Thoma et al. conducted laboratory experiments in which they generated SOA from O3/OH oxidation of biogenic and anthropogenic VOCs in an oxidation flow reactor, collected SOA filter samples, and analyzed the samples with orbitrap mass spectrometry. SOA tracers identified in the samples were added to a mass spectral database. Ambient PM2.5 samples were likewise analyzed with the same orbitrap mass spectrometry technique and were screened for the SOA tracers identified in the laboratory OFR studies. About half of the signal of the CHO-containing compounds was attributed to those SOA tracers. A principal component analysis on the ambient samples was conducted that identified clusters of compounds corresponding to different atmospheric processes.

Overall this paper describes an innovative analytical approach that can aid in the interpretation of atmospheric oxidative aging processes and source apportionment. I support eventual publication of the paper after consideration of my comments below.

1. Can the authors explicitly comment on whether application of their aerosolomics data base is limited to PM2.5 samples that are analyzed specifically with a Thermo Fisher Orbitrap mass spectrometer and the related Compound Discovery software? I think that this is the case, but if I am wrong, it would be useful to clarify how/where else it can be applied.

2. **L75**. In addition to the irradiance, the external OH reactivity (OHR) also significantly influences the OH exposure (e.g. Li et al., 2015; Peng et al., 2015; Rowe et al., 2020). 50 ppb SO_2 was used in offline calibration experiments to constrain the OH exposure (Fig. S1). The corresponding OHR in the SO_2 calibration experiments is 50 ppb * 2.5e10 molec/cm3 * 9e-13 cm3/molec/s = 1 s^{-1}. However, the OHR in the SOA studies was most likely considerably larger than 1 s^{-1}. Consequently, the OH exposure was probably lower than what was suggested from the calibration data due to OH suppression. For example, in the α-pinene OFR experiment, ~83.9 $\mu g\ m^{-3}$ SOA was generated; assuming an SOA yield of approximately 0.3 (e.g. Lambe et al., 2015) and complete consumption of the α-pinene, the initial α-pinene concentration was approximately 83.9/0.3 = 280 $\mu g\ m^{-3} \approx 51$ ppb, with a corresponding OHR ~ 51*2.5e10*5.33e-11 = 68 s^{-1}. At these conditions, along with the experimental conditions provided in Section 2.1, I estimate that the corresponding OH exposure was approximately $1*10^{11}$ molec/cm^{-3}-sec using the OFR254 OH exposure estimation equation introduced by Peng et al. (2016) and published online here: https://sites.google.com/site/pamwiki/estimation-equations?authuser=0. This calculated OH exposure value is about 10 times lower than that obtained from their SO_2 calibration. This analysis should be applied to the other OFR experiments as well, and a column should be added to Table S1 with the corresponding calculated OH_{exp} for comparison purposes. While the overall conclusions of the paper will remain unchanged, the (likely) lower photochemical age may provide higher confidence in applying the SOA tracers in the authors' aerosolomics database to ambient PM2.5 samples that may also be subject to lower aging timescales.

3. **L228**: The authors state: "Furthermore, ion source dimerization is a known phenomenon that hinders the unambiguous identification of atmospheric dimers, or leads to misinterpretation of results from direct-injection HESI." Can they provide references for this statement?

4. **L264**: The formulas that are listed for terpenylic acid and MBTCA are incorrect: terpenylic acid should be $C_8H_{12}O_4$, MBTCA should be $C_8H_{12}O_6$.

5. **L307** – **L316**: I find the discussion of the "unassigned" signals vague and unsatisfactory. It seems clear from the text that 53.5% of the signal in the ambient samples does not correspond to the SOA tracers obtained from the laboratory OFR studies. However, all that is stated is that "isoprene as well as other monoterpenes to be promising candidates closing this gap". Further, Figure 7 conflates "unassigned" with "unidentified", because information about compound MW, carbon number, and O/C is shown here, though it is difficult to interpret from the figure. At the least, they should list the formulas of the "few compounds with high signal intensities [that] remain unassigned" in the text, and perhaps add a table to the supplement with the corresponding information that is shown in Figure 7, rather than interpreting the result without providing the data to support the hypothesis as was done here.

References

R. Li, B.B. Palm, A.M. Ortega, J. Hlywiak, W. Hu, Z. Peng, D.A. Day, C. Knote, W.H. Brune, J. de Gouw, and J. L. Jimenez. Modeling the radical chemistry in an Oxidation Flow Reactor: radical formation and recycling, sensitivities, and OH exposure estimation equation. *Journal of Physical Chemistry A*, 119, 4418–4432, doi:10.1021/jp509534k, 2015.

Z. Peng, D.A. Day, H. Stark, R. Li, B.B. Palm, W.H. Brune, and J.L. Jimenez. HOx radical chemistry in oxidation flow reactors with low-pressure mercury lamps systematically examined by modeling. *Atmos. Meas. Tech.*, 8, 4863-4890, doi:10.5194/amt-8-4863-2015, 2015.

A. T. Lambe, P. S. Chhabra, T. B. Onasch, W. H. Brune, J. F. Hunter, J. H. Kroll, M. J. Cummings, J. F. Brogan, Y. Parmar, D. R. Worsnop, C. E. Kolb, and P. Davidovits. Effect of oxidant concentration, exposure time and seed particles on secondary organic aerosol chemical composition and yield. *Atmos. Chem. Phys.*,15, 3063-3075, 2015.

Rowe, J. P., Lambe, A. T., and Brune, W. H.: Technical Note: Effect of varying the λ = 185 and 254 nm photon flux ratio on radical generation in oxidation flow reactors, Atmos. Chem. Phys., 20, 13417–13424, https://doi.org/10.5194/acp-20-13417-2020, 2020.