Supporting Information

Atmospheric Oxidation of Squalene: Molecular Study Using COBRA Modeling and High-Resolution Mass Spectrometry

David R. Fooshee[1], Paige K. Aiona[2], Alexander Laskin[3], Julia Laskin[4], Sergey A. Nizkorodov[2,*], Pierre F. Baldi[1,*]

[1]*School of Information and Computer Sciences, University of California, Irvine, CA 92697-3435*

[2]*Department of Chemistry, University of California, Irvine, California 92697, USA.*

[3]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, USA.*

[4]*Physical Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA.*

* Corresponding authors:

Pierre F. Baldi, e-mail: pfbaldi@ics.uci.edu, phone: (949) 824-5809

Sergey A. Nizkorodov, e-mail: nizkorod@uci.edu, phone: (949) 824-1262

Figure S1: An ESI mass spectrum of squalene observed before the ozone exposure. Unlike the reconstructed mass spectrum of oxidized squalene in Figure 1 of the manuscript, which is plotted as a function of the neutral mass, this figure is plotted as a function of the actual *m/z* ratio of the observed ions. As the ionization efficiency likely improves greatly with oxidation, the peak abundance of the $[Sq+Na]^+$ ion (where Sq stands for $C_{30}H_{50}$) is likely suppressed and peak abundances of the $[SqO_n+Na]^+$ (n>0) ions are likely enhanced with respect to their actual relative concentrations in the sample. The inset shows a possible structure corresponding to the $C_{30}H_{50}O_6$ product in which all 6 double bonds of squalene are oxidized, detected as $[SqO_6+Na]^+$ *m/z* 529.350.

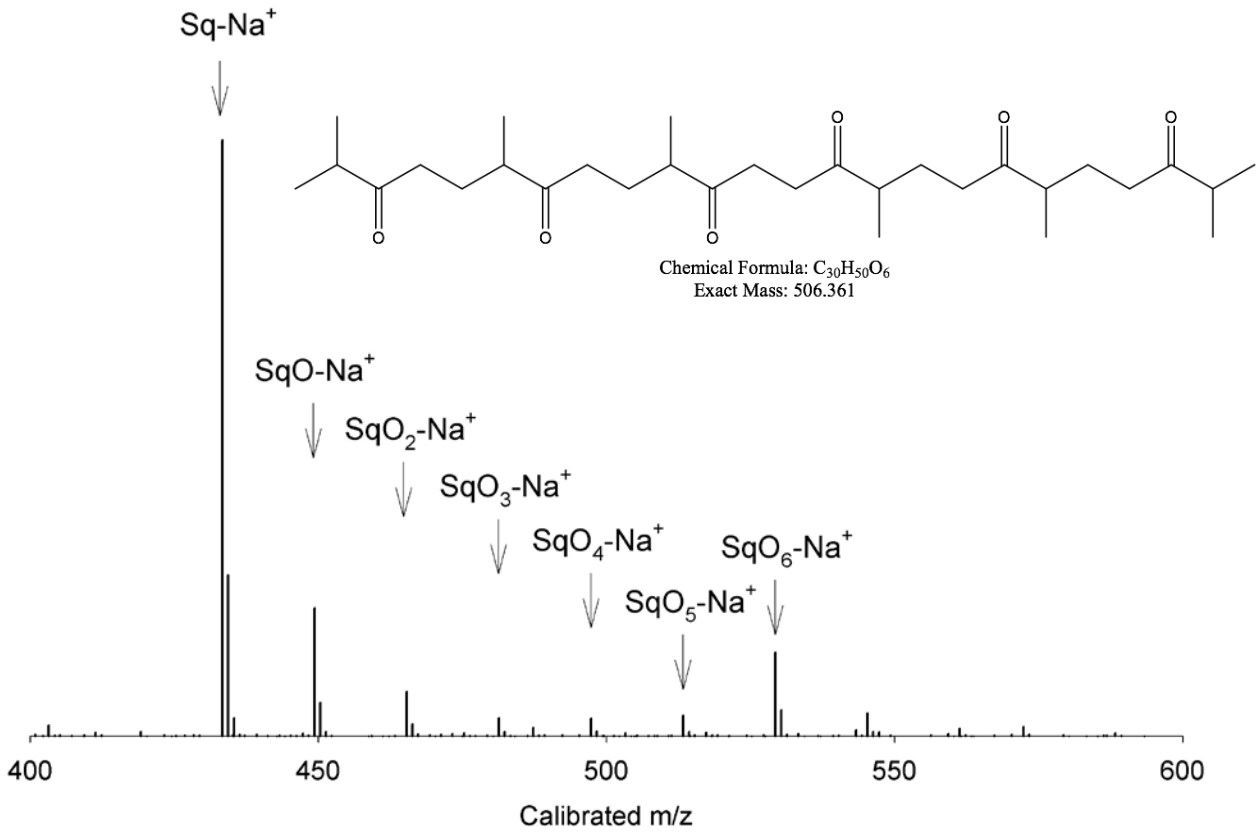
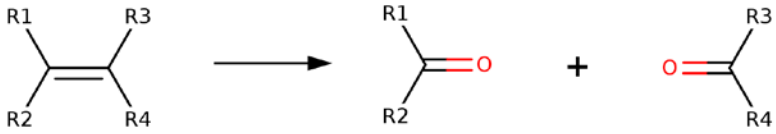
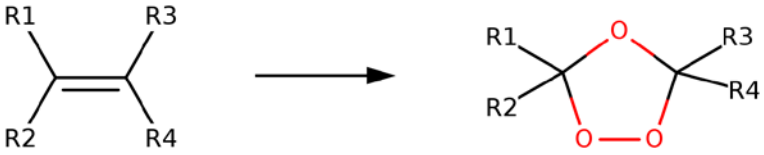
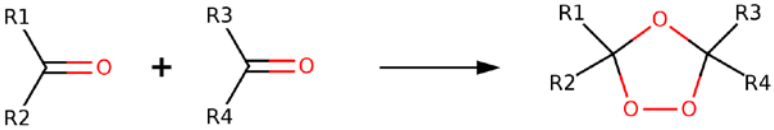
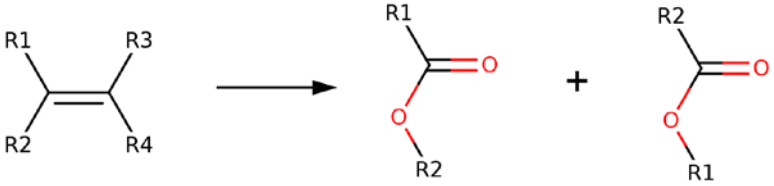
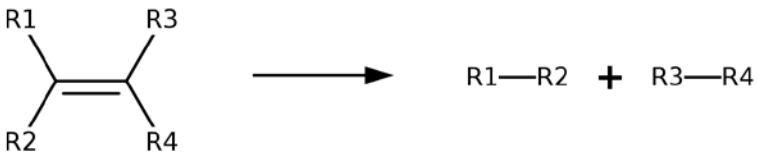
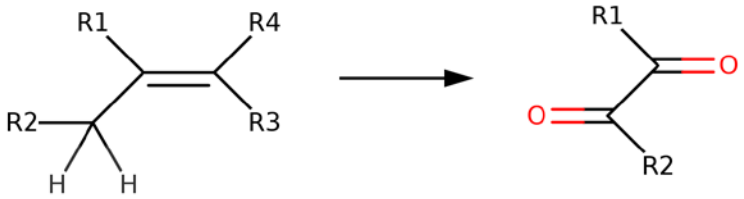
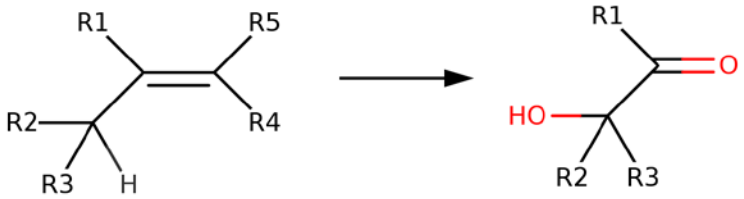
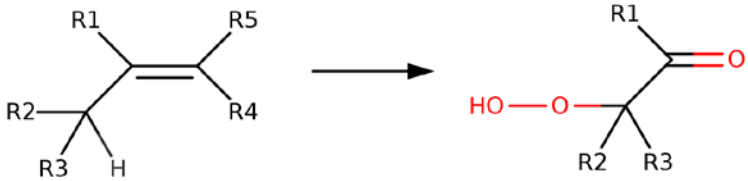
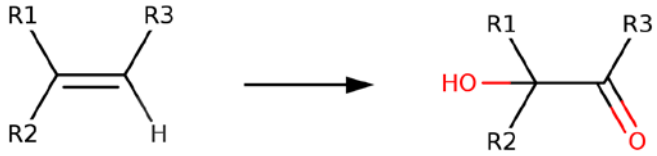
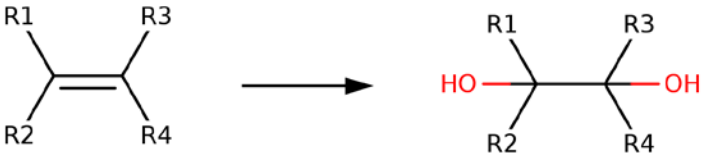
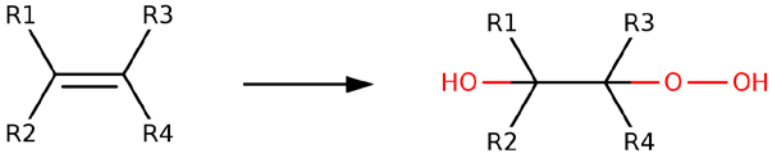


Table S1: Graphical representations of SMIRKS-encoded reaction rules used to define the squalene oxidization simulation. Chemical rules from Table 1 were combined to remove radicals and Criegee intermediates (CIs) from the product pool, while still producing all products resulting from these intermediates. Note that when rules are combined into one (e.g., "Rules 1, 6, and 7"), the combined rule does not generate the total set of products generated by all of the individual building-block rules. Instead those products are generated by prior rules in the set. Exclusion percent is a measure of the reaction's contribution to the pool of correctly predicted products.

Note	Exclusion percent	Rule
Rule 1 produces a carbonyl and a Criegee intermediate on either side of the double bond. Here we produce the carbonyl for each case, while the CIs are consumed in later rules.	3.2	
Rules 1 and 2, for Rule 1 products	1.0	
Rules 1 and 2, for Rule 1 products from different reactions	12.2	
Rules 1 and 3	0	
Rules 1 and 4	2.5	

Rules 1, 6, and 7	16.3	
Rules 1, 6, and 8	0	
Rules 1, 6, and 9	15.0	
Rules 6 and 7	0	
Rules 6 and 8	0.5	
Rules 6 and 9	5.8	

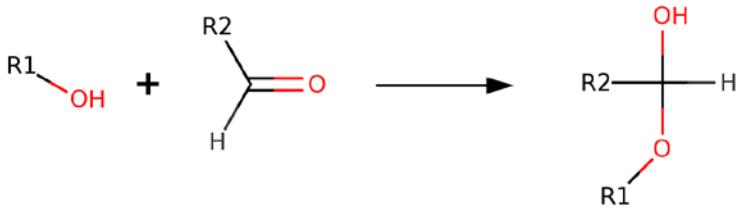
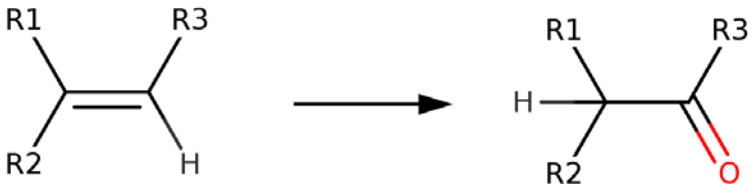
Rules 10 and 11 combined (by allowing R1 to start with an O atom)	5.8	
Rule 12	1.2	

Table S2: Prominent mass differences observed in the oxidized squalene, which correspond to $C_cH_hO_o$ compounds. Frequency refers to the number of times a particular mass difference is encountered in the list of the observed compounds. Note that differences can result from an addition of one molecular fragment and simultaneous subtraction of addition of another molecular fragment as shown in the assignment column.

Difference (Da)	frequency	c	h	o	"Assignment"
116.04734	586	5	8	3	5-hydroxy-4-oxopentanal (5-OH-4-OPA)
74.036781	582	3	6	2	Hydroxy acetone (HA)
68.062599	515	5	8	0	Isoprene (ISO)
190.08412	452	8	14	5	[5-OH-4-OPA] + [HA]
47.984745	442	0	0	3	Ozone
42.01056	433	2	2	1	[5-OH-4-OPA] – [HA]
142.09938	418	8	14	2	[5-OH-4-OPA + HA] – [Ozone] or [ISO] + [HA]
15.994915	398	0	0	1	Oxygen atom

Table S3: List of major primary and secondary products generated by ozonolysis of squalene and expected to be in condensed phase (3). All peaks were predicted by the COBRA simulation, while half of them were detected by HR-MS. Peaks not detected by HR-MS were likely oxidized by the stronger oxidation conditions used here compared with the milder conditions used in Ref. 3.

Product	**Formula**	**Detected by HR-MS**	**Predict-ed by COBRA**
geranyl acetone	$C_{13}H_{22}O$		X
1-hydroxy-6,10-dimethylundeca-5,9-dien-2-one	$C_{13}H_{22}O_2$	X	X
5,9,13-trimethyltetradeca-4,8,12-trienal	$C_{17}H_{28}O$		X
5,9,13-trimethyltetradeca-4,8,12-trienoic acid	$C_{17}H_{28}O_2$	X	X
4,9,13,17-tetramethyl-octadeca-4,8,12,16-tetraenal	$C_{22}H_{36}O$		X
4,9,13,17-tetramethyl-octadeca-4,8,12,16-tetraenoic acid	$C_{22}H_{36}O_2$	X	X
4,8,13,17,21-pentamethyl-docosa-4,8,12,16,20-pentaenal	$C_{27}H_{44}O$		X
4,8,13,17,21-pentamethyl-docosa-4,8,12,16,20-pentaenoic acid	$C_{27}H_{44}O_2$	X	X