**Atmospheric
Chemistry
and Physics**
Discussions

Interactive comment on "Technical Note: Development of chemoinformatic tools to enumerate functional groups in molecules for organic aerosol characterization" *by* G. Ruggeri and S. Takahama

Anonymous Referee #2

Received and published: 26 January 2016

Ruggeri and Takahama report on the development and application of chemoinformatic patterns relevant to the formation of organic aerosol (OA). The method uses the Open-Babel toolkit in Python to identify functional groups/molecular structures (largely specified by SMARTS) in individual molecules (specified by SMILES structures). SMARTS specifications for four sets of functional groups/molecular structures and example applications are provided. While other groups have used such methods previously, this is the first attempt that I am aware of to publish SMARTS patterns and scripts of importance for OA formation, along with the demonstration of specificity and the validation of

those patterns and scripts. The conversion of individual molecules to SMILES and to SMARTS is a time consuming and challenging process, and I believe that the community certainly could benefit from the efforts of Ruggeri and Takahama (e.g., particularly the patterns and software).

I am unsure whether Atmospheric Chemistry & Physics is the most suitable journal for this manuscript. Further, in its current form, the manuscript needs to be simplified to be an effective Technical Note. I found the diversity of datasets and example applications distracting. For example, in Fig. 6, the authors are comparing functional group distributions derived from GC /MS measurements made during the late 1990's with functional group distributions derived from FTIR measurements in 2010; also included are estimates of measured/unmeasured fractions for each instrument based on additional (and separate) publications. The authors spend one paragraph describing each of the measurements, the dominant functional groups, and the reasons for the discrepancies between the wood burning samples. In the conclusions section, the authors return to this discrepancy between the GC/MS and FTIR measurements for wood smoke, which has little to do with the overall contribution of the Ruggeri and Takahama work. It would be sufficient to demonstrate that functional group distributions can be obtained using their approach. Further comments are provided below on the inclusion of likely extraneous information that diminishes the potential contribution of this work. It is recommended that the authors consider simplifying the applications presented, and focus more discussion on validation.

Comments: p 33638, line 20-23: The description of the partitioning calculations is confusing. It is not clear what parameters were of interest, or how they were determined. The calculated pure compound vapor pressures could be used to calculate partitioning coefficients, which then could be used (with total OA) to calculate the fraction of each compound in the gas vs. particle phase based on Pankow 1994. Alternatively, calculated particle phase fractions (with total OA) could be used to calculate pure component saturation concentrations based on Donahue et al. 2006. It is suggested that

the Greek symbol zeta not be used to define the partitioning coefficient, as that symbol traditionally has been used to define mole-fraction scale activity coefficients in partitioning calculations.

page 33639/33640, discussion of Fig. 5: It appears the authors are mapping compounds measured by Rogge et al. in vehicle OA and woodsmoke OA, as well as compounds predicted by MCM. If this is correct, the detailed discussion of PMF/AMS HOA and BBOA in the text does not add anything, and only confuses what is presented in Fig. 5. Additionally, the legend should make it clearer that compounds from woodsmoke POA (and not biomass burning) and vehicle POA are plotted. It does not make sense to me that the compounds in woodsmoke POA span a higher volatility range than the compounds predicted by MCM. Particularly given that MCM is typically unable to predict measured OA loadings without invoking particle phase or aqueous phase chemistry. The authors do reference the abundance of MCM compounds in the IVOC region and absence in the LVOC, as well as particle-phase chemistry...maybe the traces are mislabeled and the green trace is MCM?

Table 1: The caption for table 1 is long and contains important information that is not thoroughly discussed in the text. It is recommended that the discussion in the text be expanded to cover the inclusion of special patterns for formaldehyde and formic acid.

Table 3 is not referenced in the text. Is it necessary?

Editorial:

It is recommended that the authors list the groups of patterns in the same order throughout the manuscript. For example, in the abstract and the introduction, SIM-POL.1 appears as the first group, but in section 2.1 it is the second group.

p. 33637, lines 15-19: Awkward. It is recommended that the authors reword this description of the test compounds. If I understand, it seems sufficient to say that compounds were selected either from and then give the list (without repeating "either" or

C11995

"selected").

p. 33639, line 10: Replace *is* with *are* (...but are shown...).

p. 33642, lines 17-19: The discussion of the contribution of peroxacyl nitrate mass vs. O:C is confusing as written. It is suggested that the authors reword these sentences to more clearly articulate that while the total mass contribution is only 26%, the total O:C contribution is 53% due to the high O:C ratio of these functional groups (or the dominant compound).

Fig. 6 caption: Was the OA fraction in gray truly unmeasured? Or was it unresolvable using the technique specified?

Interactive comment on Atmos. Chem. Phys. Discuss., 15, 33631, 2015.