Figure 2:

5
6   586   Observation of two peaks for melphalan (**I**; top panel), but only one peak for the DOCH$_3$

7   587   (**II**) and DOH (**III**) hydrolysis products. The drift time difference between the two

8
9   588   melphalan peaks is larger than the difference between DOH and DOCH$_3$.

10
11 589

590

591

592

593   Figure 3:

594   Observation of two peaks for *para*-benzocaine (**IV**) and *meta*-benzocaine (**VI**) as well as

595   aniline (**VII**). Only one peak is observed for *ortho*-benzocaine (**V**).

596
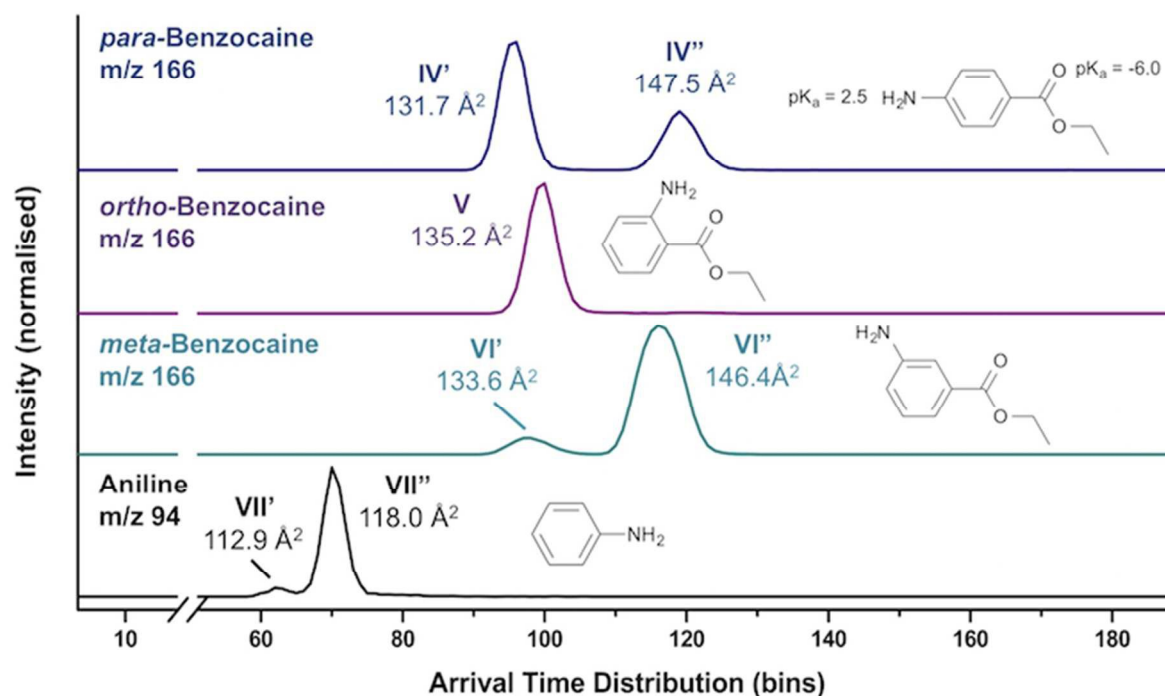
597

598   Figure 4

599   3D-visualisation of the lowest-energy structures of melphalan (**I**) after conformational

600   analysis of the protonated molecules and subsequent standard-level DFT optimisation.

601   Three possible protomers are shown: O_{CO} (left), N_{AA} (center) and N_{Cl} (right). Molecular

602   electrostatic potentials (MEPs) are also given. Red areas display negative sites (e.g.

603   electron dense) and blue areas more positive sites (e.g. protonated).
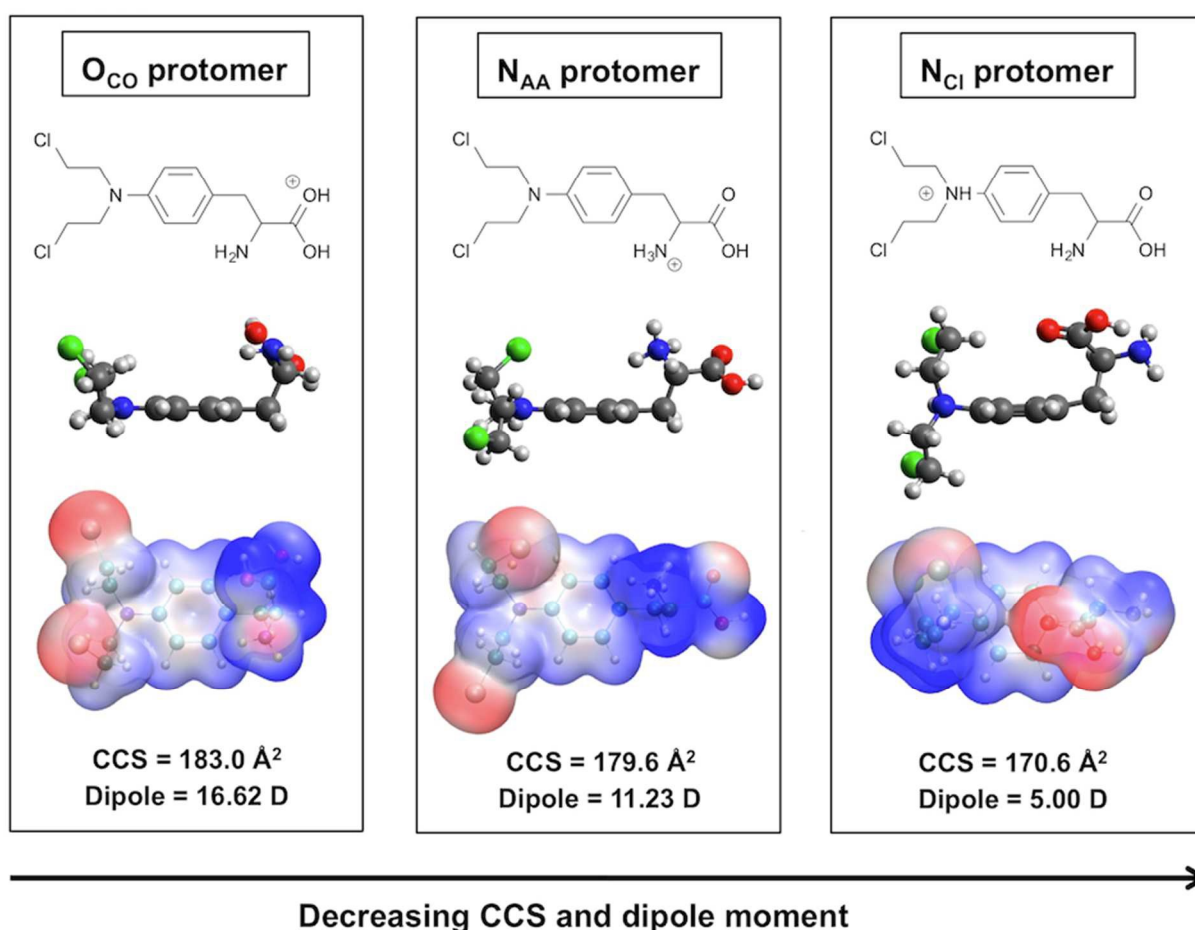
604

605

606 Figure 5

607 Visualisation of the experimental and calculated (both "standard" and "high" level) CCS

608 values (Å^2) from Table 2 and 3. Only 4 of the 7 compounds investigated (**I**, **IV**, **VI**, and **VII**)

609 are separated experimentally into protomer pairs (**I'/I''** etc.), while for all of them CCS

610 values were calculated for the 2 most plausible isoforms.

611
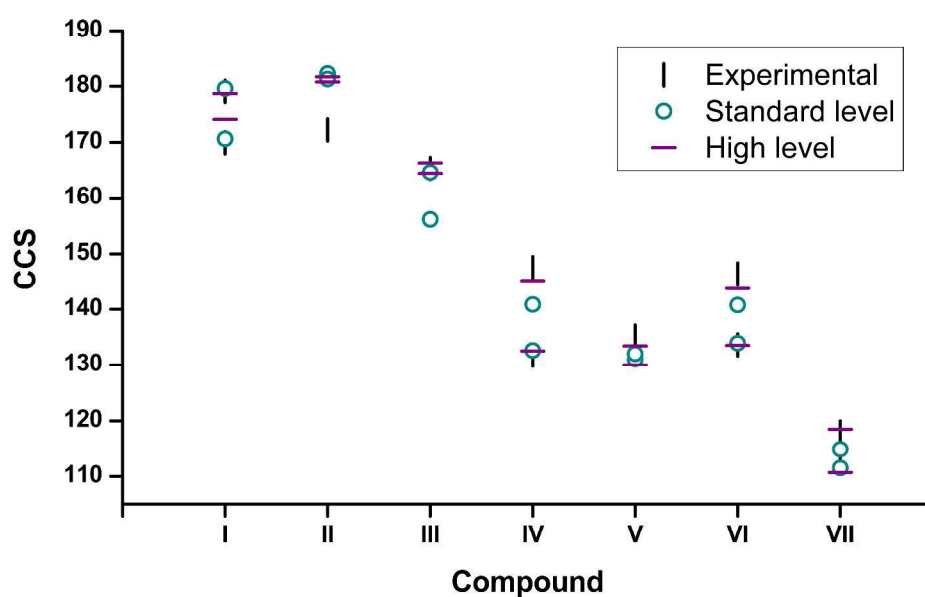
612

613

614

615   Figure 6

616   Correlation between calculated ΔCCS (Å^2) and Δ dipole moment (D) values for protomer

617   pairs, using "standard" and high-level DFT calculations and the nitrogen-based MobCal

618   software. For the 4 experimentally observed protomer pairs, the dipole moments were

619   calculated using the best-fitting DFT level (apart from melphalan, all "high" level). At "high"

620   level a linear fit (R^2=0.8784) suggests itself with aniline as an outlier, while at "standard"

621   level, many calculated (Δ)CCS deviate from the experiment and no correlation is found

622   with the calculated dipoles (R^2=0.1543).

623

624
625
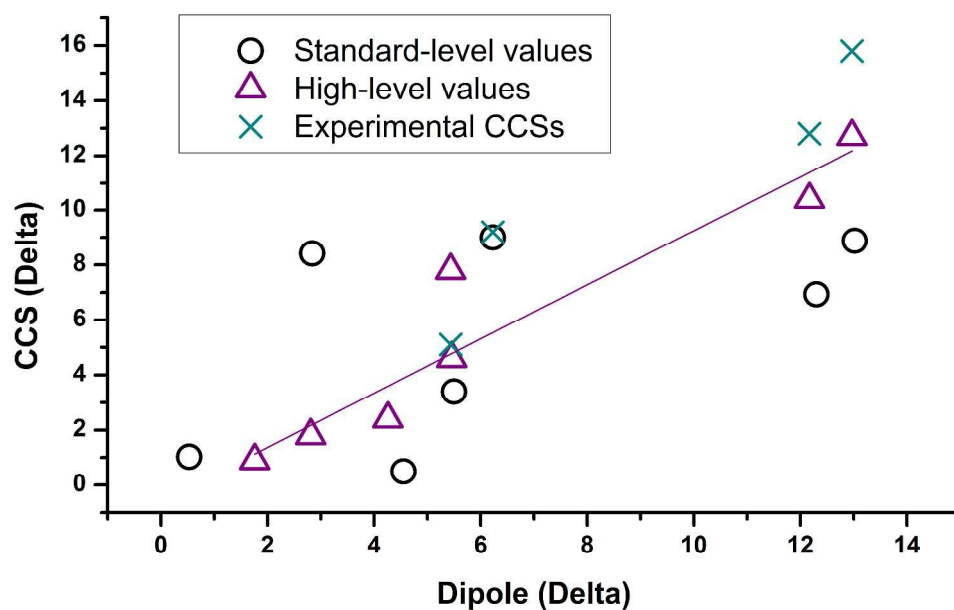
1
2
3
4   626   **<u>Tables</u>**
5
6   627
7   628   Table 1
8
9   629   Experimental CCS_{N2} values derived from TWIMS. MobCal-calculated CCS_{N2} using both
10
11   630   "standard" and high-level DFT optimizations are given for comparison. A detailed
12
13   631   overview of the calculated values can be found in Tables 2 and 3, together with calculated
14
15   632   energies, Boltzmann weights and dipole moments.

Analyte	Observed signal(s)	Protonation site	$CCS_{exp.}$ ($Å^2$) (Synapt G2 HDMS)	$CCS_{calc.}$ ($Å^2$)	
				Standard level	**High level**
Melphalan	**I'**	N_{Cl}	<u>169.9 (± 1.5)</u>	<u>170.6</u>	174.1
(Mel)	**I"**	N_{AA}	<u>179.1 (± 0.9)</u>	<u>179.6</u>	178.7
Dimethoxymelphalan	**II**	N_{OCH3}	172.2 (± 0.9)	182.3	181.7
($DOCH_3$)		N_{AA}		181.3	180.8
Dihydroxymelphalan	**III**	N_{OH}	<u>165.3 (± 0.6)</u>	164.6	<u>166.3</u>
(DOH)		N_{AA}		156.1	<u>164.4</u>
para-benzocaine	**IV'**	O_{CO}	<u>131.7 (± 0.8)</u>	132.6	<u>132.5</u>
	IV"	N_{NH2}	<u>147.5 (± 0.6)</u>	140.9	<u>145.2</u>
ortho-benzocaine	**V**	O_{CO}	135.2 (± 0.3)	131.0	<u>131.0</u>
		N_{NH2}		132.0	<u>133.4</u>
meta-benzocaine	**VI'**	O_{CO}	<u>133.6 (± 1.1)</u>	133.9	<u>133.5</u>
	VI"	N_{NH2}	146.4 (± 0.8)	140.8	<u>143.8</u>
Aniline	**VII'**	ring (*para*-)	<u>112.9 (± N/A)</u>	111.5	<u>110.7</u>
	VII"	N_{NH2}	<u>118.0 (± 2.8)</u>	114.9	<u>118.5</u>

<u>Underlined</u> values represent best matching calculated and experimental CCS values

633

634

635   Table 2

636   Overview of the 5 lowest-energy melphalan structures for each protomer. Structures were
637   acquired after conformational analysis of the protonated molecule, followed by
638   standard-level DFT optimization. The global E ranking and energies (relative to the
639   lowest-energy structure) give an indication of which protonation sites are most favored in
640   the gas-phase. Note the significantly different dipole moments for the N_{AA}/O_{CO} and N_{Cl}
641   structures.

Protonation site	E-ranking (relative)	E-ranking (global)	Rel. E (kcal/mol)	Boltzmann weight (%)	Dipole (Debye)	$CCS_{calc.}$ (Å^2)
$N_{AA}{}^{†*}$	<u>1</u>	<u>1</u>	<u>0.0000</u>	<u>10.87</u>	<u>11.23</u>	<u>179.6</u>
	2	2	0.0220	10.47	11.49	177.3
	3	3	0.0878	9.37	11.38	181.0
	4	4	0.1908	7.87	12.66	179.2
	5	5	0.1995	7.76	12.26	176.6
N_{Cl}	<u>1</u>	<u>17</u>	<u>1.0718</u>	<u>1.78</u>	<u>5.00</u>	<u>170.6</u>
	2	23	1.2669	1.28	7.55	170.5
	3	30	2.6268	0.13	7.72	170.1
	4	33	3.1156	0.06	4.77	172.4
	5	37	3.6565	0.02	4.70	177.5
O_{CO}	<u>1</u>	<u>151</u>	<u>31.0944</u>	<u>0.00</u>	<u>16.62</u>	<u>183.0</u>
	2	152	34.4133	0.00	18.33	182.3
	3	153	34.4195	0.00	18.13	185.6
	4	154	35.4411	0.00	22.65	184.1
	5	155	37.1824	0.00	16.86	181.2

<u>Underlined</u> values represent calculated CCS values for lowest-energy structures
† Expected protonation site in solution (i.e. based on pK_a)
* Favoured protonation site in the gas phase (i.e. based on Boltzmann weights)

642

643

1
2
3
644   Table 3

645   Overview of the lowest-energy structure for each protomer of dimethoxy- and

646   dihydroxymelphalan (DOCH$_3$ and DOH). Structures were acquired after conformational

647   analysis of the protonated molecule and both "standard" and high-level DFT optimization.

648   Note that, unlike melphalan, N$_{AA}$ is the least favored protomer. Similar dipole moments

649   are observed for each protonation site. The global energy ranking is given for each

650   compound and per DFT optimization level.

Analyte	Protonation site	DFT opt. level	E-ranking (per level)	Relative E (kcal/mol)	Boltzmann weight (%)	Dipole (Debye)	CCS$_{TM,N2}$ (Å^2)	CCS$_{exp.}$ (Å^2)
Mel (I)	N$_{AA}$†*	Standard	1	0.0000	10.87	11.23	179.6	179.1 (± 0.9)
		High	1	0.0000	8.23	10.55	178.7	
	N$_{Cl}$	Standard	17	1.0718	1.78	5.00	170.6	169.9 (± 1.5)
		High	5	0.2027	5.85	5.09	174.1	
DOCH$_3$ (II)	N$_{AA}$†	Standard	126	7.5470	0.00	8.50	181.3	172.2 (± 0.9)
		High	119	5.7718	0.01	9.78	180.8	
	N$_{OCH3}$*	Standard	1	0.0000	15.76	7.97	182.3	
		High	1	0.0000	11.98	8.02	181.7	
DOH (III)	N$_{AA}$†	Standard	26	4.7214	0.01	3.73	156.1	165.3 (± 0.6)
		High	34	5.3294	0.01	9.80	166.3	
	N$_{OH}$*	Standard	1	0.0000	35.99	6.57	164.6	
		High	1	0.0000	36.68	6.99	164.4	

† Expected protonation site in solution (i.e. based on pK$_a$)
* Favoured protonation site in the gas phase (i.e. based on Boltzmann weights)

651

652

653 Table 4

654 Overview of the lowest-energy structure for each of the benzocaine and aniline
655 protomers, after conformational analysis of the protonated molecule and both "standard"
656 and high-level DFT optimization. For all molecules (apart from *ortho*-benzocaine)
657 significantly different CCS values are observed for both protomers. Standard-level DFT
658 optimization tends to underestimate CCS values, but similar dipole moments are
659 observed at both levels anyway.

660

Analyte	Protonation site	DFT opt. level	E-ranking (per level)	Relative E (kcal/mol)	Boltzmann weight (%)	Dipole (Debye)	$CCS_{TM,N2}$ ($Å^2$)	$CCS_{exp.}$ ($Å^2$)
para-benzocaine (**IV**)	O_{CO}*	Standard	1	0.0000	67.22	2.73	132.0	131.7 (± 0.8)
		High	1	0.0000	74.22	2.62	132.5	
	N_{NH2}†	Standard	5	11.6453	0.00	15.75	140.9	147.5 (± 0.6)
		High	5	11.9007	0.00	15.59	145.2	
ortho-benzocaine (**V**)	O_{CO}	Standard	5	9.9485	0.00	0.99	131.0	135.2 (± 0.3)
		High	3	9.9485	0.00	1.04	131.0	
	N_{NH2}†*	Standard	1	0.0000	68.58	5.54	131.5	
		High	1	0.0000	74.59	5.30	133.4	
meta-benzocaine (**VI**)	O_{CO}*	Standard	1	0.0000	45.62	0.96	133.9	133.6 (± 1.1)
		High	1	0.0000	57.02	0.87	133.5	
	N_{NH2}†	Standard	5	1.3987	4.30	13.25	140.8	146.4 (± 0.8)
		High	5	2.4900	0.85	13.05	143.8	
Aniline (**VII**)	ring (*para*-)*	Standard	1	0.0000	99.22	1.67	111.5	112.9 (± N/A)
		High	1	0.0000	90.60	1.62	110.7	
	N_{NH2}†	Standard	2	2.9687	0.68	7.18	114.9	118.0 (± 2.8)
		High	2	1.3535	9.36	7.07	118.5	

† Expected protonation site in solution (i.e. based on pK$_a$)
* Favoured protonation site in the gas phase (i.e. based on Boltzmann weights)

661

662

663   Table 5

664   Comparison of CCS values calculated for structures with or without their charge

665   distribution. Apart from melphalan (**I**), the effect of the charge distribution is essential in

666   order to calculate a distinct CCS for each of the compound's protomers.

Analyte	Protonation site	$CCS_{calc.}$ ($Å^2$)	
		Charge distr.	No charge distr.
Mel	N_{Cl}	170.6	149.0
(**I**)	N_{AA}	179.6	155.9
$DOCH_3$	N_{OCH3}	181.7	161.8
(**II**)	N_{AA}	180.8	161.5
DOH	N_{OH}	166.3	144.4
(**III**)	N_{AA}	164.4	142.0
para-benzocaine	O_{CO}	132.5	109.2
(**IV**)	N_{NH2}	145.2	109.4
ortho-benzocaine	O_{CO}	131.0	105.2
(**V**)	N_{NH2}	133.4	106.1
meta-benzocaine	O_{CO}	133.5	109.4
(**VI**)	N_{NH2}	143.8	108.8
Aniline	ring (*para*-)	110.7	76.6
(**VII**)	N_{NH2}	118.5	76.6

667