Supporting Information for

Critical contribution of chemically diverse carbonyl molecules to the oxidative potential of atmospheric aerosols

Feifei Li\textsuperscript{1,2}, Shanshan Tang\textsuperscript{3}, Jitao Lv\textsuperscript{1,2}, Shiyang Yu\textsuperscript{1,2}, Xu Sun\textsuperscript{1,4}, Dong Cao\textsuperscript{1}, Yawei Wang\textsuperscript{1,2,5}, Guibin Jiang\textsuperscript{1,2,5}

\textsuperscript{1}State Key Laboratory of Environmental Chemistry and Eco-toxicology, Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
\textsuperscript{2}College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
\textsuperscript{3}Collaborative Center for Physics and Chemistry, Institute of International Innovation, Beihang University, Yuhang District, Hangzhou 311115, China
\textsuperscript{4}Beijing Urban Ecosystem Research Station, State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
\textsuperscript{5}School of Environment, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

Correspondence to: Jitao Lv (jtlv@rcees.ac.cn) and Yawei Wang (ywwang@rcees.ac.cn)
List of supporting information

Texts
Text S1: Chemicals.
Text S2: Solid phase extraction procedure.
Text S3: ESI-FT-ICR MS analysis.
Text S4: Reduction of the carbonyl group by NaBH$_4$.
Text S5: Supplement statistical analysis.

Tables
Table S1: Detailed information on representative PM$_{2.5}$ samples.

Figures
Figure S1: UV-Vis spectra during SRNOM reduction by NaBH$_4$.
Figure S2: Derivatization reaction scheme of PFBHA and carbonyl molecules.
Figure S3: Reproducibility of DTT measurements of samples.
Figure S4: Validation plot of PLSR model for carbonyl molecules significantly positively correlated with DTT$_{OC}$.
Figure S5: PCA score plots for pro-oxidative carbonyls.
Figure S6: Multivariate OPLS-DA (1 + 1 + 0) models for the pro-oxidative carbonyls.
Figure S7: Daily PM$_{2.5}$ concentration at the downtown site, suburban site, and montane site in Beijing during each sampling season.
Figure S8: The proportion of molecules in seven compounds was estimated based on the normalized molecular intensities.
Figure S9: Seven compounds differ in the percentage distribution of all organic, non-carbonyl, and carbonyl molecules based on the normalized molecular intensities.
Figure S10: The proportion of each group in all organic, non-carbonyl, and carbonyl molecules was estimated based on the normalized molecular intensities.
Figure S11: Differences in the percentage distribution of elemental composition based on the normalized molecular intensities in all organic, non-carbonyl, and carbonyl molecules.
Figure S12: DTT consumption rates of organic in PM$_{2.5}$ over Beijing at the downtown site, suburban site, and montane site.
Figure S13: The magnitude-weighted average parameters of carbonyl molecules.
Figure S14: The difference in the percentage distribution based on the normalized molecular intensities of carbonyl molecules in winter, winter Olympics, and summer samples for the seven compound groups.
Figure S15: The van Krevelen diagram of carbonyl molecules from different elemental compositions.
Figure S16: Spearman correlation matrix of carbonyl molecule characteristic parameters and DTT$_{OC}$.
Figure S17: OPLS-DA score plots for normalized intensity of individual pro-oxidative carbonyls.
Text S1: Chemicals

Ultrapure water (18.2 MΩ·cm) was supplied by a Milli-Q water 100 purification system. Acetonitrile and methanol (HPLC grade) were obtained from Fisher. O-(2,3,4,5,6-Pentafluorobenzyl)hydroxylamine hydrochloride (PFBHA, ≥ 99%, CAS 57981-02-9) obtained from Sigma-Aldrich and Sodium borohydride (NaBH₄, ≥ 98%, CAS 16940-66-2) obtained from Sinopharm Chemical Reagent Co. LTD.

Dithiothreitol (DTT, ≥ 99.4%, Biotechnology grade, Amresco), Diethylenetriaminepentaacetic acid (DTPA, ≥ 99%, CAS 67-43-6, Sigma-Aldrich), 5,5-Dithiobis(2-nitrobenzoic acid) (DTNB, > 99%, CAS 69-78-3, J&K Scientific) and phosphate buffer saline (PBS, 0.1M, pH 7.2-7.4, Solarbio) were used to measure the OP of PM₂.₅ samples.

Text S2: Solid phase extraction procedure

The water-soluble fraction of PM₂.₅ was desalted and purified using Varian Bond Elute PPL cartridges (200 mg per 3 mL). Briefly, the cartridges were rinsed with 3 mL of methanol (MS grade) and 9 ml of ultrapure water (pH 2) prior to use. The samples were acidified to pH 2 with HCl (32%, ultrapure) to increase the extraction efficiency for organic acids and phenols, and 20 mL were passed through the cartridges by gravity at a flow rate of approximately 2 mL min⁻¹. Cartridges were rinsed with three volumes of 0.01 M HCl for removal of salts, dried with a stream of N₂, and immediately extracted with two volumes of methanol (MS grade). The extracted WEOM samples were stored at -4 °C in the dark, and the WISOM was concentrated with N₂ to 1 mL and then stored at -4 °C in the dark.

Text S3: ESI-FT-ICR MS analysis

Samples for ESI FT-ICR-MS analysis were continuously infused into the ESI unit by syringe infusion at a flow rate of 120 μL h⁻¹. The ESI needle voltage was set to −3.8 kV. All the samples were analyzed in negative ionization mode with broadband detection. Ions accumulated in a hexapol ion trap for 0.2 s before being introduced into the ICR cell. The lower mass limit was set to m/z = 120 Da and the upper mass limit to m/z = 920 Da. 300 mass spectra were averaged per sample. The spectra were externally calibrated with 10 mM of sodium formate solution in 50% isopropyl alcohol using a linear calibration and then internally recalibrated using an in-house reference mass list. After internal calibration, the mass error was < 0.5 ppm over the entire mass range.

Based on previous studies, the following sample molecular parameters were calculated:

\[ \text{DBE} = 1 + \frac{2C-H+N}{2} \] (Lv et al., 2016)

\[ \text{Almod} = \frac{(1 + C - \frac{O}{2} - \frac{H}{2} - S)/(C - \frac{O}{2} - N - S)}{\text{Wang et al., 2021}} \]

\[ \text{NOSC} = 4 - [(4C + H - 3N - 20 - 2S)/C] \] (Zhang et al., 2023)

\[ \overline{OS_C} \approx 2 \frac{O}{2} - \frac{H}{2} \] (Kroll et al., 2011)
Where C, H, N, O, and S represent the numbers of carbon, hydrogen, nitrogen, oxygen, and sulfur atoms in each formula.

The magnitude-weighted average parameters such as C, H, O, N, S, MW, O/C, H/C, NOSC, Almod, and DBE for each sample can be determined by the following formula:

\[(M)_W = \left[ \sum_i I_i \times (M)_i \right] / \sum_i I_i\]

where \(I_i\) and \((M)_i\) are the relative abundance and M value of peak \(i\), respectively. The relative abundance is calculated as the abundance of the individual peak divided by the maximum abundances in a given spectrum. (Lv et al., 2016)

**Text S4: Reduction of the carbonyl group by NaBH₄**

The aqueous solution of SRNOM was de-oxygenated by nitrogen purging for 15 min, and an excess of NaBH₄ was added under nitrogen protection and reacted in the dark. (Phillips and Smith, 2014, 2015)

Changes in light absorption during reduction were monitored using UV 3600 (Shimadzu, Japan) to determine the mass and time of NaBH₄ required for the carbonyl group to be completely reduced. The mass ratios of NaBH₄ to SRNOM were set to 1:1, 10:1, and 20:1, and the UV-vis absorbances after the reactions of 0.2H, 2H, and 12H were recorded, respectively (Figure S3). It can be found that the absorbances of the mass ratios of 10:1 and 20:1 were the same after the reaction 12H, so NaBH₄ at 10 times the sample mass completely consumed the carbonyl group in the sample after the reaction 12H.

**Text S5: Supplement statistical analysis**

The pro-oxidative carbonyls are screened by the following steps. Carbonyl molecules with more than 50% detection rate in all samples were first filtered as common molecules (1314), and the correlation between the normalized molecular intensity of common carbonyl molecules and \(\text{DTT}_\text{OC}\) was analyzed by Spearman rank correlation. Due to the large dataset, the use of the FDR method can reduce the false positive results during multiple hypothesis testing in large datasets. The p-values in Spearman's rank correlation results were corrected by FDR and labeled as \(p_{\text{fdr}}\). A total of 705 carbonyl molecules were screened for significant positive correlation with \(\text{DTT}_\text{OC}\) based on \(p_{\text{fdr}} < 0.05\) and Spearman's \(r > 0\). Next, the PLSR model was used to select the key variables among oxidation potential-related carbonyl molecules to avert the influence of collinearity in multivariate correlation analysis. The criterion for screening important variables was set as a VIP > 1, which screened a total of 380 pro-oxidative carbonyls. The results of 999-time permutation tests indicated that the PLSR model was valid and non-overfit (Figure S4).

The samples were categorized into winter, Winter Olympics, and summer groups according to the sampling time, and no outliers were detected by the quality assessment of the sample variants by the PCA model (Figure S5). Further multivariate OPLS-DA (1 + 1 + 0) models explained the seasonal differences in the abundance of pro-oxidative carbonyls (Figure S17), and the model was validated by permutation tests (n=200) (Figure S6).
| Sampling site | Sampling number | Sampling season | Data       | TOC (mg/L) | \( \text{DTT}_m \) (pmol/min/ngC) | WSOM \( \text{DTT}_m \) (pmol/min/m²) | \( \text{WISOM} \) \( \text{DTT}_
abla \) (pmol/min/m²) | Number of carbonyl molecules | Number of non-carbonyl molecules | Number of all molecules |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mountain site</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-2</td>
<td>Winter</td>
<td>23 Jan. 2022</td>
<td>24 Jan. 2022</td>
<td>323.20</td>
<td>37.81</td>
<td>178.37</td>
<td>104.45</td>
<td>1768</td>
<td>8633</td>
<td>10401</td>
</tr>
<tr>
<td>M-3</td>
<td>Winter</td>
<td>31 Jan. 2022</td>
<td>1 Feb. 2022</td>
<td>42.94</td>
<td>55.41</td>
<td>17.36</td>
<td>16.58</td>
<td>1585</td>
<td>5784</td>
<td>7369</td>
</tr>
<tr>
<td>M-4</td>
<td>Olympics</td>
<td>9 Feb. 2022</td>
<td>10 Feb. 2022</td>
<td>199.10</td>
<td>43.72</td>
<td>127.06</td>
<td>73.52</td>
<td>1795</td>
<td>8032</td>
<td>9827</td>
</tr>
<tr>
<td>M-6</td>
<td>Olympics</td>
<td>8 Feb. 2022</td>
<td>18 Feb. 2022</td>
<td>115.27</td>
<td>64.88</td>
<td>54.58</td>
<td>22.03</td>
<td>1764</td>
<td>8345</td>
<td>10109</td>
</tr>
<tr>
<td>M-7</td>
<td>Summer</td>
<td>10 Jul. 2022</td>
<td>27 Jul. 2022</td>
<td>130.07</td>
<td>7.96</td>
<td>7.40</td>
<td>3.57</td>
<td>1230</td>
<td>7119</td>
<td>8349</td>
</tr>
<tr>
<td>M-8</td>
<td>Summer</td>
<td>16 Jul. 2022</td>
<td>24 Jul. 2022</td>
<td>166.13</td>
<td>10.70</td>
<td>25.40</td>
<td>8.61</td>
<td>1396</td>
<td>8442</td>
<td>9838</td>
</tr>
<tr>
<td>M-9</td>
<td>Summer</td>
<td>11 Jul. 2022</td>
<td>20 Jul. 2022</td>
<td>166.33</td>
<td>8.97</td>
<td>21.31</td>
<td>11.58</td>
<td>1612</td>
<td>7810</td>
<td>9422</td>
</tr>
<tr>
<td>S-1</td>
<td>Winter</td>
<td>19 Jan. 2022</td>
<td>20 Jan. 2022</td>
<td>189.30</td>
<td>57.98</td>
<td>160.19</td>
<td>29.81</td>
<td>1825</td>
<td>7572</td>
<td>9397</td>
</tr>
<tr>
<td>S-2</td>
<td>Winter</td>
<td>23 Jan. 2022</td>
<td>24 Jan. 2022</td>
<td>488.70</td>
<td>40.18</td>
<td>286.58</td>
<td>159.05</td>
<td>1306</td>
<td>8839</td>
<td>10145</td>
</tr>
<tr>
<td>S-3</td>
<td>Winter</td>
<td>31 Jan. 2022</td>
<td>1 Feb. 2022</td>
<td>41.30</td>
<td>84.16</td>
<td>50.73</td>
<td>14.40</td>
<td>1809</td>
<td>7083</td>
<td>8892</td>
</tr>
<tr>
<td>S-4</td>
<td>Olympics</td>
<td>9 Feb. 2022</td>
<td>10 Feb. 2022</td>
<td>328.10</td>
<td>52.44</td>
<td>251.07</td>
<td>89.09</td>
<td>1579</td>
<td>7821</td>
<td>9400</td>
</tr>
<tr>
<td>S-5</td>
<td>Olympics</td>
<td>6 Feb. 2022</td>
<td>14 Feb. 2022</td>
<td>82.57</td>
<td>58.75</td>
<td>35.40</td>
<td>17.16</td>
<td>1769</td>
<td>7501</td>
<td>9270</td>
</tr>
<tr>
<td>S-6</td>
<td>Olympics</td>
<td>8 Feb. 2022</td>
<td>18 Feb. 2022</td>
<td>93.19</td>
<td>57.88</td>
<td>78.72</td>
<td>43.25</td>
<td>1714</td>
<td>7138</td>
<td>8852</td>
</tr>
<tr>
<td>S-8</td>
<td>Summer</td>
<td>16 Jul. 2022</td>
<td>24 Jul. 2022</td>
<td>213.23</td>
<td>11.54</td>
<td>35.14</td>
<td>19.89</td>
<td>1676</td>
<td>8227</td>
<td>9903</td>
</tr>
<tr>
<td>D-1</td>
<td>Winter</td>
<td>19 Jan. 2022</td>
<td>20 Jan. 2022</td>
<td>162.20</td>
<td>52.87</td>
<td>166.87</td>
<td>101.65</td>
<td>1686</td>
<td>7104</td>
<td>8790</td>
</tr>
<tr>
<td>D-3</td>
<td>Winter</td>
<td>31 Jan. 2022</td>
<td>1 Feb. 2022</td>
<td>85.70</td>
<td>45.42</td>
<td>56.81</td>
<td>33.85</td>
<td>1290</td>
<td>6222</td>
<td>7512</td>
</tr>
<tr>
<td>D-4</td>
<td>Olympics</td>
<td>9 Feb. 2022</td>
<td>10 Feb. 2022</td>
<td>263.40</td>
<td>51.40</td>
<td>197.59</td>
<td>122.90</td>
<td>1640</td>
<td>7699</td>
<td>9339</td>
</tr>
<tr>
<td>D-6</td>
<td>Olympics</td>
<td>8 Feb. 2022</td>
<td>18 Feb. 2022</td>
<td>74.79</td>
<td>71.00</td>
<td>77.50</td>
<td>42.01</td>
<td>1379</td>
<td>7894</td>
<td>9273</td>
</tr>
<tr>
<td>D-7</td>
<td>Summer</td>
<td>10 Jul. 2022</td>
<td>27 Jul. 2022</td>
<td>134.07</td>
<td>12.57</td>
<td>12.04</td>
<td>8.94</td>
<td>1354</td>
<td>7687</td>
<td>9041</td>
</tr>
<tr>
<td>D-8</td>
<td>Summer</td>
<td>16 Jul. 2022</td>
<td>24 Jul. 2022</td>
<td>230.73</td>
<td>11.29</td>
<td>37.21</td>
<td>20.04</td>
<td>1770</td>
<td>8227</td>
<td>9997</td>
</tr>
</tbody>
</table>
Figure S1. Derivatization reaction scheme of PFBHA and carbonyl molecules.
Figure S2. Reproducibility of DTT measurements of samples. (a) DTT measurements of PM$_{2.5}$ samples and backgrounds, using the downtown site on January 27, 2022, as an example. (b) DTT measurements of a water-soluble organic carbon (WSOM) quality control sample, which was prepared containing a mixture of all WSOM samples. (c) DTT measurements of a water insoluble organic carbon (WISOM) quality control sample, which was prepared containing a mixture of all WISOM samples.
Figure S3. UV-Vis spectra during SRNOM reduction by NaBH₄.
Figure S4. Validation plot of PLSR model for carbonyl molecules significantly positively correlated with DTT$_{OC}$, n = 999, R$^2$ = (0.0, 0.404), Q$^2$ = (0.0, -0.197).
Figure S5. PCA score plots for pro-oxidative carbonyls in (a) winter versus summer, (b) winter versus winter Olympics, and (c) winter Olympics versus summer samples.
Figure S6. Multivariate OPLS-DA (1 + 1 + 0) models for the pro-oxidative carbonyls were validated by permutation test (n = 200). (a) Validation plot of OPLS-DA model obtained from carbonyl molecules in winter and summer samples, R² = (0.0, 0.019), Q² = (0.0, -0.574); (b) Validation plot of OPLS-DA model obtained from carbonyl molecules in winter and Winter Olympic samples, R² = (0.0, 0.301), Q² = (0.0, -0.602); (c) Validation plot of OPLS-DA model obtained from carbonyl molecules in Winter Olympics and summer samples, R² = (0.0, 0.225), Q² = (0.0, -0.58).
Figure S7. Daily PM$_{2.5}$ concentration at the downtown site, suburban site, and montane site in Beijing during each sampling season. The red dots represent the day on which PM$_{2.5}$ concentrations exceeded 80% of that sampling season. The yellow dots represent the day on which PM$_{2.5}$ concentrations are at 40-60% of that sampling season. The green dots represent the day on which PM$_{2.5}$ concentrations are below 20% of that sampling season.
Figure S8. The proportion of molecules in seven compounds was estimated based on the normalized molecular intensities. Organic molecules detected in FT-ICP MS were distinguished into seven compounds based on H/C and O/C. (Group 1: lipids-like, Group 2: proteins-like, Group 3: lignins-like, Group 4: carbohydrates-like, Group 5: unsaturated hydrocarbons, Group 6: aromatic structures, Group 7: tannins-like).
Figure S9. Seven compounds differ in the percentage distribution of all organic, non-carbonyl, and carbonyl molecules based on the normalized molecular intensities (Kruskal-Wallis tests. Differences between groups were considered statistically significant when $p < 0.05$, with $0.01 < p < 0.05$ marked by *; $p < 0.01$ marked by **; and $p < 0.001$ marked by ***).
Figure S10. The proportion of each group in all organic, non-carbonyl, and carbonyl molecules was estimated based on the normalized molecular intensities. Organic molecules detected in FT-ICP MS were divided into four groups according to the elemental composition, namely CHO, CHON, CHOS, and CHONS.
Figure S11. Differences in the percentage distribution of elemental composition based on the normalized molecular intensities in all organic, non-carbonyl, and carbonyl molecules (Kruskal-Wallis tests, Differences between groups were considered statistically significant when $p < 0.05$, with $0.01 < p < 0.05$ marked by *; $p < 0.01$ marked by **; and $p < 0.001$ marked by ***).
Figure S12. (a) Volume-normalized DTT consumption rate (DTTv) and mass-normalized DTT consumption rate (DTTvm) of organic matter in PM$_{2.5}$ at the downtown site, suburban site, and montane site (Kruskal-Wallis tests. Differences between groups were considered statistically significant when $p < 0.05$, with $0.01 < p < 0.05$ marked by *, $p < 0.01$ marked by **, and $p < 0.001$ marked by ***). (b) DTTv of water-soluble organic carbon (WSOM) and organic carbon (WISOM) for representative samples. (c) DTT consumption rate of WSOM normalized by mass concentration of TOC (DTToc).
Figure S13. The magnitude-weighted average values of molecular weight (MWw), O/C, H/C, the nominal carbon oxidation states (NOSC), double bond equivalence (DBE), and modified aromaticity index (Almod) of carbonyl molecules (Kruskal-Wallis tests, Differences between groups were considered statistically significant when $p < 0.05$, with $0.01 < p < 0.05$ marked by *, $p < 0.01$ marked by **; and $p < 0.001$ marked by ***).
Figure S14. The difference in the percentage distribution based on the normalized molecular intensities of carbonyl molecules in winter, winter Olympics, and summer samples for the seven compound groups (Group1: lipids-like, Group2: proteins-like, Group3: lignins/CRAMs-like, Group4: carbohydrates-like, Group5: unsaturated hydrocarbons, Group6: aromatic structures, Group7: tannins-like).
Figure S15. The van Krevelen diagram of carbonyl molecules from different elemental compositions (Group 1: lipids-like, Group 2: proteins-like, Group 3: lignins/CRAMs-like, Group 4: carbohydrates-like, Group 5: unsaturated hydrocarbons, Group 6: aromatic structures, Group 7: tannins-like).
Figure S16. Spearman correlation matrix of carbonyl molecule characteristic parameters and DTT consumption rate of WSOM normalized by mass concentration of TOC (DTToC). Colors from blue to red represent changes in the Spearman correlation coefficient from −1 to 1.
Figure S17. OPLS-DA score plots for normalized intensity of individual pro-oxidative carbonyls in (a) winter and winter Olympics, (b) winter and summer, and (c) winter Olympics and summer samples.
References


