SUPPLEMENTARY INFORMATION

THE SAPRC ATMOSPHERIC CHEMICAL MECHANISM GENERATION SYSTEM (MECHGEN)

William P. L. Carter[1], Jia Jiang[2], Zhizhao Wang[1], and Kelley C. Barsanti[3]

[1] College of Engineering Center for Environmental Research and Technology (CE-CERT), University of California, Riverside, California 92521

[2] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139

[3] Atmospheric Chemistry Observations & Modeling Lab, NSF National Center for Atmospheric Research (NCAR), P.O. Box 3000, Boulder, CO 80301

Contents

List of Tables

List of Figures

List of Schemes

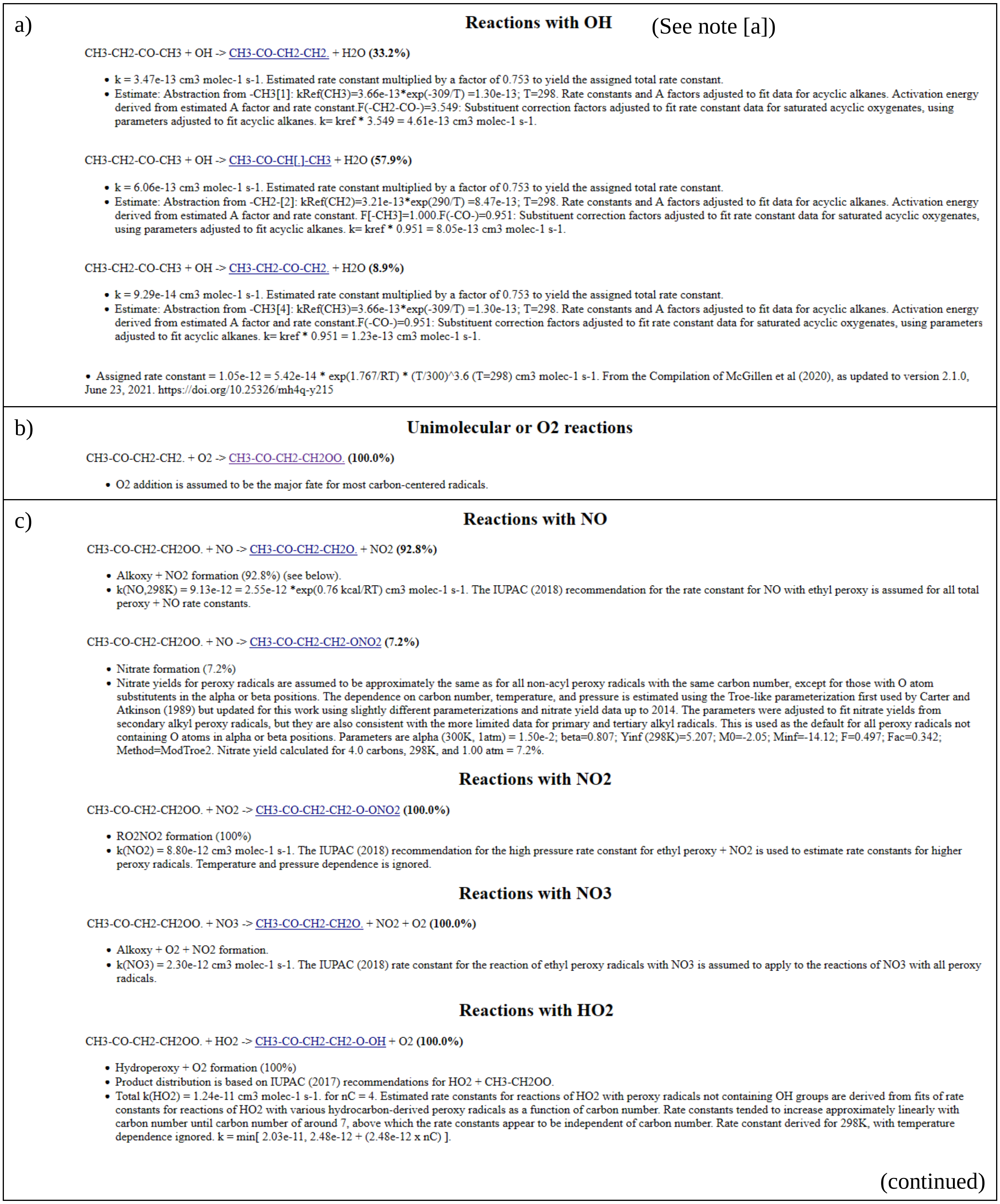
S1. Additional output examples

Figure S1 shows examples of results of single step mechanism generations for the reaction of methyl ethyl ketone with OH radicals and for representative intermediates that are formed. Figure S2 through Figure S4 show examples of MechGen output for 1,3-butadiene after the full mechanism generation operation. Figure S2 shows the explicit mechanism, Figure S3 shows the minimally reduced mechanism, and Figure S4 shows the estimated product yields. The output is similar for both web and terminal access, though the hyperlinks for creating and reacting organic products are only present with web access.

S2. Derivation of Standard Environmental Conditions

The standard environmental conditions included with MechGen were derived from results of multi-day dynamic box model simulations of urban air pollution scenarios, with continuous daytime emissions of pollutants and variable inversion heights. The SAPRC box modeling software (Carter, 1994b) and an earlier version of the SAPRC-22 mechanism (Carter, 2023) were employed. Although the inputs were similar to those for the "averaged conditions" scenario used by Carter (1994b, a) to derive various ozone reactivity scales, the total reactive organic gas (ROG) inputs were reduced to more closely represent pollution levels in current ozone episodes. In addition, there were no background or aloft pollutants. Instead, the simulation was carried out for four simulated days with continuous ROG and NO_x emissions, with pollutants carried over from day to day and with pollutants diluted by approximately a factor of 5 each night due to assumed variation of the mixing height. Only data from the second and fourth day were used to derive the averages for the standard environments representing daytime conditions, and only data from the second and third night were used for the nighttime standard environment. Details concerning the inputs, and also the times used to derive the average oxidant concentrations for the standard environments, are summarized in Table S1. The input files were created by editing the input files for the Carter (1994b, a) reactivity scenarios that are available with the SAPRC box modeling software and files (Carter, 2024).

Figure S1. Screenshots of the MechGen web interface as a result of (a) creating methyl ethyl ketone with OH, and (b-d) selecting and reacting a radical on the list from a previous reaction.



a)

Reactions with OH

(See note [a])

CH3-CH2-CO-CH3 + OH -> CH3-CO-CH2-CH2. + H2O **(33.2%)**

- k = 3.47e-13 cm3 molec-1 s-1. Estimated rate constant multiplied by a factor of 0.753 to yield the assigned total rate constant.
- Estimate: Abstraction from -CH3[1]: kRef(CH3)=3.66e-13*exp(-309/T) =1.30e-13; T=298. Rate constants and A factors adjusted to fit data for acyclic alkanes. Activation energy derived from estimated A factor and rate constant.F(-CH2-CO-)=3.549: Substituent correction factors adjusted to fit rate constant data for saturated acyclic oxygenates, using parameters adjusted to fit acyclic alkanes. k= kref * 3.549 = 4.61e-13 cm3 molec-1 s-1.

CH3-CH2-CO-CH3 + OH -> CH3-CO-CH[.]-CH3 + H2O **(57.9%)**

- k = 6.06e-13 cm3 molec-1 s-1. Estimated rate constant multiplied by a factor of 0.753 to yield the assigned total rate constant.
- Estimate: Abstraction from -CH2-[2]: kRef(CH2)=3.21e-13*exp(290/T) =8.47e-13; T=298. Rate constants and A factors adjusted to fit data for acyclic alkanes. Activation energy derived from estimated A factor and rate constant. F[-CH3]=1.000.F(-CO-)=0.951: Substituent correction factors adjusted to fit rate constant data for saturated acyclic oxygenates, using parameters adjusted to fit acyclic alkanes. k= kref * 0.951 = 8.05e-13 cm3 molec-1 s-1.

CH3-CH2-CO-CH3 + OH -> CH3-CH2-CO-CH2. + H2O **(8.9%)**

- k = 9.29e-14 cm3 molec-1 s-1. Estimated rate constant multiplied by a factor of 0.753 to yield the assigned total rate constant.
- Estimate: Abstraction from -CH3[4]: kRef(CH3)=3.66e-13*exp(-309/T) =1.30e-13; T=298. Rate constants and A factors adjusted to fit data for acyclic alkanes. Activation energy derived from estimated A factor and rate constant.F(-CO-)=0.951: Substituent correction factors adjusted to fit rate constant data for saturated acyclic oxygenates, using parameters adjusted to fit acyclic alkanes. k= kref * 0.951 = 1.23e-13 cm3 molec-1 s-1.

- Assigned rate constant = 1.05e-12 = 5.42e-14 * exp(1.767/RT) * (T/300)^3.6 (T=298) cm3 molec-1 s-1. From the Compilation of McGillen et al (2020), as updated to version 2.1.0, June 23, 2021. https://doi.org/10.25326/mh4q-y215

b)

Unimolecular or O2 reactions

CH3-CO-CH2-CH2. + O2 -> CH3-CO-CH2-CH2OO. **(100.0%)**

- O2 addition is assumed to be the major fate for most carbon-centered radicals.

c)

Reactions with NO

CH3-CO-CH2-CH2OO. + NO -> CH3-CO-CH2-CH2O. + NO2 **(92.8%)**

- Alkoxy + NO2 formation (92.8%) (see below).
- k(NO,298K) = 9.13e-12 = 2.55e-12 *exp(0.76 kcal/RT) cm3 molec-1 s-1. The IUPAC (2018) recommendation for the rate constant for NO with ethyl peroxy is assumed for all total peroxy + NO rate constants.

CH3-CO-CH2-CH2OO. + NO -> CH3-CO-CH2-CH2-ONO2 **(7.2%)**

- Nitrate formation (7.2%)
- Nitrate yields for peroxy radicals are assumed to be approximately the same as for all non-acyl peroxy radicals with the same carbon number, except for those with O atom substitutents in the alpha or beta positions. The dependence on carbon number, temperature, and pressure is estimated using the Troe-like parameterization first used by Carter and Atkinson (1989) but updated for this work using slightly different parameterizations and nitrate yield data up to 2014. The parameters were adjusted to fit nitrate yields from secondary alkyl peroxy radicals, but they are also consistent with the more limited data for primary and tertiary alkyl radicals. This is used as the default for all peroxy radicals not containing O atoms in alpha or beta positions. Parameters are alpha (300K, 1atm) = 1.50e-2; beta=0.807; Yinf (298K)=5.207; M0=-2.05; Minf=-14.12; F=0.497; Fac=0.342; Method=ModTroe2. Nitrate yield calculated for 4.0 carbons, 298K, and 1.00 atm = 7.2%.

Reactions with NO2

CH3-CO-CH2-CH2OO. + NO2 -> CH3-CO-CH2-CH2-O-ONO2 **(100.0%)**

- RO2NO2 formation (100%)
- k(NO2) = 8.80e-12 cm3 molec-1 s-1. The IUPAC (2018) recommendation for the high pressure rate constant for ethyl peroxy + NO2 is used to estimate rate constants for higher peroxy radicals. Temperature and pressure dependence is ignored.

Reactions with NO3

CH3-CO-CH2-CH2OO. + NO3 -> CH3-CO-CH2-CH2O. + NO2 + O2 **(100.0%)**

- Alkoxy + O2 + NO2 formation.
- k(NO3) = 2.30e-12 cm3 molec-1 s-1. The IUPAC (2018) rate constant for the reaction of ethyl peroxy radicals with NO3 is assumed to apply to the reactions of NO3 with all peroxy radicals.

Reactions with HO2

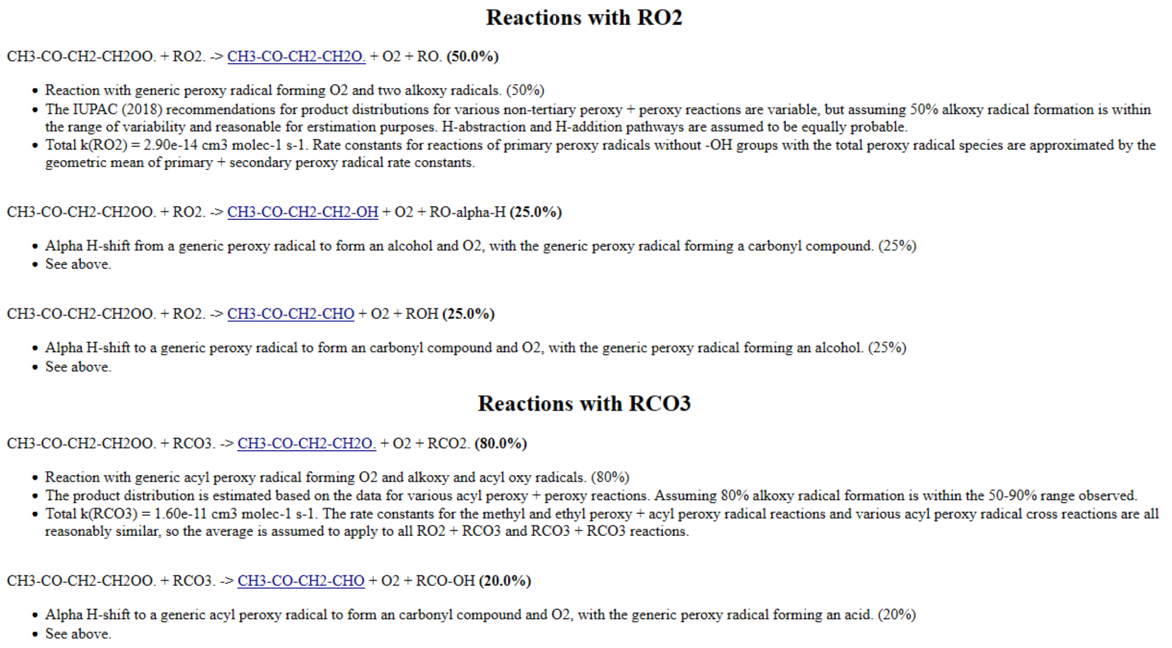
CH3-CO-CH2-CH2OO. + HO2 -> CH3-CO-CH2-CH2-O-OH + O2 **(100.0%)**

- Hydroperoxy + O2 formation (100%)
- Product distribution is based on IUPAC (2017) recommendations for HO2 + CH3-CH2OO.
- Total k(HO2) = 1.24e-11 cm3 molec-1 s-1. for nC = 4. Estimated rate constants for reactions of HO2 with peroxy radicals not containing OH groups are derived from fits of rate constants for reactions of HO2 with various hydrocarbon-derived peroxy radicals as a function of carbon number. Rate constants tended to increase approximately linearly with carbon number until carbon number of around 7, above which the rate constants appear to be independent of carbon number. Rate constant derived for 298K, with temperature dependence ignored. k = min[2.03e-11, 2.48e-12 + (2.48e-12 x nC)].

(continued)

Figure S1 (continued)

c) cont



Reactions with RO2

CH3-CO-CH2-CH2OO. + RO2. -> CH3-CO-CH2-CH2O. + O2 + RO. **(50.0%)**

- Reaction with generic peroxy radical forming O2 and two alkoxy radicals. (50%)
- The IUPAC (2018) recommendations for product distributions for various non-tertiary peroxy + peroxy reactions are variable, but assuming 50% alkoxy radical formation is within the range of variability and reasonable for erstimation purposes. H-abstraction and H-addition pathways are assumed to be equally probable.
- Total k(RO2) = 2.90e-14 cm3 molec-1 s-1. Rate constants for reactions of primary peroxy radicals without -OH groups with the total peroxy radical species are approximated by the geometric mean of primary + secondary peroxy radical rate constants.

CH3-CO-CH2-CH2OO. + RO2. -> CH3-CO-CH2-CH2-OH + O2 + RO-alpha-H **(25.0%)**

- Alpha H-shift from a generic peroxy radical to form an alcohol and O2, with the generic peroxy radical forming a carbonyl compound. (25%)
- See above.

CH3-CO-CH2-CH2OO. + RO2. -> CH3-CO-CH2-CHO + O2 + ROH **(25.0%)**

- Alpha H-shift to a generic peroxy radical to form an carbonyl compound and O2, with the generic peroxy radical forming an alcohol. (25%)
- See above.

Reactions with RCO3

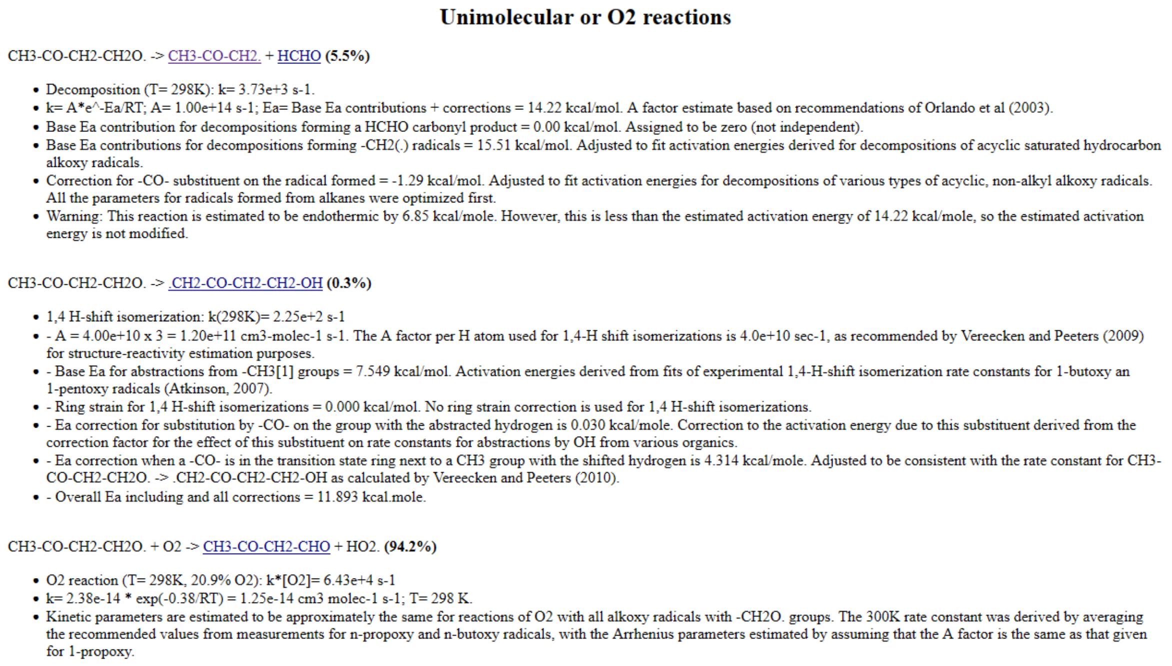
CH3-CO-CH2-CH2OO. + RCO3. -> CH3-CO-CH2-CH2O. + O2 + RCO2. **(80.0%)**

- Reaction with generic acyl peroxy radical forming O2 and alkoxy and acyl oxy radicals. (80%)
- The product distribution is estimated based on the data for various acyl peroxy + peroxy reactions. Assuming 80% alkoxy radical formation is within the 50-90% range observed.
- Total k(RCO3) = 1.60e-11 cm3 molec-1 s-1. The rate constants for the methyl and ethyl peroxy + acyl peroxy radical reactions and various acyl peroxy radical cross reactions are all reasonably similar, so the average is assumed to apply to all RO2 + RCO3 and RCO3 + RCO3 reactions.

CH3-CO-CH2-CH2OO. + RCO3. -> CH3-CO-CH2-CHO + O2 + RCO-OH **(20.0%)**

- Alpha H-shift to a generic acyl peroxy radical to form an carbonyl compound and O2, with the generic peroxy radical forming an acid. (20%)
- See above.

d)



Unimolecular or O2 reactions

CH3-CO-CH2-CH2O. -> CH3-CO-CH2. + HCHO **(5.5%)**

- Decomposition (T= 298K): k= 3.73e+3 s-1.
- k= A*e^-Ea/RT; A= 1.00e+14 s-1; Ea= Base Ea contributions + corrections = 14.22 kcal/mol. A factor estimate based on recommendations of Orlando et al (2003).
- Base Ea contribution for decompositions forming a HCHO carbonyl product = 0.00 kcal/mol. Assigned to be zero (not independent).
- Base Ea contributions for decompositions forming -CH2(.) radicals = 15.51 kcal/mol. Adjusted to fit activation energies derived for decompositions of acyclic saturated hydrocarbon alkoxy radicals.
- Correction for -CO- substituent on the radical formed = -1.29 kcal/mol. Adjusted to fit activation energies for decompositions of various types of acyclic, non-alkyl alkoxy radicals. All the parameters for radicals formed from alkanes were optimized first.
- Warning: This reaction is estimated to be endothermic by 6.85 kcal/mole. However, this is less than the estimated activation energy of 14.22 kcal/mole, so the estimated activation energy is not modified.

CH3-CO-CH2-CH2O. -> .CH2-CO-CH2-CH2-OH **(0.3%)**

- 1,4 H-shift isomerization: k(298K)= 2.25e+2 s-1
- - A = 4.00e+10 x 3 = 1.20e+11 cm3-molec-1 s-1. The A factor per H atom used for 1,4-H shift isomerizations is 4.0e+10 sec-1, as recommended by Vereecken and Peeters (2009) for structure-reactivity estimation purposes.
- - Base Ea for abstractions from -CH3[1] groups = 7.549 kcal/mol. Activation energies derived from fits of experimental 1,4-H-shift isomerization rate constants for 1-butoxy an 1-pentoxy radicals (Atkinson, 2007).
- - Ring strain for 1,4 H-shift isomerizations = 0.000 kcal/mol. No ring strain correction is used for 1,4 H-shift isomerizations.
- - Ea correction for substitution by -CO- on the group with the abstracted hydrogen is 0.030 kcal/mole. Correction to the activation energy due to this substituent derived from the correction factor for the effect of this substituent on rate constants for abstractions by OH from various organics.
- - Ea correction when a -CO- is in the transition state ring next to a CH3 group with the shifted hydrogen is 4.314 kcal/mole. Adjusted to be consistent with the rate constant for CH3-CO-CH2-CH2O. -> .CH2-CO-CH2-CH2-OH as calculated by Vereecken and Peeters (2010).
- - Overall Ea including and all corrections = 11.893 kcal.mole.

CH3-CO-CH2-CH2O. + O2 -> CH3-CO-CH2-CHO + HO2. **(94.2%)**

- O2 reaction (T= 298K, 20.9% O2): k*[O2]= 6.43e+4 s-1
- k= 2.38e-14 * exp(-0.38/RT) = 1.25e-14 cm3 molec-1 s-1; T= 298 K.
- Kinetic parameters are estimated to be approximately the same for reactions of O2 with all alkoxy radicals with -CH2O. groups. The 300K rate constant was derived by averaging the recommended values from measurements for n-propoxy and n-butoxy radicals, with the Arrhenius parameters estimated by assuming that the A factor is the same as that given for 1-propoxy.

[a] The web system shows the products formed as hyperlinked, so users can select them to see how they react. Terminal output is similar but there are no hyperlinks.

Figure S2. MechGen output showing results of a full mechanism generation operation with default options for the reactions of 1,3-butadiene.

```
Full set of reactions of CH2=CH-CH=CH2 with OH, O3, and NO3

Rxn     k      Fac  Weight  Reaction
1   6.30e-11   95%   95.0%  CH2=CH-CH=CH2 + OH -> .aCH2-aCH-aCH[.]-CH2-OH {*OHadd}
2   3.31e-12    5%    5.0%  CH2=CH-CH=CH2 + OH -> CH2=CH-CH(CH2.)-OH {*OHadd}
3   3.15e-18   50%   50.0%  CH2=CH-CH=CH2 + O3 -> CH2=CH-CHO + CH2OO {*O3Ole}
4   3.15e-18   50%   50.0%  CH2=CH-CH=CH2 + O3 -> HCHO + CH2=CH-CHOO {*O3Ole}
5   5.50e-14   50%   50.0%  CH2=CH-CH=CH2 + NO3 -> .aCH2-aCH-aCH[.]-CH2-ONO2 {*NO3add}
6   5.50e-14   50%   50.0%  CH2=CH-CH=CH2 + NO3 -> CH2=CH-CH(CH2.)-ONO2 {*NO3add}
7             100%   95.0%  .aCH2-aCH-aCH[.]-CH2-OH {*OHadd} -> .aCH2-aCH-aCH[.]-CH2-OH
8              65%   61.7%  .aCH2-aCH-aCH[.]-CH2-OH + O2 -> CH2=CH-CH[OO.]-CH2-OH
9              35%   33.3%  .aCH2-aCH-aCH[.]-CH2-OH + O2 -> .OOCH2-CH=CH-CH2-OH
10  8.64e-12   95%   58.5%  CH2=CH-CH[OO.]-CH2-OH + NO -> CH2=CH-CH[O.]-CH2-OH + NO2
11  4.80e-13    5%    3.2%  CH2=CH-CH[OO.]-CH2-OH + NO -> CH2=CH-CH(ONO2)-CH2-OH
12  2.30e-12  100%   61.7%  CH2=CH-CH[OO.]-CH2-OH + NO3 -> CH2=CH-CH[O.]-CH2-OH + NO2 + O2
13  1.61e-11  100%   61.7%  CH2=CH-CH[OO.]-CH2-OH + HO2 -> CH2=CH-CH(CH2-OH)-O-OH + O2
14  1.28e-11   80%   49.4%  CH2=CH-CH[OO.]-CH2-OH + RCO3 -> CH2=CH-CH[O.]-CH2-OH + O2 + RCO2.
15  3.20e-12   20%   12.3%  CH2=CH-CH[OO.]-CH2-OH + RCO3 -> CH2=CH-CO-CH2-OH + O2 + RCO-OH
16  2.45e+8   100%  169.6%  CH2=CH-CH[O.]-CH2-OH -> CH2=CH-CHO + .CH2-OH
17  1.16e+0   100%   33.3%  .OOCH2-CH=CH-CH2-OH -> HO-aCH[.]-aCH-aCH[.]-CH2-O-OH
18  8.47e-12   93%   30.8%  .OOCH2-CH=CH-CH2-OH + NO -> .OCH2-CH=CH-CH2-OH + NO2
19  6.59e-13    7%    2.4%  .OOCH2-CH=CH-CH2-OH + NO -> HO-CH2-CH=CH-CH2-ONO2
20  1.61e-11  100%   33.3%  .OOCH2-CH=CH-CH2-OH + HO2 -> HO-CH2-CH=CH-CH2-O-OH + O2
21             92%   30.6%  HO-aCH[.]-aCH-aCH[.]-CH2-O-OH + O2 -> HCO-CH=CH-CH2-O-OH + HO2.
22              8%    2.6%  HO-aCH[.]-aCH-aCH[.]-CH2-O-OH + O2 -> HO-CH=CH-CH[OO.]-CH2-O-OH
23  2.32e+7   100%   30.8%  .OCH2-CH=CH-CH2-OH -> HO-CH2-aCH[.]-aCH-aCH[.]-OH
24             92%   28.4%  HO-CH2-aCH[.]-aCH-aCH[.]-OH + O2 -> HCO-CH=CH-CH2-OH + HO2.
25              8%    2.4%  HO-CH2-aCH[.]-aCH-aCH[.]-OH + O2 -> HO-CH=CH-CH[OO.]-CH2-OH
26            100%    5.0%  CH2=CH-CH(CH2.)-OH {*OHadd} -> CH2=CH-CH(CH2.)-OH
27            100%    5.0%  CH2=CH-CH(CH2.)-OH + O2 -> CH2=CH-CH(CH2OO.)-OH
28  4.85e-3   100%    5.0%  CH2=CH-CH(CH2OO.)-OH -> .aCH2-aCH-aC[.](OH)-CH2-O-OH
29  8.64e-12   95%    4.7%  CH2=CH-CH(CH2OO.)-OH + NO -> CH2=CH-CH(CH2O.)-OH + NO2
30  4.80e-13    5%    0.3%  CH2=CH-CH(CH2OO.)-OH + NO -> CH2=CH-CH(OH)-CH2-ONO2
31  2.30e-12  100%    5.0%  CH2=CH-CH(CH2OO.)-OH + NO3 -> CH2=CH-CH(CH2O.)-OH + NO2 + O2
32  1.61e-11  100%    5.0%  CH2=CH-CH(CH2OO.)-OH + HO2 -> CH2=CH-CH(OH)-CH2-O-OH + O2
33  1.60e-11  100%    5.0%  CH2=CH-CH(CH2OO.)-OH + RCO3 -> CH2=CH-CH(CH2O.)-OH + O2 + RCO2.
34            100%    5.0%  .aCH2-aCH-aC[.](OH)-CH2-O-OH + O2 -> CH2=CH-CO-CH2-O-OH + HO2.
35            100%   14.7%  CH2=CH-CH(CH2O.)-OH + O2 -> CH*=CH-O-CH=CH* + HO2. + H2O
36  8.77e-9   100%    2.6%  HO-CH=CH-CH[OO.]-CH2-O-OH -> HO-CH=CH-CH(CH[.]-O-OH)-O-OH
37  2.51e+1   100%    2.6%  HO-CH=CH-CH[OO.]-CH2-O-OH -> .OOCH2-CH(CH=CH-OH)-O-OH
38  8.64e-12   95%    2.5%  HO-CH=CH-CH[OO.]-CH2-O-OH + NO -> HO-CH=CH-CH[O.]-CH2-O-OH + NO2
39  4.80e-13    5%    0.1%  HO-CH=CH-CH[OO.]-CH2-O-OH + NO -> HO-CH=CH-CH(ONO2)-CH2-O-OH
40  2.30e-12  100%    2.6%  HO-CH=CH-CH[OO.]-CH2-O-OH + NO3 -> HO-CH=CH-CH[O.]-CH2-O-OH + NO2 +
O2
41  1.61e-11  100%    2.6%  HO-CH=CH-CH[OO.]-CH2-O-OH + HO2 -> HO-CH=CH-CH(CH2-O-OH)-O-OH + O2
42  1.60e-11  100%    2.6%  HO-CH=CH-CH[OO.]-CH2-O-OH + RCO3 -> HO-CH=CH-CH[O.]-CH2-O-OH + O2 +
RCO2.
43            100%    2.6%  HO-CH=CH-CH(CH[.]-O-OH)-O-OH -> HCO-CH(CH=CH-OH)-O-OH + OH
44  4.85e-3   100%    2.6%  .OOCH2-CH(CH=CH-OH)-O-OH -> HO-aCH[.]-aCH-aC[.](CH2-O-OH)-O-OH
45  7.05e+2   100%    2.6%  .OOCH2-CH(CH=CH-OH)-O-OH -> HO-CH=CH-CH[OO.]-CH2-O-OH
46  8.64e-12   95%    2.5%  .OOCH2-CH(CH=CH-OH)-O-OH + NO -> .OCH2-CH(CH=CH-OH)-O-OH + NO2
47  4.80e-13    5%    0.1%  .OOCH2-CH(CH=CH-OH)-O-OH + NO -> HO-CH=CH-CH(CH2-ONO2)-O-OH
48  2.30e-12  100%    2.6%  .OOCH2-CH(CH=CH-OH)-O-OH + NO3 -> .OCH2-CH(CH=CH-OH)-O-OH + NO2 + O2
49  1.61e-11  100%    2.6%  .OOCH2-CH(CH=CH-OH)-O-OH + HO2 -> HO-CH=CH-CH(CH2-O-OH)-O-OH + O2
50  1.60e-11  100%    2.6%  .OOCH2-CH(CH=CH-OH)-O-OH + RCO3 -> .OCH2-CH(CH=CH-OH)-O-OH + O2 +
RCO2.
51  2.40e+9   100%    7.7%  HO-CH=CH-CH[O.]-CH2-O-OH -> HCO-CH=CH-OH + .CH2-O-OH
52            100%    2.6%  HO-aCH[.]-aCH-aC[.](CH2-O-OH)-O-OH -> HO-CH=CH-CO-CH2-O-OH + OH
53  1.33e+11  100%    7.7%  .OCH2-CH(CH=CH-OH)-O-OH -> HO-aCH[.]-aCH-aCH[.]-O-OH + HCHO
54  8.64e-12   95%    2.3%  HO-CH=CH-CH[OO.]-CH2-OH + NO -> HO-CH=CH-CH[O.]-CH2-OH + NO2
55  4.80e-13    5%    0.1%  HO-CH=CH-CH[OO.]-CH2-OH + NO -> HO-CH=CH-CH(ONO2)-CH2-OH
56  2.30e-12  100%    2.4%  HO-CH=CH-CH[OO.]-CH2-OH + NO3 -> HO-CH=CH-CH[O.]-CH2-OH + NO2 + O2
57  1.61e-11  100%    2.4%  HO-CH=CH-CH[OO.]-CH2-OH + HO2 -> HO-CH=CH-CH(CH2-OH)-O-OH + O2
58  1.60e-11  100%    2.4%  HO-CH=CH-CH[OO.]-CH2-OH + RCO3 -> HO-CH=CH-CH[O.]-CH2-OH + O2 + RCO2.
59  2.45e+8   100%    7.1%  HO-CH=CH-CH[O.]-CH2-OH -> HCO-CH=CH-OH + .CH2-OH
60            100%   50.0%  .aCH2-aCH-aCH[.]-CH2-ONO2 {*NO3add} -> .aCH2-aCH-aCH[.]-CH2-ONO2
```

(continued)

Figure S2 (continued)

```
61             100%   50.0%  CH2=CH-CH(CH2.)-ONO2 {*NO3add} -> CH2=CH-CH(CH2.)-ONO2
62              65%   32.5%  .aCH2-aCH-aCH[.]-CH2-ONO2 + O2 -> CH2=CH-CH[OO.]-CH2-ONO2
63              35%   17.5%  .aCH2-aCH-aCH[.]-CH2-ONO2 + O2 -> .OOCH2-CH=CH-CH2-ONO2
64             100%   50.0%  CH2=CH-CH(CH2.)-ONO2 + O2 -> CH2=CH-CH(CH2OO.)-ONO2
65  8.64e-12    95%   47.4%  CH2=CH-CH(CH2OO.)-ONO2 + NO -> CH2=CH-CH(CH2O.)-ONO2 + NO2
66  4.80e-13     5%    2.6%  CH2=CH-CH(CH2OO.)-ONO2 + NO -> CH2=CH-CH(ONO2)-CH2-ONO2
67  2.30e-12   100%   50.0%  CH2=CH-CH(CH2OO.)-ONO2 + NO3 -> CH2=CH-CH(CH2O.)-ONO2 + NO2 + O2
68  1.24e-11   100%   50.0%  CH2=CH-CH(CH2OO.)-ONO2 + HO2 -> CH2=CH-CH(ONO2)-CH2-O-OH + O2
69  1.28e-11    80%   40.0%  CH2=CH-CH(CH2OO.)-ONO2 + RCO3 -> CH2=CH-CH(CH2O.)-ONO2 + O2 + RCO2.
70  3.20e-12    20%   10.0%  CH2=CH-CH(CH2OO.)-ONO2 + RCO3 -> CH2=CH-CH(CHO)-ONO2 + O2 + RCO-OH
71  1.74e+6     96%  132.5%  CH2=CH-CH(CH2O.)-ONO2 -> .aCH2-aCH-aCH[.]-ONO2 + HCHO
72  6.43e+4      4%    4.9%  CH2=CH-CH(CH2O.)-ONO2 + O2 -> CH2=CH-CH(CHO)-ONO2 + HO2.
73  8.64e-12    95%   30.8%  CH2=CH-CH[OO.]-CH2-ONO2 + NO -> CH2=CH-CH[O.]-CH2-ONO2 + NO2
74  4.80e-13     5%    1.7%  CH2=CH-CH[OO.]-CH2-ONO2 + NO -> CH2=CH-CH(ONO2)-CH2-ONO2
75  2.30e-12   100%   32.5%  CH2=CH-CH[OO.]-CH2-ONO2 + NO3 -> CH2=CH-CH[O.]-CH2-ONO2 + NO2 + O2
76  1.24e-11   100%   32.5%  CH2=CH-CH[OO.]-CH2-ONO2 + HO2 -> CH2=CH-CH(CH2-ONO2)-O-OH + O2
77  1.28e-11    80%   26.0%  CH2=CH-CH[OO.]-CH2-ONO2 + RCO3 -> CH2=CH-CH[O.]-CH2-ONO2 + O2 + RCO2.
78  3.20e-12    20%    6.5%  CH2=CH-CH[OO.]-CH2-ONO2 + RCO3 -> CH2=CH-CO-CH2-ONO2 + O2 + RCO-OH
79  4.83e+5     93%   82.9%  CH2=CH-CH[O.]-CH2-ONO2 -> CH2=CH-CHO + .CH2-ONO2
80  3.74e+4      7%    6.4%  CH2=CH-CH[O.]-CH2-ONO2 + O2 -> CH2=CH-CO-CH2-ONO2 + HO2.
81  2.66e-2    100%   17.5%  .OOCH2-CH=CH-CH2-ONO2 -> HO-O-CH2-aCH[.]-aCH-aCH[.]-ONO2
82  8.47e-12    93%   16.2%  .OOCH2-CH=CH-CH2-ONO2 + NO -> .OCH2-CH=CH-CH2-ONO2 + NO2
83  6.59e-13     7%    1.3%  .OOCH2-CH=CH-CH2-ONO2 + NO -> O2NO-CH2-CH=CH-CH2-ONO2
84  2.30e-12   100%   17.5%  .OOCH2-CH=CH-CH2-ONO2 + NO3 -> .OCH2-CH=CH-CH2-ONO2 + NO2 + O2
85  1.24e-11   100%   17.5%  .OOCH2-CH=CH-CH2-ONO2 + HO2 -> HO-O-CH2-CH=CH-CH2-ONO2 + O2
86  1.60e-11   100%   17.5%  .OOCH2-CH=CH-CH2-ONO2 + RCO3 -> .OCH2-CH=CH-CH2-ONO2 + O2 + RCO2.
87             100%   17.5%  HO-O-CH2-aCH[.]-aCH-aCH[.]-ONO2 -> HCO-CH=CH-CH2-O-OH + NO2
88  4.65e+5     88%   45.0%  .OCH2-CH=CH-CH2-ONO2 -> HO-CH2-aCH[.]-aCH-aCH[.]-ONO2
89  6.43e+4     12%    6.2%  .OCH2-CH=CH-CH2-ONO2 + O2 -> HCO-CH=CH-CH2-ONO2 + HO2.
90             100%   45.0%  HO-CH2-aCH[.]-aCH-aCH[.]-ONO2 -> HCO-CH=CH-CH2-OH + NO2
91              14%    6.8%  CH2=CH-CHOO {*O3Ole} -> CH2=CH-CHOO[syn]
92              28%   13.8%  CH2=CH-CHOO {*O3Ole} -> CH2=CH-CHOO[anti]
93              59%   29.5%  CH2=CH-CHOO {*O3Ole} -> .CH2-CHO + HCO.
94             100%    7.7%  HO-aCH[.]-aCH-aCH[.]-O-OH -> HCO-CH=CH-OH + OH
95             100%  132.5%  .aCH2-aCH-aCH[.]-ONO2 -> CH2=CH-CHO + NO2
96             100%   13.8%  CH2=CH-CHOO[anti] -> CH2=CH-CH*-O-O*
97             100%    6.8%  CH2=CH-CHOO[syn] -> .CH2-CHO + HCO.
98             100%   36.2%  .CH2-CHO + O2 -> HCO-CH2OO.
99  2.56e+0    100%   36.2%  HCO-CH2OO. -> CO[.]-CH2-O-OH
100 9.13e-12   100%   36.2%  HCO-CH2OO. + NO -> HCO-CH2O. + NO2
101            100%   36.2%  CO[.]-CH2-O-OH + O2 -> CO[OO.]-CH2-O-OH
102 9.30e+5     94%   33.9%  HCO-CH2O. -> HCHO + HCO.
103 6.43e+4      6%    2.3%  HCO-CH2O. + O2 -> HCO-CHO + HO2.
104 2.18e+4    100%   36.2%  CO[OO.]-CH2-O-OH -> .OOCH2-CO-O-OH
105 9.13e-12   100%   36.2%  .OOCH2-CO-O-OH + NO -> .OCH2-CO-O-OH + NO2
106 2.30e-12   100%   36.2%  .OOCH2-CO-O-OH + NO3 -> .OCH2-CO-O-OH + NO2 + O2
107 1.08e-11    85%   30.8%  .OOCH2-CO-O-OH + HO2 -> HO-O-CH2-CO-O-OH + O2
108 1.91e-12    15%    5.4%  .OOCH2-CO-O-OH + HO2 -> .OCH2-CO-O-OH + OH + O2
109 1.28e-11    80%   29.0%  .OOCH2-CO-O-OH + RCO3 -> .OCH2-CO-O-OH + O2 + RCO2.
110 3.20e-12    20%    7.2%  .OOCH2-CO-O-OH + RCO3 -> HCO-CO-O-OH + O2 + RCO-OH
111 2.16e+4     25%   26.9%  .OCH2-CO-O-OH -> CO[.]-O-OH + HCHO
112 6.43e+4     75%   80.1%  .OCH2-CO-O-OH + O2 -> HCO-CO-O-OH + HO2.
113             42%   21.0%  CH2OO {*O3Ole} -> CH2OO
114             17%    8.4%  CH2OO {*O3Ole} -> HCO. + OH
115             18%    9.0%  CH2OO {*O3Ole} -> CO2 + H2
116             18%    9.0%  CH2OO {*O3Ole} -> CO + H2O
117              5%    2.6%  CH2OO {*O3Ole} -> CO2 + H. + H.
118            100%  176.8%  .CH2-OH + O2 -> HCHO + HO2.
119            100%    7.7%  .CH2-O-OH -> HCHO + OH
120            100%   82.9%  .CH2-ONO2 -> HCHO + NO2
121            100%   78.6%  HCO. + O2 -> CO + HO2.
122            100%   26.9%  CO[.]-O-OH -> CO2 + OH
123            100%    5.2%  H. -> HO2.
```

Figure S3. MechGen output showing the minimally reduced processed mechanism for the reactions of 1,3-butadiene under default conditions.

```
.ACT
13-BUTDE     ! CH2=CH-CH=CH2
ACROLEIN     ! CH2=CH-CHO
VOC-0014     ! CH2=CH-CH*-O-O*
ORG-4682     ! CH2=CH-CH(ONO2)-CH2-OH
ORG-1004     ! CH2=CH-CH(CH2-OH)-O-OH
ORG-1048     ! CH2=CH-CO-CH2-OH
ORG-0212     ! HCO-CH=CH-CH2-O-OH
ORG-4248     ! HCO-CH=CH-CH2-OH
VOC-0001     ! HO-CH2-CH=CH-CH2-ONO2
VOC-0002     ! HO-CH2-CH=CH-CH2-O-OH
VOC-0003     ! CH2=CH-CO-CH2-O-OH
FURAN        ! CH*=CH-O-CH=CH*
ORG-4894     ! CH2=CH-CH(OH)-CH2-ONO2
ORG-5118     ! CH2=CH-CH(OH)-CH2-O-OH
VOC-0004     ! HCO-CH(CH=CH-OH)-O-OH
ORG-0974     ! HCO-CH=CH-OH
ORG-5115     ! HO-CH=CH-CH(ONO2)-CH2-O-OH
ORG-4934     ! HO-CH=CH-CH(CH2-O-OH)-O-OH
VOC-0006     ! HO-CH=CH-CO-CH2-O-OH
VOC-0005     ! HO-CH=CH-CH(CH2-ONO2)-O-OH
VOC-0007     ! HO-CH=CH-CH(ONO2)-CH2-OH
ORG-4941     ! HO-CH=CH-CH(CH2-OH)-O-OH
VOC-0009     ! CH2=CH-CH(CHO)-ONO2
ORG-0120     ! CH2=CH-CH(ONO2)-CH2-ONO2
VOC-0008     ! CH2=CH-CH(ONO2)-CH2-O-OH
ORG-4309     ! CH2=CH-CO-CH2-ONO2
VOC-0010     ! CH2=CH-CH(CH2-ONO2)-O-OH
VOC-0013     ! HCO-CH=CH-CH2-ONO2
VOC-0011     ! O2NO-CH2-CH=CH-CH2-ONO2
VOC-0012     ! HO-O-CH2-CH=CH-CH2-ONO2
GLYACID      ! HCO-CO-O-OH
ORG-0549     ! HO-O-CH2-CO-O-OH

.STS
13-BUTDEr1   ! CH2=CH-CH[OO.]-CH2-OH
13-BUTDEr2   ! .OOCH2-CH=CH-CH2-OH
13-BUTDEr3   ! CH2=CH-CH(CH2OO.)-OH
13-BUTDEr9   ! HCO-CH2OO.
13-BUTDEr7   ! CH2=CH-CH[OO.]-CH2-ONO2
13-BUTDEr8   ! .OOCH2-CH=CH-CH2-ONO2
13-BUTDEr6   ! CH2=CH-CH(CH2OO.)-ONO2
13-BUTDEr4   ! {0.034_.OOCH2-CH(CH=CH-OH)-O-OH}{0.966_HO-CH=CH-CH[OO.]-CH2-O-OH}
13-BUTDEr5   ! HO-CH=CH-CH[OO.]-CH2-OH
13-BUTDEr10  ! .OOCH2-CO-O-OH

.RXN
R) 1.12e-11 -1.053 0.00  ;13-BUTDE + OH = #.62 13-BUTDEr1 + #.33 13-BUTDEr2 +
                          #.05 13-BUTDEr3
R) 1.34e-14 4.537  0.00  ;13-BUTDE + O3 = #.5 ACROLEIN + #.14 VOC-0014 +
                          #.36 13-BUTDEr9 + #.09 H2O + #.08 OH + #.5 HO2 +
                          #.54 CO + #.12 CO2 + #.5 HCHO + #.21 HCHO2 +
                          #.09 H2
R) 1.10e-13              ;13-BUTDE + NO3 = #.32 13-BUTDEr7 + #.18 13-BUTDEr8 +
                          #.5 13-BUTDEr6
R) 9.13e-12              ;13-BUTDEr1 + NO = #.95 ACROLEIN + #.05 ORG-4682 +
                          #.95 HO2 + #.95 NO2 + #.95 HCHO
R) 2.30e-12              ;13-BUTDEr1 + NO3 = ACROLEIN + O2 + HO2 + NO2 + HCHO
R) 1.61e-11              ;13-BUTDEr1 + HO2 = ORG-1004 + O2
R) 1.60e-11              ;13-BUTDEr1 + RCO3 = #.8 ACROLEIN + #.2 ORG-1048 +
                          O2 + #.8 HO2 + #.8 HCHO
R) 1.71e+7 9.778  0.00   ;13-BUTDEr2 = #.92 ORG-0212 + #.08 13-BUTDEr4 +
                          #.92 HO2
R) 9.13e-12              ;13-BUTDEr2 + NO = #.86 ORG-4248 + #.07 13-BUTDEr5 +
                          #.07 VOC-0001 + #.86 HO2 + #.93 NO2
```

(continued)

Figure S3 (continued)

```
R) 1.61e-11               ;13-BUTDEr2 + HO2 = VOC-0002 + O2
R) 3.81e+9 16.219 0.00    ;13-BUTDEr3 = VOC-0003 + HO2
R) 9.13e-12               ;13-BUTDEr3 + NO = #.95 FURAN + #.05 ORG-4894 +
                           #.95 H2O + #.95 HO2 + #.95 NO2
R) 2.30e-12               ;13-BUTDEr3 + NO3 = FURAN + O2 + H2O + HO2 + NO2
R) 1.61e-11               ;13-BUTDEr3 + HO2 = ORG-5118 + O2
R) 1.60e-11               ;13-BUTDEr3 + RCO3 = FURAN + O2 + H2O + HO2
R) 8.48e-9                ;13-BUTDEr4 = VOC-0004 + OH
R) 8.81e-12               ;13-BUTDEr4 + NO = #.95 ORG-0974 + #.05 ORG-5115 +
                           #.95 OH + #.95 NO2 + #.95 HCHO
R) 2.22e-12               ;13-BUTDEr4 + NO3 = ORG-0974 + O2 + OH + NO2 + HCHO
R) 1.55e-11               ;13-BUTDEr4 + HO2 = ORG-4934 + O2
R) 1.55e-11               ;13-BUTDEr4 + RCO3 = ORG-0974 + O2 + OH + HCHO
R) 1.65e-4                ;13-BUTDEr4 = VOC-0006 + OH
R) 3.10e-13               ;13-BUTDEr4 + NO = #.95 ORG-0974 + #.05 VOC-0005 +
                           #.95 OH + #.95 NO2 + #.95 HCHO
R) 7.82e-14               ;13-BUTDEr4 + NO3 = ORG-0974 + O2 + OH + NO2 + HCHO
R) 5.47e-13               ;13-BUTDEr4 + HO2 = ORG-4934 + O2
R) 5.44e-13               ;13-BUTDEr4 + RCO3 = ORG-0974 + O2 + OH + HCHO
R) 9.13e-12               ;13-BUTDEr5 + NO = #.95 ORG-0974 + #.05 VOC-0007 +
                           #.95 HO2 + #.95 NO2 + #.95 HCHO
R) 2.30e-12               ;13-BUTDEr5 + NO3 = ORG-0974 + O2 + HO2 + NO2 + HCHO
R) 1.61e-11               ;13-BUTDEr5 + HO2 = ORG-4941 + O2
R) 1.60e-11               ;13-BUTDEr5 + RCO3 = ORG-0974 + O2 + HO2 + HCHO
R) 9.13e-12               ;13-BUTDEr6 + NO = #.91 ACROLEIN + #.03 VOC-0009 +
                           #.05 ORG-0120 + #.03 HO2 + #1.86 NO2 + #.91 HCHO
R) 2.30e-12               ;13-BUTDEr6 + NO3 = #.96 ACROLEIN + #.04 VOC-0009 +
                           O2 + #.04 HO2 + #1.96 NO2 + #.96 HCHO
R) 1.24e-11               ;13-BUTDEr6 + HO2 = VOC-0008 + O2
R) 1.60e-11               ;13-BUTDEr6 + RCO3 = #.77 ACROLEIN + #.23 VOC-0009 +
                           O2 + #.03 HO2 + #.77 NO2 + #.77 HCHO
R) 9.13e-12               ;13-BUTDEr7 + NO = #.88 ACROLEIN + #.07 ORG-4309 +
                           #.05 ORG-0120 + #.07 HO2 + #1.83 NO2 + #.88 HCHO
R) 2.30e-12               ;13-BUTDEr7 + NO3 = #.93 ACROLEIN + #.07 ORG-4309 +
                           O2 + #.07 HO2 + #1.93 NO2 + #.93 HCHO
R) 1.24e-11               ;13-BUTDEr7 + HO2 = VOC-0010 + O2
R) 1.60e-11               ;13-BUTDEr7 + RCO3 = #.74 ACROLEIN + #.26 ORG-4309 +
                           O2 + #.06 HO2 + #.74 NO2 + #.74 HCHO
R) 1.71e+7 12.011 0.00    ;13-BUTDEr8 = ORG-0212 + NO2
R) 9.13e-12               ;13-BUTDEr8 + NO = #.82 ORG-4248 + #.11 VOC-0013 +
                           #.07 VOC-0011 + #.11 HO2 + #1.74 NO2
R) 2.30e-12               ;13-BUTDEr8 + NO3 = #.88 ORG-4248 + #.12 VOC-0013 +
                           O2 + #.12 HO2 + #1.88 NO2
R) 1.24e-11               ;13-BUTDEr8 + HO2 = VOC-0012 + O2
R) 1.60e-11               ;13-BUTDEr8 + RCO3 = #.88 ORG-4248 + #.12 VOC-0013 +
                           O2 + #.12 HO2 + #.88 NO2
R) 6.02e+7 10.052 0.00    ;13-BUTDEr9 = 13-BUTDEr10
R) 2.55e-12 -0.755 0.00   ;13-BUTDEr9 + NO = HO2 + NO2 + #.94 CO + #.94 HCHO +
                           #.06 GLYOXAL
R) 2.55e-12 -0.755 0.00   ;13-BUTDEr10 + NO = #.75 GLYACID + #.25 OH +
                           #.75 HO2 + NO2 + #.25 CO2 + #.25 HCHO
R) 2.30e-12               ;13-BUTDEr10 + NO3 = #.75 GLYACID + O2 + #.25 OH +
                           #.75 HO2 + NO2 + #.25 CO2 + #.25 HCHO
R) 1.27e-11               ;13-BUTDEr10 + HO2 = #.85 ORG-0549 + #.11 GLYACID +
                           O2 + #.19 OH + #.11 HO2 + #.04 CO2 + #.04 HCHO
R) 1.60e-11               ;13-BUTDEr10 + RCO3 = #.8 GLYACID + O2 + #.2 OH +
                           #.6 HO2 + #.2 CO2 + #.2 HCHO
```

Figure S4. MechGen output showing estimated product yields in various environments for the reactions of 1,3-butadiene, derived using default mechanism generation options and environments.

```
Explicit products formed when 13-BUTDE CH2=CH-CH=CH2 reacts with OH, O3, and NO3 under the
following conditions:

Condition Description
Mid NOx   Mid NOx standard urban conditions (near EBIR)
High NOx  High NOx urban conditions (near MIR)
Low NOx   Low NOx downwind conditions (NOx = MOIR NOx/10)
Night     Nighttime conditions for multi-day, mid-NOx scenario

Products listed in descending order of maximum yield. Yields < 0.05% not shown.

Mid NOx   High NOx  Low NOx   Night     Explicit product or reacted
98.0%     97.6%     97.6%     6.7%      OH reacted
1.9%      2.2%      2.3%      6.4%      O3 reacted
0.1%      0.2%      0.1%      86.9%     NO3 reacted

63.97%    78.14%    49.30%    117.28%   NO2
89.31%    93.43%    78.80%    13.06%    HO2.
67.42%    82.55%    51.98%    0.07%     NO-loss
5.00%      -        17.48%    82.13%    O2
57.16%    61.01%    47.01%    59.51%    HCHO
54.74%    58.34%    45.00%    58.91%    CH2=CH-CHO
26.48%    17.03%    28.71%    13.59%    HCO-CH=CH-CH2-O-OH
3.74%      -        13.83%    0.95%     CH2=CH-CH(CH2-OH)-O-OH
3.10%     11.98%    0.87%     2.61%     HCO-CH=CH-CH2-OH
 -         -         -        8.13%     CH2=CH-CH(ONO2)-CH2-O-OH
 -         -         -        5.28%     CH2=CH-CH(CH2-ONO2)-O-OH
4.39%     4.80%     3.46%     0.74%     H2O
4.22%     4.60%     3.26%     0.17%     CH*=CH-O-CH=CH*
1.05%     1.37%     1.25%     3.45%     CO
2.96%     3.17%     2.37%      -        CH2=CH-CH(ONO2)-CH2-OH
 -         -         -        2.92%     CH2=CH-CH(CHO)-ONO2
 -         -         -        2.69%     CH2=CH-CO-CH2-ONO2
2.23%     2.33%     1.83%     0.13%     HCO-CH=CH-OH
2.32%     1.70%     2.18%     1.21%     OH
0.47%     0.45%     0.52%     1.48%     HCO-CO-O-OH
0.40%     0.46%     0.49%     1.35%     CH2OO
0.38%     0.40%     0.44%     1.21%     CO2
0.29%      -        1.02%      -        CH2=CH-CH(OH)-CH2-O-OH
0.26%     1.01%     0.07%      -        HO-CH2-CH=CH-CH2-ONO2
0.26%     0.30%     0.32%     0.88%     CH2=CH-CH*-O-O*
 -         -         -        0.68%     HO-O-CH2-CH=CH-CH2-ONO2
0.17%     0.20%     0.21%     0.58%     H2
0.14%      -        0.56%      -        HO-CH=CH-CH(CH2-O-OH)-O-OH
0.15%      -        0.43%     0.12%     CH2=CH-CO-CH2-O-OH
 -         -        0.13%     0.38%     HO-O-CH2-CO-O-OH
 -         -         -        0.36%     HCO-CH=CH-CH2-ONO2
0.24%      -        0.31%      -        HO-CH2-CH=CH-CH2-O-OH
0.09%      -        0.28%     0.15%     CH2=CH-CO-CH2-OH
0.23%     0.26%     0.18%      -        CH2=CH-CH(OH)-CH2-ONO2
0.11%     0.07%     0.09%      -        HO-CH=CH-CH(ONO2)-CH2-O-OH
 -        0.05%      -         -        HO-CH=CH-CH(ONO2)-CH2-OH

0.08%     0.05%     0.03%     0.03%     Carbon balance check (should be ~0)

0.76%     0.38%     1.74%     0.10%     VBS6      (C[sat] = 1000.0 ug/m3) (mass yield)
138.35%   132.77%   147.23%   151.37%   VBS7      (volatile) (mass yield)

0.04%     0.02%     0.08%     0.00%     SOA mass yield from VBS when atm PM = 50 ug/m3
0.06%      -        0.16%      -        SOA mass yield from products for atm PM = 50 ug/m3
```

The calculated concentration-time results for the species used to define the standard environments are shown in Figure S5, as well as the average concentrations used to define the standard environments and the intervals that were averaged. The red lines show results from the earlier version of the SAPRC-22 mechanism (derived from output of an earlier version of MechGen not using standard environments) that were used to calculate the averages, while the black lines show results from the current SAPRC-22 mechanism (Carter, 2023). The predictions of the two mechanisms are very close except for the O_3, OH, and HO_2 for the high NO_x scenario, where the current mechanism is somewhat more reactive. (If only a red line is seen, then both mechanisms give the same predictions.) Higher NO_x scenarios tend to be more sensitive to mechanism changes. The environments were not recalculated when the mechanism was updated because the SAPRC-22 mechanisms, derived in part from MechGen outputs, are affected by the standard environments, though the dependence is quite indirect and weak. In any case, the dependence of the results on mechanism changes is quite minor compared to variability of actual environments and uncertainties in the representativeness of the scenarios employed.

Table S2 summarizes the average concentrations of the atmospheric oxidants used to define the three daytime and one nighttime urban environments. The table also shows the fractions of peroxy radicals that undergo bimolecular reactions with NO_x and peroxy species in these environments. As expected, most of the peroxy radicals react with NO or HO_2 during the daytime, while the reactions with NO dominate in the high NO_x environment and decrease to ~75% in the lowest NO_x environment. In the nighttime environment, ~60% of the peroxy radicals react with NO_3, while the remaining ~40% react primarily with HO_2 and acyl peroxy radicals to a similar extent. Consumption of peroxy radicals through reaction with other non-acyl peroxy radicals is minor in all these environments, according to the predictions of SAPRC-22 (Carter, 2023), which are based on estimated mechanisms and assignments generated by MechGen.

Note that the standard daytime scenarios utilize constant photolysis rates for direct overhead sunlight, derived from the actinic fluxes used in the model simulations discussed above. This may somewhat overestimate the relative importance of photolysis reactions when estimating relative reaction rates or deriving yields, and probably should be revised as new environments are developed when MechGen is updated. Additionally, there are no environments representing rural or remote conditions, but MechGen gives advanced users the capability of specifying conditions of their own environments (Carter et al., 2025). Although the current system does not allow users to modify the actinic fluxes directly, they could specify environments with reduced light intensity by inputting an "HV" value less than one.

S3. Additional Information on Mechanism Processing Procedures

The sections below give additional information concerning procedures used during and following full explicit mechanism generation procedures. Scheme S1 and Scheme S2 give the algorithm used for full mechanism generation. The initial processing of interconverting intermediates is discussed in Section S3.1, the procedures for deriving minimally reduced mechanisms is discussed in Section 0, and the derivation of product yields is discussed in Section S3.3. These procedures are described in text and also given in various "schemes" showing the procedures as pseudo-programming codes. Note that these schemes are intended to highlight the essential features of the algorithms employed and omit some of the details involved when the procedures are actually implemented in the software.

S3.1. Initial Processing for Interconverting Intermediates

Some explicit mechanisms generated by MechGen include intermediates that react to ultimately reform themselves. The process for deriving overall product yields in given environments, as discussed in Section S3.3, requires that these sets of "cyclic" intermediates be identified at the start, or the procedure

Table S1. Inputs used in the dynamic box model simulations to derive average pollutant concentrations used to define the standard environments provided with MechGen.

Input Type	Input Values or Description
Location and date for photolyses	Los Angeles, June 21, all times solar daylight
Start and end times	Start at 6:00 AM solar daylight time, end at 9:00 PM on day 4
Emissions start and end times	6:00 AM - 9:00 PM each day, constant flux
ROG emissions input fluxes [a]	3.85 micromoles C /M^2 /day
NO_x emissions input fluxes [a] (varied with scenario)	High NO_x (approx. MIR) 0.87 micromoles N / M^2 /day High O_3 (approx. MOIR) 0.57 Mid NO_x (approx. EBIR) 0.37 Low NO_x (1/10 approx MOIR) 0.06
ROG speciation	75% anthropogenic and 25% biogenic by carbon, as used when deriving lumped mixtures for SAPRC-22 (Carter, 2023)
NO_x speciation	75% NO, 25% NO_2, no HONO
Background species	None except for [CO] = 2 ppm and [CH_4] = 2 ppm
Daily inversion height schedule	6:00 AM - 8:00 AM Constant at 100 M 8:00 AM - 3:00 PM Increase linearly from 100 M to 540 M 3:00 PM - 9:00 PM Constant at 540 M 9:00 PM - 6:00 AM Declines linearly from 540 M to 100 M [b]
Temperature and humidity	300K, constant; [H_2O] = 2.08e4 (~50%RH), constant
Calculation of photolysis rates	Used the actinic fluxes employed when calculating the Carter (1994a,b) reactivity scenarios, as provided by Jeffries for summer conditions at 640 meters (unpublished results). These give an NO_2 photolysis rate of 0.72 min^{-1} for direct overhead sun for the SAPRC-22 mechanism employed.
Averaging times for concentrations in environments	Daytime: 9:00 AM - 5:00 PM, days 2 and 4 Nighttime: 10:00 PM - 5:00 AM, nights 2 and 3 (Mid NO_x only)

[a] Emissions are given in flux units because the mixing height (and therefore mixing volume) varies throughout the day. The ROG input is set as one-quarter that used in the "averaged conditions" scenario based on the model scenarios used to calculate the Carter (1994a) reactivity scales. This was adjusted to obtain ozone levels closer to those in current high O_3 urban conditions in the United States.

[b] The reacting mixture is diluted with background air as the inversion height declines at night.

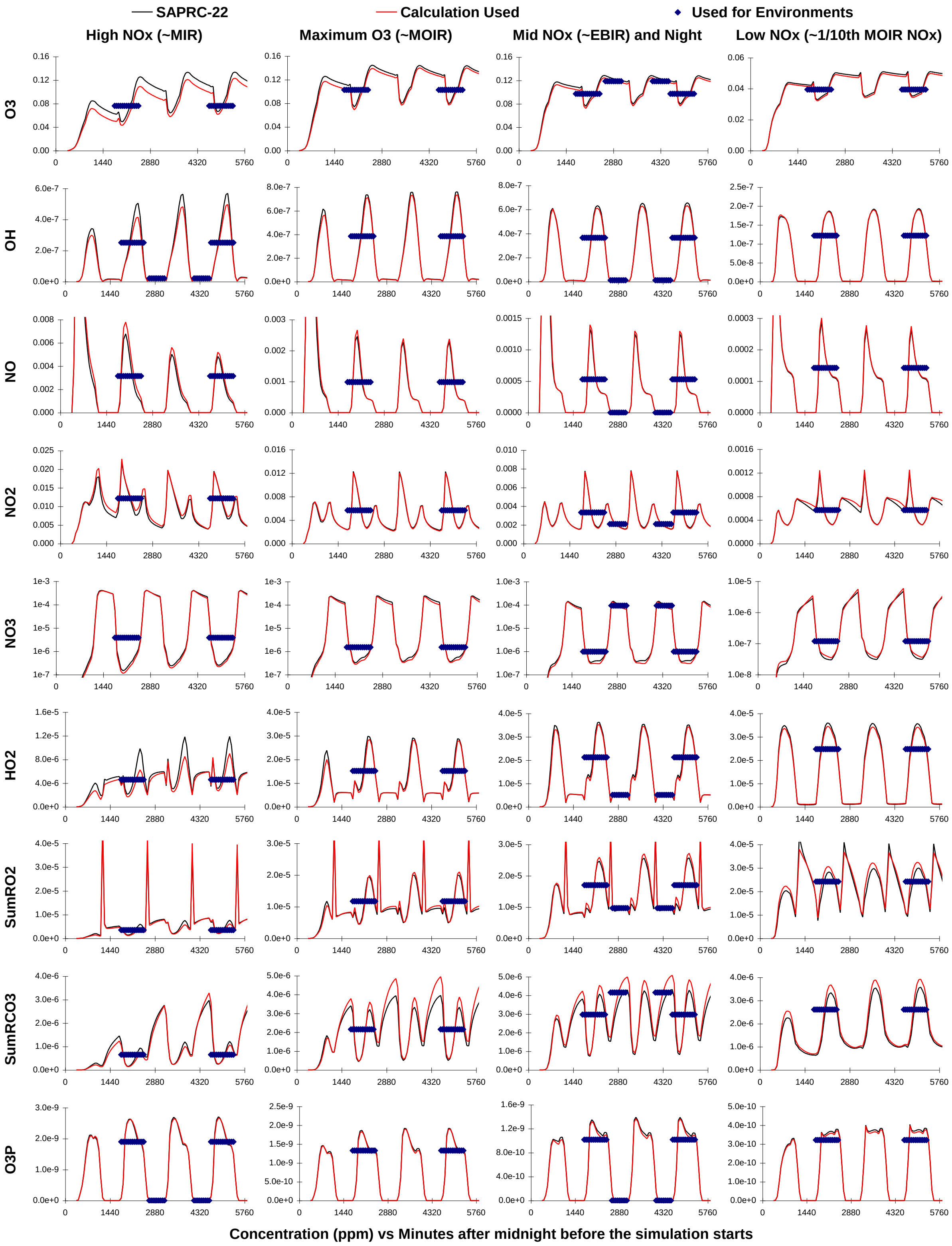


Figure S5. Concentration-time plots of model simulations used to derive oxidant concentrations for the standard environments. Averages used for the daytime and nighttime environments are also shown, indicating the time periods when the averages were derived.

Table S2. Concentrations of species and parameters used to define the standard sets of environmental conditions used in this work.

Scenario	High NOx (MIR)	High O_3 (MOIR)	Standard (EBIR)	Low NOx	Nighttime (EBIR)
Default use [a]	MG		PY, S	MG	MG
Species	Concentrations (molec / cm^3)				
OH	7.51e+6	1.17e+7	1.11e+7	3.69e+6	2.86e+5
O3	1.78e+12	2.38e+12	2.26e+12	9.20e+11	2.90e+12
O^3P	5.50e+4	3.76e+4	2.84e+4	8.85e+3	~0
NO	9.58e+10	3.00e+10	1.59e+10	4.12e+9	7.43e+5
NO_2	3.15e+11	1.50e+11	8.63e+10	1.32e+10	4.88e+10
NO_3	8.54e+6	1.25e+7	9.92e+6	1.25e+6	2.25e+9
HO_2	~0	4.15e+8	6.00e+8	7.17e+8	1.27e+8
Total RO_2's	~0	3.01e+8	4.59e+8	6.69e+8	2.35e+8
Total RCO_3's	~0	4.94e+7	7.30e+7	7.27e+7	1.04e+8
Other inputs	Values				
hν [b]	1	1	1	1	0
T, H_2O [c]	T = 300K; H_2O is present [d]				
Reaction time [e]	6 hours				
Reactant	Fraction of peroxy radical bimolecular reaction with reactant				
NO	~100%	98%	93%	76%	0.08%
HO_2	~0	2%	6%	22%	22%
NO_3	0.002%	0.01%	0.01%	0.01%	59%
Total RCO_3's	~0	0.3%	0.8%	2%	19%
Total RO_2's	~0	0.002%	0.005%	0.02%	0.04%

[a] Usage codes: MG: used for mechanism generation to determine reactions that can be neglected; PY: used to derive product yields for deriving mechanisms of lumped model species; S: used for comparisons of product yields with different lumping methods. It has no default use if blank, but its use can be selected.

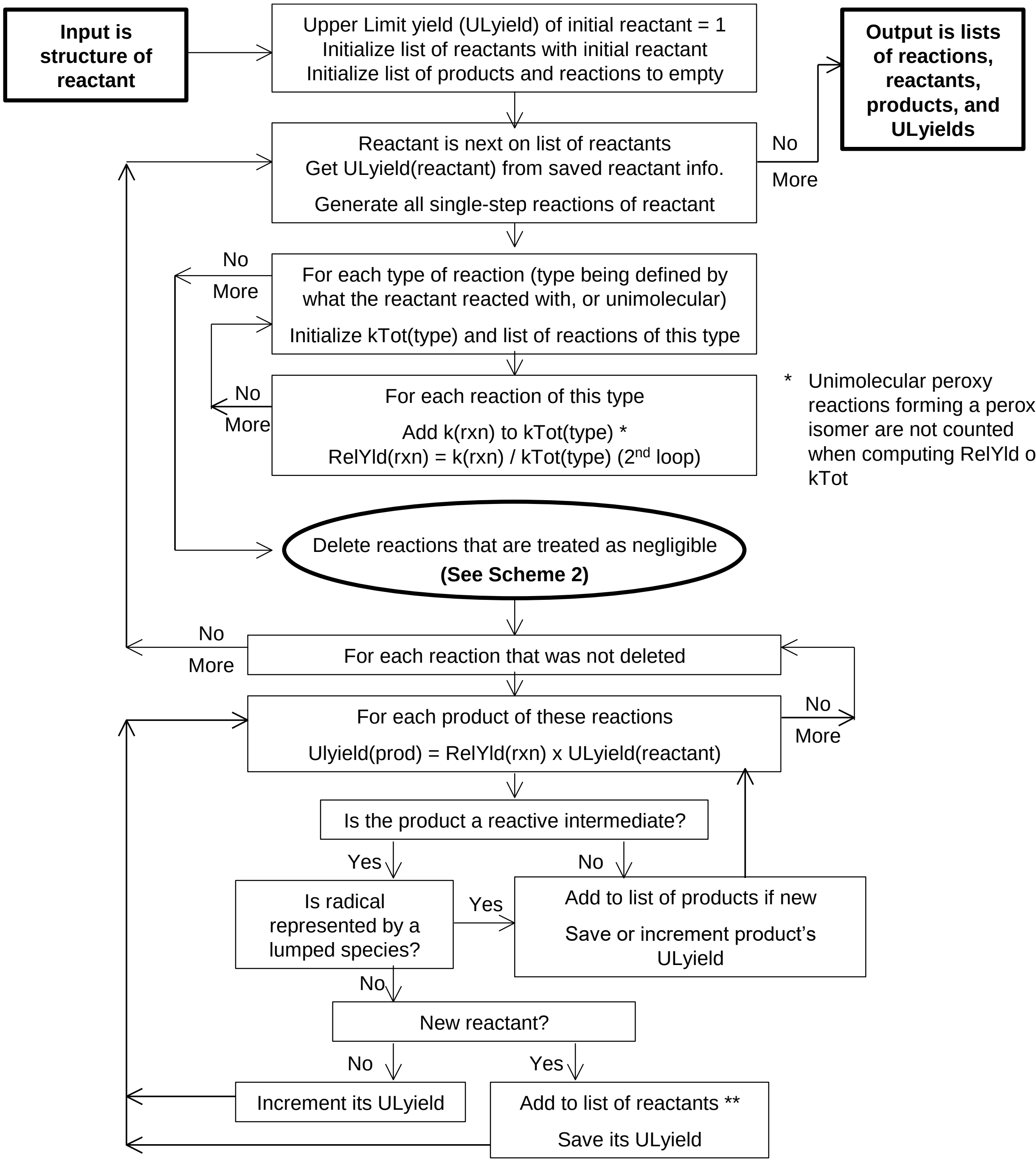
[b] The photolysis rates used for the environments are those calculated for direct overhead sunlight, multiplied by the "HV" value assigned to the environment. The actinic fluxes used are the same as used in the box model calculations as indicated in Table S1.

[c] These are properties of the reactor, not the environment. These are used at mechanism generation time and cannot be modified during subsequent processing. See the User Manual that is available at the MechGen web site (Carter, 2025).

[d] H_2O is a Boolean parameter that is used to determine the fate of unreactive Criegee intermediates (Carter et al., 2025). There are no other reactions generated by MechGen that depend on [H_2O].

[e] This is only used when deriving product yields for lumped model species mechanisms to determine how much product can be formed when the starting compounds have insufficient time to react completely. This is a property of the environment.

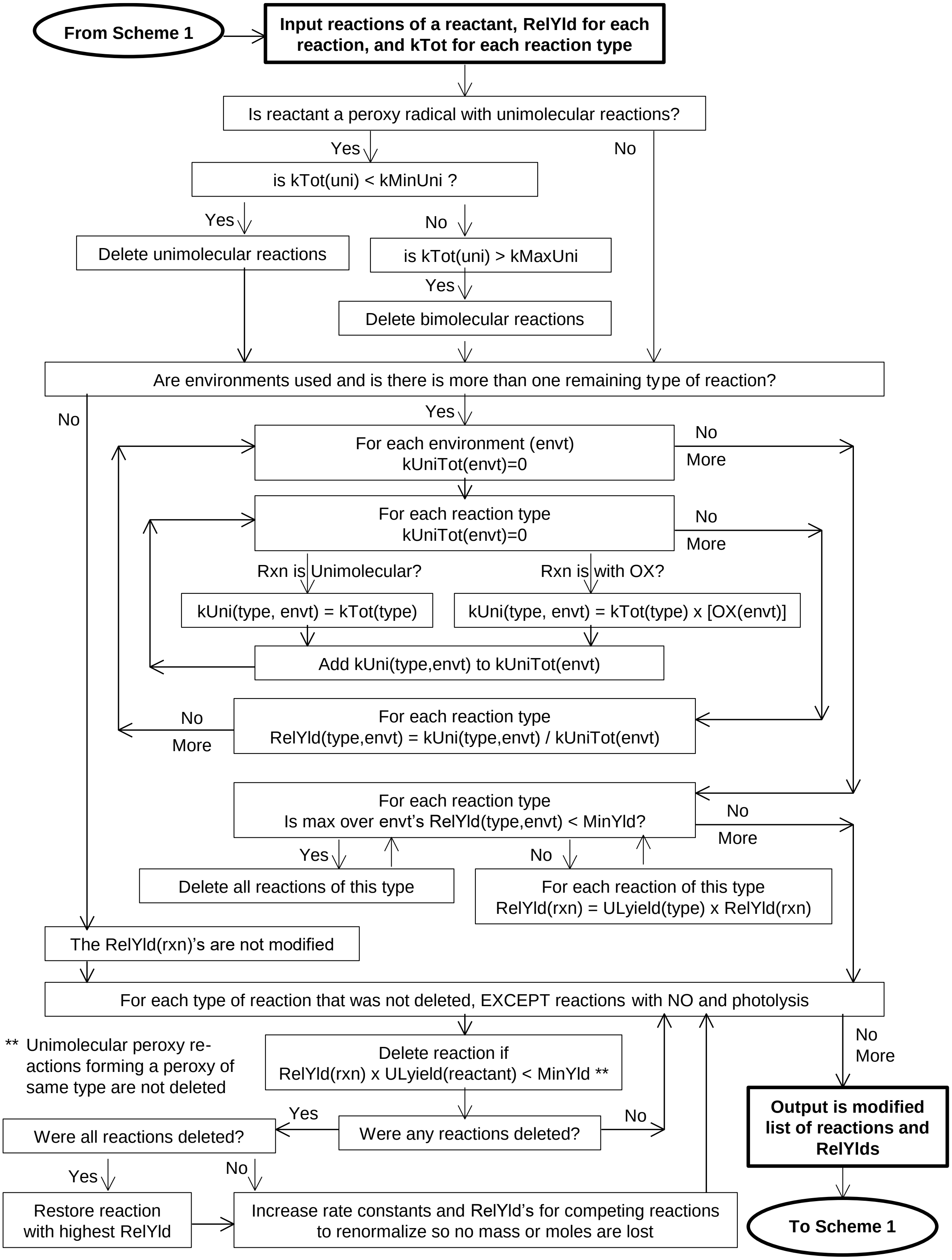
Scheme S1. Algorithm used to for full mechanism generation. The portion used to delete negligible reactions is shown in Scheme S2.



* Unimolecular peroxy reactions forming a peroxy isomer are not counted when computing RelYld or kTot

** Reactants are inserted into list in following order: (1) most carbons; (2) most hydrogens; (3) highest U.L.yield, (4) first formed. This reduces chance that a reactant is formed after it is reacted.

Scheme S2. Portion of the algorithm used to delete reactions considered to be negligible. The full algorithm in shown in Scheme S1.



will not converge. However, the occurrence of interconverting intermediates does not prevent the derivation of minimally reduced processed mechanisms, as discussed in Section 0 (and also Section 3.2 of the main manuscript), as long as interconverting intermediates are retained in the reduced mechanism. Nevertheless, we found that identifying these cyclic sets is more straightforward during the initial processing of the explicit generated mechanism, rather than using the minimally reduced processed mechanisms where many reactions are combined, causing some loss of information.

The only types of intermediates that MechGen currently predicts undergo interconversions are hydroperoxy substituted peroxy radicals, and phenoxy and phenyl peroxy radicals. All of these undergo environment-dependent bimolecular reactions and thus cannot be removed during the minimal reduction step. In addition, some interconversions involving hydroperoxy-substituted peroxy radicals occur at rates that are orders of magnitude faster than the competing reactions that consume the interconverting radicals (Vereecken and Nozière, 2020). This results in the kinetic differential equations for such mechanisms exhibiting mathematical "stiffness", making them difficult to integrate using current numerical methods without requiring excessive computation time. These fast reactions cannot be removed by replacing them with the products formed because they do not represent net sinks for the interconverting intermediates. Furthermore, the slower loss reactions must not be eliminated by application of the MinYld tests incorporated in MechGen, or there will be no loss processes for these radicals in the mechanism. This issue is dealt with during mechanism generation by not considering the rates of hydroperoxy-peroxy interconversion reactions when estimating relative yields. This eliminates the problem of treating necessary consumption reactions as negligible, but does not solve the mathematical stiffness problem.

MechGen deals with this stiffness problem as part of the initial processing step where cycles of interconverting intermediates are identified. Each set of interconverting intermediates is classified as either "slow" or "fast", depending on the ratios of the interconversion rate constants relative to the estimated total unimolecular or pseudo-unimolecular rate constants for loss reactions. The slow sets are saved as properties of the reactant object so they can be used later when calculating yields in specified environments, but are not considered otherwise during this initial processing step. The fast sets are dealt with by assuming that the interconverting radicals are in approximate equilibrium, and replacing the individual radicals in the mechanism by a lumped radical model species representing the sum of their concentrations. Interconversion reactions of these radicals can then be deleted from the mechanism, and the loss reactions of the individual radicals can be replaced with reactions of this lumped radical, with rate constants multiplied by the equilibrium fraction of the reacting radical in the lumped total. The criteria used to distinguish between slow and fast sets, and the method used to estimate the equilibrium fractions, are discussed below.

Members of cyclic sets where the total interconversion rate constants exceed the estimated sums of unimolecular + pseudo-unimolecular rate constant for consumption reactions by a factor of 30 or more are classified as fast, and the rest are classified as slow. Sets where all members are fast are classified as fast sets, while those where all are slow are classified as slow sets. Sets where some members are fast and some are slow are split up into fast and slow sets. In this case, the lumped radical representing the members of the fast set is added to the slow set when deriving product yields.

Calculating equilibrium ratios for fast sets of interconverting radicals is straightforward if there are only two members of the set, but cases with three or more members can occur during multi-generation mechanism derivations where peroxy intermediates with more than one hydroperoxy group are predicted to be formed. In such cases, the equilibrium ratios are calculated using the following differential equations, with the steady state approximation being employed for the radicals:

$$dR_i/dt = F_i + \Sigma_{j\neq i}\, k_{j,i}\, R_j - [\Sigma_{j\neq i}\, k_{i,j} + k'_i\,]\, R_i \approx 0$$

$$F_i = -\Sigma_{j\neq i}\, k_{j,i}\, R_j + [\Sigma_{j\neq i}\, k_{i,j} + k'_i\,]\, R_i$$

$$F_i = \Sigma_j\, A_{j,i}\, R_j, \text{ where } A_{i,j} = -k_{j,i} \text{ if } i\neq j, \text{ and } A_{i,i} = \Sigma_{j\neq i}\, k_{i,j} + k'_i$$

Here, R_i refers to the concentrations of the rapidly interconverting radicals, F_i is the rate of formation of R_i from the other reactions in the mechanism, $k_{i,j}$ is the rate constant for the conversion of R_i to R_j, and $k'i$ is the total unimolecular or pseudo-unimolecular rate constant for all other reactions of R_i. This can be cast into matrix form as

$$\underline{F} = \underline{A}\; \underline{R}$$

where $\underline{F}$ and $\underline{R}$ are vectors of the F_i's and R_i's, respectively, and $\underline{A}$ is a matrix derived from the rate constants as indicated above. The R_i values can then be obtained by inverting the matrix $\underline{A}$.

$$\underline{R} = \underline{A}^{-1}\; \underline{F}$$

Note that the quantities of interest are the *Relative* R_i values, or R_i / [$\Sigma_j R_j$], which will depend primarily on the $k_{i,j}$'s. Although at least one nonzero value of the k'_i's must be used or the matrix cannot be inverted, the relative R_i values are quite insensitive to the F_i and k'_i values, as long as $\Sigma_{j\neq i}\, k_{i,j} >> k'i$, which implies that $\Sigma_{j\neq i}\, k_{i,j}\, R_j >> F_i$ if the R_i's are in steady state. Simulations integrating the kinetic differential equations were used to verify that this is the case. This is important because the F_i and k'_i may depend on the environment where the mechanism is applied, and this pre-processing step is intended to be environment-independent (other than temperature and O_2 concentration). In order to process the mechanism, we assume that all F_i values are unity, since the relative R_i values only depend on the relative F_i's, and since the results are not sensitive to the values of F_i used. Therefore,

$$\text{Relative } R_i = Ri \,/\, [\, \Sigma_j R_j \,] = \Sigma_j A^{-1}_{j,i} \,/\, [\, \Sigma_j \Sigma_k A^{-1}_{k..j} \,]$$

This assumption of equilibrium and environment-independence is an approximation and the actual ratios of interconverting radicals during actual simulations will depend, to some extent, on the environment and relative rates of formations of the individual radicals. However, if the steady state approximation is appropriate then the total concentrations of the interconverting radicals should still be correct under this treatment, and errors introduced by assuming equilibrium and environment independence should be small compared to uncertainties in the $k_{i,j}$ values, which are the major factors affecting the relative R_i values, and which are all estimated.

Scheme S3 shows the pseudo-code used for this pre-processing step. The inputs are the lists of reactions and species in the explicit mechanism. The rate constant ratio used to distinguish between "fast" and "slow" cyclic sets, 30 by default, is a property of the lumping object that the MechGen programmer can modify. The output consists of the lists of fast and slow sets of interconverting intermediates, which are saved as properties of the reactant for subsequent use (for the slow sets) or for diagnostic purposes (for the fast sets), and the mechanism and species lists where the intermediates in fast sets are represented by lumped model species. If no fast sets are found, the output reaction and species lists are the same as input.

S3.2. Derivation of Minimally Reduced Mechanisms

The overall process of deriving minimally reduced mechanisms from the generated explicit mechanism is discussed in Section 3.2 of the main manuscript. The procedure employed is outlined in Scheme S4. The starting point is the output of the explicit mechanism after being processed to remove fast interconversions as discussed in the previous section and shown in Scheme S3. The list of species produced during the mechanism generation process includes information about which species are reacting

Scheme S3 Pseudo-code showing procedure used to identify sets of interconverting intermediates in explicit generated mechanisms and to apply the equilibrium approximation for rapidly interconverting sets.

Starting point is lists of reactions and species in the explicit mechanism

```
Loop over explicit reactions
    If this is a unimolecular reaction of a peroxy radical with an assigned rate constant (not fast), then
        If not on list of potentially cyclic peroxy radicals, then
            Add reactant to that list
            Set kUni = k(RO2+NO) * [NOeff]
                        Initialize two lists – one for sets of interconverting radicals (referred to as
                        "mates"), and one for their interconverting reactions
                Set kCyc to 0
        Add rate constant to kUni
        If product is a peroxy radical, then
            Add product to list of "mates" of radical
            Add reaction to the interconverting reaction list for this radical
            Add rate constant to total kCyc
    Else if this is a phenoxy + O3 reaction, then
        Add reactant and the phenyl peroxy product to list of "phenyl" sets
End loop over explicit reactions

Remove potentially cyclic radicals that have no "mates" from list

Use "mates" data and the cyclic radical list to form sets of interconverting peroxy radicals
Delete sets that have only one member

Initialize "slow" sets with the "phenyl" sets, if any

Initialize "fast" sets to empty

Loop over each cyclic peroxy set
    Loop over each member of the set
        If kCyc/(kUni-kCyc) > 30, then
            Add this to the "fast" subset
        Else add this to the "slow" subset
    End loop over members
        If the "fast" subset has more than one member, then
            Create a model species lumping all the members of the "fast" subset
            Add this new species to the list of intermediates that react
            Remove "fast" subset members from list of intermediates that react
        Add the "fast" subset to the "fast" sets
    If the "slow" subset is not empty, then
        If the “fast” subset is not empty, then
                        Add the lumped model species or the only member for the "fast" subset to the
                        "slow" subset
            Add this expanded "slow" subset to the "slow" cyclic sets
        Else if the "slow" subset has more than one member
            Add the "slow" subset to the total "slow" cyclic sets
End loop over all cyclic peroxy sets
```

Save the list of "slow" cyclic sets as a property of the reactant

(Continued)

Scheme S3 (continued)

```
If no "fast" cyclic sets, then
    Return lists of explicit reactions and species without modifications
    Finished with procedure

Loop over each "fast" cyclic set
    Initialize an N x N matrix A, where N is the number of members of the set to zeros
    Loop over each member "i" of the set
        Set A(i, i) = kUni(i)
        Loop through each reaction in list of interconverting reactions for this radical "i"
            Let "j" be the member of the set formed in this reaction
            Set A(i, j) = - (rate constant for this reaction)
        End loop over interconverting reactions
    End loop over members
    Invert A to get Ainv
    For each member "i" of the set
        Calculate Fac(i) = $\Sigma_j$ Ainv(j, i), where "j" refers to the other members of the set
        Normalize Fac(i) by dividing it by $\Sigma_j$ Fac(j)
        Associate this list of factors with this set
End loop over "fast" cyclic sets

Loop over each explicit reaction
    If the reactant is a member of a "fast" cyclic set, then
        If the reaction is on the list of interconverting reactions for that radical, then
            Delete this reaction
        Else multiply the rate constant by the "Fac" value for this member of that set
    If the reaction has a product that is in a "fast" cyclic set, then
        Replace this product with the lumped model species for that set
End loop over reactions

Save the list of "fast" cyclic sets and their relative composition factors as a property on the
reactant

Return lists of modified reactions and species

Finished with procedure
```

Scheme S4. Pseudo-code showing procedure used to derive the minimally processed reduced mechanism from the explicit mechanism modified if applicable as indicated in Scheme S3.

```
**Starting point is lists of reactions and species in the explicit mechanism after being processed to replace rapidly interconverting radicals with reactions of lumped cyclic radical sets.**

Loop over each explicit reaction after being processed
    If the reaction is with $O_2$, then
        Change it to pseudo-unimolecular using [$O_2$] assigned to the reactor
    If the last reaction processed had different reactants (excluding $O_2$), then
        Create a new processed reaction for these reactants, with no products
        Initialize rate constant for these reactants to zero
    If "LostMoles" is a product, then
        Replace it with atom number counter species, with yields = atom number in the reactant
    Add the rate constant for this reaction to the total rate constant for this processed reaction [a]
    Add each product to the processed reaction with the yield multiplied by the rate constant
End loop over explicit reactions

Initialize an empty list of replacement reactions for each intermediate with only unimolecular or $O_2$ reactions, excluding peroxy intermediates (replaced species)

Loop over each processed reaction
    Divide all product yields by the total rate constant
    If the reactant is on the list of intermediates with only unimolecular or $O_2$ reactions, then
        Remove this from the list of processed reactions
End loop over processed reactions

Loop until no changes made [b]
    Set "changed" to false
    Loop over each replaced species
        Loop over each product of this species
            If product is a species to be replaced, then
                Replace product with its products
                Multiply yield of its products by the yield of the product being replaced
                Set "changed" to true
        End loop over products
    End loop over replacement species
Loop ends when "changed" still false

Loop over each remaining processed reaction
    Loop over each product of reaction
        If product is a replaced species, then
            Replace this product by products in replacement reaction
            Multiply yield of products by yield of product being replaced
    End loop over products
End loop over processed reactions

**Save list of processed reactions and non-replaced species as a property on the reactant**
**Finished with procedure**
```

[a] Use the rate constant for the temperature of the reactor, or the quantum yield if a photolysis reaction.

[b] This will converge because the interconverting intermediates are not included in these substitutions.

intermediates, and which of those react only by unimolecular or O_2 reactions. The latter includes those that react so fast that their rate constants are not specified.

The output of this procedure is the lists of reactions and species in the minimally reduced processed mechanism. The species list includes the initial reactant and all stable products from the explicit mechanism, as well as the intermediates that undergo bimolecular reactions, which are mainly peroxy radicals. Most of these intermediates are represented explicitly, except for lumped model species used to represent rapidly interconverting groups. The species list also provides the structures within those groups and their relative equilibrium fractions derived as discussed above.

S3.3. Procedure for Estimating Product Yields

The procedure for estimating overall product yields for a given environment is shown in Scheme S5. It involves first converting all the reactions in the processed mechanism into pseudo-unimolecular reactions with rate constants calculated for the conditions of the environment. Then, all the reacting intermediates in the reactions forming them are replaced with the products of the pseudo-unimolecular reactions of these intermediates. Bimolecular reactions are converted into unimolecular by multiplying the rate constant by the concentration of the co-reactant in the standard environment specified, removing the co-reactant as a reactant, and combining all the pseudo-unimolecular reactions of the initial reactant or the intermediate so there is only reaction for each. Photolysis reactions are treated as unimolecular reactions with rate constants derived for the spectral distribution of the light source that is assigned to the reactor and environments, using the sum of overall quantum yields assigned to the reactions, when applicable.

The pseudo-unimolecular reactions derived for the initial reactant are used as the starting point, and all of its products that are intermediates are replaced by the products of the unimolecular reactions of the intermediates. Before this can be done, the intermediates formed in the reactions of the intermediates also need to be replaced by the products that they form. This can be done in an iterative manner if there are no sets of cyclic intermediates, though several iterations may be required if the reactions are not generated in the optimum order. As discussed in Section 3.3 of the main manuscript, the mechanism generation procedure attempts to minimize cases where reactions forming reacting intermediates are generated after the reactions of the intermediates, though this is not always successful.

This procedure will not converge if there are sets of cyclic intermediates, so it needs to be modified to account for them. As discussed above, "fast" cyclic intermediates are replaced by non-cyclic lumped intermediates before preparing the processed mechanism, but the "slow" cyclics still remain in the processed mechanism. However, these are identified during the initial processing discussed above, and are used as inputs to the yield derivation procedure discussed below. As a first step, only the non-cyclic intermediates are removed as the reactions of the initial reactant, with the formation of the cyclics and their reactions being retained in the pseudo-unimolecular mechanism. The non-cyclic intermediates formed in the reactions of the cyclic intermediates are removed from the unimolecular reactions of the cyclics in the same way as for the starting reactant. The resulting pseudo-unimolecular mechanism then represents only the products formed from the starting reactant and from each of the cyclic intermediates, with all the non-cyclic intermediates removed.

The cyclic intermediates are then removed from this mechanism in a stepwise manner, by processing each set of interconverting intermediates one at a time. Strictly speaking, they could all be processed in a single step, but this would require inversions of sometimes large matrices that are mostly diagonal, requiring more computer processing time than necessary. The ordering of cyclic sets being processed does not affect the results and is essentially arbitrary, determined by the order in which they are discovered in the pre-processing step. The other cyclic intermediates that have not yet been removed are

Scheme S5. Pseudo-code showing procedure used to derive overall product yields for a given environment.

```
Input is:
    List of reactions and reactants in the processed mechanism
    Concentrations of atmospheric oxidants in the environment used
    Photolysis frequencies for each type of photolysis reaction in the environment
    List of interconverting intermediates found during initial processing (Scheme 2)

Initialize empty lists of reactants and reactions of reactants

First loop over each processed reaction
    If this is a photolysis reaction, then
        Set kUni = (photolysis frequency for type) x (overall quantum yield for the reaction)
    Else
        Set k as the rate constant for reactor temperature and pressure
        If this is a bimolecular reaction, then
            Set kUni = k x (concentration of co-reactant in the environment)
        Else Set kUni = k
    Replace the rate constant for the reaction with kUni (but keep the co-reactant, if bimolecular)
    If the reactant is not on the list of organic reactants, then
        Add this reactant to the list of reactants
        Add its overall reaction to list of reactions of reactants with zero rate constant and no products
        Add kUni to the rate constant for the overall reaction of the reactant
End first loop over processed reactions

Second loop over each processed reaction
    Find reactant on list of reactants and identify its overall reaction
    Set Fac = kUni (for this processed reaction) / kUni (for the overall reaction of reactant)
    Define Rct2 as the co-reactant if bimolecular, "uni" if unimolecular, or "HV" if photolysis
    If reactant is the initial reactant, then
        Add "<rct2>-rct" to the list of products of the overall reaction, with yield of Fac
    Else
        Add "<rct2>-loss" to the list of products of the overall reaction, with yield of Fac
    Loop over each product of the processed reaction
        If the product is on the list of reactants and is NOT a cyclic radical, then
            Add this product to the list of radical products of the overall reaction, with yield x Fac
        Else
            Add this product to the products of the overall reaction, with yield x Fac
    End loop over products
End second loop over processed reactions

Loop over each reactant in a list containing the initial reactant + cyclic intermediates
    Get list of radical products formed in the overall reaction of the reactant

    Loop until list of radical products is empty
        Get the first radical from the radical products list
        Set Yld is the yield of this radical product in the overall reaction
        Loop over each non-radical product on the list
            Add the product to the products of the overall reaction being processed, with yield x Yld
        End loop over non-radical products
        Loop over each radical product on the list
            Add this radical to the list of the radical products of the overall reaction, with yield x Yld
        End loop over radical products
    Loop ends if no more radical products [a]
End loop over initial reactant + cyclic intermediates list
```

(Continued)

Scheme S5 (continued)

```
If no cyclic intermediates, then
    Return kUni, products and yields of the overall reaction of the initial reactant
    Finished with procedure

Loop over each set of cyclic intermediates
    Initialize N x N matrix A to all zeros, where N is the number of cyclic intermediates in this set
    First Loop over each cyclic intermediate in this set (matrix index "i")
        Set list of non-radical products to empty
        Set A(i, i) = kUni (rate constant for its overall reaction of intermediate)
        Loop over each product in its overall reaction
            If the product is another cyclic intermediate in this set, then
                Set A(i, j) = - kUni x (yield of the cyclic intermediate "j" from this intermediate "i")
            Else
                Add product and its yield to list of non-radical products for the intermediate
        End loop over products
    End first loop over cyclic intermediates in this set

    Invert matrix A

    Second loop over each cyclic intermediate in this set (matrix index "i")
        Set Fac = $\Sigma_j$ [ $A^{-1}$ (i, j) x (Yield of cyclic intermediate "j" in the overall reaction of initial reactant)]
        Loop over each non-cyclic intermediate product of overall reaction of this cyclic intermediate
            Add product to the products of the overall reaction of the initial reactant, with yield x Fac
        End loop over products
    End second loop over cyclic intermediates
End loop over sets of cyclic intermediates

Return kUni, products and yields of the overall reaction of the initial reactant
Finished with procedure
```

[a] This may take a number of iterations, but should converge if cyclics were identified properly.

treated as final products during these steps, and the products they form are not considered until their interconverting reactions are processed.

The procedure used for removing a given set of interconverting cyclic radicals is as follows. Let R_i and R_j refer to interconverting radicals, P_k refer to other products (including interconverting radicals in other sets that have not yet been removed), k_i be the total pseudo-unimolecular rate constant for reactions of R_i, $y_{i,j}$ be the yield of cyclic radical j from the overall reactions of radical i, $y_{i,k}$ be the yield of product P_k from radical i, k_0 be the total pseudo-unimolecular rate constant for the initial reactant, $y_{0,i}$ be the yield of R_i in the initial reaction, and $y_{0,k}$ be the yield of product P_k in the initial reaction due to the reactions of the cyclic intermediates. Note that the values of k_0, k_i, $y_{i,j}$, $y_{i,k}$ $y_{0,i}$, are all known at the start of the procedure, and the values of $y_{0,k}$ need to be derived. If we apply the steady-state approximation on the cyclic intermediates, then

$$dR_i/dt = k_0\, y_{0,i} + \Sigma_{j\neq i}\, k_j\, y_{j,i}\, R_j - k_i\, R_i \approx 0$$

$$k_0\, y_{0,i} = \Sigma_j\, A_{j,i}\, R_j, \text{ where } A_{i,j} = -\, k_i\, y_{j,i} \text{ if } i\neq j, \quad \text{and} \quad A_{i,i} = k_i$$

or in matrix form

$$\underline{F} = \underline{A}\ \underline{R}$$

where $\underline{F}$ and $\underline{R}$ are vectors of the k_0 $y_{0,i}$'s and R_i's, respectively, and $\underline{A}$ is the matrix derived from the rate constants and cyclic radical yields as indicated above. The R_i values can then be obtained by inverting the matrix $\underline{A}.\underline{R} = \underline{A}^{-1}\ \underline{F}$

or $$R_i = k_0 \Sigma_j A^{-1}_{i,j}\, y_{0,j}$$

This can be used to obtain yields of products P_k from the reactions of the cyclic radicals from

$$dP_k/dt = k_0\, y_{0,k} = \Sigma_i\, y_{i,k}\, R_i = k_o\, \Sigma_i\, [\, y_{i,k}\, \Sigma_j\, A^{-1}_{i,j}\, y_{0,j}\,]$$

or $$y_{0,k} = \Sigma_i\, [\, y_{i,k}\, \Sigma_j\, A^{-1}_{i,j}\, y_{0,j}\,]$$

This is used to update the products and product yields in the pseudo-unimolecular reactions of the initial reactant, which in turn is used as the starting point for the next set of cyclic radicals. Once this is completed, the procedure returns the list of products and yields of the pseudo-unimolecular reaction of the initial reactant, as well as the total pseudo-unimolecular rate constant. The yields are taken as the yields of the products when the initial reactant reacts in the environment. These also include "yields" giving the relative amounts the initial reactant reacts with the various atmospheric oxidants or unimolecularly, and also relative amounts of NO, HO_2, NO_3, total peroxy, and total acyl peroxy radicals that react with the peroxy intermediates. Note that if the mechanism was generated with lumping in effect, the yields output will include yields of lumped radicals in the mechanism, if any, but not the products formed from the reactions of these lumped radicals.

S4. Volatility and Condensed Fraction Estimates

The example mechanisms discussed in this paper include estimates of equilibrium fractions of products in the particle phase. There are derived by summing up the yields of each organic product, multiplied by the equilibrium fraction of the product in the particle phase in a specified environment. The latter can be estimated using equilibrium partitioning theory (Pankow, 1994), as implemented by Donahue et al. (2006), using the following equations:

Fraction in particle phase = $[1 + C^*(\text{compound, T}) / C_{OA}(\text{environ})\ ]^{-1}$ (unitless)
C*(compound, T) = Effective saturation concentration of the compound ($\mu g/m^3$)
= $P_{vap}(\text{compound, T}) \times \text{Mwt(compound)} \times \gamma \times \text{UF} / (\text{R x T})$
C_{OA}(environ) = Organic aerosol concentration in the environment ($\mu g/m^3$)
P_{vap}(compound, T) = Vapor pressure of the compound at temperature T (atm)
Mwt(compound) = Molecular weight of the compound (g/mol)
γ = Activity coefficient (unitless) -- approximated as 1.0 (unitless)
T(environ) = Temperature of the environment (deg K) (deg K)
R = Gas constant = $8.21\text{x}10^{-5}$ (m^3 atm /degK mol)
UF = Unit conversion factor = 10^{-6} (g/μg)

The inputs related to the environment include the temperature (T) and the organic aerosol concentration (C_{OA}), and the examples in the following section use T = 298 K and C_{OA} = 50 $\mu g/m^3$. The compound-dependent inputs are the molecular weight (Mwt) and the vapor pressure (P_{vap}) at the temperature of the environment. The vapor pressure is estimated from the structure of the molecule as discussed below.

MechGen currently has the capability of estimating heats of formation using either the SIMPOL (Pankow and Asher, 2008) or EVAPORATION (Compernolle et al., 2011) methods. The SIMPOL method is used by default and in the examples discussed in this work because it provides estimates for a wider range of compounds. We may add the capability of using additional methods, e.g., Nannoolal et al (2008), in the future. Further discussion of vapor pressure and SOA estimation methods is beyond the scope of this work.

S5. Additional Information on Representative Compounds Studied

S5.1. Results for α-Pinene

The effects of varying the MinYld parameter and environment options for mechanisms generated for α-pinene on yield changes relative to the reference mechanism for this compound for the mid NO_x urban environment is shown in Figure 9 of the main manuscript. The effects of the corresponding yield changes for other representative urban environments are compared to those with the mid NO_x environments in Figure S6. The trends are similar for the urban daytime environments, but the total yield changes are lower at night, especially for SOA precursors. There is no NO present at night so there is no NO consumed in the nighttime environment.

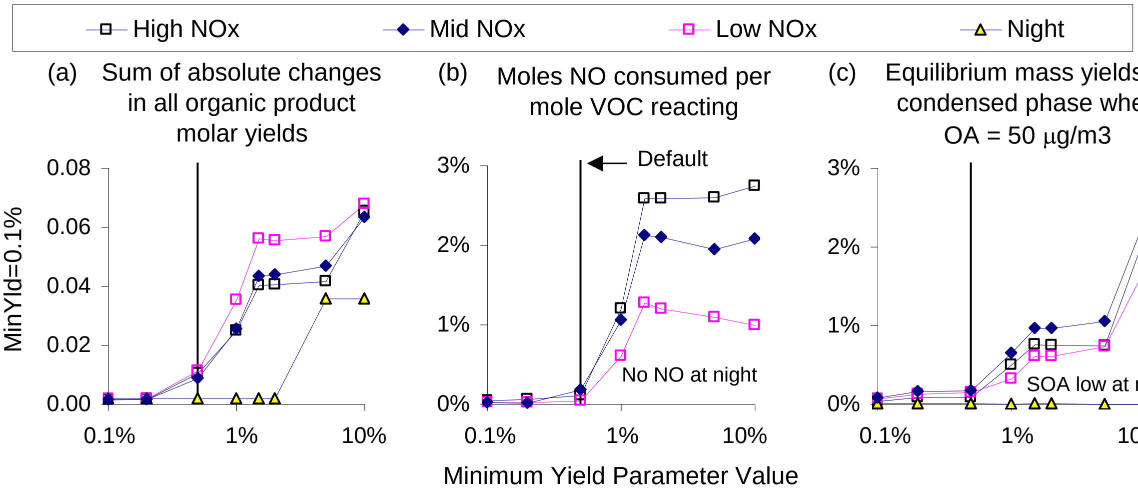


Figure S6. Effects of varying the MinYld parameter on total yield changes predicted for various environments for α-pinene mechanisms derived using the default environment options.

S5.2. Results for Other Representative Compounds

Table S3 lists the representative compounds for which mechanisms were derived as discussed in Section 3.4 of the main manuscript, and gives mechanism derivation results for various mechanism generation options. These include the number of reactions in the explicit mechanisms, the number of stable products, the numbers of intermediates in the minimally reduced processed mechanisms, and the differences in product yields in the mid-NO_x environment compared to the large reference mechanisms. The default MinYld of 0.5% and the default environment options were used except as indicated on the column headings.

Figure S7 shows the ratios of numbers stable products and reacting intermediates to the numbers of reactions in the explicit mechanisms against the mechanism size for the compounds and mechanisms listed in Table S3. Figure S8 shows the ratios of numbers of reacting intermediates or reactions in the minimally processed mechanisms to their corresponding numbers in the explicit mechanism for these same cases. Figure S9 shows yield differences for the high NO_x environment against mechanism size. In all these figures, the mechanism size is measured by the log of the number of explicit reactions, though the plots would be qualitatively similar if other measures of mechanism size were used.

Table S3. List of representative compounds used to provide examples of results of single generation mechanism derivations using three sets of mechanism generation options.

Compound	Default Options				No Environments				MinYld = 0.1% [a]			
	Explicit Rxns	Processed		Yield Diff [b]	Explicit Rxns	Processed		Yield Diff [b]	Explicit Rxns	Processed		Yield Diff [b]
		Prods	Int's			Prods	Int's			Prods	Int's	
Normal Alkanes												
Propane	26	11	3	0.005	46	15	4	-	37	13	4	-
n-Butane	39	14	4	0.001	52	18	4	0.001	41	14	4	0.001
n-Pentane	74	26	8	0.003	102	33	8	0.001	94	34	9	0.001
n-Hexane	138	50	14	0.003	183	62	14	0.002	161	57	15	0.001
n-Heptane	182	67	18	0.004	250	84	19	0.004	248	85	23	0.001
n-Octane	195	75	18	0.017	256	91	18	0.016	363	121	34	0.001
n-Nonane	271	103	25	0.010	356	127	25	0.010	496	165	46	0.001
n-Decane	347	130	32	0.010	467	164	33	0.009	661	221	61	0.001
n-Undecane	411	156	38	0.010	567	206	40	0.010	847	285	78	0.002
n-Dodecane	498	189	46	0.017	684	241	49	0.016	1187	401	108	0.003
n-Tridecane	621	235	58	0.014	945	334	69	0.012	1676	565	151	0.003
n-Tetradecane	781	292	73	0.020	1163	407	85	0.019	2280	766	204	0.003
n-Pentadecane	997	370	93	0.022	1489	518	110	0.020	3108	1046	277	0.003
n-Hexadecane	1140	422	106	0.024	1732	600	127	0.023	3902	1308	345	0.005
Representative C_8 Compounds												
2,2,3,3-Methyl Butane	63	21	7	0.001	80	28	7	-	67	22	7	-
2-Methyl Heptane	1057	315	77	0.006	1507	441	87	0.005	1823	510	125	0.002
Ethylcyclohexane	1586	490	130	0.008	5532	1553	267	0.006	5554	1458	352	0.001
trans-2-octene	783	237	60	0.015	1634	443	109	0.013	2442	687	199	0.002
3,5-Octadiene	661	193	39	0.014	1491	382	71	0.013	1361	385	91	0.002
1-Ethyl cyclohexene	989	290	79	0.012	3976	1071	252	0.009	4463	1227	367	0.007
Octanal	399	136	39	0.013	1158	359	76	0.010	981	306	85	0.007
2-Octanone	268	97	27	0.031	399	137	31	0.016	397	136	37	0.001
n-Butyl Butyrate	447	129	37	0.005	602	169	37	0.005	672	179	50	0.002
m-Xylene	101	29	7	0.001	123	37	7	-	108	31	7	0.001
2,4-Dimethyl Phenol	107	39	7	0.024	115	42	7	0.024	184	56	14	-
Representative Terpenes												
a-Pinene	1717	488	139	0.013	4270	1193	256	0.012	5982	1597	457	0.003
b-Pinene	3852	1071	288	0.020	12064	3121	650	0.016	12418	3154	902	0.006

[a] Uses default environment options.

[b] Sum of absolute differences in product yields derived for these options and the mid NO_x urban standard environment, compared to yields derived for MinYld=0.01%. A "-" indicates that the difference is less than 0.0005.

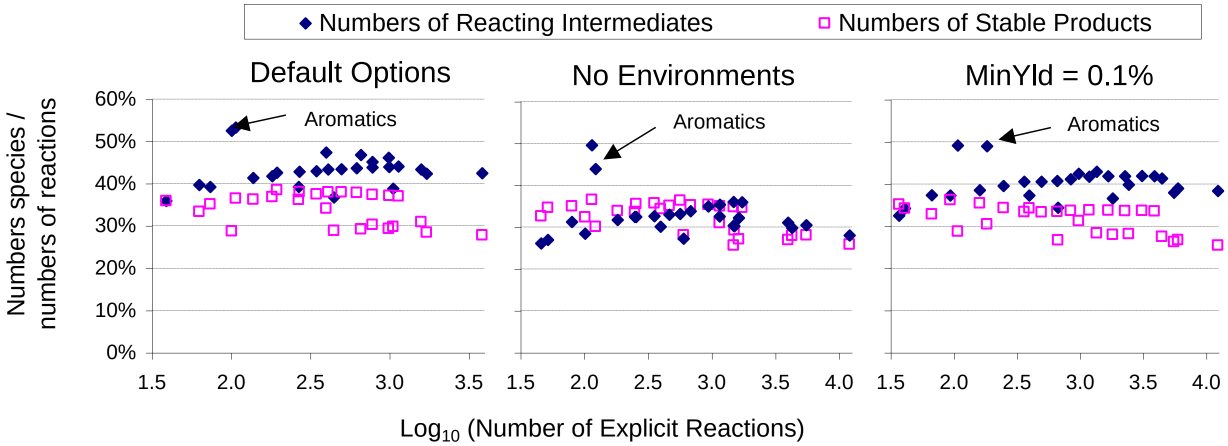


Figure S7. Ratios of the numbers of intermediates and stable products relative to numbers of reactions against a measure of mechanism size for the explicit generated mechanisms for the 27 representative compounds summarized in Table S3.

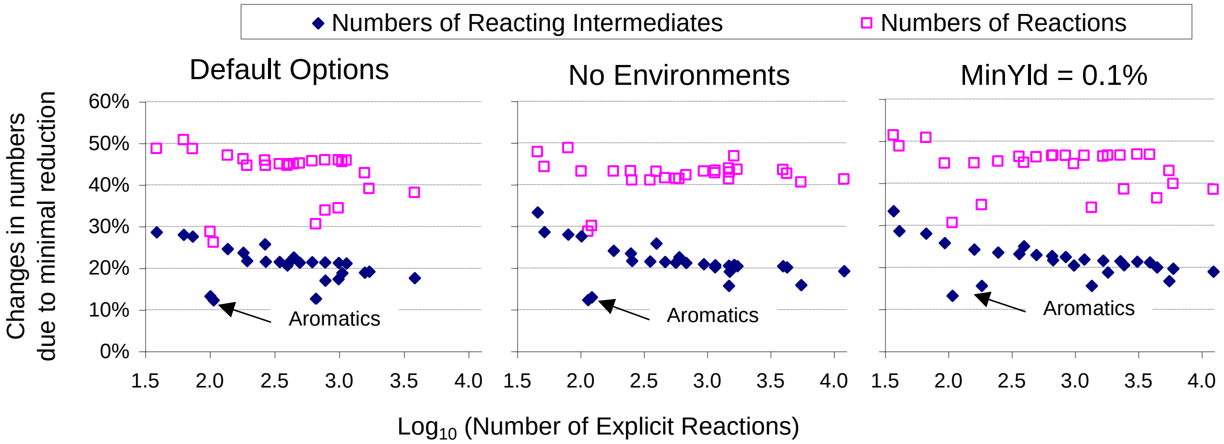


Figure S8. Ratios of the numbers of reactions and intermediates in the minimally processed reduced mechanisms to the numbers of explicit reactions against the log of the numbers of explicit reactions for the mechanisms used to create Table S3.

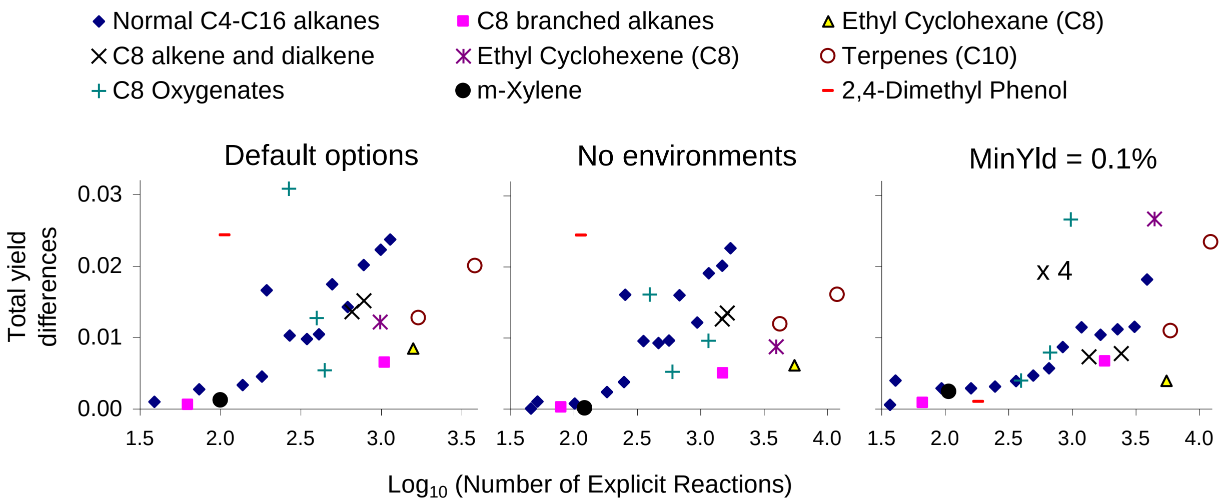


Figure S9. Total yield differences for the high NO_x environment for compounds listed in Table S3 against the numbers of explicit reactions in their mechanisms. These are shown for mechanisms generated using the default options and using no environments.

S6. Additional Information on Multi-Generation Mechanisms

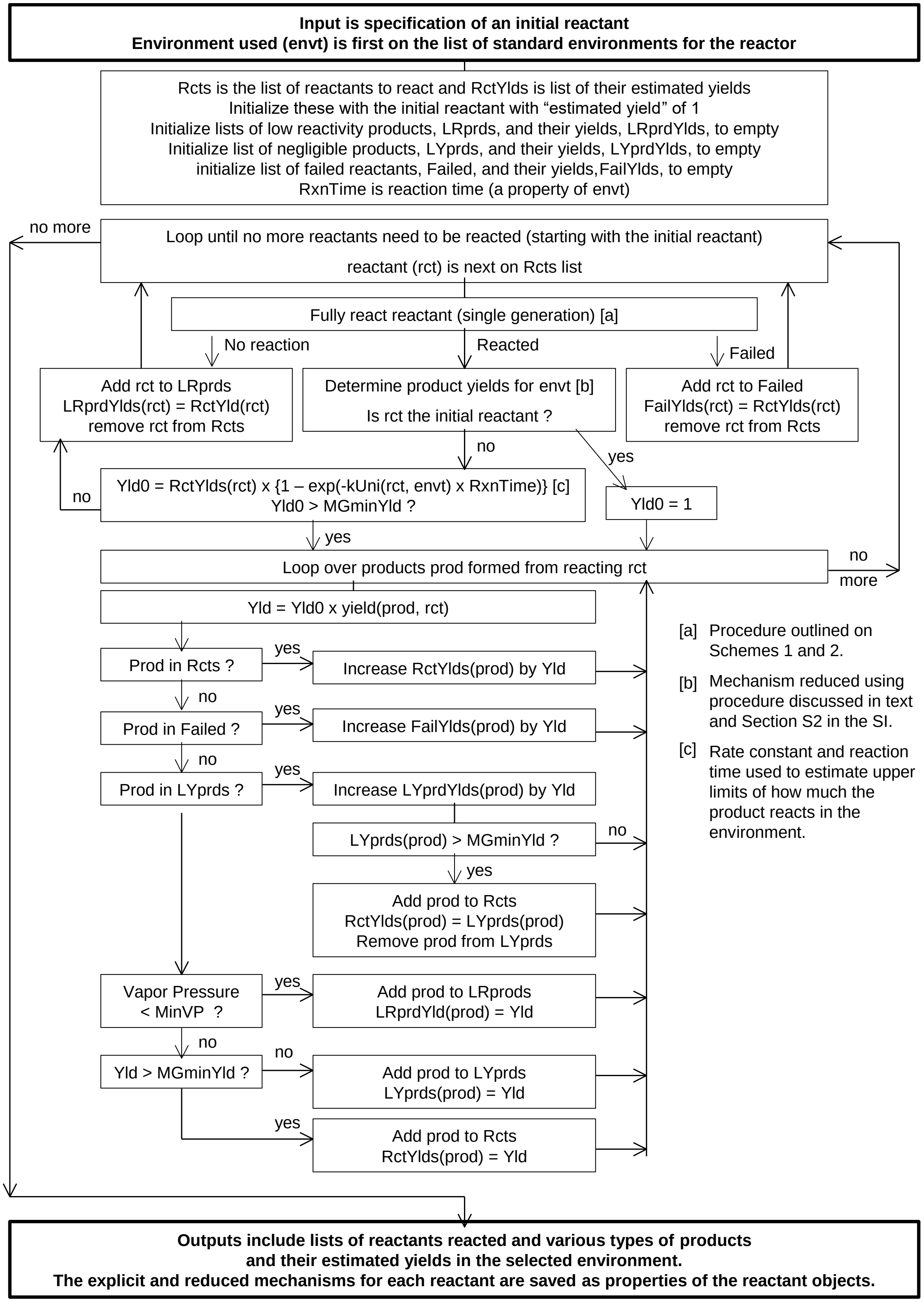
S6.1. Procedure

Scheme S6 shows a diagram of the algorithm used to generate multi-generation mechanisms for a given reactant, given the reaction structure, a standard environment indicating the conditions where it reacts, the MGminYld parameter giving the lowest estimated yield, and the MinVP parameter giving the lowest estimated vapor pressure for compounds to react. The results consist of lists of reacting and unreacting products and the minimally processed reduced mechanisms of the starting compound and its non-negligible multi-generation products that were derived from their explicit mechanisms. The types of information that can be obtained concerning the reactants, products, and reactions in multi-generation mechanisms are listed in Table S4. The methods for obtaining this information are described in the User Manual that is available at the MechGen web site (Carter, 2025).

S6.2. Representative results

Figure S10 gives concentration-time results for the highest concentration structural counter species or C_1 products for the compounds where multi-generation mechanisms were derived, as listed in Table 9 of the main manuscript. These mechanisms were all generated using a MGminYld of 0.001% and the mid NO_x urban standard environment. Plots of effects of varying the MinYld parameter on yields of the various functional groups shown for α-pinene are shown on Figure S11.

Scheme S6. Diagram of the algorithm used to generate multi-generation mechanisms for a given reactant and environment.



[a] Procedure outlined on Schemes 1 and 2.

[b] Mechanism reduced using procedure discussed in text and Section S2 in the SI.

[c] Rate constant and reaction time used to estimate upper limits of how much the product reacts in the environment.

Table S4. Information that can be obtained concerning the reactants, products, and reactions following multi-generation mechanism derivations.

Item	Description
	General reactant and product information
Identity, Smiles	MechGen structure and SMILES strings
Name	Permanent or temporary name used for compound
Type	Compound type and type code number
nC, etc.; Mwt	Numbers of C, H, O, N, S, F, Cl, I, Br, Si, and P atoms, and molecular weight
VP	Estimated vapor pressure at the temperature of the reactor [a]
	Information applicable to this particular multi-generation mechanism and environment
Moles	Approximate yield of reactant or product as moles per mole of initial reactant used during the process of mechanism generation [b]
Generation and formed from	Generation of reactions where this first formed in non-negligible yields, and reactants forming this reactant and reactants forming it
Reacted with	List of atmospheric oxidants that this reactant reacted with, including "uni" to indicate unimolecular reactions, when applicable
kUni	Pseudo-unimolecular rate constant for consumption of the reactant in the environment used for multi-generation mechanism derivation
NegMoles, C, O, etc.	Moles, atom numbers, and mass of products that are formed with yields less than MGminYld during multi-generation mechanism derivation
Numbers of reactions and products	Total numbers of reactants and products in the explicit and reduced mechanisms for the compounds
	Structural Information about organic reactants and products [c]
SepDBs, ConjDBs, CumDBs	Numbers of separated, conjugated or cumulated double bonds
Tbonds	Number of triple bonds
NArings, AroRngs	Numbers of non-aromatic or aromatic rings
Subst2	Number of groups in molecule bonded to two or more non-carbon centered groups
OHs, Acids, Pacid	Numbers of non-acid or acid OH groups
OOHs, Pacid	Numbers of non-acid or acid hydroperoxy groups
EtherO, EsterO	Number of ether or ester oxygens
OOs	Number of peroxy groups (excluding hydroperoxides or peroxy acids)
CHOs	Number of aldehyde and formate groups
COs	Number of ketone groups (carbonyls not in acid, peroxyacid, or ester groups)
ONO2s	Number of nitrate groups, excluding peroxynitrates
PANs, OONO2	Numbers of acyl peroxy (-CO-O-ONO2), or non-acyl peroxynitrate groups
NO2s, ONOs	Numbers of nitro or nitrite groups

Table S4 (continued)

Item	Description
Aminos, Amides	Number of amino or amide (-CO-NHx) groups
	Generated Mechanisms
Explicit mechanisms for all reacted compounds	The explicit mechanisms of all the reacting compounds can be output in a single file. Structures of the reacting compounds are listed at the top of the file, but the products are not listed. Temperature dependences or files giving absorption cross sections and quantum yields are given with the reactions where applicable.
Processed mechanisms for individual reacted compounds [d]	Minimally reduced processed mechanisms can be output separately for each reacting compound. Model species names are assigned to each reactant, intermediate, and these are listed with the structures being represented. This includes low yield products that are formed in reactions but not reacted.
Entire processed multi-generation mechanism [d]	This gives a merged processed mechanism including the reactions of all the compounds reacted. The format is similar to that given for the individual compounds except that low yield products are represented using counter species for each atom and their total mass, rather than individually [e]. The reactions can be output either with a single rate constant for each reaction representing default reactor conditions, or by giving the temperature dependence and names of photolysis assignment files for reactions where they are applicable.
Pseudo-unimolecular multi-generation mechanisms [d]	This gives a mechanism consisting of a single pseudo-unimolecular reaction of each reacting species, with rate constants calculated for the default temperature and light spectrum of the reactor and the oxidant concentrations in the standard environment used, and with product yields indicating the yields of the products in the environment used to derive the mechanism. These mechanisms are much smaller than the entire processed mechanisms because they need no intermediate species, but are applicable only for calculating reactant and stable product concentrations as a function of time when reacting in that environment.

[a] Estimated by either the SIMPOL.1 (Pankow and Asher, 2008) or EVAPORATION (Compernolle et al., 2011) methods, as selected by the user. SIMPOL is the default used in the examples in this paper.

[b] These are approximate estimates used to determine whether the products are formed in high enough yields to be included in the mechanism. The actual amount formed in the environment may be higher because some products may also be formed from reactions of subsequently reacted products.

[c] Some compounds may have more than one such group. These count the numbers of groups rather than the numbers of compounds.

[d] These files are output in a format that can be used as input to the SAPRC box modeling programs (Carter, 2024). However, the entire processed multi-generation mechanisms for all but the smallest compounds are too large to be modeled using the SAPRC software, but the software can handle the pseudo-unimolecular mechanisms.

[e] This is because tens of thousands of low yield products can be predicted for multi-generation mechanisms of medium or larger sized molecules. Summary information about the atoms and mass being represented should be more useful for most applications because although the yields of each individual compound be negligible, the sum of their yields may not be.

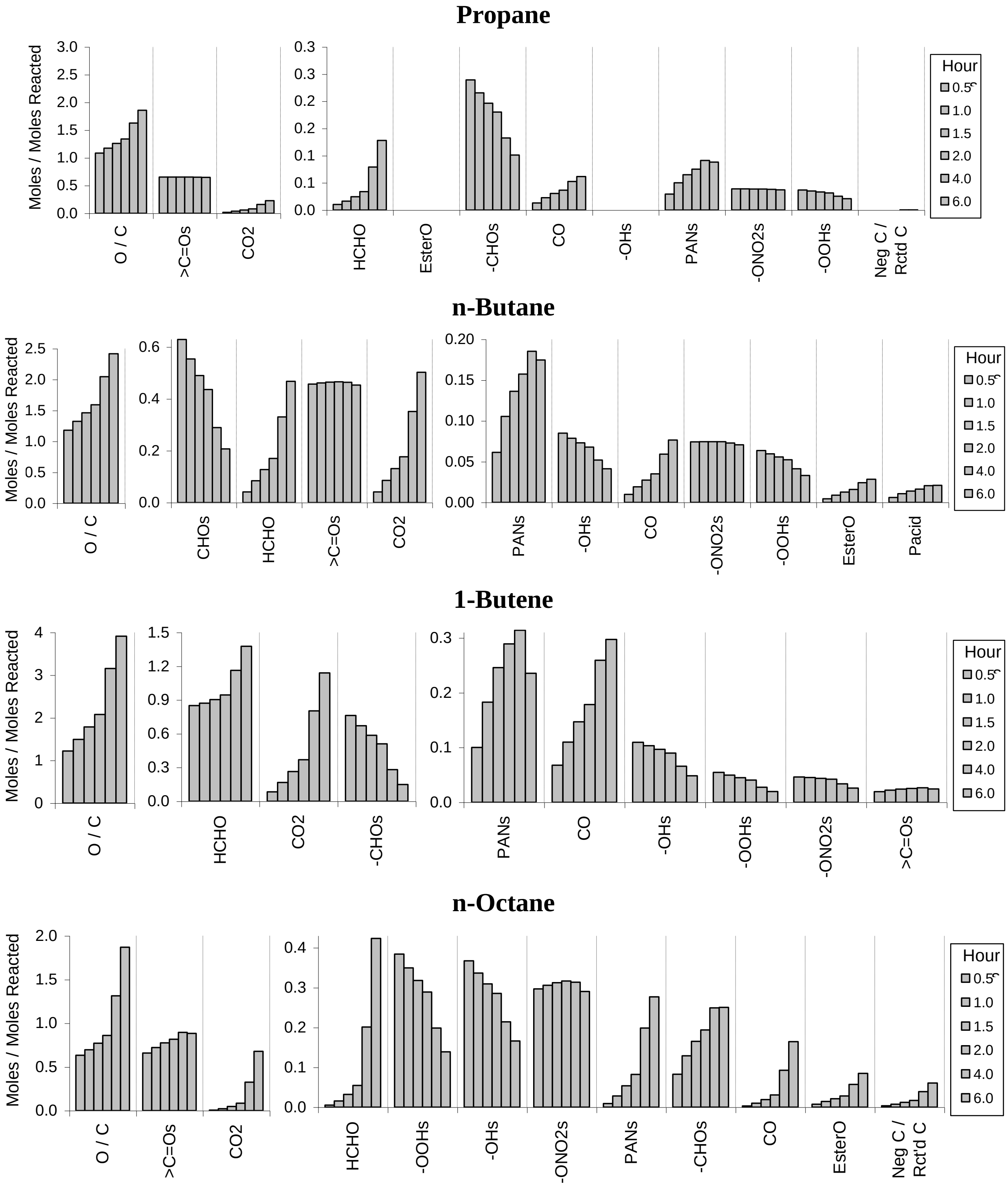
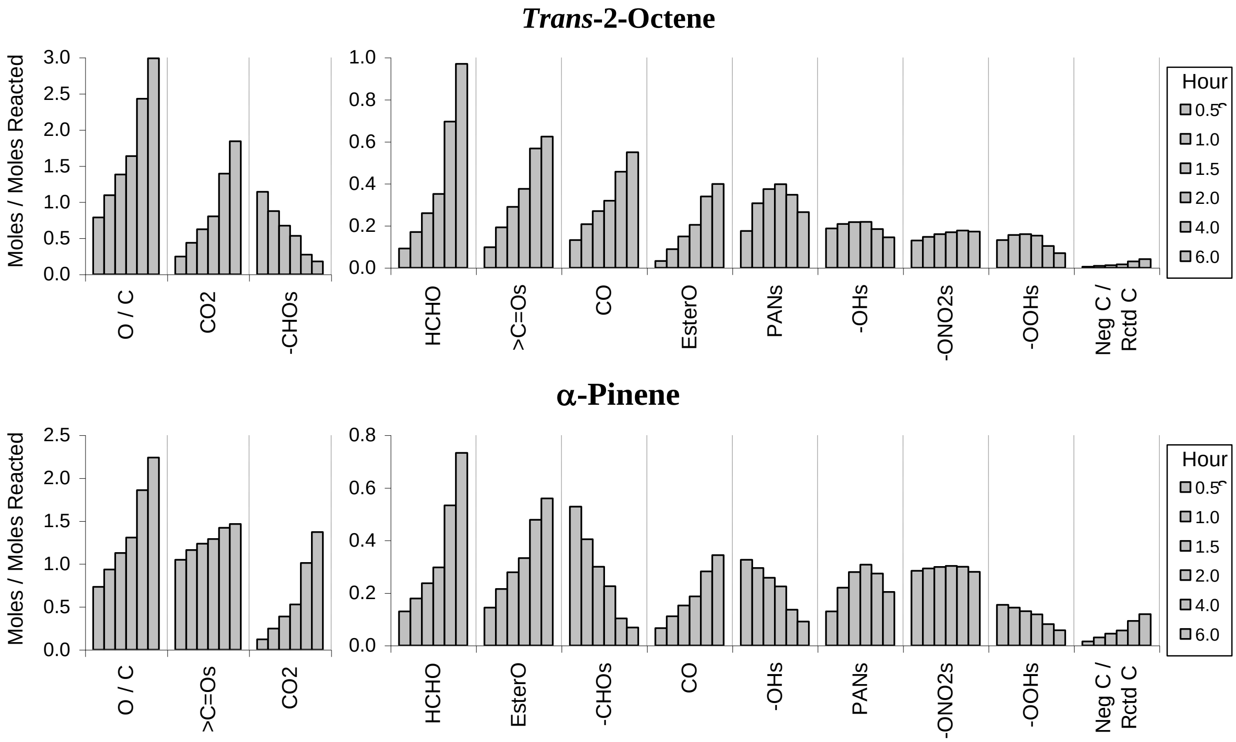


Figure S10. Moles of various types of products or functional groups formed as a function of time for 6-hour simulations using multi-generation mechanisms for several example compounds. Products are ordered from high to low maximum yields.


***Trans*-2-Octene**

α-Pinene

Figure S10 (continued). Moles of various types of products or functional groups formed as a function of time for 6-hour simulations using multi-generation mechanisms for several example compounds. Products are ordered from high to low maximum yields.

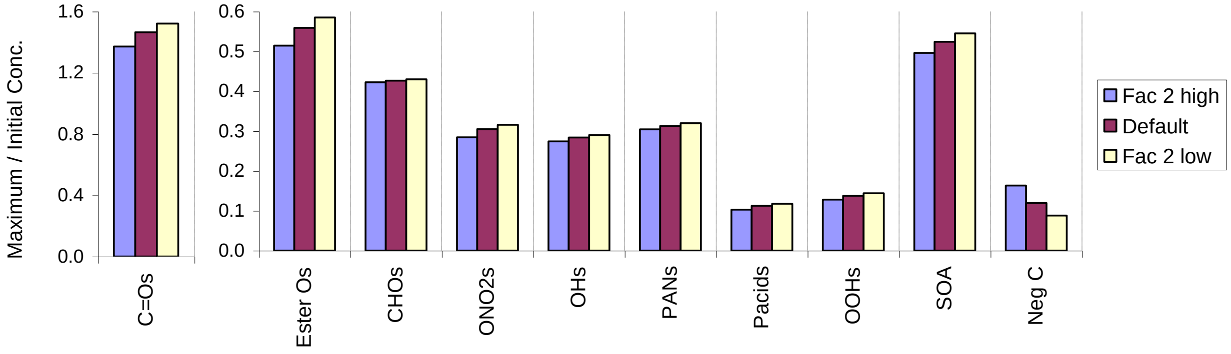


Figure S11. Effects of varying the MGminYld parameter on yields of products or functional groups for the simulations using multi-generation mechanisms for α-pinene.

S7. References

Carter, W. P. L.: Calculation of Reactivity Scales Using an Updated Carbon Bond IV Mechanism, Report Prepared for Systems Applications International Under Funding from the Auto/Oil Air Quality Improvement Research Program, Available at https://doi.org/10.5281/zenodo.13787754, 1994a.

Carter, W. P. L.: Development of Ozone Reactivity Scales for Volatile Organic Compounds, Journal of Air and Waste Management, 44, 881–899, https://doi.org/10.1080/1073161X.1994.10467290, 1994b.

Carter, W. P. L.: Documentation of the SAPRC-22 Mechanisms. Available at https://doi.org/10.5281/zenodo.12601488. September 9, 2023.

Carter, W. P. L.: SAPRC-07 and SAPRC-11 Chemical Mechanisms, Test Simulations, and Environmental Chamber Simulation Files. Web site at https://www.cert.ucr.edu/~carter/SAPRC/SAPRCfiles.htm. June 22, 2024.

Carter, W. P. L.: "SAPRC Mechanism Generation System for the Atmospheric Reactions of Volatile Organic Compounds in the Presence of NOx,". Web site at https://mechgen.cert.ucr.edu or https://intra.engr.ucr.edu/~carter/MechGen/, February 26, 2025.

Carter, W. P. L., Jiang, J., Orlando, J. J., and Barsanti, K. C.: Derivation of atmospheric reaction mechanisms for volatile organic compounds by the SAPRC mechanism generation system (MechGen), Atmos. Chem. Phys., 25, 199–242, https://doi.org/10.5194/acp-25-199-2025, 2025.

Compernolle, S., Ceulemans, K., and Müller, J.-F.: EVAPORATION: a new vapour pressure estimation methodfor organic molecules including non-additivity and intramolecular interactions, Atmospheric Chemistry and Physics, 11, 9431–9450, https://doi.org/10.5194/acp-11-9431-2011, 2011.

Nannoolal, Y., Rarey, J., and Ramjugernath, D.: Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions, Fluid Phase Equilibria, 269, 117–133, https://doi.org/10.1016/j.fluid.2008.04.020, 2008.

Pankow, J. F. and Asher, W. E.: SIMPOL.1: a simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds, Atmospheric Chemistry and Physics, 8, 2773–2796, https://doi.org/10.5194/acp-8-2773-2008, 2008.

Vereecken, L. and Nozière, B.: H migration in peroxy radicals under atmospheric conditions, Atmospheric Chemistry and Physics, 20, 7429–7458, https://doi.org/10.5194/acp-20-7429-2020, 2020.