# Supporting Information for Technical Note: Relating functional group measurements to carbon types for improved model-measurement comparisons of organic aerosol composition

Satoshi Takahama[1] and Giulia Ruggeri[1]

[1]ENAC/IIE Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland

*Correspondence to:* Satoshi Takahama (satoshi.takahama@epfl.ch)

## S1 Chemoinformatic tools

To construct $\mathbf{Y}$, $\boldsymbol{\Theta}$, and $\mathbf{X}$ (Section 2.2), we use the APRL-SSP program with a minor modification. Ruggeri and Takahama (2016) showed that there can be a one-to-one correspondence between functional groups (FGs) and non-carbon atoms such that the mapping is unambiguous. Past constraints specified that all atoms must be accounted for by at least one FG:

$$\bigcup_{j \in \mathcal{J}} \{a : a \in \mathcal{A}_i, a \in \mathcal{A}_j\} = \mathcal{A}_i \quad \forall\, i \in \mathcal{M}\,, \tag{S1}$$

and that all non-carbon atoms $\mathcal{A}_i \setminus \mathcal{C}_i$ cannot be matched by more than one group:

$$\bigcap_{j \in \mathcal{J}} \{a : a \in \mathcal{A}_i \setminus \mathcal{C}_i, a \in \mathcal{A}_j\} = \varnothing \quad \forall\, i \in \mathcal{M}\,. \tag{S2}$$

Polyfunctional carbon atoms were not included in their validation at the time, but we now impose an additional condition that each FG includes the definition for only one carbon atom (with exceptions noted in Appendix S2):

$$\left| \{a : a \in \mathcal{C}_i, a \in \mathcal{A}_j\} \right| = 1 \qquad \forall\, i \in \mathcal{M}, j \in \mathcal{J}\,. \tag{S3}$$

$|\cdot|$ denotes the cardinality of the set. To satisfy this new condition, we split the `C=O-O` group into three separate groups (`R2C=O-O`, `RHC=O-O`, `H2C=O-O`) as carbon was double counted otherwise. This step is inconsequential from a mass perspective, but important for fulfilling the relationship (equation 3) for the complete APIN MCM mechanism. The corresponding patterns have been updated in the APRL-SSP repository.

## S2 Generalization of carbon types

We note two generalizations to the carbon type descriptions introduced in Section 2.2 that can be considered. First, these carbon types focus on the functionality of each carbon, but do not consider its complete bonding environment (e.g., configuration to other carbon atoms). For instance, carbon atoms defined by functionalization only by hydroperoxide, alkoxy radicals, and

peroxy radicals can differ according to whether the carbon is sp$^3$ or sp$^2$-bonded to other carbon atoms. Hydroxyl groups in phenols are differentiated from alcohols in similar instances, but we have not made this distinction for these three groups as nomenclature for them are not common and also does not affect our analysis. It is possible to define SMARTS patterns to make the differentiation in other applications where carbon type representations are useful.

5     Second, there are several FGs included in pattern development by Ruggeri and Takahama (2016) and present in photooxidation products of 1,3,5-trimethylbenzene (Bloss et al., 2005; Ruggeri et al., 2016) in which two carbon atoms are associated with a single FG: anhydride, ester, and peroxide. In this case, equation S3 would permit 2 carbon atoms for these exceptional groups. To accommodate these groups, the carbon type formulation can be extended to "carbon units" consisting of either one or two atoms, or a correction factor can be introduced to the currently proposed formulation. In the latter approach, the

10   carbon-group matrix $\theta_{kj}$ can be replaced by $\tilde{\theta}_{kj} = \theta_{kj}\gamma_j$ and group-carbon matrix $\phi_{jk}$ replaced by $\tilde{\phi}_{kj} = \phi_{kj}\gamma_j^{-1}$, where $\gamma$ is a coefficient is a correction factor to complete the FG and carbon balances of equations 3 and 4, respectively. $\gamma_j = 0.5$ for these two-carbon FGs, and $\gamma_j = 1$ for the rest. All equalities expressed in this manuscript would hold exactly, except for carbon type oxidation state (equation 8) that is only approximately true for ester groups since one carbon atom is double-bonded to oxygen while the other is only singly bonded to another oxygen atom. However, the overall oxidation state estimate of equation 7 still

15  holds when summed over each molecule that contains both carbon atoms of the ester group.

    When combined with information regarding the carbon skeleton, these carbon types can provide another origin for derivation of chemical basis sets. In Figure S1, molecular abundances for gas and aerosol phases are depicted using carbon types and $n_C$. Together with Figure 2 and definitions in Section 2.2, each of the common basis set dimensions (O, C, H, OS$_C$) used in the aerosol community can be derived. When neighboring interactions among groups are desired, these carbon types can form the

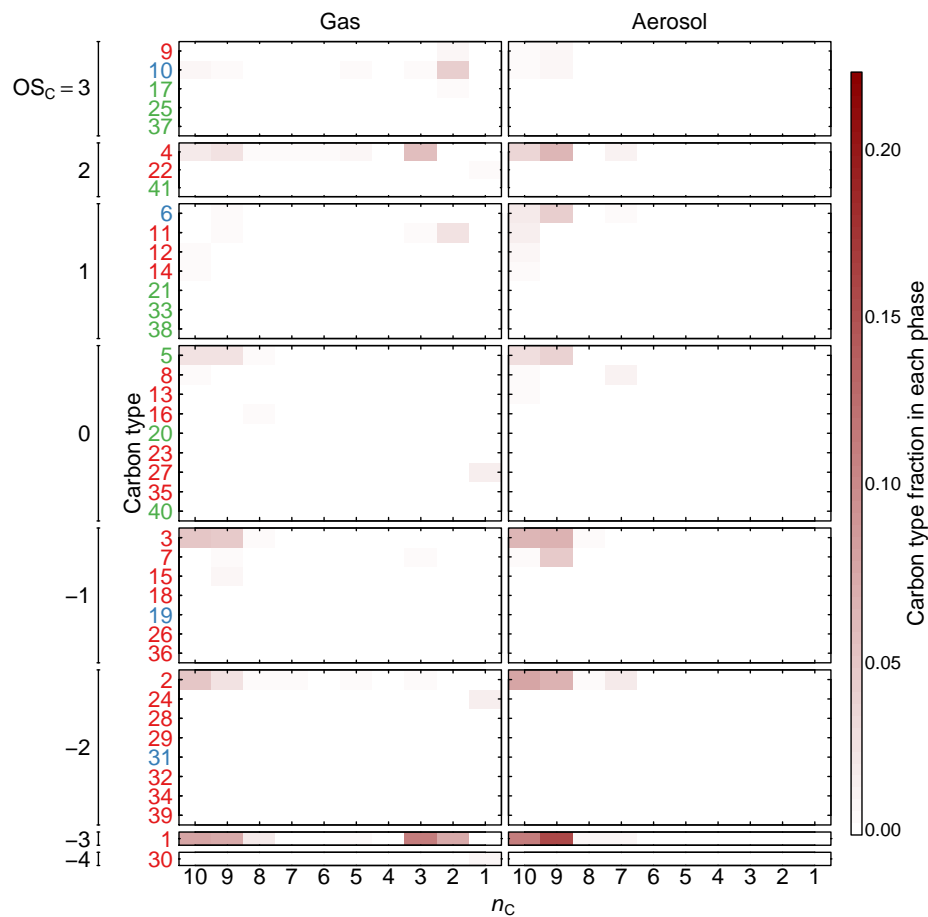20  basis of multi-carbon unit representations as hinted above.

**Figure S1.** Molecular abundance at $t_{\mathrm{max\,SOA}}$ described in terms of their carbon types and number of carbon atoms. The carbon abundance in each grid cell is normalized by the total molar abundance of its phase (gas or aerosol). The colors for the carbon types are the same as in Figure 2

# References

Bloss, C., Wagner, V., Jenkin, M. E., Volkamer, R., Bloss, W. J., Lee, J. D., Heard, D. E., Wirtz, K., Martin-Reviejo, M., Rea, G., Wenger, J. C., and Pilling, M. J.: Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons, *Atmospheric Chemistry and Physics*, 5, 641–664, doi:10.5194/acp-5-641-2005, 2005.

5  Ruggeri, G. and Takahama, S.: Technical Note: Development of chemoinformatic tools to enumerate functional groups in molecules for organic aerosol characterization, *Atmospheric Chemistry and Physics*, 16, 4401–4422, doi:10.5194/acp-16-4401-2016, 2016.

Ruggeri, G., Bernhard, F. A., Henderson, B. H., and Takahama, S.: Model-measurement comparison of functional group abundance in $\alpha$-pinene and 1,3,5-trimethylbenzene secondary organic aerosol formation, *Atmospheric Chemistry and Physics Discussions*, 2016, 1–34, doi:10.5194/acp-2016-46, 2016.