The SAPRC Atmospheric Chemical Mechanism Generation System (MechGen)

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

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

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

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

*Corresponding author: William P. L. Carter, email: william.carter@ucr.edu

ABSTRACT

MechGen is a software system designed to derive gas-phase reaction mechanisms for reactive organic compounds under atmospherically relevant conditions for use in chemical models and for data analysis and interpretation. It has been used to derive versions of the SAPRC mechanisms used in airshed models, with SAPRC-22 being the most recent. MechGen derives fully explicit mechanisms for many types of organic compounds and their oxidation products when they react in the atmosphere in the presence of oxides of nitrogen and other pollutants, and then uses these explicit mechanisms to derive reduced or lumped mechanisms more suitable for use in airshed models. This paper gives an overview of the system, describes the procedures it uses to generate explicit and reduced mechanisms, and presents several types of applications. The assignments and estimates used to derive individual chemical reactions and assign rate constants are discussed in a separate companion paper. The system is publicly accessible for generating explicit mechanisms for single compounds and viewing associated documentation using a web-based interface. A separate terminal login is available for deriving mechanisms for multiple compounds, multi-generation mechanisms, and portions of lumped mechanisms for airshed models, as well as for system programming and management. MechGen is designed to accommodate updates to the chemical estimates and assignments that it uses to reflect our evolving knowledge of and ability to estimate atmospheric reactions of organic compounds.

1.   INTRODUCTION

1.1.   Background

35

36

37   Many hundreds of volatile organic compounds (VOCs) are emitted into the lower atmosphere,
38   from both anthropogenic and biogenic sources. Once emitted, they can undergo reactions to form
39   oxidized organic products, including gas-phase toxics, criteria pollutants, and secondary organic aerosol
40   (SOA). For example, in the presence of nitrogen oxides (NO_x), these reactions can generate radicals that
41   react to form ozone (O_3) and oxidized nitrogen compounds that affect air quality. The atmospheric
42   reaction mechanisms for most of these compounds are complex, particularly for larger molecules that
43   involve a large number of reactive intermediates and form a large number of oxidized organic products
44   that can continue to react in the atmosphere. In addition, in most cases these mechanisms involve
45   reactions with rate constants that have not been measured and must be estimated. Because of the
46   complexity, for practical reasons it is necessary either to greatly simplify the mechanisms for most VOCs,
47   use extensive lumping or reduction in VOC representations, or use an automated chemical mechanism
48   generation system to generate estimated mechanisms.

49   An overview of automated generation of reaction mechanisms was presented by Green (2019),
50   which included references to systems that have been developed. A notable example is the RMG system of
51   Green and co-workers (RMG, 2025), originally developed to model combustion systems (Gao et al.,
52   2016) but later extended to apply to other areas such as liquid-phase systems and heterogeneous catalysis
53   (Liu et al., 2021). RMG has an extensive set of tools, links to kinetics and other databases, and has
54   extensive documentation (RMG, 2025). Kirchner (2005) discussed an atmospheric chemical mechanism
55   generator program called CHEMATA, but that work focused on its application to condensed mechanisms,
56   and literature searches revealed no subsequent work with that system. The most comprehensive systems
57   specifically designed to predict detailed mechanisms for lower tropospheric air pollution models, are the
58   Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) system
59   (Aumont et al., 2005) and the SAPRC mechanism generation system (MechGen) that is the subject of this
60   paper.

61   MechGen was developed for deriving the portions of the SAPRC-90 through SAPRC-16
62   atmospheric chemical mechanisms (Carter, 1990, 2000, 2010a, b, 2016; Carter and Heo, 2012, 2013;
63   Venecek et al., 2018) that concerned reactions of C_{2+} organic compounds. The first version, used for
64   SAPRC-90 (Carter, 1990), generated mechanisms only for alkanes, using the procedures and estimates
65   documented by Carter and Atkinson (1985). When SAPRC-99 was being developed, the mechanism
66   generation system was re-written and extended to cover a much wider range of acyclic and monocyclic
67   compounds, including monoalkenes, alcohols, ethers, esters, aldehydes, ketones and organic nitrates in
68   addition to alkanes (Carter, 2000). That version was used to determine the net effects of these compounds
69   in the presence of NO_x that was incorporated into SAPRC-99, though reactions in the absence of NO_x
70   were not generated. A number of updates to the system were made when SAPRC-99 was updated to
71   SAPRC-07 (Carter, 2010a, b), including the ability to generate mechanisms for a wider variety of
72   compounds; however, it remained limited to generation of mechanisms in the presence of NO_x, and
73   aromatics were still not supported. The system was further updated for use in the development of the
74   SAPRC-16 (Carter, 2016; Venecek et al., 2018) mechanism and unpublished updated versions,
75   incorporating capabilities to generate reactions in the absence of NO_x, reactions of aromatics,
76   autoxidation reactions of peroxy radicals, and other enhancements. Further updates were made for the
77   current version, which has been used in the latest of the SAPRC mechanisms, SAPRC-22 (Carter, 2023).

78   The chemical basis of MechGen and the methods it uses to assign or estimate rate constants or
79 mechanisms are documented by Carter et al. (2025a). The current, and first publicly documented, version
80 of MechGen, v1.1, is available online, along with a comprehensive User Manual and a quick start guide.
81 These are available at the MechGen website (Carter, 2025) and GitHub page
82 (https://github.com/SAPRC/MechGen), and archived versions can also be downloaded from these
83 websites. This paper describes the software system and the procedures it uses for mechanism generation
84 and processing.

1.2.   Chemical Systems Represented

86   MechGen is capable of generating fully explicit mechanisms for the atmospheric reactions of
87 most types of organic compounds emitted into or formed in the lower atmosphere, and of the intermediate
88 radicals they form. While temperature-dependent rate constants are provided for many reactions, for
89 others the rate constants or branching ratios are only applicable for conditions representing the
90 troposphere, i.e., around 298 K and 1 atm. MechGen is not currently designed for estimating mechanisms
91 for combustion modeling or for the low temperature and pressures characteristic of the upper troposphere.

92   Table 1 lists the types of stable compounds whose reactions can be generated, and shows the
93 types of initial atmospheric reactions that can be generated for them. Table 2 lists the types of reactions
94 that are generated, including reactions of intermediate radicals as well as reactions of stable compounds.
95 The reactions of stable compounds include H-atom abstractions by OH, NO_3, and Cl radicals, additions to
96 double bonds by these radicals and by O_3 and O^3P, and reactions by photolysis. The types of radicals
97 generated include carbon-centered radicals that in most cases react primarily with O_2; peroxy radicals that
98 in most cases react with NO, NO_2, NO_3, HO_2, or other peroxy radicals and in many cases also have
99 unimolecular reactions; alkoxy radicals that can react with O_2 or by various types of unimolecular
100 reactions; and excited and stabilized Criegee intermediates formed in reactions of alkenes with O_3.
101 Measured rate constants or branching ratios derived from measured product yields are used when data are
102 available. In the absence of such data, which is typically the case, rate constants or branching ratios are
103 estimated using various structure-activity relationships (SARs) or other estimation methods. The details
104 of the types of reactions and how rate constants or branching ratios are assigned or estimated are given by
105 Carter et al. (2025a).

1.3.   System Software and Access

107   The current version of MechGen is incorporated into an online Multi-User Object Oriented
108 (MOO) system that was originally developed as a programmable text-based virtual reality system (Curtis,
109 1997; Fox, 2004; Wikipedia, 2025). Features of the MOO object-oriented programming language, which
110 is similar to Python, made it well suited for mechanism generation applications compared to the original
111 Fortran version used for SAPRC-90 (Carter and Atkinson, 1985; Carter, 1990). The MOO system also
112 has online access capabilities that make it relatively straightforward to permit multiple users to access it
113 online simultaneously.

114   Since MechGen is written in an object-oriented language, the system employs various software
115 "objects" that include properties defining their characteristics and programs or subroutines controlling
116 their operations. The MechGen software objects referenced in this paper include <u>reactant</u> objects that
117 represent molecules or radicals that can be reacted; <u>group</u> objects that represent portions of the molecule
118 that determine how they react; <u>reactor</u> objects that are assigned to each user login that provides a user
119 interface and gives mechanism generation options; <u>environment</u> objects that contain concentrations of
120 atmospheric species that can be used to determine product yields from mechanisms; <u>multi-generation</u>
121 <u>mechanism</u> objects that control multi-generation mechanism derivations for a given reactant; and <u>lumping</u>

objects that control the full mechanism generation process and whether and how lumped mechanisms are also produced. The MechGen system employs other types of objects, but a discussion of these details of the software system is beyond the scope of this paper.

MechGen is accessible online via a web interface available at the MechGen website (Carter, 2025) or through a terminal interface such as Telnet (Postel and Reynolds, 1983), with instructions available at the MechGen and MechGen GitHub sites. Terminal access provides the highest degree of system capability, while web access is more user-friendly for basic operations and is better suited for providing information about the system and the reactions being generated. Due to the high computational demand, resource-intensive operations such as deriving multi-generation mechanisms, or generating mechanisms for large or for large numbers of molecules, are not currently available using the online system. To access the full capabilities of MechGen, users need to install the system locally by following the instructions provided in the User Manual.

1.4. Overview of Documentation

The methods used by the current version of MechGen to derive or estimate rate constants and mechanisms for gas-phase atmospheric reactions of organic are described by Carter et al. (2025a). While that paper describes the chemical basis for the mechanisms MechGen generates, it does not explain the system itself or its functionality. This paper fills that gap by discussing the following specific capabilities of MechGen:

- Deriving unimolecular or bimolecular reactions of individual compounds and radicals under atmospheric conditions, using assigned or estimated rate constants and products formed (Section 2);

- Deriving explicit single-generation mechanisms of a selected compound by reacting the compound and all the rapidly reacting intermediates formed, but not reacting the stable, non-radical products (Section 3.1);

- Deriving minimally reduced mechanisms that give the same predictions as the explicit mechanism at a selected temperature, but with about 1/3 the numbers of reactions and 1/2 the numbers of species (Section 3.2);

- Estimating yields of products formed when reacting a compound under selected environmental conditions (Section 3.3);

- Deriving explicit multi-generation mechanisms of a selected compound by fully reacting the compound and all stable reactive products formed in non-negligible yields (Section 4); and

- Optionally driving lumped mechanisms that use a limited number of lumped model species to represent chemically similar compounds, and by using various approaches to reduce the numbers of reactive intermediates in the mechanisms (Section 5).

Additional details about the methods and algorithms are also given in the Supplement. Details concerning how to work with MechGen and a description of features not covered here are given in the User Manual that is available at the MechGen website (Carter, 2025). The MechGen GitHub site (https://github.com/SAPRC/MechGen) also contains the User Manual and downloadable software and files used by MechGen (see Section 7.3). The examples shown here and the Supplement are based on the current chemistry assignments and methods discussed by Carter et al. (2025a). The system is expected to operate similarly when the chemistry assignments and estimates are updated in the future.

163 Table 1.　　　　Types of stable compounds whose reactions are supported by MechGen.

Type of Compound	Reactions [a]
Alkanes	OH, NO$_3$
Alkenes (including those with multiple double bonds)	OH, O$_3$, NO$_3$, O^{3}P
Alkynes	OH, NO$_3$
Aromatic hydrocarbons	OH
Aldehydes, ketones, hydroperoxides, organic nitrates	OH, NO$_3$, hν
Alcohols, ethers, esters, other oxygenates	OH, NO$_3$
Unsaturated aldehydes, ketones, hydroperoxides, and nitrates	OH, O$_3$, NO$_3$, O^{3}P, hν
Phenols	OH, NO$_3$
Furans and other unsaturated oxygenates	OH, O$_3$, NO$_3$, O^{3}P
Amines	OH, NO$_3$
Peroxynitrates	Unimolecular, OH, NO$_3$, hν
Bi- and polyfunctional compounds	OH, O$_3$, NO$_3$, O^{3}P, hν, as applicable

164　[a] See Carter et al. (2025a) for more detail and types of compounds not supported.
165
166
167
168 Table 2.　　　　Summary of types of reactions supported by MechGen.

Reactant(s) [a]	Type of reactions [b]
VOC + OH	H-atom abstraction. Addition to double, aromatic, and triple bonds
VOC + O$_3$	Addition to double bonds and to N atoms in amines
VOC + NO$_3$	H-atom abstraction. Addition to double bonds
VOC + O^{3}P	Addition to double bonds
VOC + hν	Usually breaking the weakest bond, but depends on the compound
VOC Uni.	Unimolecular decompositions of peroxynitrates
Carbon-centered Radicals	O$_2$ addition. H-abstraction from α-OH groups, unimolecular reactions of excited OH adducts, unimolecular decompositions
Peroxy or Acyl peroxy Radicals	Reactions with NO forming the corresponding alkoxy radical and organic nitrate; reactions with NO$_2$, NO$_3$, HO$_2$ or other peroxy radicals; and unimolecular reactions
Alkoxy Radicals	α-H abstraction by O$_2$ forming the corresponding carbonyl compound. β-scission and H-elimination decompositions. H-shift and ester rearrangement isomerizations
Criegee intermediates	Decompositions, stabilization, or rearrangements.

169　[a] "VOC" indicates any stable volatile organic compound supported by the system.

170　[b] See Carter et al. (2025a) for more detail concerning types of reactions.

171

2.   REACTANT SPECIFICATION AND SINGLE-STEP REACTIONS

2.1.   Creation and Specification of Reactants

Reactants are created as objects within MechGen by specifying their structures. The structure of an organic reactant or radical is specified by giving the "groups" in the molecule or radical, and indicating the groups each are bonded to, and the type of bond. As listed in Table 3, groups are parts of molecules that are treated as units in the system, and contain no more than one carbon, nitrogen, or halogen atom, but can also contain one or several hydrogen or oxygen atoms. The leading or trailing "-", "=", "#", or "-a" indicate groups that can bond to neighboring groups by single, double, triple bonds, or aromatic/allylic bonds. Any groups with such designations can bond to other groups with such designations to form chains. Mechanisms are generated and rate constants are estimated based on the groups present, the groups they are bonded to, and in some cases on groups elsewhere in the molecule. Groups can be radical or non-radical, with reactants containing a radical group referred to as radicals or intermediates. Examples of reactant designations are given in Table 4, and additional discussion on reactant designation and creation by specifying structures is available in the User Manual.

MechGen also includes objects representing elementary species that can react with reactants (e.g., OH or O_3), react with radicals (e.g., O_2 or NO), or be formed in reactions (e.g., H_2O or CH_4). These are listed in Table 5.

2.2.   Single Step Reactions

Single step mode consists of generating only the initial reaction(s) of a specified compound or intermediate, without reacting the resulting products or intermediates. When reacting stable compounds, it is necessary to specify the type of reaction (e.g., whether unimolecular or a bimolecular reaction with a specified oxidant), whereas all possible reactions are automatically generated for radicals. Generating reactions in single step mode is useful to obtain information on how the system estimates a compound's reactions, since the results show not only the products formed and the rate constants or branching ratios derived, but also documentation on the estimates or assignments used during reaction generation.

The algorithm for generating single step reactions is shown in Figure 1, and examples of output are shown in Figure S1 in the Supplement. Note that Figure S1 shows not only the generated reactions but also text documenting how the reactions were derived. Reaction types and derivation methods are discussed by Carter et al. (2025a), and the specific procedures shown in Figure 1 depend on whether the reactant is a stable compound or an intermediate, and also on the type of intermediate. Assigned reactions and rate constants are used for reactants for which assignments are made, but for most reactants, reactions and rate constants have to be estimated using methods discussed by Carter et al. (2025a). In some cases, assignments are made for rate constants but not for mechanisms or branching ratios; in those cases, the reactions are generated using the general estimation methods, but the rate constants output are the assigned rather than the estimated values.

Some radicals are estimated to undergo unimolecular or O_2 reactions so rapidly that it is not necessary to estimate their total rate constants to predict their atmospheric fates. For example, most alkyl radicals are consumed only by reaction with O_2, and many "explicit" mechanisms simply replace them with the peroxy radicals they form, resulting in these alkyl radicals not appearing in the generated mechanism. MechGen does include these radicals and their reactions, though it flags them as being "fast" and only outputs their branching ratios, when applicable.

213   For photolysis reactions, the system outputs a "photolysis set" name that gives absorption cross
214   sections and optionally quantum yields as a function of wavelength, and also a wavelength-independent
215   quantum yield if the quantum yields are not given in the photolysis set. The photolysis sets referenced by
216   MechGen are those used for the SAPRC-22 mechanism as listed and documented by Carter et al. (2025a),
217   and these files can be downloaded with other files needed to implement that mechanism (Carter 2024).
218   Single step reaction output also includes the calculated photolysis frequency for a representative solar
219   light intensity and spectrum, but this should not be considered part of the mechanism because it depends
220   on the light environment.

221   In the case of peroxy radicals, the system first derives unimolecular reactions if they are estimated
222   to be non-negligible for this radical, then derives the bimolecular reactions with NO, NO_2, NO_3, HO_2,
223   generic alkyl peroxy radicals (RO_2), and generic acyl (RCO_3) peroxy radicals in that order. Deriving all
224   the possible organic peroxy + peroxy reactions is impractical due to the large number of peroxy radicals
225   formed in realistic photooxidation systems and the fact that the other peroxy radicals present are unknown
226   at the time of mechanism generation. Instead, MechGen treats peroxy radicals as reacting with the totals
227   of all alkyl and all acyl peroxy radicals. The products from each reaction type include those formed from
228   the subject radical, plus counter species used to represent the type of co-products formed from the sum of
229   peroxy radicals it reacts with. Those counter species can be either generic alkoxy radicals, designated
230   "RO." or "RCO2" [representing $RC(O)OO\cdot$]; or carbonyl or alcohol H-transfer products, designated "RO-
231   alpha-H" and "ROH" or "RCO-OH" (Carter et al., 2025a). They are included as products of these
232   reactions during reaction generation for tracking if desired, but are deleted as products when the
233   mechanisms are reduced or lumped.

234   The user-modifiable parameters that affect the single step reaction generation process are shown
235   in Table 6. The temperature affects the thermal rate constants. The pressure affects rate constants assigned
236   to falloff reactions and also estimated nitrate yields in the reactions of peroxy radicals with NO, and in
237   some cases, fractions of excited radicals that are stabilized. The presence of water affects estimated
238   reactions of low reactivity stabilized Criegee intermediates, as indicated in Figure 1 (Carter et al., 2025a).
239   When users select single step operations, all estimated reactions are displayed, allowing users to see
240   which reactions are considered and how their rate constants are estimated. However, negligible reactions
241   are removed when the outputs of single step reaction operations are processed when called during full
242   mechanism generation, as discussed in Section 3. In addition, reactions of non-acyl peroxy radicals with
243   NO_2 forming alkyl peroxynitrates (shown in Figure S1c) are not included during full mechanism
244   generation by default because these peroxynitrates are assumed to rapidly decompose at temperatures of
245   interest (Carter et al., 2025a). These reactions can be included by either increasing the kFastUni parameter
246   (see Table 8), or decreasing the reactor temperature such that the estimated rate constant is less than
247   kFastUni."

248   **2.3.   User Modifiable Mechanism Assignments**

249   As indicated above, MechGen uses assigned rate constants and branching ratios when deriving
250   single-step mechanisms whenever possible, and otherwise uses estimates and SARs when no assignments
251   are available. While updates will be made as new data become available, the assignments will not always
252   be up to date. In addition, some users may wish to employ MechGen to derive mechanisms using recently
253   measured or newly derived rate constants and branching ratios, or to explore how varying uncertain
254   estimates impact mechanisms. To address this, MechGen allows users to add custom (or "user")
255   mechanism assignments, which can supplement or replace the default settings. The steps involved in
256   making or managing user assignments are described in the User Manual.

257

| 258 | Table 3. | List of groups and group designations used to specify C_{2+} organic reactants whose |
| 259 | | reactions can be generated using MechGen. |

Type	Group designations [a]			
Non-Radical Groups				
Alkane	-CH3	-CH2-	-CH()-	-C()()-
Alkene	=CH2	=CH-	=C()-	=C=
Alkyne	#CH	#C-		
Aromatic or Allylic [b]	-aCH-	-aC()-		
Oxygenate	-CHO	-CO-	-OH	-O-
Nitrate, nitro, nitroso	-ONO2	-NO2	-NO	
Amino and amine-oxy	-NH2	-NH-	-N()-	-N[O]()-
Imine	=NH	=N-		
Criegee intermediates	CH2OO-			
	-C[OO]-syn-	-CHOO[syn]-	-CHOO[anti]	-syn-C[OO]--
Halogen	-F	-Cl	-Br	-I
Radical or Reactive Intermediate Groups				
Carbon centered radicals	-CH2.	-CH[.]-	-C[.]()-	=CH.
	=C[.]-	-pC[.]-	-CO.	
Allylic radicals [b]	-aCH2.	-aCH[.]-	-aC[.]()-	
Peroxy radicals	-CH2OO.	-CH[OO.]-	-C[OO.]()-	
	-CO[OO.]	=CHOO.	=C[OO.]-	-pC[OO.]-
Alkoxy radicals	-CH2O.	-CH[O.]-	-C[O.]()-	=CH[O.]
	=C[O.]-	-CO2.	-pC[O.]-	
Criegee intermediates [c]	-CHOO {excited}	-C[OO]- {excited}		
Radicals from amines	-NH.	-N[.]-	-NH2[O]	-NH[O]-
Carbenes	-CH[..]	-C[..]-		

260 [a] Bond types are indicated by "-" (single or aromatic), "=" (double), or "#" (triple). The number of
261   bonds shown indicate if it is bonded to one or two or more groups. The "()" notation is not part of the
262   group designation itself, but is used to indicate that the group is bonded to a 3rd or 4th group. An "a"
263   or "p" prefix in the group name indicates that it has alternating single and double bonds. The "p"
264   prefix is used for phenyl, phenoxy or phenyl peroxy radicals. Some groups can have different bond
265   designations depending on their location in the molecule, e.g., "CH3-" or "-C[.]=" rather than "-CH3"
266   and "=C[.]-". See examples in Table 4.

267 [b] Allylic groups are used in reactants with two or more resonance structures involving adjacent double
268   bonds and radical centers. Allylic radical groups are used for the portions where that the radical center
269   is located in at least some of the resonance structures. Aromatic groups are used for the portions
270   where the radical center is not located in any resonance structures.

271 [c] The "{excited}" designation is given at the end of the structure designation and indicates how the
272   excited intermediate was formed. This designation is also used following full reactant structure
273   designations to indicate excited adducts formed when radicals add to double bonds, but in this case
274   the excitation is associated with the entire molecule and not a single group.

275　Table 4.　　　Examples of designations of selected representative compounds or intermediates.

Compound	MechGen Structures [a]
Stable Compounds	
propene	CH2=CH-CH3; CH3-CH=CH2
methyl acetylene	CH#C-CH3; CH3-C#CH
3,3-dimehtyl pentane	CH3-CH2-C(CH3)(CH3)-CH2-CH3; CH3-C(CH3)(CH2-CH3)-CH2-CH3
trans-3-methyl-2-pentene	CH3-^CH=C(vCH3)-CH2-CH3; CH3-CH2-C(^CH3)=CH-^CH3 [b]
cyclopropane	CH2*-CH2-CH2*; *CH2-CH2-*CH2 [c]
bicyclo [1.1.1] heptane	CH2*1-CH*2-CH2-CH*1-CH2*2; CH*12-CH2-CH(CH2*1)-CH2*2 [c]
toluene	CH3-aC*-aCH-aCH-aCH-aCH-aCH* [d]; CH3-C*=CH-CH=CH-CH=CH*
naphthalene	aC*12-aCH-aCH-aCH-aCH-aC*1-aCH-aCH-aCH-aCH*2
2-propyl nitrate	CH3-CH(CH3)-ONO2; CH3-CH(ONO2)-CH3
carbitol acetate	CH3-CH2-O-CH2-CH2-O-CH2-CH2-O-CO-CH3
Intermediates	
2-propyl radical	CH3-CH[.]-CH3
2-propyl peroxy radical	CH3-CH[OO.]-CH3
methyl allyl radical	CH3-aCH[.]-aCH-aCH2 [d]; CH3-CH[.]-CH=CH2; CH3-CH=CH-CH2.
OH + benzene adduct	HO-CH*-aCH[.]-aCH-aCH[.]-aCH-aCH[.]*
Excited Criegee int. from O3 + alkene	CH3-CH2-C[OO]-CH3 {*O3Ole}; CH3-C[OO]-CH2-CH3 {*O3Ole}
Stabilized Criegee int. with *syn* -CH3	CH3-syn-C[OO]-CH2-CH3; CH3-CH2-C[OO]-syn-CH3

276　[a] The first MechGen structure codes given are those generated by the system. If subsequent structures
277　　　are given, they are alternatives that can be used to create the same compound.

278　[b] The symbols "v" and "^" are used to indicate "cis" or "trans" configurations, analogous to "\" and "/"
279　　　in Smiles notation.

280　[c] The symbol "*" is used to indicate ring closure. "*1", "*2", etc used for multiple ring structures,
281　　　analogous to "1", "2", etc. in Smiles notation.

282　[d] The symbol "aC" used to indicate carbon centers where with single and double bonds are in
283　　　resonance, as in aromatic species or allylic radicals.

284 Table 5. List of elementary species used by MechGen.

Type	Group designation				
Non-reacting products [a]	H2 HCHO	H2O NH3	HONO HCl	HNO3	CH4
Species that react with VOCs	OH Cl.	O3 Uni [b]	NO3	O3P	hν [b]
Species that react with radicals	O2	NO	NO2	RO2. [c]	RCO3. [c]
Counter species [c]	RO.	ROH	RCO2.	RCO-OH	RO-alpha-H [d]

285 [a] These species are treated as unreactive by MechGen.

286 [b] Pseudo-species used to indicate photolysis or unimolecular reactions of stable compounds.

287 [c] Generic species used as reactants (RO2., RCO3.) or products (RO., etc.) in peroxy + peroxy reactions.

288 [d] "RO-alpha-H" refers to the carbonyl formed when an α-hydrogen on an alkoxy radical is abstracted.
289
290
291
292 Table 6. List of parameters that affect single-step mechanism generation.

Parameter (code)	Affects	Default	Note
Temperature (T)	Thermal rate constants and O_2 concentration	298K	
Pressure (P)	Pressure-dependent rate constants and O_2 level	1 atm	
O_2 fraction (O2)	Percentage of O_2 in reacting environment	20.95%	[a]
Light Source (light)	Actinic flux values used to calculate rate constants for photolysis reactions in output displays	As used by Carter (1994)	[b]
Water present (H2O)	Parameter indicating whether water is present.	H_2O is present	[c]

293 [a] Affects output displays and estimates of product yields, but not reactions in explicit generated
294 mechanisms. However, this affects the full mechanism generation process discussed in Section 3.

295 [b] The actinic fluxes are used to calculate rates of photolysis reactions given the absorption cross
296 sections and quantum yields assigned to the photolysis reactions. The photolysis frequencies are not
297 part of the mechanism but are included as outputs of single step mechanism generation and are also
298 used to derive overall product yields for various environments, as discussed in Section 3.3 and to
299 derive multi-generation mechanisms as discussed in Section 4.

300 [c] Currently applicable only to low reactivity Criegee intermediates (Carter et al., 2025a).
301
302
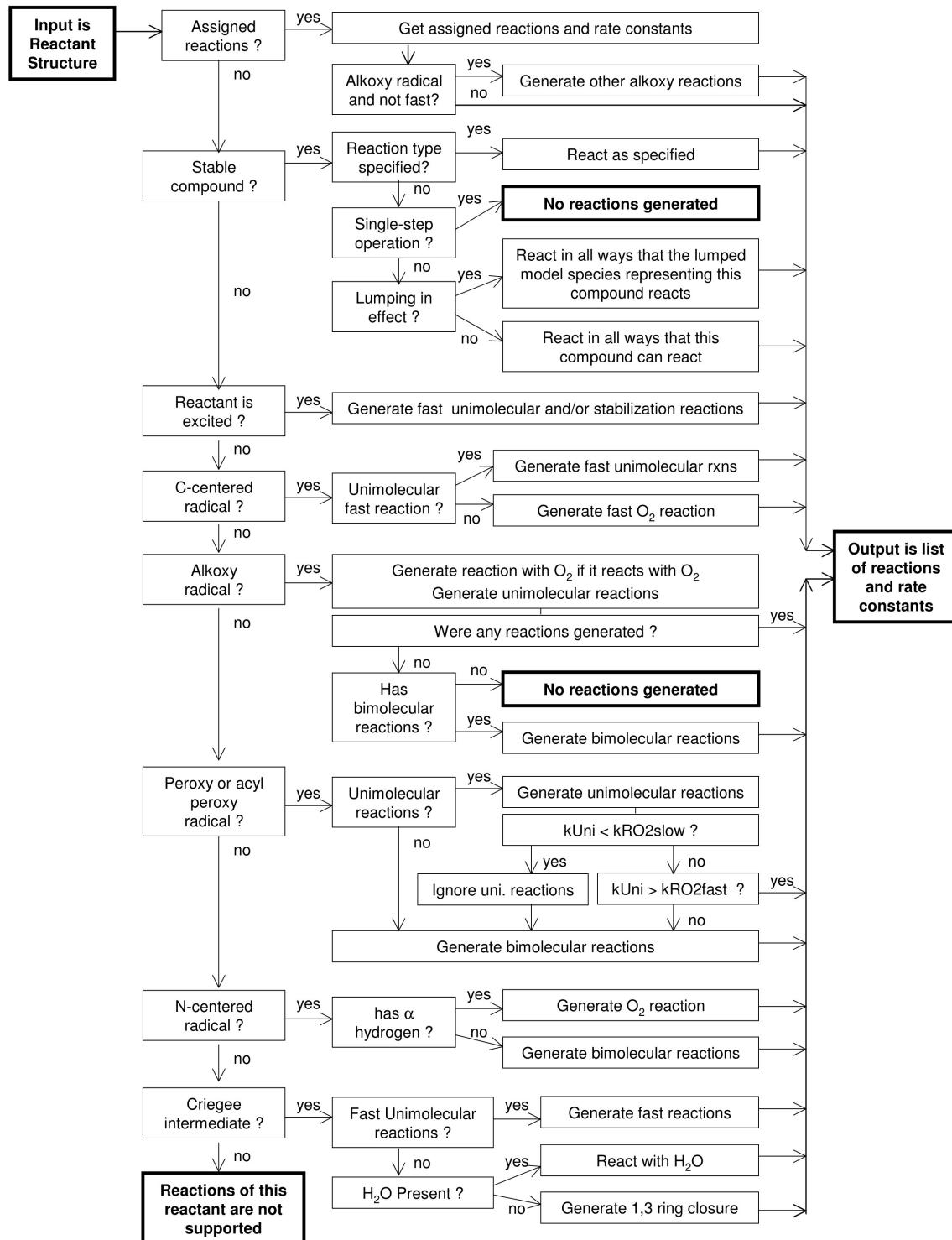
303

304   Figure 1.   Diagram of algorithm used to generate single-step reactions.

305

3. FULL MECHANISM DERIVATION

3.1. Description of Process

Full mechanism derivation for a compound involves first reacting it through all, or a designated subset, of its possible initial atmospheric reactions; and then reacting any reactive intermediates that are formed. Stable products formed are not reacted, so the mechanism output reflects only a single generation of reactions, as opposed to multi-generation mechanism derivations discussed in Section 4. Reactive intermediates are defined as products that have radical or intermediate groups (see Table 3) or non-radical compounds that undergo unimolecular reactions with rate constants greater than a specified value, which is 0.0167 s^{-1} (1 min^{-1}) by default. The result is a sequence of reactions that were generated and their rate parameters, lists of intermediates involved and products formed, and information about the products and intermediates that may be useful.

In order to provide the capability to generate explicit mechanisms of appropriate sizes for specific modeling applications, and to estimate relative yields of products formed under various conditions, MechGen provides standard environments that specify concentrations of atmospheric species that can be used for these purposes. The currently available standard environments are summarized in Table 7, with additional information given in Section S2 of the Supplement.

The currently available standard environments listed in Table 7 all represent urban conditions, but users can define new environments as described in the User Manual to include a wider range of conditions. The default use of these environments is indicated in Table 7, with three environments representing the range of urban conditions with regard to NO_x availability, and an additional environment representing nighttime conditions. By default, the "Standard" or mid-NO_x environment is not used for mechanism generation, as reactions important under those conditions would also be important under either high or low NO_x conditions, if not both. The mid NO_x environment is useful, however, for determination of relative product yields under urban conditions that are equally sensitive to both VOC and NO_x. It is also the default environment used when deriving multi-generation mechanisms (see Section 4).

The user-modifiable options controlling full mechanism derivation are summarized in Table 8, and the algorithm employed is shown in Scheme S1 of the Supplement, which outlines the overall process, and Scheme S2, which details the portion that determines which reactions are deleted. A key option affecting the sizes of generated mechanisms is the "MinYld" parameter, which determines which competing reactions in the single-step reactions can be neglected. A related option concerns whether and which environments are used during mechanism generation to estimate upper limit yields for competing bimolecular reactions. The "MinYld" test is implemented by assigning an "upper limit yield" for the initial reactant as unity, and estimating upper limit yields of intermediates whose reactions are generated as the product of the assigned or estimated upper limit yield of the reactant forming it, multiplied by the upper limit relative yield of the intermediate in the reactions of the reactant forming it. Upper limit yields are also derived for stable products formed in the generated mechanisms, for the purpose of estimating their approximate importance in the mechanism once it is generated. Reactions forming products with estimated upper yields less than MinYld are neglected, unless there is only one reaction of the reactant. When reactions are neglected, the rate constants and yields of the competing reactions are adjusted to ensure no mass or moles are lost (see Scheme S2). Upper limit yields are also derived for non-reacting products, but this is used for information purposes only.

348   The relative yields of products from reactions of the initial reactant or reacting intermediate can
349   depend on the environment. MechGen assumes constant temperature, pressure, and O_2 levels when doing
350   full mechanism generations. As a result, relative yields of products from reactants that undergo only
351   unimolecular reaction or with O_2 are treated as environment independent. However, environments
352   significantly affect yields of products from reactants that undergo other bimolecular reactions. In such
353   cases, either a single environment must be specified to calculate yields for the MinYld test, or upper limit
354   relative yields must be estimated. If no environment is specified, upper limit yields are derived by
355   assuming that each type of bimolecular reaction is equally important. If more than one environment is
356   used for mechanism generation, relative yields are determined for each environment, and the highest yield
357   among these environments is then used to derive upper limit yields for the MinYld test. Note that the total
358   of upper limit yields for competing reactions can exceed 100% unless only one environment is used.

359   As shown in Scheme S2, the MinYld test is not used for reactions of peroxy radicals with NO and
360   for photolysis reactions of initial reactants. The only reactions with NO that may be generated are
361   reactions of peroxy radicals forming alkoxy radicals or organic nitrates. Since the formation of alkoxy
362   radicals is always relatively more important, neglecting the nitrate-forming reaction using MinYld test
363   will result in underestimation of radical and NO_x sink processes without significantly reducing
364   mechanism size. Similarly, the MinYld test is not used for photolysis reactions because MechGen does
365   not generate many competing such reactions and deleting any of them may affect predictions of radical
366   sources without significantly reducing the size of the mechanism. Generally, it is unimolecular reactions
367   that tend to have many competing processes where some may be negligible according to the MinYld test.

368   For a given MinYld value, assuming no environments gives the largest mechanisms that are
369   applicable to a full range of conditions, while assuming a single environment gives the smallest
370   mechanisms that are suitable for that environment, but they may not be as accurate when applied to other
371   conditions. Using an appropriate set of environments produces mechanisms of intermediate size that may
372   be optimum if the environments used represent the range of conditions where the mechanism may be
373   used. As indicated in Table 8, the default is to use three environments that represent the range of urban
374   conditions. Use of additional environments may be appropriate if the mechanisms are to be used for more
375   remote scenarios.

376   An example of a mechanism derived using a full mechanism generation process is shown in
377   Figure S2 of the Supplement, with a representative subset displayed in Figure 2. These figures show
378   reactions generated for 1,3-butadiene under default conditions. Note that designations such as
379   "{*OHadd}" or "{*O3ole}" indicate excited intermediates. This output includes the rate constants at the
380   default temperature of 298 K ("k"), the branching ratio for various competing reactions ("Fac"), the
381   estimated upper limit weighting factor to determine which reactions can be neglected ("Weight"), and the
382   reactants and products involved of the reactions. If a rate constant is not listed, the reaction is assumed to
383   be fast and the reactant to be in steady state, making the fate of the reactant independent of the rate
384   constant. In such cases, the branching ratios for competing fast routes are given in the "Fac" column.

385   The counter species such as "RCO2." and "RCO-OH", shown in Figures S1 and S2, are used for
386   reactions of peroxy radicals with other peroxy radicals, as indicated in Table 5. The output shown in
387   Figure S2 does not include the assigned or estimated Arrhenius parameters for calculating rate constants
388   at different temperatures for some reactions, but this information can be obtained by selecting other
389   output formats for the reactions as discussed in the User Manual.

390   The numbers of reacting intermediates, stable products, and reactions for mechanisms derived for
391   representative compounds using various mechanism generation options are discussed in Section 3.4 and
392   ratios of these are shown in Figure S7 in the Supplement. In most cases, the ratios of the numbers of

393 reacting intermediates in the explicit mechanisms are about 30-50% the numbers of reactions, tending to
394 be smaller when no environments or lower MinYld values are used, with no clear dependence on
395 mechanism sizes. The ratios appear to be higher for the representative aromatics than for other types of
396 compounds, at least for these examples. The ratios of numbers of stable products to numbers of reactions
397 tend to be in the 30-40% range and are less variable than the ratios for intermediates.

3.2. Derivation of Minimally Reduced Mechanisms

399   The numbers of reactions and intermediates in generated mechanisms can be significantly
400 reduced by combining parallel reactions of the same reactants and by eliminating intermediates that
401 always rapidly form the same product(s) regardless of the environment. Combining parallel reactions
402 gives product yields derived from the ratios of the rate constants for individual reactions to the total rate
403 constant, and reduces the numbers of reactions, but does not affect the numbers of species in the
404 mechanism. The numbers of predicted stable products cannot be reduced without further lumping, but
405 many reacting intermediates whose fates do not depend on environmental conditions can be removed by
406 assuming steady state so that they can be replaced by the products they form. This results in reducing the
407 total number of intermediates by factors of 3 or higher (see below).

408   The intermediates removed by this process include species that undergo only unimolecular
409 reactions or reactions with O_2, H_2O, or by stabilization, meaning their products do not depend on the
410 environment if O_2, temperature, water content, and pressure are assumed constant. These include all
411 carbon-centered radicals, all excited intermediates, almost all alkoxy and nitrogen-centered radicals, and
412 many Criegee intermediates. These all react fast enough on the time scale involved with atmospheric
413 modeling that they can be assumed not to build up in concentration, so replacing them by their products
414 should not significantly affect model predictions.

415   However, this process does not eliminate peroxy or acyl peroxy radicals that do not undergo fast
416 unimolecular reactions, since they are consumed primarily or at least significantly by bimolecular
417 reactions involving NO_x and other peroxy species, whose relative concentrations depend on the reacting
418 environment. Nevertheless, as discussed below, this results in the removal of most reactive intermediates
419 in the explicit generated mechanisms.

420   Vereecken and Nozière (2020) calculated that most hydroperoxy-substituted peroxy radicals
421 should interconvert to other such radicals with relatively high rate constants, and these are incorporated
422 into the estimated mechanisms produced by MechGen (Carter et al., 2025a). These interconversion
423 reactions are not considered when estimating whether competing reactions may be negligible because
424 they are not net sink processes for such radicals. However, including these rapid interconversion reactions
425 in the processed or lumped mechanisms used in models can potentially cause numerical "stiffness"
426 problems in model simulations and also needs to be taken into account when product yields are estimated,
427 as discussed in Section 3.3.

428   Therefore, as part of this initial processing, rapidly interconverting hydroperoxy-peroxy radicals
429 are identified and represented in the processed mechanism by a lumped radical species representing the
430 sets of interconverting radicals that are assumed to be in equilibrium. Reactions forming these radicals are
431 represented as forming this lumped radical species, and reactions of these radicals are replaced by
432 reactions of this lumped species, forming products of the interconverting radicals with yields multiplied
433 by the equilibrium fraction of the radical derived from the rate constants involved. This process is
434 discussed in Section S3.1 of the Supplement. This removes these rapid interconversion reactions from the
435 mechanism, while retaining their effects on overall product yields.

436   Mechanisms processed this way are referred as "minimally reduced" or "processed" mechanisms.
437   They should give essentially the same predictions in models as the explicit generated mechanisms as long
438   as the temperature and O_2 levels are constant at the defaults set for the reactor when the mechanism was
439   generated (see Table 6), and as long as the steady state approximation is appropriate for the reacting
440   intermediates removed during the minimal reduction process. This has been verified in test calculations
441   using representative chemical systems.

442   MechGen automatically derives a minimally processed mechanism after any successful full
443   mechanism generation process. These are used as the input when deriving product yields, deriving multi-
444   generation mechanisms, or when deriving lumped mechanisms using various approaches, as discussed in
445   the following sections. The algorithm employed to derive minimally reduced mechanisms from the
446   explicit generated mechanisms is outlined in Section S3.2 of the Supplement.

447   The minimal reduction process tends to reduce the numbers of reactions by factors of 2 - 2.5, and
448   the numbers of reacting intermediates by factors of 3 - 5. This is shown in Figure S8 in the Supplement,
449   which shows plots of these ratios against a measure of the mechanism sizes for the mechanisms of the 27
450   representative compounds and mechanism generation options discussed in Section 3.4. The extent of
451   reduction in numbers of reactions is variable and has no clear dependence on the mechanism size, but the
452   reduction in numbers of intermediates is less variable and clearly decreases with mechanism size. Again,
453   the two representative aromatics appear to be outliers in this regard, having much greater reductions in
454   numbers of intermediates. This may be related to the fact that the aromatics tend to have the larger
455   numbers of intermediates than comparably sized mechanisms for other compounds (see Figure S7).

456   Representative examples of MechGen output for a minimally processed mechanism are shown in
457   Figure 3 and Figure S3, which use the explicit mechanism from Figure 2 or S2 as the starting point.
458   Figure S3 shows the complete output for the products of the reactions of 1,3-butadiene in the default
459   standard environment, while Figure 3 shows the first parts of the species and reaction listings. The format
460   aligns with that used by the SAPRC modeling software (Carter, 2024), which is similar to formats used
461   by other modeling software systems. The top part of the figures shows the names used for reactants and
462   products in the mechanisms and their corresponding structures. Intermediates are named using the name
463   of the reactant followed by "-" and a sequence number, while stable species use various naming methods.
464   Compounds that are important in emissions or the atmosphere are generally given 2-8 character SAPRC
465   names based on abbreviating their actual names; compounds that may not be as important but where
466   permanent names are or have been needed when developing SAPRC mechanisms are given permanent
467   "ORG-nnnn" names; and others are given temporary "VOC-nnnn" names that are applicable only to this
468   processed mechanism. Additionally, processed mechanisms using actual structures strings to name the
469   reactants and products can also be output, as discussed in the User Manual.

470   **3.3.   Estimation of Product Yields**

471   The relative yields of products formed in the generated mechanisms can be determined if the
472   conditions of the environment where the compound reacts are specified and assumed to be constant.
473   Derivations of product yields for specified environments are useful to provide a means for assessing
474   which types of products are important and how their yields vary with environmental conditions. It is also
475   useful for improving the efficiency and optimizing the size of multi-generation mechanisms, as discussed
476   below in Section 4.

477   The algorithm MechGen uses to derive relative product yields from a processed mechanism is
478   described in Section S3.3 of the Supplement. It uses in part the fact that most radical intermediates can be
479   ordered such that intermediates are formed only by reactions of intermediates above it on the list. This

480 assumption is generally true for mechanisms derived using MechGen, with the exceptions of
481 interconversion of phenoxy and phenyl peroxy radicals due to their reactions with O_3 and NO,
482 respectively, and to interconversions of hydroperoxy-substituted peroxy radicals due to rapid H-shift
483 reactions, as discussed by Carter et al (2025a) and Vereecken and Nozière (2020). The most rapid of the
484 hydroperoxy-peroxy interconversions are removed by employing an equilibrium approximation, but some
485 such interconversions are too slow for this approximation to be appropriate, as discussed in Section S3.1.
486 For such cases, a matrix inversion procedure derived from the steady-state equations for the pseudo-first-
487 order reactions of the interconverting radicals is employed, as discussed in Section S3.3.

488 Figure S4 in the Supplement shows an example of product yield output corresponding to the
489 processed mechanism shown in Figure 3 and S3 and the explicit mechanism in Figure S2. Selected
490 portions of this output are shown in Figure 4. This product yield output consists of three parts:

- 491 • The first part summarizes the environments used and, for the web system, gives links to obtain
  492 their oxidant concentrations.

- 493 • The second part shows the fractions of the compound that react with the various oxidants, and is
  494 not shown if compound reacts with only one reactant. For this example, most of the reaction is
  495 with OH, except in the nighttime environment where reaction with NO_3 dominates.

- 496 • The third part shows the product yields for the various environments, sorted in descending order
  497 by average yield in the environments, with organic products hyperlinked in the web output so
  498 they can be readily created and reacted if desired, and with inorganic products such as NO_2 (e.g.,
  499 from peroxy + NO reactions), HO_2, O_2, and H_2O included. The entry for "NO-loss" shows how
  500 many moles of NO are consumed by reaction with peroxy radicals, which is directly responsible
  501 for O_3 formation when organics react in the presence of NO_x.

502 If the reactant forms non-volatile or semi-volatile products, the product yield output will also
503 include mass percent yields of various volatility bins. If a representative atmospheric organic PM level is
504 specified as a reactor option (50 μg m^{-3} is the default), the product yield output will also include total
505 SOA yields corresponding to that organic PM level. This is discussed in Section S4 of the Supplement.
506 The total mass yield of non-volatile products is also output. Note that these mass bin yields will sum up to
507 more than 100% because of the oxygen and nitrate groups added to the molecules during the oxidation
508 process.

509 3.4. Effects of Varying Full Mechanism Derivation Options

510 The mechanisms derived using the full mechanism operation depend on the values of the MinYld
511 parameter and the use of environments, both of which affect which reactions and intermediates are
512 neglected due to low relative yields. The effects of different choices in this regard are discussed in this
513 section. This was examined by varying MinYld and environment options when deriving mechanisms for
514 the representative compound α-pinene, and by comparing results for a more limited set of choices for 27
515 representative compounds of various types (listed in Table S3 in the Supplement). These include the
516 series of n-alkanes from propane through hexadecane, various C_8 alkanes, alkenes, oxygenates, and
517 aromatics, and two terpenes.

518 Figure 5 shows the effects of varying the MinYld parameter and environment options on the
519 numbers of explicit reactions, stable products, and peroxy intermediates in mechanisms generated for α-
520 pinene for various environment options. As expected, the number of reactions increases as the MinYld is
521 decreased, with no tendency to level off at either the high or low MinYld range. As also expected, using

522   no environments resulted in the largest mechanisms, while using the three default environments (shown in
523   Table 7) resulted in about three times fewer reactions in the case of α-pinene, and using only one
524   environment reduces the number of reactions by almost an additional factor of two. This is because
525   estimated upper limit yields are highest if no environments are used and increase with the numbers of
526   environments if more than one is used. Note that the numbers of peroxy radicals are important because
527   they determine the minimum size of the processed mechanisms, even if product lumping is employed.

528   Although removing reactions through the MinYld test does not affect mass balance because they
529   are replaced by competing reactions determined to be more important, these removals will change
530   predicted product yields, reducing the potential accuracy of the mechanism. This can be assessed by
531   assuming that a mechanism derived with very small MinYld values and no environments can approximate
532   the mechanism derived without deletions. Comparing product yields predicted by a mechanism generated
533   with a given set of options with those predicted by this reference mechanism gives an indication of the
534   effects of these deletions on predictions. The total changes in product yields are quantified as:

535 $$\text{Total yield change (Set, Envt)} = \sum_{\text{Products}} \left| \text{Yield(Product, Set, Envt) - Yield(Product, Ref)} \right|$$

536   where "Set" refers to a mechanism derived for a compound using the mechanism generation options being
537   considered, "Envt" is a representative environment, "Ref" refers to reference mechanism for the
538   compound derived with the lowest MinYld and no environments, and "Yield" is the molar yield of
539   product per mole VOC reacted under the conditions of that environment. The summation includes all
540   organic products with yields greater than 0.0005 moles per mole VOC reacted and also amounts of NO
541   consumed, and NO_2, HO_2 and OH radicals formed. Yields for products less than 0.0005 moles per mole
542   VOC reacted are insignificant individually and thus excluded in the summation, though not necessarily
543   insignificant collectively. The values for the total yield change can range from zero to a maximum of the
544   total moles of products formed per mole reactant, which is usually more than one.

545   Figure 6a shows the total yield changes predicted for the mid-NO_x standard urban environment
546   for all the α-pinene mechanisms shown in Figure 5, as a function of the MinYld parameter. The yields
547   predicted by the reference mechanism derived using a very low MinYld value of 0.01% and no
548   environments (with ~26,000 explicit reactions) are used as the standard. As expected, yield changes
549   increase as the value of MinYld increases, though the dependence on environment options is relatively
550   small except at higher MinYld values. The total changes exceed 0.05 moles product per mole α-pinene
551   reacted when MinYld > 1%, but are around 0.01 moles or less when the default MinYld value of 0.5% is
552   used. Decreasing MinYld to 0.1% resulted in changes being decreased to 0.003 moles or less.

553   Figure 6b shows the effects of MinYld on moles of NO consumed per mole of α-pinene reacted,
554   and Figure 6c shows effects on the yields of organic products in the condensed phase, assuming
555   equilibrium and atmospheric aerosol levels of 50 μg m^{-3} (as discussed in Section S4 of the Supplement).
556   The effects on NO consumption reflect effects on O_3 formation because reactions of intermediates with
557   NO are the processes responsible for O_3 formation in the lower atmosphere (e.g., see Finlayson-Pitts and
558   Pitts Jr., 1999). The amount of organic products predicted to be in the condensed phase assuming
559   equilibrium provides an indication of the effect of the VOC's reaction on SOA formation, though this
560   should be considered to be highly approximate because effects on SOA depend on environmental
561   conditions, estimates of vapor pressure are uncertain, and the equilibrium approximation is an
562   oversimplification. Figure 6b and c indicate that, for α-pinene at least, the apparent effects on both O_3 and
563   SOA are relatively small, being only 0.3% or less when the default MinYld parameter is used.

564   Figure 6 shows effects of MinYld on product yield changes for α-pinene under mid-NO$_x$ urban
565   conditions. The effects on yield changes for other standard environments incorporated in MechGen are
566   shown in Figure S6 in the Supplement. The total yield changes for the daytime environments are similar,
567   though the effect on NO consumption is lower under lower NO$_x$ conditions because there is less NO
568   consumed. The total yield changes are much less under nighttime conditions, presumably because fewer
569   reactions tend to be important. Note that no NO is present in the nighttime scenario, and MechGen
570   predicts that α-pinene forms much lower yields of low-volatility products at night compared to daytime.

571   The effects of varying MinYld and environment options for other compounds are shown in Figure
572   7. This shows effects on (a) numbers of explicit reactions, and (b) total yield changes, for the mid-NO$_x$
573   environment for 27 representative compounds of various types (listed in Table S3 in the Supplement).
574   Plots showing ratios of numbers of reacting intermediates or stable products to numbers of reactions for
575   these compounds with different mechanism generation options are shown on Figure S7. As expected, the
576   mechanism size increases with the size of the molecules, but the type of molecule is also important, as
577   indicated by the large variability for the representative C$_8$ compounds. Highly branched compounds and
578   aromatics tend to have the smallest generated mechanisms, whereas cyclic alkenes, such as terpenes, tend
579   to have the largest. The effects on product yield tend to correlate with mechanism size, though the results
580   are variable and the correlation is relatively weak (~50%) except at the lowest MinYld level.

581   Table 7.   Standard environments available for determinations of which reactions are negligible
582   during mechanism generation and for estimation of product yields.

Scenario	High NO_x	Standard (mid NO_x)	Low NO_x	Nighttime
Description [a]	Maximum VOC reactivity (MIR)	Equal sensitivity of O_3 to VOC and NO_x (EBIR)	NO_x levels 1/10 those giving maximum O_3 (MOIR NO_x / 10)	EBIR scenario at nighttime
Default use [b]	Gen, PY	PY, MGen	Gen, PY	Gen, PY
Component concentrations (molec/cm^3)				
OH	7.51e+6	1.11e+7	3.69e+6	2.86e+5
O_3	1.78e+12	2.26e+12	9.20e+11	2.90e+12
O^3P	5.50e+4	2.84e+4	8.85e+3	~ 0
NO	9.58e+10	1.59e+10	4.12e+9	7.43e+5
NO_2	3.15e+11	8.63e+10	1.32e+10	4.88e+10
NO_3	8.54e+6	9.92e+6	1.25e+6	2.25e+9
HO_2	~ 0	6.00e+8	7.17e+8	1.27e+8
Total RO_2's	~ 0	4.59e+8	6.69e+8	2.35e+8
Total RCO_3's	~ 0	7.30e+7	7.27e+7	1.04e+8

583   [a] See Section S2 in the Supplement for more information on these scenarios. All these represent urban
584   conditions similar to those used to derive the (Carter, 1994) reactivity scales except that the total
585   VOC levels were reduced to more closely represent modern ambient conditions.

586   [b] Gen = used to determine negligible reaction during full mechanism generation; PY = used for
587   determination of product yields; MGen = used for multi-generation mechanism derivations discussed
588   in Section 4.

589 Table 8. List of parameters that affect full mechanism generation.

Parameters	Affects	Default
Parameters that affect all full mechanism generation operations		
Minimum yield (MinYld)	Reactions with estimated maximum yields below this value are excluded during full mechanism generation. See Scheme 2 in the Supplement. [a]	0.5%
Environmental conditions	Sets of environmental conditions to determine upper limit yields involving bimolecular reactions and for displaying product yields. Users can select which environments, if any, are used for mechanism generation.	High NOx Low NOx Nighttime
Lumping	Set of lumping options that control generation and processing for lumped mechanisms. See Section 5.	No Lumping
kFastUni	Minimum unimolecular rate constant for a non-radical product to be treated as an intermediate and reacted.	1.67×10^{-2} s^{-1}
kRO2slow	Minimum rate constant for unimolecular reactions of peroxy radicals not to be neglected.	1.0×10^{-3} s^{-1}
kEqConv	Peroxy radical interconversion rate constant above which the interconverting radicals can be treated as in equilibrium when the mechanisms are processed.	30 s^{-1}
kRO2fast	Total unimolecular rate constant for peroxy radicals above which bimolecular reactions are ignored. [a]	50 s^{-1}
Explicit radical list	Organic radicals that are treated as final products and not reacted during "fully react" operations.	None (list empty)
Parameters that affect only multi-generation operations. See Section 4		
Reacting environment	Environment used to estimate product yields to determine if the organic product should be reacted.	Standard (mid NO$_x$)
MGminYld	Minimum estimated yield for a product formed in a multi-generation operation to react.	0.01%
MinVP	Minimum estimated vapor pressure for a product formed in a multi-generation operation to react.	1.0×10^{-13} atm.
RxnHours	Reaction time used to estimate amounts of products reacted in the environment for the purpose of determining if its products may be non-negligible.	6 hours
Parameter affecting estimates of SOA formation. See Section S4 in the Supplement.		
PM	Concentration of organic particle matter in μg/m^3	50 μg/m^3

590 [a] These tests are not used when hydroperoxy-substituted peroxy radicals rapidly isomerize to form an
591   another such radical that can isomerize back to the initial radical, resulting in no net loss of the
592   radical, making competing reactions potentially non-negligible.

```
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}
...
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
...
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
...
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.
...
```

Figure 2.   Portions of MechGen output showing representative results of a full mechanism generation operation for the reactions of 1,3-butadiene with default options. The full output is shown in Figure S2 in the Supplement.

```
.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
...

.STS
13-BUTDE-1    ! CH2=CH-CH[OO.]-CH2-OH
...

.RXN
R) 1.12e-11 -1.053 0.00  ;13-BUTDE + OH = #.62 13-BUTDE-1 + #.33 13-BUTDE-2 +
                          #.05 13-BUTDE-3
R) 1.34e-14 4.537  0.00  ;13-BUTDE + O3 = #.5 ACROLEIN + #.14 VOC-0014 +
                          #.36 13-BUTDE-8 + #.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-BUTDE-6 + #.18 13-BUTDE-7 +
                          #.5 13-BUTDE-5
R) 9.13e-12              ;13-BUTDE-1 + NO = #.95 ACROLEIN + #.05 ORG-4682 +
                          #.95 HO2 + #.95 NO2 + #.95 HCHO
R) 2.30e-12              ;13-BUTDE-1 + NO3 = ACROLEIN + O2 + HO2 + NO2 + HCHO
R) 1.61e-11              ;13-BUTDE-1 + HO2 = ORG-1004 + O2
R) 1.60e-11              ;13-BUTDE-1 + RCO3 = #.8 ACROLEIN + #.2 ORG-1048 +
                          O2 + #.8 HO2 + #.8 HCHO
...
```

Figure 3.   Portions of MechGen output showing the minimally reduced processed mechanism for the reactions of 1,3-butadiene with default options. The complete output is shown in Figure S3 in the Supplement.

```
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
...
0.06%      -          0.16%      -          SOA mass yield from products for atm PM = 50 ug/m3
```
                        (lower yield products not shown)

601   Figure 4.   Portions of MechGen output showing estimated product yields the reactions of 1,3-
602   butadiene with OH, O_3 and NO_3 radicals for various environments, derived using default
603   mechanism generation options and environments.

604
605
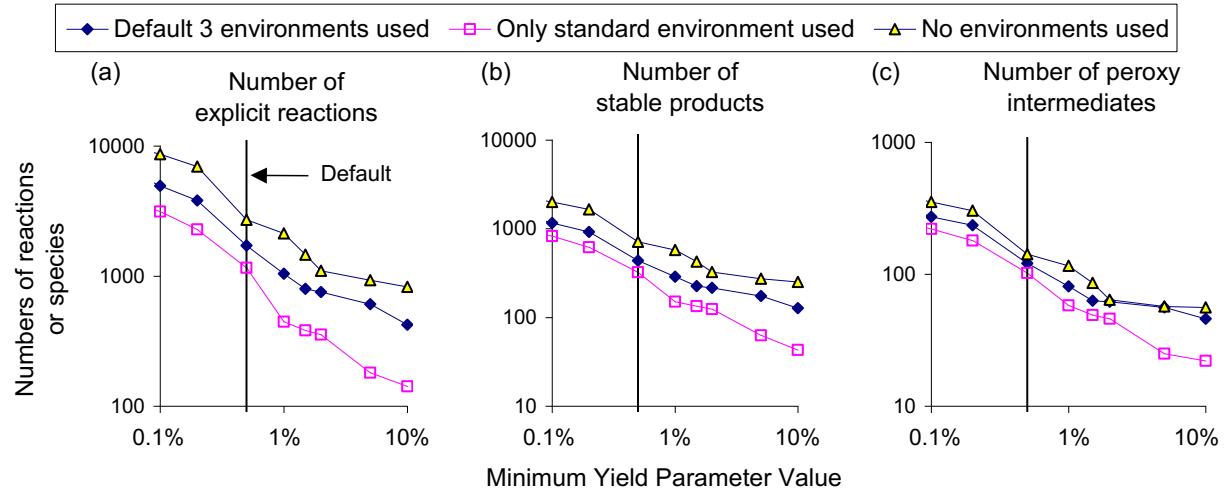
606

607   Figure 5.   Effects of varying the MinYld parameter and environment options on the numbers of
608   explicit reactions in full mechanisms generated for α-pinene.

22
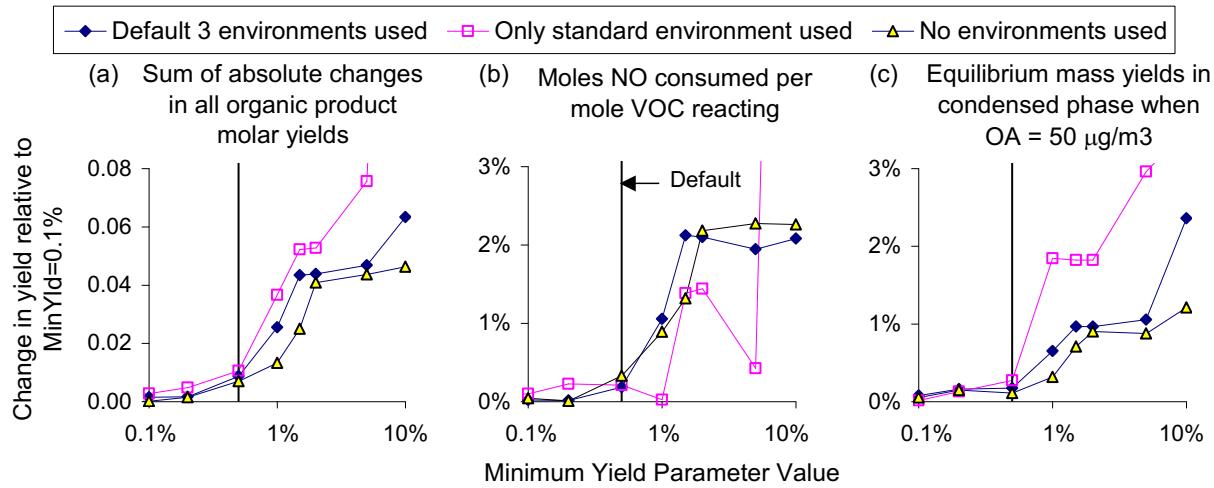
Figure 6. Effects of varying the MinYld parameter and environment options on relative changes in yields of selected groups of products from α-pinene when reacted under mid-NO$_x$ urban environmental conditions.
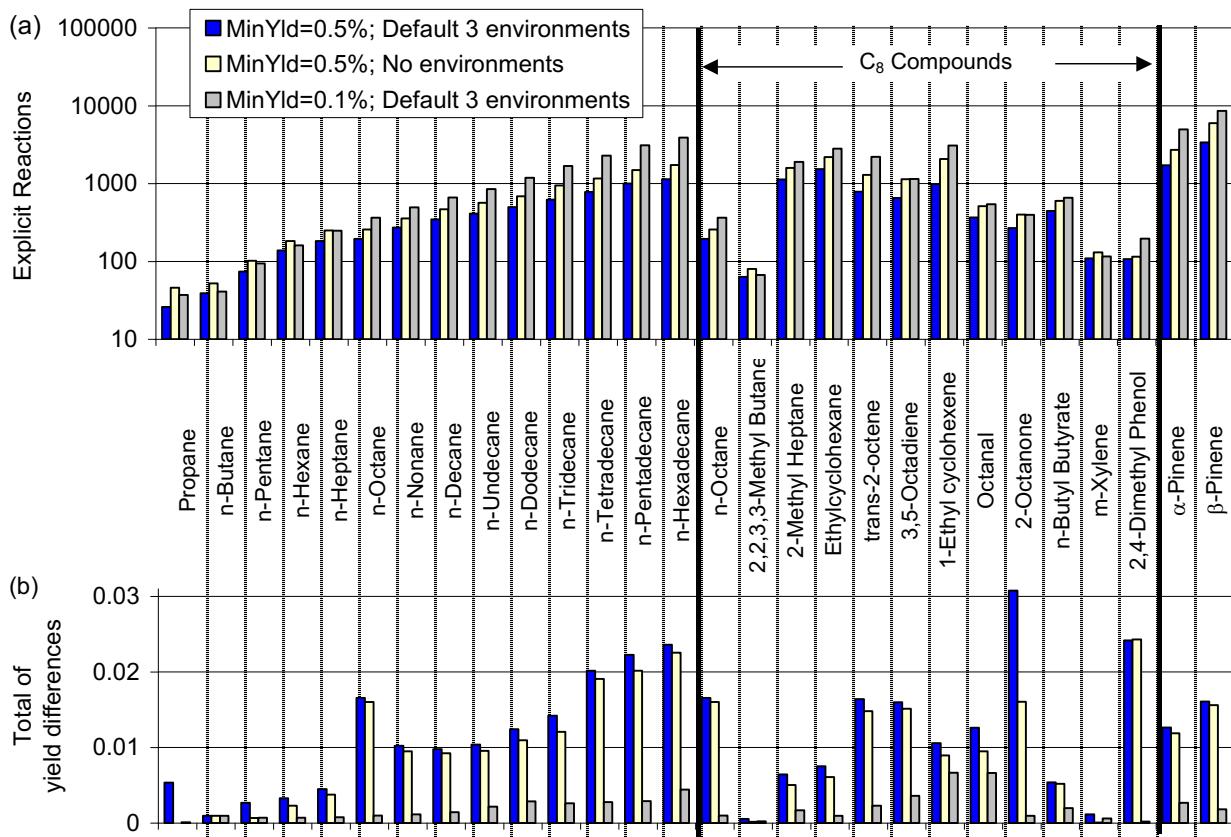
Figure 7. Numbers of explicit reactions and sum of differences in product yields relative to the largest generated mechanisms for the various mechanism generation options.

4. DERIVATION OF MULTI-GENERATION MECHANISMS

4.1. Description of Process and Outputs

The full mechanism generation process discussed above produces only a single generation mechanism because it does not react the stable products formed. MechGen can also derive multi-generation mechanisms where all reactive products formed in non-negligible yields, including those formed after reacting multiple generations, are reacted. It involves carrying out full mechanism generation for a selected starting compound, then determining the stable products formed, then generating mechanisms for products in non-negligible yields, then repeating the procedure until no reactive, non-negligible products remain. This would be relatively straightforward, except that reacting all products regardless of their yields would result in unmanageably large mechanisms, dominated by reactions and species of negligible importance under conditions of interest. Therefore, much of the multi-generation derivation by MechGen focuses on determining which products are in fact non-negligible and which don't have to be reacted.

The procedures used for deriving multi-generation mechanisms are described in the User Manual. They involve accessing the system using the terminal interface, creating a software object to control the process for a specified compound and to store the results, and then using it to carry out the operations and transmit the results. These objects have several parameters that control which products are formed in sufficiently high yields to react and which can be treated as reactive. One of these is the specification of an environment used to estimate product yields and another is the "MGminYld" parameter that determines which yields are treated as negligible. In addition, to determine which product can be treated as reactive, a "MinVP" parameter gives the minimum vapor pressure for product compound to react in the gas phase, and a "RxnHours" parameter determines the time for compounds to react and form products. These options can be changed by the users, but by default, the environment used for the multi-step mechanism generation is the first one listed to derive product yields as discussed Section 3.3. The default values of MGminYld, MinVP, and RxnHours are 0.01%, 10^{-13} atm, and 6 hours, respectively (Table 8). These need to be specified to minimize the numbers of totally negligible reactions and processes in the multi-generation mechanisms that are derived.

The algorithm MechGen uses to derive multi-generation mechanisms is summarized below, and a flowchart showing the process is given in Scheme S6 in the Supplement. The process starts by generating full single-generation mechanisms for the initial reactant and for all products formed in non-negligible yields. Each mechanism generation is followed by deriving minimally reduced mechanisms, estimating product yields for the selected environment, and determining what products are formed in sufficiently high yield to react. The estimated upper limit yields are derived based on:

$$\text{Yield of initial reactant} = \text{Reactivity correction for initial reactant} = 100\%$$

$$\text{kPUni} = \text{kUni} + \text{kh}\nu + \Sigma_{OX}\, k_{OX} \times [OX]^{environment}$$

$$\text{Reactivity correction for reacting product} = 1 - e^{-\text{kPuni} \times \text{RxnTime}}$$

$$\text{Est. yield of product} = \text{Est. yield of reactant forming it} \times \text{Reactivity correction for reactant forming the product} \times \text{Yield of product from reactant}$$

24

655 where "kPUni" is the pseudo-unimolecular rate constant for the reactions of a reactant in the environment,
656 "kUni" and "khv" are the unimolecular and photolysis rate constants for the reactant (if applicable), "OX"
657 and "k_{OX}" refer to all oxidants with which the compound reacts and their rate constants, "RxnTime" is
658 derived from "RxnHours" to be consistent with the units used for kPuni, and "Est." refer to upper limit
659 estimates. The yields of products from reactants in the environment are derived as discussed in Sections
660 3.3, and the kPUni values are derived by multiplying the rate constants for the initial reactions multiplied
661 by the concentration of the oxidant with which it reacts in the environment, if applicable. The reactivity
662 correction factors are calculated as fractions in the environment and are used to account for the fact that
663 slower reacting compounds form lower amounts of products during the reaction time in the environment.
664 The formula for calculating fractions reacted, derived from integrating the kinetic differential equation, is

665 $$\text{Factor Reacted (Reactant, Environment)} = 1 - e^{-kPuni(\text{Reactant, Environment}) \text{ x RxnTime}}.$$

666 This gives near 100% fractions reacted for rapidly reacting compounds, and fractions that are
667 approximately proportional to the kPuni for slowly reacting compounds. Note that the estimated upper
668 limit yields tend to decrease as the number of generations of reactions increases.

669 The standard environments used (or not used) when carrying out the full, single-generation
670 mechanism derivations for individual compounds, as discussed in Section 3.1, do not necessarily need to
671 be the same as the one required to carry out the multi-generation process. If standard environments are
672 used during the single-generation mechanism derivation process for individual compounds, they should
673 also represent the environment used for the multi-generation process. This is assured if the environment
674 used for the multi-generation process is one of those used for the single-generation process, of if no
675 environments are used for the single-generation process. By default, the three standard environments used
676 for single-generation processes incorporate the conditions of the default standard environment for multi-
677 generation derivations, as indicated in Table 7.

678 The results of this process consist of lists of reacting species, their reactions and rate constants,
679 and lists of low reactivity, low volatility, or low yield products that are not reacted. As discussed in the
680 User Manual, these results can be transmitted to users in several types of files. These include: (1) files
681 containing lists of reactants, non-negligible products, and counter species representing atom numbers and
682 total masses of low yield products and their approximate yields; (2) files giving information about the
683 reactants and products in terms of groups in the molecules; (3) files with listings of the explicit and
684 processed reactions of the reacting compounds; and (4) files that can be used as input to prepare the
685 mechanisms for model simulations using SAPRC box modeling software (Carter, 2024). The model
686 preparation input files come in two forms: one containing the minimally reduced processed mechanisms
687 for all the reactants, and the other giving pseudo-unimolecular reactions of the reactants that are
688 applicable only to the conditions of the environment used to generate the mechanism. The specific types
689 of information that can be obtained are listed in Table S4 of the Supplement.

690 **4.2.   Examples of Results**

691 To illustrate the results of multi-generation mechanism derivation operations and their
692 dependence on selected options, mechanisms were derived for several example compounds ranging from
693 propane to α-pinene using a standard set of options, and additional multi-generation mechanisms were
694 derived for α-pinene using differing sets of options. The compounds and mechanism derivation options
695 are summarized in Table 9, which gives the sets of options, the numbers of explicit reactions, information
696 on lost or low yield carbon, and estimated SOA formation for the mechanisms. In the cases where the
697 compound or MGminYld parameter was varied, the single-generation mechanisms for individual
698 reactants were derived using the three default environments listed in Table 7 and the default mechanism

699 generation options given in Table 8. In the cases where the environments were varied, the single-
700 generation mechanisms were derived using the stated environment only, with defaults used for the other
701 parameters. Table 9 shows that, as expected, the results depend significantly on the MGMinYld parameter
702 and the multi-generation environment used.

703   The pseudo-unimolecular reaction output was used to prepare input for the SAPRC box modeling
704 software (Carter, 2024) to carry out model simulations for the evolution of product concentrations and
705 negligible atom counter species over time, under conditions of the environment used to derive the multi-
706 generation mechanism. The use of this output is equivalent to assuming that the atmospheric oxidant
707 levels are constant at the levels associated with the environment, as given in Table 7, and at constant light
708 intensity with photolysis rates calculated for the light associated with the reactor. The model simulations
709 were carried out for 6 hours of simulated time, with the starting reactant being initially present at 1 ppm
710 mixing ratio, with no dilution or variation in reaction conditions or oxidant species. The calculated
711 concentrations were used in conjunction with separate MechGen output giving structural information
712 about the compounds (Carter et al., 2025a) for the purpose of computing total yields of compounds with
713 different structures (structural counter species), and total carbons in compounds that were not reacted
714 because of estimated low yields or reactivities.

715   Figure 8 presents concentration-time results for the highest yield structural counter species or C_1
716 products for α-pinene, and these results for all representative compounds listed in Table 9 are given in
717 Figure S10 of the Supplement. These mechanisms were all generated using a MGminYld value of 0.001%
718 and the mid NO_x urban standard environment, which are the defaults. These figures also show the O/C
719 ratios for the reactant and products as a function of time. The structural counter species are listed and
720 briefly described in Table S4 in the Supplement. Concentrations relative to the amount of starting
721 material reacted at the times are shown at 30-minute intervals up to 2 hours of simulation time, then at 2-
722 hour intervals after that. The evolution of the ratio of total carbon in low yield unreacted compounds
723 (NegC) to total reacted carbon is also shown in Figure 8 or S10, except for cases when they are near zero.

724   Figure 8 and S10 show that the O/C ratios increase with time, as expected due to ongoing
725 oxidation. The yields of the C_1 products also increase significantly and make relatively high contributions
726 to the total products formed by the end of the simulations. Relative concentrations of other types of
727 products tend to go down with time after the first half hour, particularly more reactive products such as
728 aldehydes. As expected, the distribution of products significantly depends on the type of compound. Note
729 that these simulations treat formaldehyde as a non-reacting product, while in fact a significant fraction of
730 it would react to form CO and CO_2 by the end of the simulations.

731   Examples of effects of using different environments on multi-generation mechanisms derived for
732 α-pinene are shown in Table 9 and Figure 9. In those examples, the single-generation mechanisms were
733 derived using only the subject environment rather than the three default environments, which is the reason
734 that the number of reactions derived for α-pinene in these calculations is smaller than in those where the
735 compound or MGminYld parameter was varied (see Figure 5), though the overall product yields are
736 essentially the same. As discussed earlier, the standard environment represents intermediate NO_x
737 conditions, and thus includes both high and low NO_x reactions. Therefore, as shown in Table 9, the
738 mechanism derived for it has more reactions and reactants than those derived separately for high and low
739 NO_x conditions. On the other hand, fewer types of reactions are predicted to be important in the nighttime
740 environment, so significantly smaller mechanisms are derived. As shown in Figure 9, the concentrations
741 of compounds with hydroperoxide groups are highest under low NO_x conditions and the yields of N-
742 containing compounds are lower, though the yields of nitrates or peroxyacylnitrates (PANs) are not as
743 strongly affected as one might expect based on the relative NO_x levels in the environments. The yields of
744 compounds with ester, alcohol, or peroxy acid groups are significantly lower under nighttime conditions,

745 though the predicted levels of the other compounds are generally within the ranges predicted for the
746 daytime environments. Further discussion of the mechanistic implications of these results is beyond the
747 scope of this paper, but this is worthy of additional study.

748 The effects of varying the MGminYld parameter on products or counter species in the α-pinene
749 mechanisms are shown in Figure S11 in the Supplement. Increasing MGminYld raises yields of neglected
750 compounds (reflected by NegC), resulting in a slight decrease in yields of most products, with the yields
751 decreasing by ~10% when MinYld is increased from 0.25% to 1%. This means that yields in multi-
752 generation mechanisms are necessarily lower limits unless very low values of MGminYld are used to
753 minimize the amount of untreated carbon (NegC).

754
755
756 Table 9.　　　　Summary of multi-generation mechanism derivation examples.

Compound	Options [a]		Mech Size		Neg /Rct'd C [c]	SOA [d]
	MGminYld	Environ [b]	Reactions	Reacted		
Propane			299	37	0.03%	~0
n-Butane		Mid NO$_x$ urban (default)	1203	109	0.2%	0.01%
1-Butene	0.01% (default)		1381	126	0.3%	0.02%
n-Octane			50235	1466	6%	7%
Trans-2-octene			24570	945	4%	5%
α-Pinene			82375	1695	12%	52%
		High NOx	20819	1252	8%	36%
α-Pinene	0.01%	Mid NO$_x$	56446	1683	12%	52%
		Low NOx	44933	1337	9%	51%
		Night	5961	95	0.3%	7%
	0.02%		56952	1088	16%	49%
α-Pinene	0.01%	Mid NO$_x$	82375	1695	12%	52%
	0.005%		116426	2587	9%	54%

757 [a] All other options were held at defaults shown in Table 8.

758 [b] Environment(s) used when deriving the multi-generation mechanism.

759 [c] Ratio of moles carbon in compounds that are not reacted due to low yields divided by moles
760 　　carbon in the reacting starting compound after reacting for RxnTime = 6 hours.

761 [d] SOA = total moles of all low-volatility products multiplied by their estimated fractions in the
762 　　particle phase when total atmospheric SOA levels were set at 50 µg m^{-3} (see Section S4).

763
764
765
766
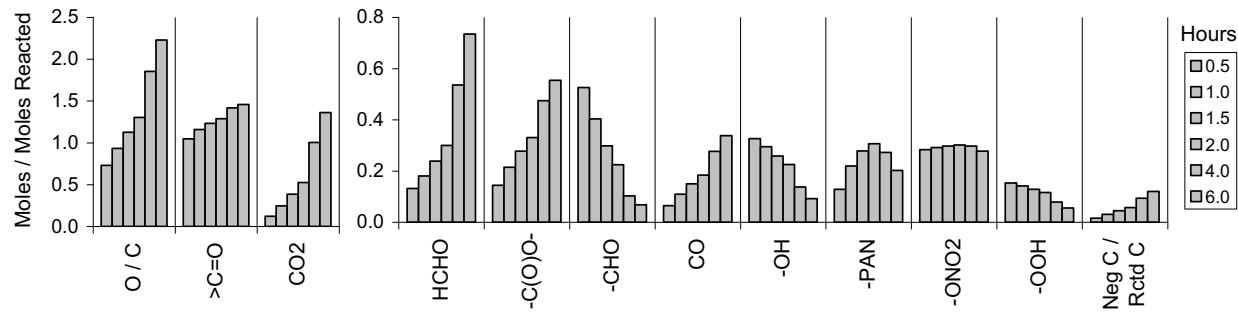
767
768 Figure 8.   Moles of various types of products or functional groups formed as a function of time for
769   6-hour simulations using multi-generation mechanisms for α-pinene. Products are
770   ordered from high to low maximum yields.

771
772
773
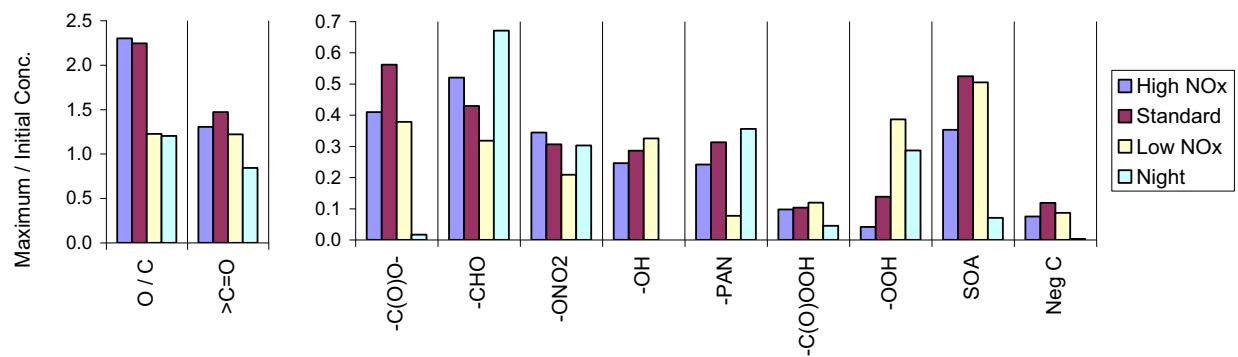
774
775 Figure 9.   Effects of varying the environment on maximum molar concentrations of various types of
776   products or functional groups for the simulations using multi-generation mechanisms for
777   α-pinene.

778
779

5. DERIVATION OF LUMPED MECHANISMS

Though smaller than the explicit mechanisms, the minimally reduced mechanisms are still too large for use in practical modeling applications involving complex atmospheric mixtures. Further reduction is necessary, both by using a limited number of lumped model species to represent chemically similar compounds, and by using various approaches to reduce the numbers of reactive intermediates in the mechanisms. MechGen can optionally be used to derive lumped mechanisms for organics that are consistent with the lumping approaches used in recent SAPRC mechanisms (Carter, 2010a, b, 2023; Carter and Heo, 2013). As indicated in Table 8, the lumping approach must be specified as one of the options used for full mechanism generation. The default lumping option is "explicit" or no lumping, which involves generating all reactions expected to be non-negligible and then preparing only minimally reduced processed mechanisms from the results, with no additional processing. This was used for all the examples discussed in this paper. Other lumping options currently available in MechGen are briefly summarized below, with additional details given in the User Manual, as well as relevant SAPRC mechanisms papers and reports. A more complete discussion of lumping algorithms is beyond the scope of the present paper and will be presented in more detail in a subsequent paper (see also Carter, 2023).

SAPRC-11 lumping is the approach used in the SAPRC-11 mechanism (Carter and Heo, 2013), and is essentially the same as that used for SAPRC-07 (Carter, 2010a, b) except as discussed by Carter and Heo (2013). All organic products are represented by one of ~50 lumped or explicit model species, and all peroxy intermediates are represented by ~30 chemical operators that represent products formed and how their yields may change with NO_x conditions. (An "operator" in this discussion refers to a model species that does not correspond directly to a chemical species but is added to represent overall effects of various reactions [e.g., Carter, 2020a,b]). In addition, all acyl peroxy, phenoxy, and phenyl peroxy intermediates are represented using lumped radical model species in all current SAPRC mechanisms. Note that the peroxy operator method as originally implemented for SAPRC-07 and SAPRC-11 does not provide for peroxy radical isomerization reactions, so strictly speaking they are not compatible with mechanisms currently generated using MechGen, which includes such reactions when predicted. This issue is addressed by assigning an "effective NO" concentration to take competitions between unimolecular and bimolecular reactions into account when determining yields of operators representing the products formed, as discussed by Carter (2023). The default effective NO for SAPRC11 lumping is 0.5 ppb, reflecting intermediate NO_x urban conditions.

SAPRC-22 lumping is the approach used in the "standard" version of SAPRC-22 mechanism (Carter, 2023). This approach represents organic products using ~110 lumped or explicit model species, with the numbers of lumped species increased in part to better represent NO_x recycling and SOA formation potentials. Peroxy radicals are represented using a modified version of the operator approach used for SAPRC-07 and -11, in which operator yields are calculated at both low and high effective NO levels, and an extra operator species is added for each reaction where the yields are different at different NO levels, so the lumped mechanism is applicable for a wide range NO conditions. This lumping approach was used to derive the SAPRC-22 mechanism (Carter, 2023), using procedures discussed in the User Manual.

If SAPRC-22 lumping is selected, users have the option to generate SAPRC-22-compatible lumped mechanisms that represent selected specific organic compounds explicitly in models, rather than having them lumped as in the standard SAPRC-22 mechanism. The products formed in the reactions of these explicitly represented compounds will then be lumped to be compatible with SAPRC-22 and these lumped reactions can be output along with the rest of the SAPRC-22 mechanism to be used as input to the SAPRC model preparation programs (Carter, 2024). These expanded versions of SAPRC-22 can then be

825   used in airshed models to evaluate the effects of the selected compounds on O_3 or other measures of air
826   quality, or to allow comparisons of their predicted concentrations with observations.

827   If any type of SAPRC lumping is selected, the mechanism generation process will exclude the
828   reactions of acyl peroxy, phenoxy, or phenyl peroxy radicals, as these are all represented by lumped
829   model species, so their explicit reactions are not used when deriving the lumped mechanism. In addition,
830   if a non-explicit lumping method is employed, the types of initial reactions the organic compounds
831   undergo are determined by the types of initial reactions of the lumped model species used to represent
832   them, which is part of the input used to define the lumping approach, as discussed in the User Manual. In
833   some cases, this is a subset of the types of initial reactions generated when the default "explicit" (no
834   lumping) option is in effect.

6. DISCUSSION

Because of the complexity of the atmospheric reactions of most organic compounds, automated mechanism generation systems such as MechGen provide an important link from chemical kinetic and mechanistic data and theories to model predictions of chemical transformations in the atmosphere. As discussed by Kaduwela et al. (2015) and Ervens et al. (2024), the stages of mechanism development for models that provide this link consist of (1) compiling and evaluating available basic mechanistic data; (2) developing SARs and other estimation methods needed to develop complete mechanisms; (3) deriving detailed mechanisms that incorporate these data and estimation methods; (4) reducing the detailed mechanisms so they can be used in various modeling applications; and (5) evaluating the predictions of the mechanisms against observations. MechGen is a tool that can provide the link between mechanistic data and theory and detailed mechanisms, and can also be used for mechanism reduction. The development of the more recent SAPRC mechanisms for airshed models (Carter, 2010a, b, 2023; Carter and Heo, 2013) was undertaken with this approach in mind, using MechGen to derive the SAPRC lumped mechanisms from the explicit mechanisms for the many compounds that are represented.

Detailed mechanisms derived by MechGen can also be used in analyses of laboratory (e.g., Li et al., 2022) or ambient data, providing an alternative or supplement to the Master Chemical Mechanism (MCM) (Bloss et al., 2005; Jenkin et al., 1997, 2003; Saunders et al., 2003), which has been widely used for this purpose. MechGen offers the flexibility of being applicable to any compound within the scope of its predictive capabilities, whereas MCM, although comprehensive, is a manually-developed mechanism that covers a finite number of compounds. Another alternative without this limitation is the GECKO-A mechanism generation system (Aumont et al., 2005), which also has some online functionality (https://geckoa.lisa.u-pec.fr/, https://www2.acom.ucar.edu/modeling/gecko).

One advantage of using MechGen is its relatively user-friendly online interface that allows new users to easily examine predictions for single-step reactions or complete mechanisms for selected compounds without extensive training. It also offers terminal-based access that can be obtained using accounts created using the online system to use features of the system not available to web users, and users can download and install their own copies of this software to use full features, including resource-intensive operations like multi-generation mechanism derivation. This is discussed in the User Manual that is available at the MechGen web site (Carter, 2025). MechGen output can serve as input to the SAPRC box modeling software (Carter, 2024) to carry out model simulations with either minimally reduced detailed mechanisms or extended versions of the SAPRC-22 mechanism (Carter, 2023) with selected compounds represented explicitly. It can also be used for comparing its predictions with those of GECKO-A or incorporated in MCM to assess differences in predictions that indicate areas where additional experimental, theoretical, or SAR development work are needed.

One of the reasons that MechGen has not been as widely used as MCM or GECKO-A for applications other than SAPRC mechanism development is that it has not been described or documented in the peer-reviewed literature until now, including the continuous revisions that it has undergone. Most researchers have either not been aware of MechGen or have not had a stable version to cite if they do use it. An additional factor that may inhibit its use as an alternative to MCM is that MechGen outputs mechanisms in the format used by the SAPRC model simulation software, which is not as widely used as other modeling software systems. This paper, along with the recently published paper of Carter et al. (2025a), addresses its lack of adequate documentation. Future updates to MechGen will be documented and made available as separate versions to ensure reproducibility. We are developing software converting MechGen mechanism output to other formats and have already released a converter for the F0AM modeling system (Wolfe et al, 2016), which is available from the MechGen GitHub repository.

880 MechGen, along with GECKO-A and MCM, represent the current state of the science in
881 atmospheric reactions of organics in the lower troposphere, though they all require updates in some
882 respects to be consistent with recent advances in the continuously evolving science in this area. They all
883 incorporate different estimates or assumptions regarding some uncertain processes, though they also have
884 many assumptions and estimates in common that may not necessarily be correct. The assignments and
885 estimates incorporated in MechGen that need to be updated are discussed by Carter et al. (2025a).
886 Although GECKO-A is also continually being updated, it does not yet incorporate H-shift isomerization
887 (including autooxidation) reactions of peroxy radicals, which are important for many compounds such as
888 monoterpenes (Li et al., 2022). On the other hand, GECKO-A has a more detailed representation of
889 peroxy + peroxy and photolysis reactions. Comparing MechGen predictions with those of GECKO-A and
890 MCM can help identify areas where further experimental, theoretical, or SAR development work is most
891 needed, though this does not rule out the possibility that they all share assumptions or estimates for these
892 highly complex and uncertain mechanisms that may in fact be incorrect.

893 The atmospheric chemistry of organics and development of SARs or estimation methods
894 continues to be an active area of research, and MechGen and other mechanism generation systems need to
895 be periodically updated to continue to represent the state of the science and thus continue to be
896 appropriate tools for developing mechanisms for models. Updating the estimates for the many types of
897 chemical systems that need to be represented, testing the associated software changes, and evaluating the
898 predictions against available data, is a time-consuming process, making it essentially impossible for any
899 system to be completely up to date at any point of time. This is one reason why comparisons of
900 predictions of different mechanism generation systems or mechanisms are important. Carter et al. (2025a)
901 discuss the many types of estimates and points out the areas where updates are needed and planned.

902 MechGen is designed to accommodate updates to the underlying chemistry assignments and
903 SARs. Modifying chemical assignments or parameters used in existing SARs is straightforward and
904 discussed in the User Manual, but adding or deleting types of reactions or adding or changing the
905 structures of SARs will require programming changes. Since implementing such updates would result in
906 new versions of the system that will give different predictions, it is important that older versions be
907 archived and made available so that previous work that used these versions can be reproduced or
908 evaluated. This was not the case for previous versions of MechGen, but will be for newer versions going
909 forward.

910 The development and maintenance of MechGen has largely been the work of the primary author
911 (Carter), who is nearing full retirement. Unique features and strengths of MechGen (including its
912 integration with other SAPRC tools and models) have led to a growing base of MechGen users, and as
913 this user base has grown, so has the need for a collaborative team to carry it into the future. Steps to
914 ensure the sustainability of this resource thus far have included the publication of the chemical basis of
915 the MechGen system (Carter et al., 2025a), this paper describing the software system, the development of
916 a GitHub page (https://github.com/SAPRC/MechGen), publications describing the use of MechGen
917 beyond the derivation of SARPC mechanisms (Jiang et al., 2020; Li et al., 2022), and dissemination of
918 knowledge through the growing user community. In addition, the MechGen is written in a programming
919 language (MOO) that is no longer widely used or maintained, so conversion to a more actively supported
920 platform would enhance ongoing maintenance and collaborative development in the future. The
921 development of collaborative teams, including those working on the chemical mechanism estimates and
922 databases, as well as the software, will ensure that MechGen remains a valuable, adaptable, and robust
923 mechanism generation system well into the future.

SUPPLEMENTARY INFORMATION

The Supplementary Information published with this manuscript contains additional information and details on procedures discussed this paper. These include the following: additional output examples (Section S1); derivation of the standard environments currently available in MechGen (Section S2); additional details concerning procedures used (Section S3); the method used to estimate fractions of semi-volatile products in the condensed phase (Section S4); additional information regarding multi-generation mechanisms (Section S5); and additional information on the results with the representative compounds (Section S6).

USER MANUAL

A complete User Manual describing the operation of the MechGen system using both the web and terminal interfaces is available with the MechGen source code (see below) and at the MechGen web site (Carter, 2025) and MechGen GitHub page (https://github.com/SAPRC/MechGen/tree/master/docs). It discusses how to obtain, install, and configure the system on user's own computer. Updated versions of the user manual will be made available on the MechGen and GitHub sites when the source code or documentation has been updated. A quick start web user guide for new users is also available at the MechGen web site and the SAPRC MechGen GitHub page, where it will also be updated when appropriate.

CODE AND DATA AVAILABILITY

The MechGen source code (v1.1, last updated July, 25, 2025) and its documentation, including the full User Manual and Quick Start Guide, are available on Zenodo via Carter et al. (2025b) at https://zenodo.org/records/16622705. Detailed system information, including online access and further development, is provided on the MechGen website (https://mechgen.cert.ucr.edu/), while the GitHub repository at https://github.com/SAPRC/MechGen offers the latest software releases and installation resources, with updates archived on Zenodo.

All the source code and data used by MechGen are contained within a single MOO code database file as described in available MOO documentation (Curtis, 1997; Fox, 2004; Wikipedia, 2025). The program source code and the data can be viewed and modified by downloading and installing the MechGen database and the MOO server software as described in the User Manual. Once the system is installed and configured, the user can access the system using Telnet and log in as a systems programmer (see User Manual) and use MOO commands to view or modify the source code (Curtis, 1997; Fox, 2004).

The data used by MechGen to implement the chemical assignments and SARs are described by Carter et al. (2025a). The detailed parameters, mechanism assignments, and rate constants used are given in tables in the paper or its Supplementary Information.

ACKNOWLEDGEMENTS

REFERENCES

Aumont, B., Szopa, S., and Madronich, S.: Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach, Atmospheric Chemistry and Physics, 5, 2497–2517, https://doi.org/10.5194/acp-5-2497-2005, 2005.

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

Carter, W. P. L.: A detailed mechanism for the gas-phase atmospheric reactions of organic compounds, Atmospheric Environment. Part A. General Topics, 24, 481–518, https://doi.org/10.1016/0960-1686(90)90005-8, 1990.

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, 1994.

Carter, W. P. L.: Documentation of the SAPRC-99 Chemical Mechanism for VOC Reactivity Assessment. Available at https://doi.org/10.5281/zenodo.12600705. May 8, 2000.

Carter, W. P. L.: Development of the SAPRC-07 Chemical Mechanism, Atmospheric Environment, 44, 5324–5335, https://doi.org/10.1016/j.atmosenv.2010.01.026, 2010a.

Carter, W. P. L.: Development of the SAPRC-07 Chemical Mechanism and Updated Ozone Reactivity Scales. Available at https://doi.org/10.5281/zenodo.12601346. January 27, 2010b.

Carter, W. P. L.: Preliminary Documentation of the SAPRC-16 Mechanism. Available at: https://intra.engr.ucr.edu/~carter/SAPRC/16/S16doc.pdf. October 29, 2016.

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. and Atkinson, R.: Atmospheric chemistry of alkanes, Journal of Atmospheric Chemistry, 3, 377–405, https://doi.org/10.1007/BF00122525, 1985.

Carter, W. P. L. and Heo, G.: Development of Revised SAPRC Aromatics Mechanisms. Available at https://intra.engr.ucr.edu/~carter/SAPRC/saprc11.pdf. April 12, 2012.

Carter, W. P. L. and Heo, G.: Development of revised SAPRC aromatics mechanisms, Atmospheric Environment, 77, 404–414, https://doi.org/10.1016/j.atmosenv.2013.05.021, 2013.

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, 2025a.

Carter, W. P. L, Z. Wang, and J. Jiang, J.: SAPRC/MechGen: MechGenv1.1. https://zenodo.org/records/16622705, 2025b.

Curtis, P.: LambdaMOO Programmer's Manual, Available at https://lambda.moo.mud.org/pub/MOO/ProgrammersManual_toc.html. March 1997.

Ervens, B., Rickard, A., Aumont, B., Carter, W. P. L., McGillen, M., Mellouki, A., Orlando, J., Picquet-Varrault, B., Seakins, P., Stockwell, W. R., Vereecken, L., and Wallington, T. J.: Opinion: Challenges and needs of tropospheric chemical mechanism development, Atmos. Chem. Phys., 24, 13317–13339, https://doi.org/10.5194/acp-24-13317-2024, 2024.

Finlayson-Pitts, B. J. and Pitts Jr., J. N.: Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications, 1st ed., Elsevier, 1999.

Fox, K.: MOO-Cows FAQ, Available at https://www.moo.mud.org/moo-faq/. October 8, 2004.

Gao, C. W., Allen, J. W., Green, W. H., and West, R. H.: Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms, Computer Physics Communications, 203, 212–225, https://doi.org/10.1016/j.cpc.2016.02.013, 2016.

Green, W. H.: Chapter 5 - Automatic generation of reaction mechanisms, in: Computer Aided Chemical Engineering, vol. 45, edited by: Faravelli, T., Manenti, F., and Ranzi, E., Elsevier, 259–294, https://doi.org/10.1016/B978-0-444-64087-1.00005-X, 2019.

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol for mechanism development, Atmospheric Environment, 31, 81–104, https://doi.org/10.1016/S1352-2310(96)00105-7, 1997.

Jenkin, M. E., Saunders, S. M., Wagner, V., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds, Atmospheric Chemistry and Physics, 3, 181–193, https://doi.org/10.5194/acp-3-181-2003, 2003.

Jiang, J., Carter, W. P. L., Cocker, D. R., III, and Barsanti, K. C.: Development and Evaluation of a Detailed Mechanism for Gas-Phase Atmospheric Reactions of Furans, ACS EARTH AND SPACE CHEMISTRY, 4, 1254–1268, https://doi.org/10.1021/acsearthspacechem.0c00058, 2020.

Kaduwela, A., Luecken, D., Carter, W., and Derwent, R.: New directions: Atmospheric chemical mechanisms for the future, Atmospheric Environment, 122, 609–610, https://doi.org/10.1016/j.atmosenv.2015.10.031, 2015.

Kirchner, F.: The chemical mechanism generation programme CHEMATA—Part 1: The programme and first applications, Atmospheric Environment, 39, 1143–1159, https://doi.org/10.1016/j.atmosenv.2004.09.086, 2005.

Li, Q., Jiang, J., Afreh, I. K., Barsanti, K. C., and Cocker, D. R., III: Secondary organic aerosol formation from camphene oxidation: measurements and modeling, ATMOSPHERIC CHEMISTRY AND PHYSICS, 22, 3131–3147, https://doi.org/10.5194/acp-22-3131-2022, 2022.

Liu, M., Grinberg Dana, A., Johnson, M. S., Goldman, M. J., Jocher, A., Payne, A. M., Grambow, C. A., Han, K., Yee, N. W., Mazeau, E. J., Blondal, K., West, R. H., Goldsmith, C. F., and Green, W. H.: Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation, J. Chem. Inf. Model., 61, 2686–2696, https://doi.org/10.1021/acs.jcim.0c01480, 2021.

Postel, J. and Reynolds, J.: Telnet Protocol Specification. Available at https://datatracker.ietf.org/doc/html/rfc854, May 1983.

RMG: RMG-Reaction Mechanism Generator. Available at https://rmg.mit.edu. January 28, 2025.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmos. Chem. Phys., 3, 161–180, https://doi.org/10.5194/acp-3-161-2003, 2003.

Venecek, M. A., Cai, C., Kaduwela, A., Avise, J., Carter, W. P. L., and Kleeman, M. J.: Analysis of SAPRC16 chemical mechanism for ambient simulations, Atmospheric Environment, 192, 136–150, https://doi.org/10.1016/j.atmosenv.2018.08.039, 2018.

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.

Wikipedia: LambdaMOO. https://en.wikipedia.org/wiki/LambdaMOO, January 13, 2025.

Wolfe, G. M., M. R. Marvin, S. J. Roberts, K. R. Travis and J. Liao. Geoscientific Model Development, 9, 3309-3319, https://doi.org/10.5194/gmd-9-3309-2016, 2016.