Automatic Compound Annotation from Mass Spectrometry Data Using MAGMa

Lars Ridder,*1,2 Justin J. van der Hooft,3 and Stefan Verhoeven2
1 Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands
2 Netherlands Science Center, Science Park 140, 1098 XG Amsterdam, The Netherlands
3 Glasgow Polyomics, Garscube Estate, University of Glasgow, Bearsden, G61 1QH, United Kingdom

The MAGMa software for automatic annotation of mass spectrometry based fragmentation data was applied to 16 MS/MS datasets of the CASMI 2013 contest. Eight solutions were submitted in category 1 (molecular formula assignments) and twelve in category 2 (molecular structure assignment). The MS/MS peaks of each challenge were matched with in silico generated substructures of candidate molecules from PubChem, resulting in penalty scores that were used for candidate ranking. In 6 of the 12 submitted solutions in category 2, the correct chemical structure obtained the best score, whereas 3 molecules were ranked outside the top 5. All top ranked molecular formulas submitted in category 1 were correct. In addition, we present MAGMa results generated retrospectively for the remaining challenges. Successful application of the MAGMa algorithm required inclusion of the relevant candidate molecules, application of the appropriate mass tolerance and a sufficient degree of in silico fragmentation of the candidate molecules. Furthermore, the effect of the exhaustiveness of the candidate lists and limitations of substructure based scoring are discussed.

Please cite this article as: L. Ridder, et al., Automatic Compound Annotation from Mass Spectrometry Data Using MAGMa, Mass Spectrom (Tokyo) 2014; 3(3): S0033; DOI: 10.5702/massspectrometry.S0033

Keywords: mass spectrometry, small molecule identification, metabolomics, MS/MS fragmentation, substructures

INTRODUCTION

The identification of unknown molecules is a major bottleneck in untargeted LC-MS based metabolite profiling. In the absence of data from reference compounds or NMR experiments, fragmentation data from tandem mass spectrometry (MS/MS) may be used to derive chemical information about the unknown molecules. The interpretation of the product ion spectra by experts is labor intensive, and several academic and commercial projects are dedicated to the development of software tools to facilitate this task. For example, computer algorithms are used to derive the most likely elemental formula of a mass peak, based on accurate mass and/or isotope ratios, or to provide candidate chemical structures on the basis of MS fragmentation patterns. One strategy to achieve the latter is to search for similar spectra in mass spectral databases. However, in the absence of informative reference spectra, computational methods are needed to annotate fragment ion spectra solely on the basis of the chemical structures of potential candidate molecules. A common approach is to select the best matching chemical structures from comprehensive sets of candidate molecules. Such a procedure must assess how well the observed fragment peaks can be explained by a given candidate structure, e.g. based on substructures, or rule-based fragments. Different scoring functions based on matching the in silico fragments with observed peaks have been applied to prioritize lists of candidate structures, including different regio-isomers.

Recently, we have introduced an extended substructure-based algorithm (MAGMa) capable of handling spectral trees generated by multistage accurate MSn, resulting in hierarchical and internally consistent trees of substructures representing the subsequent fragmentation steps. The method has been used to perform a complete annotation of an LC-MSn metabolite profile of green tea, illustrating its utility at the scale of a metabolomics experiment. Although the MAGMa algorithm was developed to be able to handle MSn multistage fragmentation data with n≥2, its performance on MS/MS data was demonstrated to be comparable to other methods.

In 2012, the CASMI (Critical Assessment of Small Molecule Identification) contest was initiated, which challenged scientists to interpret a number of blinded MS/MS spectra and to provide ranked lists of solutions, both for the correct molecular formula and the correct chemical structure. This allowed identification experts and algorithm developers to objectively compare the reliability of their methods. The second edition of the CASMI contest, organized in 2013, comprised of accurate MS and MS/MS spectra for 16 compounds. The competition was aimed at finding the correct chemical structure (category 2), as well as the correct molecular formulas for 12 of these 16 challenges (category

*Correspondence to: Lars Ridder, Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands, e-mail: lars.ridder@wur.nl

© 2014 The Mass Spectrometry Society of Japan
1). Ranked lists of multiple solutions could be submitted. Here, we describe the application of the MAGMa algorithm to solve these challenges. We submitted solutions to 12 of the 16 challenges in category 2, as well as the corresponding unique molecular formulas for 8 challenges in category 1. In a retrospective analysis we also present the MAGMa results on the remaining challenges and discuss the influence of some of the MAGMa processing parameters in relation to specific challenges.

**METHODS**

**Input file preparation**

The peak lists provided by the CASMI contest were first converted to MAGMa input files. The MAGMa input format for multistage MS fragmentation data is an extension of the more commonly used peak lists, such as those provided for the CASMI challenges, enabling the description of a hierarchy of fragment peaks. After a given precursor peak, the next level MS fragment peaks can be listed between parenthesis. Parenthesis can be used recursively, but for the MS/MS data supplied by CASMI only two hierarchical levels were involved: the precursor mono-isotopic peak taken from the MS file and all fragment peaks taken from the MS/MS file. As an example, the MAGMa input for challenge 1 is shown in Fig. 1.

**MAGMa processing**

The input data was processed with MAGMa in command line mode or via the web interface (www.emetabolomics.org/magma). The allowed mass deviation was set to 10 ppm, with an absolute minimum of 0.002 Da, except for challenges 4, 5, and 6, for which the mass tolerance for retrieving candidate molecules was set to 4 ppm, in accordance with the provided technical details. MAGMa retrieved candidate molecules from a local database consisting of all PubChem compounds on 2 December 2013, up to a mass of 1200 Da and composed of the elements C, H, N, O, P, and S only, assuming [M+H]+ or [M−H]− precursor ions for positive and negative mode data, respectively. For challenges 7 and 8, candidates with the provided elemental composition were retrieved directly from the online PubChem service in SDF format and used as input for the MAGMa calculations. Also for challenge 9, where we recognized from the isotope ratios in the MS file that the molecule contained chlorine atoms, we obtained candidates from an online query on mono-isotopic mass. Since the CASMI deadline, we have added a local MAGMa database containing halogenated PubChem compounds. This extension allowed us to apply the default MAGMa processing retrospectively on the data for the halogenated compounds of challenges 9, 15, and 16.

MAGMa systematically fragments all candidate molecules one by one, as described previously. Substructures were generated by applying up to 3 bond dissociations. This set of substructures was expanded with single losses of hydroxyl groups (representing neutral loss of water) or primary amine groups (representing neutral loss of ammonia). A substructure penalty score is calculated as the sum of the penalty values for all dissociated bonds (1, 2, and 3...).
Candidate ranking

The fragment based penalty score provided by MAGMa is lowest for the most likely candidates and highest for the poorly matching candidates, which is opposite to the requirements for CASMI. For the CASMI entries we therefore provided a reversed ranking value, which was primarily based on the candidate score obtained from the substructure annotation. Equally scoring candidates were further ordered on the basis of their PubChem refscore, consistent with the “refined ranking” described previously. Candidates equal in both scores were assigned the same value corresponding to their average ranking position.

RESULTS

Solutions for twelve of the sixteen challenges in category 2 were submitted to the CASMI competition (Table 1, column 3). In six of the MAGMa solutions, the correct candidate was ranked in first position. In three other cases the correct candidate was in the top 5, while for the remaining three challenges lower rankings were obtained. For challenge 3, the ranking of the correct molecule among the first 17 candidates was the best solution among all participants.

MAGMa results for the remaining challenges 11, 12, 15, and 16

For challenges 11, 12, 15, and 16 no solutions were submitted. Here we describe the initial results as well as a number of retrospective calculations for further analysis. For challenge 11, three fragment peaks at \( m/z = 119.0538 \), \( m/z = 143.0534 \), and \( m/z = 158.0396 \) could initially not be explained by substructures even from the best scoring candidates, assuming a maximum mass error of 0.002 Da. This lead us to the assumption that the correct chemical structure was not present in the candidate list. In retrospect, two of the three tautomers of the correct candidate were present and it is more likely that the assumed maximum error of 10 ppm and 0.002 Da was too restrictive for a proper interpretation of the MS/MS spectrum. After the solutions were announced, we reprocessed the data using a minimum mass tolerance of 0.005 Da, resulting in acceptable substructure matches for the above mentioned peaks and in reasonable ranking positions of the two tautomers of demethoxycurcumin present (Table 1). Also for challenge 12, the correct candidate was in fact present in the initial PubChem candidate set, although at a low ranking position, as presented in

---

Table 1. Overview of the MAGMa results on the CASMI 2013 challenges, category 2.

<table>
<thead>
<tr>
<th></th>
<th>Compound</th>
<th>Entry candidates</th>
<th>All PubChem candidates</th>
<th>&lt;1000 candidates</th>
<th>Refsc. filter Rank</th>
<th>&lt;500 candidates</th>
<th>Refsc. filter Rank</th>
<th>&lt;200 candidates</th>
<th>Refsc. filter Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feruloyl tyramine</td>
<td>1084</td>
<td>1</td>
<td>858 &gt;6</td>
<td>1</td>
<td>422 &gt;10</td>
<td>1</td>
<td>181 &gt;11</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Feruloyl putrescine</td>
<td>631</td>
<td>3</td>
<td>856 &gt;4</td>
<td>7</td>
<td>486 &gt;6</td>
<td>1</td>
<td>185 &gt;11</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Acetyl-Gln-Leu-amide</td>
<td>370</td>
<td>17</td>
<td>370 &gt;10</td>
<td>17</td>
<td>370 &gt;10</td>
<td>17</td>
<td>85 &gt;1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4-Dihydroxochalcone</td>
<td>825</td>
<td>78</td>
<td>825 &gt;0</td>
<td>78</td>
<td>464 &gt;1</td>
<td>60</td>
<td>188 &gt;35</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Isopentolane</td>
<td>350</td>
<td>2</td>
<td>350 &gt;0</td>
<td>2</td>
<td>350 &gt;0</td>
<td>2</td>
<td>162 &gt;12</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Phosphatidy-l-6-acetyl-glucose</td>
<td>2</td>
<td>1</td>
<td>2 &gt;0</td>
<td>1</td>
<td>2 &gt;0</td>
<td>1</td>
<td>2 &gt;0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Convolvulain A3</td>
<td>17</td>
<td>17</td>
<td>17 &gt;0</td>
<td>1</td>
<td>17 &gt;0</td>
<td>1</td>
<td>17 &gt;0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>Prodelphinidin C2</td>
<td>1</td>
<td>1</td>
<td>1 &gt;0</td>
<td>1</td>
<td>1 &gt;0</td>
<td>1</td>
<td>1 &gt;0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Chlorpyrifos</td>
<td>113</td>
<td>1</td>
<td>178 &gt;0</td>
<td>1</td>
<td>178 &gt;0</td>
<td>1</td>
<td>178 &gt;0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>VAL-HIS-LEU-THR-PRO-VAL-GLU-LYS</td>
<td>20</td>
<td>20</td>
<td>20 &gt;0</td>
<td>1</td>
<td>20 &gt;0</td>
<td>1</td>
<td>20 &gt;0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>Demethoxycurcumin</td>
<td>4719</td>
<td>51</td>
<td>906 &gt;5</td>
<td>6</td>
<td>422 &gt;9</td>
<td>6</td>
<td>172 &gt;12</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Demethoxycurcumin (tautomer 1)</td>
<td>4719</td>
<td>25</td>
<td>906 &gt;5</td>
<td>4</td>
<td>422 &gt;9</td>
<td>9</td>
<td>172 &gt;12</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Baicalein</td>
<td>813</td>
<td>271</td>
<td>813 &gt;0</td>
<td>271</td>
<td>445 &gt;1</td>
<td>192</td>
<td>194 &gt;9</td>
<td>95</td>
</tr>
<tr>
<td>14</td>
<td>EST, Aloxistatin</td>
<td>207</td>
<td>42</td>
<td>207 &gt;0</td>
<td>42</td>
<td>207 &gt;0</td>
<td>42</td>
<td>184 &gt;11</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>Tetrahydroalstonine</td>
<td>1583</td>
<td>5</td>
<td>736 &gt;6</td>
<td>4</td>
<td>449 &gt;8</td>
<td>1</td>
<td>195 &gt;11</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>2-(Perfluoroxy)ethanol</td>
<td>720</td>
<td>2</td>
<td>720 &gt;0</td>
<td>2</td>
<td>382 &gt;2</td>
<td>1</td>
<td>182 &gt;12</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Olloxacinn</td>
<td>8172</td>
<td>699</td>
<td>998 &gt;8</td>
<td>78</td>
<td>447 &gt;11</td>
<td>33</td>
<td>189 &gt;18</td>
<td>12</td>
</tr>
<tr>
<td>18</td>
<td>Olloxacinc</td>
<td>8172</td>
<td>126</td>
<td>998 &gt;8</td>
<td>18</td>
<td>447 &gt;11</td>
<td>6</td>
<td>189 &gt;18</td>
<td>3</td>
</tr>
</tbody>
</table>

1 Number of candidates were reduced by filtering on refscore >5.
2 New list of candidates retrieved after implementing local database of halogenated PubChem compound.
3 Results obtained with increased number of substructures by breaking up to 5 bonds.
4 The MAGMa results obtained with the selected (<1000) candidate lists can be accessed online via http://www.emetabolomics.org/results_CASMI2013.html.
Table 1. In this case the poor ranking was due to chemical rearrangements not properly represented by substructures as discussed later.

For challenges 15 and 16 the correct molecules were not present in the initial MAGMa candidate list, because the MAGMa database of PubChem compounds did not contain fluorine containing molecules. After the CASMI deadline we have added a second MAGMa compound database with halogenated compounds from PubChem, from which additional candidates can be retrieved optionally. The results with this extended candidate sets are now included in Table 1. The input data was constructed by generating composite spectra of the MS/MS spectra obtained at the different collision energies for both challenges, i.e. peaks obtained at both energies are merged into a single spectrum, taking the highest intensity for ions present at both energies. Nominal \( m/z \) values were replaced by accurate \( m/z \) values when available from the MS spectra and otherwise ignored. For challenge 16, several peaks from the higher energy spectrum (with \( m/z=122.0408, \ m/z=178.03072, \ m/z=179.03857, \ m/z=193.04172, \ m/z=194.02574, \ m/z=199.05114, \ m/z=205.04175, \ m/z=219.05744, \) and \( m/z=221.07309 \) could not be explained by substructures generated with the standard number of bond dissociations (up to 3 bond dissociations and one further loss of an OH or NH2 groups). Because of the fused ring system, we suspected that a number of the observed fragments may involve more than 3 bond cleavages. Indeed, repeating the MAGMa analysis with up to 5 bond dissociations (without OH or NH2 losses) resulted in substructure annotation of all fragment peaks, and a corresponding improvement of the ranking (Table 1, bottom row). When the same extensive fragmentation was applied to the other challenges, the rankings did not change significantly (results not shown), except for challenge 13, where the ranking of aloxistatin deteriorated slightly from 42 to 58, and for challenge 15, where 2-(perfluorooctyl)ethanol was ranked on top instead of the second position of the full list of 720 candidates.

Reducing the number of candidate structures

Automatic methods allow the evaluation of large numbers of molecules for a given MS/MS spectrum. MAGMa, like other automatic methods, can retrieve thousands of candidate molecules from PubChem. The evaluation of such comprehensive lists of candidate molecules is not possible in a manual approach, which is often guided by literature and experience with known compound classes. Manual analysis will therefore, at least initially, be biased towards well-known molecules, whereas automatically evaluated candidate molecules from PubChem may include large numbers of synthetic molecules with limited relevance to the sample of interest. To mimic the bias of manual analysis, MAGMa reports the refscore, as explained in the Methods section, which can be considered as an indicator of the general relevance of a compound, with high values indicating the more well-known and widely studied molecules. In the initial solution submitted to the CASMI contest, we have used a threshold value of 5 to filter the large numbers of candidates initially obtained for challenges 1, 2, and 14 (Table 1). Table 1 (columns 4–7) also presents a more systematic analysis of the effect of filtering the candidate sets on the basis of the refscore. Without filtering, the total number of matching candidates in PubChem varies significantly between the different challenges, ranging from 1 to over \( 10^3 \) candidates. For each challenge we stepwise raised the threshold on the refscore until the candidates sets were reduced to \(< 1000\), \(< 500\), and \(< 200\) molecules. Figure 3 shows the filter option in the MAGMa results page obtained for challenge 1. Table 1 presents the actual numbers of candidates obtained for each challenge following this procedure, as well as the MAGMa rankings of the correct molecules within these candidate sets. All but one of the filtered candidate lists still contained the correct compound, indicating that all compounds in the CASMI 2013 challenge are relatively well-known molecules. Only for challenge 3, the correct compound, N2-acetyl glutaminyl leucinamide, was lost when the candidate list was reduced to less than 200 mol-

![Fig. 3. The MAGMa results page for challenge 1. The optional filter on the refscore column is highlighted.](image-url)
The MAGMa results obtained for all challenges, with selected (<1000) candidate lists, can be accessed online via http://www.emetabolomics.org/results_CASMI2013.html.

Chemical rearrangements

From Table 1 it can be concluded that the most difficult cases for MAGMa were challenges 4, 12, and 13. In challenge 4, the MS/MS spectrum of dihydrochalcone includes a peak due to loss of water. This water loss must consist of the carbonyl oxygen and two hydrogen atoms from the aliphatic backbone, which is not covered by substructures generated by MAGMa and results in the maximum penalty score. Consequently, a large number of molecules containing hydroxyl, epoxide or another cyclic ether, from which MAGMa does allow loss of water, obtained a better score. For baicalein (challenge 12), closer inspection of the data indicated that major MS/MS fragments involve loss of CO and CH₂O₂. These neutral losses are associated with chemical rearrangements of the heterocyclic part of the flavon structure. Such chemical rearrangements are not accounted for by MAGMa, and as a result, the substructures matched with these m/z values obtained high penalty scores. Finally, aloxistatin obtained a poor score in MAGMa because the most intense peak at m/z=246.1331 could not be matched with a substructure. This ion most likely represents an unusual rearrangement product. Entries for this challenge came only from three participants applying automatic methods on large candidate sets and none of them identified aloxistatin within the top 10. The results for challenges 4, 12, and 13 highlight a limitation of substructure-based methods by not accounting for chemical rearrangements.

To assess whether rule-based fragmentation could overcome this limitation for the above challenges we submitted the three molecules, dihydrochalcone, baicalein, and aloxistatin to one of the commercial rule-based software packages, Mass Frontier 7.0, using a fragmentation library search to predict fragments for baicalein in negative ion mode. The fragments predicted by Mass Frontier included the loss of water from dihydrochalcone as well as the rearrangement of the heterocyclic part of baicalein corresponding to loss of carbon monoxide, illustrating the potential advantage of rule-based fragmentation. However, two other fragments of baicalein, with m/z=169.0675 (C₅O₂ loss) and m/z=223.0396 (CH₂O₂ loss), as well as the m/z=246.13 fragment of aloxistatin were not predicted by Mass Frontier 7.0, emphasizing that they are difficult to annotate in an automatic way, by both substructure and rule-based methods.

Category 1 entries

Based on the MAGMa entries in category 2, we also submitted lists of unique molecular formulas to the category 1 contest. Each molecular formula was assigned the highest score (reversed ranking value) at which it was present in the category 2 entry list. Isotope ratios given in the MS data files were not used for the category 1 submissions. Based on the additional MAGMa analyses described above, Table 2 also lists the results for the remaining challenges in which we did not participate. For all challenges, except challenge 12, MAGMa identified the correct molecular formula, even if the top ranking chemical structure was not correct. This is in line with earlier findings based on other applications of the MAGMa algorithm. It indicates that accurate mass fragmentation data often provides sufficient information to pick the right molecular formula from a large set of molecules without using additional information, such as isotope peak ratios of the precursor ion, and even if the top-ranked chemical structure is not correct.

**DISCUSSION**

Although MAGMa performed well on most CASMI challenges, a number of lessons can be learned about the advantages and limitations of its application. For example, we did not participate in four challenges of the contest and the retrospective analyses provided hints to avoid such misses in future exercises.

The first obvious enhancement of MAGMa is the inclusion of halogenated compounds in the local databases of candidate molecules, which was required to obtain good solutions for challenges 15 and 16. This holds for challenge 9 as well, but in this case the need for halogenated candidate compounds had been recognized on the basis of the isotope peaks typical of chlorine and therefore relevant candidates had been retrieved from the online PubChem database already. We have now extended the MAGMa webinterface with an option to supplement the candidate sets with halogen containing compounds, by which satisfactory entries for challenges 15 and 16 could be generated. Secondly, the ofloxacin dataset of challenge 16 provided a clear example of a molecule where the default level of *in silico* fragmentation in MAGMa was insufficient. Retrospectively, up to 5 bonds had to be broken to generate the proper substructures for some of the MS/MS peaks. A third lesson learned was not to apply a too narrow mass tolerance, which caused us to incorrectly conclude from the initial results for challenge 11 that no correct molecule was present in the candidate list. These observations illustrate that although automated methods can facilitate certain aspects of the identification process, experts are still needed to properly apply these methods by integrating different factors and types of information.

After having recognized and “tackled” these basic aspects of the MAGMa processing, the rankings obtained for some
of the challenges remain unsatisfactory. Automatic methods allow the consideration of large numbers of candidate molecules, e.g. retrieved from comprehensive chemical databases such as Pubchem. On the positive site, this may yield relevant results for mass data where manual methods that consider only small numbers of candidates may result in incomplete annotation. This is illustrated by the CASMI results for challenge 3, where the correct molecule was only present in two automatic entries with larger numbers of candidates. However, another consequence of processing large candidate sets is that also molecules with limited (biological or clinical) relevance may be included. These may include homologues of the correct molecules that may not be distinguishable on the basis of MS/MS data. To deliberately bias the results towards known compounds, the MAGMa interface allows filtering of the PubChem candidates (before or after in silico fragmentation) based on "refscore," a simple statistic derived from the PubChem substances table. The filtering threshold on refscore could be stepwise increased to restrict candidate sets to less than 1000, less than 500, and less than 200 molecules. The rankings of the correct molecules improved, in absolute terms, with decreasing list sizes, indicating that this filter may help to focus the analysis towards the most relevant candidates. Although, this approach is not based on the actual experimental data, it may help expert users to do what they would do anyway, i.e., consider well-known molecules first. In the solutions we submitted for challenges 2 and 14, this filtering step improved the result in comparison to the non-filtered lists. Similarly, it improved the rankings in the analysis of challenges 11, 15, and 16. It should be noted, however, that while this filter can be useful for the annotation of known compounds, as in the case of most challenges in this context, it should be applied carefully in practice when samples also contain less common compounds. This is exemplified by challenge 3 for which the synthetic product N2-acetyl glutaminyl leucinamide had the lowest possible refscore of 1 and is lost when a filter on refscore is applied.

In addition to the above considerations, the results show that there is room for improving the substructure-based scoring. Major peaks in the spectra of baicalein and al- oxistatin result from chemical rearrangements that cannot be adequately described by substructures. Either no or very unfavorable substructures are matched with these peaks resulting in maximum penalties contributing to the overall candidate scores. In these cases, more favorable scores are obtained for alternative molecules for which these fragment peaks can be matched with simpler substructures. It indicates that there is still a need for improved computational algorithms that combine the capability to handle chemical rearrangements with the speed to evaluate large numbers of candidate molecules.

Finally, the fact that the selection of candidate molecules, prior to MS/MS based scoring, has a significant effect on the rankings might be considered in the design and evaluation of the contest. A format in which candidate selection and subsequent automatic scoring are evaluated independently may provide more detailed insight into the differences between the automated methods. One possibility could be to define a separate category in which participants are asked to apply their computational (MS/MS based) scoring algorithm to the same set of candidate molecules. This would enable a more direct comparison of the performance of different in silico fragmentation and scoring algorithms and thereby help to identify bottlenecks and provide directions for further developments.

Acknowledgements

This research was supported by the Netherlands eScience Center, grant number 660.011.302.

REFERENCES


