The GRETOBAPE Gas-phase Reaction Network: The Importance of Being Exothermic

Lorenzo Tinacci[1,2] , Simón Ferrada-Chamorro[1] , Cecilia Ceccarelli[1] , Stefano Pantaleone[2,3] , Daniela Ascenzi[4] ,
Andrea Maranzana[2] , Nadia Balucani[1,3,5] , and Piero Ugliengo[2]
[1] Univ. Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France; Lorenzo.Tinacci@univ-grenoble-alpes.fr
[2] Dipartimento di Chimica and Nanostructured Interfaces and Surfaces (NIS) Centre, Università degli Studi di Torino, via P. Giuria 7, I-10125 Torino, Italy
[3] Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, I-06123 Perugia, Italy
[4] Dipartimento di Fisica, Università di Trento, I-38123 Povo, Italy
[5] Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy

Abstract

The gas-phase reaction networks are the backbone of astrochemical models. However, due to their complexity and nonlinear impact on the astrochemical modeling, they can be the first source of error in the simulations if incorrect reactions are present. Over time, following the increasing number of species detected, astrochemists have added new reactions, based on laboratory experiments and quantum mechanics (QM) computations, as well as reactions inferred by chemical intuition and the similarity principle. However, sometimes no verification of their feasibility in the interstellar conditions, namely their exothermicity, was performed. In this work, we present a new gas-phase reaction network, GRETOBAPE, based on the KIDA2014 network and updated with several reactions, cleaned from endothermic reactions not explicitly recognized as such. To this end, we characterized all the species in the GRETOBAPE network with accurate QM calculations. We found that ∼5% of the reactions in the original network are endothermic, although most of them are reported as barrierless. The reaction network of Si-bearing species is the most impacted by the endothermicity cleaning process. We also produced a cleaned reduced network, GRETOBAPE-RED, to be used to simulate astrochemical situations where only C-, O-, N-, and S-bearing species with less than six atoms are needed. Finally, the new GRETOBAPE network, its reduced version, and the database with all the molecular properties are made publicly available. The species property database can be used in the future to test the feasibility of possibly new reactions.

Unified Astronomy Thesaurus concepts: Astrochemistry (75); Interstellar molecules (849); Chemical reaction network models (2237)

1. Introduction

Since the discovery of the first molecules in the interstellar medium (ISM) in the visible spectrum (Dunham 1937; Swings & Rosenfeld 1937; McKellar 1940), the astrochemical community has been trying to understand their origin and how they can survive the harsh interstellar conditions. With the evolution of astronomical facilities and, more specifically, the advent of radio telescopes, new and more complex molecules have been detected in various astronomical objects with an almost constant discovery rate of approximately four to six new molecules per year (McGuire 2022). To date, more than 270 molecules, composed of 19 different elements, have been detected in the ISM and circumstellar medium, and this number is steadily increasing. The list of detected interstellar species can be found in the Cologne Database for Molecular Spectroscopy (CDMS; Endres et al. 2016),[6] in the McGuire (2022) census (and the related Python package[7]), and on The Astrochymist website.[8]

Understanding how interstellar molecules are formed, destroyed, and connected, in addition to being interesting per se, can help in inferring the properties, history, and evolution of the astronomical objects in which they have been detected. Thus, in parallel with the identification of molecules and the measurements of their abundances in different astronomical objects, soon a community dedicated to developing astrochemical models began to play an important role (Herbst & Klemperer 1973; Prasad & Huntress 1980). A crucial element of astrochemical models is the gas-phase reaction network, namely the lists of the gas-phase chemical reactions with their rate constants, which describe how efficient the reactions are as a function of the temperature, and product branching ratios.

The increasing number of detected species has prompted modelers and chemists to add new reactions to the existing networks in order to reproduce the observed species and their abundances. In this way, the numbers of reactions, reactants, and products have been increasing each time a new molecule has been detected, which also led to an increase in the network size and complexity. Nowadays, the publicly available astrochemical gas-phase networks, KIDA[9] (Wakelam et al. 2015) and UMIST[10] (McElroy et al. 2013), contain about 8000 reactions involving more than 500 species.

Unfortunately, a vast majority (⩾80%) of the rates and branching ratios of the reported gas-phase reactions have not been measured or computed and are often based on approximate estimates (e.g., using the capture theory; Su & Chesnavich 1982; Herbst 2006; Woon & Herbst 2009; Loison et al. 2013),

---

[6] https://cdms.astro.uni-koeln.de/classic/molecules
[7] https://github.com/bmcguir2/astromol
[8] http://www.astrochymist.org/astrochymist_ism.html

[9] https://kida.astrochem-tools.org/
[10] http://udfa.ajmarkwick.net/

educated guesses on similarity principles, or simple chemical intuition. In addition, even when some experimental measurements are available, the estimated rate constants may have substantial uncertainties, as they are often based on experiments at room temperature (e.g., Anicich 2003). The uncertainty on the reaction rates is a well-known problem in the astrochemistry community. In parallel with the efforts of studying reactions via experiments and theoretical computations, sensitivity analysis studies have provided the impact of these uncertainties on the abundances predicted by astrochemical models (e.g., Wakelam et al. 2006; Penteado et al. 2017).

One possible and mandatory first step to improve the reliability of the gas-phase reaction networks that has never been performed systematically on the astrochemical networks is to estimate the exo/endothermicity of the gas-phase reactions, using accurate physicochemical quantum mechanical (QM) data for each species. Indeed, given the ISM conditions (temperatures usually less than $\sim$100–200 K), strongly endothermic reactions are inhibited and therefore can be excluded from the astrochemical networks without complex and time-consuming reaction transition state studies or experiments (Smith 2006, 2011). This verification is now possible, thanks to the availability of two theoretical studies that together provide reliable QM data (electronic state, electron spin multiplicity, geometry, harmonic frequencies, absolute electronic energy, and dipole moment) for the totality of the species in the KIDA network. The first one, by Woon & Herbst (2009), computed the QM data for a large number of neutral molecules involved in the KIDA reaction network. The second one, by Tinacci et al. (2021), complemented the first one and provided QM data for all the ions and the remaining neutral species of the KIDA network. Please note that the two studies have been carried out at the same electronic level of theory, so that they can be reliably used for computing the exo/endothermicity of all the reactions. In addition, Tinacci et al. (2021) created a database[11] with the above QM data for all the cations.

The KIDA and UMIST networks contain such a large number of reactions because their goal is to reproduce the observed abundances of large molecules, which usually have low (between 10^{-8} and 10^{-12}) abundances. However, in some cases, such as astrochemical models coupled with hydrodynamical simulations of cloud collapse or protoplanetary disk formation and evolution, the goal is to reliably reproduce only the most abundant species. In these cases, the use of large reaction networks is not only useless but also detrimental because it demands spareable computing time. Indeed, several authors have proposed reaction network reduction techniques in the past few years, each using different criteria and focusing on specific goals (Oppenheimer & Dalgarno 1974; Nelson & Langer 1999; Glover et al. 2010; Grassi et al. 2013, 2022).

The goal of the present article is twofold. The first goal is to produce a gas-phase reaction network "cleaned" from endothermic reactions not reported as such. To this purpose, we perform a systematic study of the endo/exothermicity of all reactions present in a new network, GRETOBAPE, which is based on the KIDA2014 one and updated through the years by our group. In the process, we also verify that each species present in the network is not a sink and that the obvious reactions with the most abundant molecular cations are present.

The second goal is to produce a reduced network from the cleaned one, which only includes H-, C-, N-, O-, and S-bearing species with not more than six atoms, to be employed when the use of a complete network is unnecessary.

The article is organized as follows. Section 2 details the adopted tools to deal with the reaction network, the cleaning procedure, and the computational methodology used to characterize all the chemical species present in the network. In Section 3 we describe the new reaction networks, GRETOBAPE and GRETOBAPE-RED, resulting from the cleaning and reduction processes, respectively, as well as a list of the obvious reactions possibly missing in the network. In Section 4 we discuss the obtained results and their impact on astrochemical models of cold molecular clouds and warm molecular outflow shock simulations. Finally, in Section 5 we present our conclusions, and in Section 6 we provide the hyperlinks to the publicly available online databases and reaction networks produced in this work.

2. Methodology

2.1. Computational Details for QM Calculation

All the species present in the astrochemical reaction network were characterized via QM calculations to obtain their optimized structure, absolute electronic energy, and zero-point energy (ZPE) correction. Most of the neutral species structures and their corresponding electronic spin multiplicity (m_s) in the ground state, a fundamental piece of information for the electronic QM calculations, are taken from Woon & Herbst (2009), while for the cations the information is taken from our previous work (Tinacci et al. 2021). For the chemical species that have not been theoretically characterized yet, we inferred here the structure and m_s using the same approach introduced in Tinacci et al. (2021). The list of those species with their chemical data is reported in the supporting material.

In order to have all species characterized at the same level of theory, i.e., so as to have a consistent method, we used the DFT M06-2X (Zhao & Truhlar 2008) coupled with the triple-ζ Dunning's correlation consistent basis set (cc-pVTZ) (Kendall 1992; Woon & Dunning 1993) for geometry optimization and harmonic frequency calculations. The electronic absolute energy was refined at a CCSD(T) (Knowles et al. 1993, for closed-shell species) or RO-CCSD(T) (Watts et al. 1993, for open-shell species) level of theory, in conjunction with an augmented triple-ζ correlation consistent basis set (aug-cc-pVTZ; Kendall 1992). All the molecular structures and properties computed with the adopted methodology were benchmarked in our previous work (Tinacci et al. 2021). All calculations were carried out with the Gaussian16 program (Frisch et al. 2016), and we kept the default values set up in the program.

Due to the importance of having accurate reaction energies for our reaction network cleaning methodology, we report a benchmark of our resulting ZPE-corrected reaction energy with respect to other published studies in Table 1. All differences are smaller than 4 kJ mol^{-1}, which corroborates the accuracy of present results. Please note that the differences are mostly due to a different treatment of the spin contamination, which we improved in the present study.

We decided not to compute the reaction energies using the standard enthalpy of formation (ΔH_f°(T), usually given at 298.15 K or 0 K); instead, we used the absolute electronic

---

Table 1
ZPE-corrected Reaction Energy Difference, i.e., $\Delta(\Delta H(0))$, between Previous Literature and Our Data

Reaction		$\Delta(\Delta H(0))$ (kJ mol^{-1})	Their Level of Theory	Reference
Reactants	Products			
$C_2H_5 + O$	$C_2H_4 + OH$	1.26	CCSD(T)/aug-cc-pVTZ//B2PYLYP/aug-cc-pVTZ	Vazart et al. (2020)
$C_2H_5 + O$	$CH_3CHO + H$	1.67		
$C_2H_5 + O$	$H_2CO + CH_3$	1.23		
$CH + CH_3OH$	$C_2H_4 + OH$	0.36	CCSD(T)/aug-cc-pVTZ//B2PYLYP/aug-cc-pVTZ	Vazart et al. (2020)
$CH + CH_3OH$	$CH_3CHO + H$	0.77		
$CH + CH_3OH$	$H_2CO + CH_3$	0.34		
$SiH + S$	$SiS + H$	4.29	CCSD(T)/aug-cc-pVTZ//B3LYP/aug-cc-pVTZ	Rosi et al. (2018)
$SiH + S_2$	$SiS + HS$	-1.85		
$Si + SH$	$SiH + S$	-3.92	MRCI/CBS	Mota et al. (2021)

Note. The minor differences with respect to Vazart et al. (2020) and Rosi et al. (2018) can be ascribed to the different optimization levels and also the spin contamination for all open-shell species, which was corrected in our case by using the restricted-open formalism.

energy. This choice minimizes the propagation error in the QM computations, which would be larger if carried out using the formation enthalpy. One way to overcome this limitation would be to compute the $\Delta H_f^\circ(T)$ by the isodesmic and homodesmic cycles (Wheeler et al. 2009; Cramer 2013), which compensates for the errors. The problem with this technique is the difficulty of automatizing it. In the past, some attempts were carried out, but they cannot be generalized to all chemical species (e.g., Cavallotti 2022).

It is worth noticing that many astrochemical species, radicals and exotic species with respect to the terrestrial chemistry, lack laboratory-measured $\Delta H_f^\circ(T)$. Therefore, limiting the control on the exo/endothermicity of the reaction to only those having laboratory experimental data would greatly reduce our capacity to "clean" the astrochemical networks.

Overall, we characterized 542 species. The whole list is publicly available at the ACO (Astro-Chemical Origins) project website.[12] In addition, a data frame (in `.csv` format) with all the properties of the studied species is available in the supporting material of this article (see Table 9).

2.2. Graph Theory for Reaction Networks

In the astrochemical community, the reaction network is usually encoded and manipulated with a matrix approach. In this work, we adopt an approach based on the *graph theory*, so that the network manipulation can be easily obtained via the NETWORKX (Hagberg et al. 2008) Python package (e.g., node connectivity, centrality, sinks, and other characteristics). Instead of using a multi-di-graph approach as done by Grassi et al. (2013), the mathematical object we used to describe the reaction network is the colored directed graph (also known as colored di-graph). Here each node has a color attribute and the connections are directed (Barabási 2013). A colored node can have different attributes depending on the color: in our case, it can be a species node or a rate constant node. In addition, in a directed graph, the edges have a direction, i.e., the adjacency (or connectivity) matrix is not symmetrical, so that the connection between nodes is oriented (e.g., Barabási 2013; Newman 2018). In Figure 1, we report the codification of a generic reaction network given in the abovementioned three

ways: the simple reaction list, the reaction network visualization, and, finally, the corresponding adjacency matrix of the colored directed graph. The Python scripts used to encode and characterize the network properties are publicly available in a GitHub repository[13] and in the supporting material (Section 6).

2.3. Network Cleaning Procedure

2.3.1. Identification of Endothermic Reactions

As explained in the Introduction, the first goal of the present work is to produce a network "cleaned" from endothermic reactions that have not been reported as such (see below), since they cannot occur in the cold ISM. To this end, we computed the reaction enthalpy ΔH of the neutral–neutral, ion–neutral, and ion–ion reactions present in the reaction network. Note that we did not consider photo- and electron-induced reactions (involving electrons, cosmic-ray particles, and UV radiation) because they entail strongly energetic processes, whose products are usually not in their fundamental electronic state.

In the ISM, molecules primarily reside in their ground electronic state and lowest vibrational level, with only the rotational levels populated, depending on the gas conditions. Since the rotational energy has a negligible contribution to the total reaction energy, we used the electronic energy and the vibrational energy at 0 K (also called ZPE correction) to evaluate the reaction enthalpy $\Delta H(0)$ of each reaction present in the reaction network. Then, an endothermic reaction has $\Delta H(0)$ larger than zero.

In addition, in the astrochemical networks the rate constants of the neutral–neutral and ion–neutral (but not ion–polar) reactions are reported following the modified Arrhenius equation (Kooij 1893; Laidler 1996):

$$k(T) = \alpha \left(\frac{T}{300 \text{ K}} \right)^\beta \exp\left(-\frac{\gamma}{T} \right), \tag{1}$$

where T is the gas temperature in kelvins and α, β, and γ are often derived as fitted parameters of either experimental or theoretical values of $k(T)$. It should be emphasized that γ does not always have an actual physical meaning. In other words, γ is often not a real activation barrier. That said, it can be safely

---

[12] https://aco-itn.oapd.inaf.it/home

[13] https://github.com/TinacciL/GreToBaPe_Cleaning

Reaction Mechanism **Reaction Network** **Adjacency matrix**
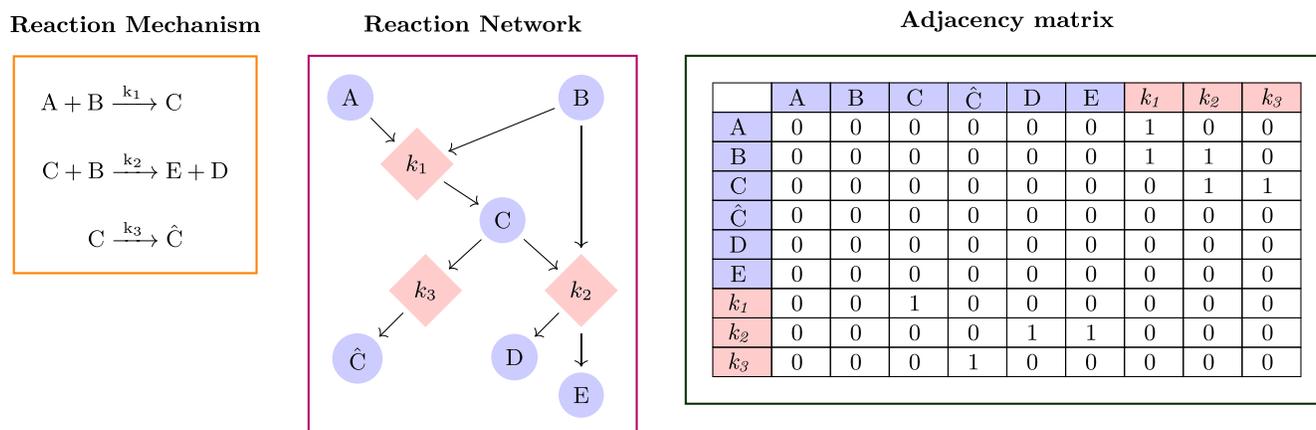
Figure 1. Three different ways to visualize and work with a reaction network: reaction list (left panel); colored directed graph (or colored di-graph), where circles correspond to chemical species, while squares indicate rate constants (middle panel); and nonsymmetric adjacency (or connectivity) matrix representation of the colored di-graph (right panel).

assumed that if the reaction has a barrier, it is lower than or equal to γ. Therefore, in order to consider that a reaction has been incorrectly inserted in the reaction network, we apply the additional criterion that $\Delta H(0) > \gamma R$, where γ is multiplied by the constant R to have it in kilojoules per mole.

Finally, our computations have an intrinsic uncertainty, i.e., the accuracy of the electronic structure QM calculations, which we conservatively evaluate to be $10\ \mathrm{kJ\ mol^{-1}}$. Therefore, when retaining or excluding a reaction from the network, we consider a $10\ \mathrm{kJ\ mol^{-1}}$ threshold when γ is both a real activation barrier and the result of an analytical fit.

In summary, a reaction is considered *correctly* included in the network if it satisfies the following criteria:

1. if the reaction is not encoded with the modified Arrhenius equation (i.e., the ion−pol formula for barrierless reactions), $\Delta H(0) \leqslant 10\ \mathrm{kJ\ mol^{-1}}$; and
2. if the reaction is encoded with the modified Arrhenius equation, $\Delta H(0) \leqslant \gamma R + 10\ \mathrm{kJ\ mol^{-1}}$.

Reactions that do not satisfy the two above criteria are excluded.

2.3.2. The Domino Effect of Removing Endothermic Reactions

In order for the network to be meaningful, each species must have at least one destruction route (i.e., loss) and one formation route (i.e., production).

Removing reactions from the network can affect other species if they are not formed or destroyed through other reactions than those removed. In other words, if a species becomes a source (species with no formation routes) or a sink (species with no destruction routes), it is also removed from the network. As a consequence, the reactions involving the removed species are also removed.

Therefore, the removal of one or more initial reactions triggers a domino effect, causing the removal of other reactions and species. Our graph theory approach, described in Section 2.2, allows us to easily follow the domino effect and identify the reactions and species to be removed.

2.4. Network Reduction Procedure

As stated in the Introduction, the second goal of this work is to produce a reduced network from the cleaned one, which only includes H-, C-, N-, O-, and S-bearing species with at most six atoms, to be used when the larger network is unnecessary. To this end, we used the following criteria and steps:

1. We only consider C-, N-, O-, and S-bearing species, plus Fe and Mg atoms, because their ions can be important positive charge carriers.
2. We only consider molecules with six or less atoms (either reactant or product).
3. The removal of larger molecules may cause the domino effect, explained in Section 2.3.2, which can lead to the removal of other reactions and species also with less than six atoms.
4. Methanol is the only exception to rules 2 and 3. In the gas phase, it is formed by reactions involving molecules with more than six atoms, so it should be removed, based on rule 3. However, methanol is an important component of the interstellar grain mantles, a major carbon carrier, believed to be mainly formed on the grain surfaces and released into the gas phase by thermal and nonthermal effects. In practice, if the model contains surface reactions that form methanol, it could be injected into the gas phase by, e.g., the sublimation or sputtering of the grain mantles. Once in the gas phase, methanol can react with species having less than six atoms (HCO^+, H_3^+, OH, etc.). For this reason, we keep methanol in the reduced network. If no methanol is produced by surface chemistry, then its abundance would be zero because no gas-phase reaction would form it so that it is not a source or a sink.
5. We remove all anions. We had two reasons for that. The first one is that, usually, they are not crucial species in studies that necessitate a reduced network. The second one is that there is still a debate on their rate of formation (e.g., Lara-Moreno et al. 2019b, 2019a).

3. Results

3.1. Original Astrochemical Network

The reaction network from which we start is based on the KIDA2014 network (Wakelam et al. 2015), updated following studies from our and other groups, as follows: Hoyermann et al. (1981), DeFrees et al. (1985), Meot-Ner et al. (1986),
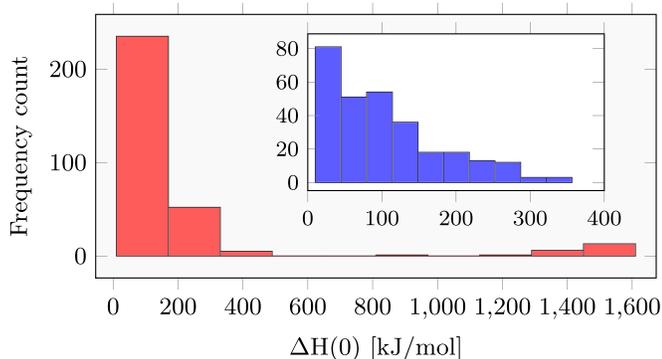
Figure 2. Histogram of the enthalpy $\Delta H(0)$ (red) of the endothermic reactions removed from the original network after the cleaning process. The inset (blue) shows the zoom-in of the 10–400 kJ mol^{-1} interval. The bin width is computed following the Freedman–Diaconis estimator (Freedman & Diaconis 1981).

Dóbé et al. (1991), Anicich (2003), Hamberg et al. (2010), Cheikh (2012), Lawson et al. (2012), Fournier (2014), Loison et al. (2014), Balucani et al. (2015), Neufeld et al. (2015), Vazart et al. (2016), Loison et al. (2016), Urso et al. (2016), Codella et al. (2017), Skouteris et al. (2017), Fontani et al. (2017), Balucani et al. (2018), Gao et al. (2018), Rosi et al. (2018), Skouteris et al. (2018), Sleiman et al. (2018), Ascenzi et al. (2019), Ayouz et al. (2019), Lara-Moreno et al. (2019a), Skouteris et al. (2019), Urso et al. (2019), Balucani (2020), Codella et al. (2020), Vazart et al. (2020), Blázquez et al. (2020), Mancini & Aragão (2021), and data, not yet published, courteously provided by Luca Mancini on some P-bearing reactions. We also add the recombination reactions of F^+ and P^+ with free electrons, which are not present in the KIDA astrochemical reaction network, although the second one is in the UMIST database. These data are taken from the database of Badnell (2006) and given in the usual Arrhenius–Kooij format (Equation (1)).

In summary, the precleaned original network comprises 499 species and 7240 reactions. The list is reported in the supporting material (Section 6: file GRETOBAPE-PRE.DAT).

Note that only 5768 of the 7240 reactions have thermodynamic data. The remaining 1472 reactions do not because they either contain species for which the electronic structures are not available (236 reactions[14]) or are recombination (640 reactions) or cosmic-ray particles and UV radiation (596 reactions).

3.2. New Cleaned Network: GRETOBAPE

In this section, we introduce the new reaction network obtained applying the criteria of Section 2.3 to the network just described in the previous section. It is important to emphasize that we only tested the endothermicity of the original network, which is based on the products listed in KIDA and the works cited above. We did not try to correct reactions a posteriori comparing the products in other astrochemical networks, notably UMIST (but see Section 3.2.3), where sometimes exothermic products of the reactions exist, for the reasons that will become clearer at the end of this section.

Table 2
List of the Species Removed from the Original Network after the Cleaning Process

Species		
SiC_4H	SiC_6H	SiC_8H
SiC_2CH_3	SiC_3H_5	H_2CSiCH
$H_2C_6N^+$	$H_2C_8N^+$	$H_2C_{10}N^+$
PNH_3^+	C_4H_5	

3.2.1. Overview of the Removed Species and Reactions

Among the 5768 reactions that have thermodynamic data (see Section 3.1) and for which we could carry out the endothermicity verification, 306 do not satisfy the two criteria described in Section 2.3, namely ~5% of the studied reactions.[15] Figure 2 shows the distribution of their endothermic energy. The majority of these reactions have endothermicities lower than about 150 kJ mol^{-1}, while about 140 reactions have values even higher.

The list of the removed reactions based on the endothermicity criteria is reported in the supporting material (Section 6: file GRETOBAPE-ENDO.DAT). In addition, we supply the list of 47 reactions that are endothermic but with a reaction enthalpy lower than 10 kJ mol^{-1} and for which more accurate calculations should be carried out before removing them for the network (Section 6: file GRETOBAPE-ENDO-0-10KJMOL. DAT). Finally, because of the domino effect, described in Section 2.3.2, 11 species and 23 additional reactions are removed, leading to a total of 329 removed reactions (Section 6: files GRETOBAPE-ENDO.DAT plus GRETOBAPE-DOMINO.DAT).

The full list of deleted species is reported in Table 2. Table 3 provides an overview of the removed reactions, grouped according to each reaction class. The ion–neutral reactions are the most affected by the cleaning process in number (140), while the cation–anion reactions are the most affected in percentage (7.7%). Table 4 lists the number of species in the final cleaned network classified with their elements.

In summary, the new cleaned network GRETOBAPE contains 6911 reactions and 488 species.

3.2.2. Description of the Removed Reactions

The 329 removed reactions have the following properties:

1. A total of 65 reactions are not encoded with the modified Arrhenius equation: either they do not satisfy criterion 1 listed in Section 2.3.1, namely $\Delta H(0) \leqslant 10$ kJ mol^{-1}, or they are removed because of the domino effect.
2. A total of 219 reactions are encoded via the Arrhenius–Kooji equation and have $\gamma = 0$: either they do not satisfy criterion 2 listed in Section 2.3.1, namely $\Delta H(0) \leqslant 10$ kJ mol^{-1}, or they are removed because of the domino effect.

---

[14] Reactions involving the species Fe, Fe^+, $C_2H_7^+$, $C_9H_3N^+$, or bimolecular reactions having as a product an electron or photon.

[15] Please note that the reaction $H + O^+ \rightarrow H^+ + O$ is endothermic by 17 kJ mol^{-1} according to our computations. However, the reaction has been studied by other authors (e.g., Stancil et al. 1999) and found to be exothermic. The reason for our erroneous result is that the ionization energy of O is a critical case that pushes the computational calculations to their limits and a difficult case for QM calculations owing to the correlation errors even if we are using one of the best available methods.

Table 3
Summary of the Number of Reactions in the Original and Final GRETOBAPE Networks

	Neutral–Neutral	Ion–Neutral	Cation–Anion	Recombination	Unimolecular	Total
Precleaning	1007	3596	1401	640	596	7240
Endothermic	58	140	108	⋯	⋯	306
Domino effect	2	⋯	⋯	9	12	23
Postcleaning	947	3456	1293	631	584	6911

Note. The reactions are listed according to their class. Each row presents the number of reactions after each step of the cleaning procedure (Section 2.3). "*Precleaning*": original network (Section 3.1). "*Endothermic*": reactions removed because they are endothermic with $\Delta H(0) > 10$ kJ mol^{-1}. "*Domino effect*": reactions removed because of the domino effect (Section 2.3.2). "*Postcleaning*": final cleaned network.

3. A total of 45 reactions are encoded via the Arrhenius–Kooji equation and have $\gamma \neq 0$: either they do not satisfy criterion 2 listed in Section 2.3.1, namely $\Delta H (0) \leqslant \gamma R + 10$ kJ mol^{-1}, or they are removed because of the domino effect.

Overall, the cleaning process strongly affects the silicon chemistry, as will be shown by the modeling simulations of Section 4.1. Indeed, out of a total number of 329 removed reactions, 48 involve Si-bearing species (GRETOBAPE-DOM-INO-ENDO-SI). In practice, about 15% of the total removed reactions regard Si-bearing species, and about 9% of the original reactions involving Si-bearing species (526) are removed. Likewise, 6 out of 11 removed species are Si-bearing species (see Table 2), which represents ∼10% of the total original Si-bearing species (61).

Figure 3 shows a graph representation of the reactions involving Si-bearing species and present in the original network, together with the removed and detected species. Figure 4 presents the same reaction network showing the deleted reactions only. First, the six removed Si-bearing species (all containing more than six atoms) are only formed and deleted by one reaction, so that their removal did not introduce a domino effect on other species. As can be seen from Figures 3 and 4, these species are in fact at the border of the reaction network, meaning that they have only a minor role in the overall Si reaction network. Second, all the removed reactions involving Si-bearing species have endothermic formation routes.

Digging into the KIDA database to understand their literature origin, we find that almost all these reactions are introduced not because of specific experiments or calculations but based on educated guesses. Specifically, their rate constants are estimated via the capture theory approach described in Herbst (2006) and Woon & Herbst (2009) for ion–neutral reactions. For example, the formation route of SiC$_4$H, SiC$_6$H, and SiC$_8$H is assumed to be Si + C$_x$H$_2$ → SiC$_x$H + H, and their constant rates are estimated.

Furthermore, looking in detail at the deleted reactions involving Si-bearing species (Figure 4), most of them are reactions that increase the complexity of the molecules, namely, they increase their number of atoms. This is a general characteristic of the cleaning process: the most affected class of reactions are those that increase the number of atoms of at least one reactant rather than those that reduce it. This result can be explained by the fact that the excess of energy of "destructive" reactions is generally much easier to guess than that of formation reactions.

Another source of endothermic reactions comes from the cation–anion reactions that involve carbon chain anions, whose data were predominantly taken from Harada & Herbst (2008).

Final remarks. It is of paramount importance to emphasize that the network GRETOBAPE is the result of the removal of endothermic reaction channels. Consequently, some loss/production channels have disappeared from the network, some of which are essential for the production of detected interstellar species. The list of removed reactions needs, therefore, a dedicated systematic (and time-consuming) revision, as will be shown in Section 3.2.3.

3.2.3. Comparison with the UMIST Database

As already emphasized at the beginning of Section 3.2, the original network on which we tested the endothermicity of the reactions is based on the KIDA2014 network, and therefore the 329 reaction channels removed refer to the KIDA database. In this subsection, we discuss some deleted reactions in comparison with the data reported in UMIST, in which sometimes a more detailed bibliography is reported. It is also worth mentioning that our scope here is not to substitute the precious work done by database maintainers, but rather to highlight some identified problems.

Reaction Si + C$_2$H$_2$ → SiC$_2$H + H. UMIST reports two products, SiC$_2$ + H$_2$ and SiC$_2$H + H, with the second one three times faster than the first. They assumed this branching ratio based on the study at 15 K by Canosa et al. (2001) of the global rate and based on Kaiser & Gu (2009) for the products. However, the analysis by Kaiser & Gu (2009) shows that the SiC$_2$H + H is indeed endothermic.

Reaction S$^+$ + CH$_4$ → H$_3$CS$^+$ + H. This reaction is crucial for the formation of the organo-sulfur products, and we deleted it since it is endothermic, despite the fact that it is considered as barrierless in KIDA and UMIST (with the same parameters). UMIST reports the reference of the experimental data by Smith et al. (1981). These authors assumed that H$_3$CS$^+$ is formed, which is impossible because S$^+$ is a quadruplet while H$_3$CS$^+$ is a triplet, and it would be impossible energetically for the spin conservation rules. Indeed, Yu et al. (2020) found that the corrected product seen by Smith et al. (1981) is H$_2$CSH$^+$ (singlet) and not H$_3$CS$^+$, as corroborated by our calculations that found the channel producing H$_2$CSH$^+$ to be exothermic. Please note that to reintroduce the reaction with the correct reaction channel (H$_2$CSH$^+$) would require revising all the reactions involving H$_3$CS$^+$ and H$_2$CSH$^+$, which is beyond the scope of the present work.

Reaction CN$^+$ + CO$_2$ → NO + C$_2$O$^+$. Experiments sometimes provided wrong products, as in the case of the channel CN$^+$ + CO$_2$ → NO + C$_2$O$^+$, which is reported in both KIDA

Table 4
Summary of the 488 Species Contained in the Cleaned Network GRETOBAPE, Together with the Elements Composing Them

Element	Number of Species Containing n Times the Element											Total
	1	2	3	4	5	6	7	8	9	10	11	
H	123	81	53	38	21	12	7	1	0	0	0	336
He	3	0	0	0	0	0	0	0	0	0	0	336
C	100	66	57	34	21	19	20	17	19	12	2	367
N	115	8	0	0	0	0	0	0	0	0	0	123
O	93	24	0	0	0	0	0	0	0	0	0	117
F	7	0	0	0	0	0	0	0	0	0	0	7
Na	6	0	0	0	0	0	0	0	0	0	0	6
Mg	3	0	0	0	0	0	0	0	0	0	0	3
Si	53	0	0	0	0	0	0	0	0	0	0	53
S	38	7	0	0	0	0	0	0	0	0	0	45
P	38	0	0	0	0	0	0	0	0	0	0	38
Cl	10	0	0	0	0	0	0	0	0	0	0	10
Fe	2	0	0	0	0	0	0	0	0	0	0	2

and UMIST and taken from the experimental work by Raksit et al. (1984). However, this channel is endothermic by more than 138 kJ mol^{-1}, as found by our computations (which are also validated by thermochemical data[16]) considering either of the two linear isomers, CCO^+ and COC^+. In this case, therefore, the problem is due not to the possible isomer structure but to the experiment itself. Indeed, another experiment by McEwan et al. (1983) did not find the NO + C_2O^+ channel but two other channels, also found by Raksit et al. (1984), i.e., CN + CO_2+ and CO + NCO^+, both of them exothermic.

3.2.4. Missing Destruction Reactions

In cold molecular clouds, with few exceptions, neutral species are mainly destroyed by the most abundant cations, namely H_3^+, HCO^+, H_3O^+, He^+, and H^+. Therefore, every neutral species should have at least one destruction channel for each of the above cations. Table 5 lists all the neutral species having no destruction channels with the aforementioned cations in the cleaned reaction network. Please note that some reaction channels were removed during the cleaning process. We mark them with a filled circle. The Table 5 list is meant to identify the reactions that need experiments and/or QM computations. In addition, our database of absolute electronic energies, ZPE, and m_s (Tinacci et al. 2021) can be used to test the possible reaction channels of formation and/or destruction routes.

As an illustrative example, it is worth mentioning the destruction reactions involving SiO. The reaction of SiO with $H_3O^+ \rightarrow H_2O + HSiO^+$ has an enthalpy of +152.1 kJ mol^{-1}. A similar endothermicity is found for the reaction SiO + $HCO^+ \rightarrow CO + HSiO^+$. However, if the product of both reactions is not $HSiO^+$, as reported in the KIDA database, but its constitutional isomer $SiOH^+$, as reported in UMIST, the reactions become respectively exothermic of -112.2 and -208.7 kJ mol^{-1}. However, since in KIDA $HSiO^+$ is produced from H_2SiO, which is in turn produced from H_3SiO^+, and $HSiO^+$ is involved in a dozen reactions (see Figure 3), a simple editing may lead to wrong results. A more detailed study on the reliability of the $SiOH^+$ isomer as a product is postponed to a future work by our group focusing on the improvement of the silicon reaction network (clearly necessary, as highlighted in Section 3.2.2).

---

[16] https://atct.anl.gov/

3.2.5. Missing Formation Reactions

The aim of the present work is to provide a reaction network as reliable as possible. Identifying important missing reactions is, therefore, an aspect of this goal. In addition to the missing destruction reactions described in the previous subsection, here we list obvious missing formation reactions, namely those forming detected species. To this scope, we used the recent McGuire (2022) census of the detected species in the ISM and extragalactic medium. Table 6 reports the detected species that do not have a formation route in our cleaned network GRETOBAPE. Most of these molecules are also missing from commonly used databases (e.g., KIDA). Hopefully, Table 6 can be used to motivate future works and strengthen the interdisciplinary collaboration between chemists and astronomers.

3.3. Reduced Network: GRETOBAPE-RED

Following the criteria described in Section 2.4, the reduced network GRETOBAPE-RED contains 204 chemical species from the 488 existent in the (cleaned) network GRETOBAPE. In addition, the removal of reactions associated with pruned species reduces the size of the network from 7240 reactions to 2810. Compared to GRETOBAPE, GRETOBAPE-RED has less than half the amount of reactions and species, implying a significant decrease in computation time when used in any chemical code. This performance boost is seen in the tests we performed on the reduced network using the KROME package (see Section 4), where the computing time of a typical gas-phase run decreases from 14 to 3 s using the complete and the reduced networks, respectively.

Due to the criteria imposed on the creation of this network, the main discrepancy we expect between the complete and the reduced networks concerns the chemistry of medium-sized (between three and six atoms) carbon chains. This is due to removing potentially important destruction routes, i.e., reactions between these medium-sized carbon chains, to form larger ones when pruning out C-bearing species with more than six atoms. Nonetheless, a complete assessment of the reliability of this reduced network is discussed in the following section, following detailed modeling of a cold case and a warm case, respectively.

4. Astrochemical Implications

This section illustrates the impact of the new reaction networks on the predicted species abundances, obtained using

Figure 3. Original Si-bearing reaction network, visualized as a directed colored graph. The reactions are represented as solid lines and the species as nodes (differently from the representation described in Section 2.2, where reactions and species are both represented by nodes). The lines are colored based on the $\log_{10}$ of the reaction rate constants calculated at 90 K and multiplied by the reactant densities at 5×10^3 yr in the "warm molecular outflow shock" model described in Section 4.1.2, using the original (precleaning) network. All rates larger than 10^{-10} are red, and those less than 10^{-25} are violet. The nodes are colored based on the graph distance of that species from Si^+ (highlighted in black). The species removed from the original network are marked by red crosses, while those detected in the ISM are marked by blue squares. Please note that the digital format of the figure is vectorial and can be zoomed in to better look at the details.

the original reaction network versus the new cleaned GRETO-BAPE and the reduced GRETOBAPE-RED networks, respectively. To this end, we run two time-dependent gas-only astrochemical codes: MYNAHOON, a modified version of the publicly available code NAHOON (Wakelam et al. 2012),[17] and KROME, a publicly available package (Grassi et al. 2014). MYNAHOON is used to evaluate the difference in the species abundances predicted with the original and new GRETOBAPE reaction networks in a typical cold molecular cloud (Section 4.1.1) and a typical warm molecular outflow shock (Section 4.1.2),

respectively. KROME is used to test the reliability of the reduced network GRETOBAPE-RED (Section 4.2) in a large parameter space.

4.1. Predicted Abundances with the Original versus the Cleaned Reaction Network GRETOBAPE

In order to assess the difference between the obtained clean reaction network GRETOBAPE and the original one, we performed two simulations: the first one simulates a typical cold molecular cloud (Section 4.1.1), and the second one simulates a typical warm molecular outflow shock (Section 4.1.2).

---

[17] Briefly, the modifications concern a more friendly usage of the code, in input and output, and not the core of the code itself.

Figure 4. Same as Figure 3, but showing only the reactions removed from the cleaning process.

4.1.1. Cold Molecular Cloud

We adopted the parameters of a typical molecular cloud, namely temperature (gas and dust) equal to 10 K, H nuclei number density equal to 2×10^4 cm^{-3}, visual extinction A_v equal to 20 mag, and cosmic-ray ionization rate for H$_2$ ζ equal to 3×10^{-17} s^{-1}. The initial elemental abundances are as follows (from Jenkins 2009, with the heavy metals depleted by a factor 100): He/H = 9.0×10^{-2}, C/H = 2.0×10^{-4}, O/H = 2.6×10^{-4}, N/H = 6.2×10^{-5}, S/H = 8×10^{-8}, Si/H = 8×10^{-9}, Fe/H = 3×10^{-9}, Mg/H = 7×10^{-9}. Finally, we started with the conditions of a pseudotranslucent cloud, namely with hydrogen in the molecular form and the other elements in atomic (ionic or neutral) form, and we let the chemistry evolve up to 1×10^7 yr.

Figure 5 shows the ratio of the predicted abundances obtained with the cleaned GRETOBAPE and original reaction networks as a function of time, for the species where this ratio is larger than 3 or smaller than 0.3 at any time $\geqslant 10^3$ yr of the simulation and whose abundances are larger than 10^{-13} at steady state. The figure also shows the abundances of these species, to further appreciate the impact. Table 7 reports the minimum and maximum ratios in the time intervals 10^4–10^5 yr, 10^5–10^6 yr, and at steady state.

Overall, 26 species are significantly affected by the reaction network cleaning process, meaning that their abundances have changed by more than a factor three during the evolution using the two reaction networks. The affected species represent only 5% of the species in the clean network GRETOBAPE, which is a reassuring result.

The species impacted by the reaction network cleaning process can be divided into three groups, where we only consider those whose abundance is larger than 10^{-13} during the evolution:

Table 5
List of Neutral Species with Missing Reactions with the Most Abundant Cations in the Cleaned Network GRETOBAPE

Species	H^+	He^+	H_3^+	H_3O^+	HCO^+
NH				★	
OH				★	
HF	★			★	★
NaH				★	
C_2				★	
MgH				★	
CN					★
CO	★			★	★
N_2	★			★	
NO				★	★
PH				★	
O_2				★	
HS				★	
HCl				★	★
SiN				★	★
SiO				•	•
NS				★	
PO				★	
CCl	★		★	★	★
ClO	★		★	★	★
SO				★	
SiS				★	
S_2				•	
H_2O				★	
HCO				★	
HNO				★	
PH_2				★	
O_2H	★	★	★	★	★
CCN				★	
NaOH				★	
CCO				•	
OCN	★		★	★	★
HNSi				•	
CO_2				★	
N_2O				★	•
HCS				★	★
NO_2	★	★		★	★

Species	H^+	He^+	H_3^+	H_3O^+	HCO^+
HPO				★	
OCS				★	
SiO_2	★			★	
SO_2	★			★	★
CH_3				★	
H_2CN	★	★	★	★	★
SiH_3				★	★
HOOH	★	★	★	★	★
$SiCH_2$				★	
H_2CS				★	
C_3N	★				
$l\text{-}SiC_3$				★	
C_3P	★				
CH_4				★	
CH_2NH				★	
CH_3O	★			★	
SiH_4				★	
H_2CCN				★	★
NH_2CN				★	
CH_2PH				•	•
HCOOH				★	
c-HCCHSi				★	
C_5				★	
C_4N	★	★		★	
SiC_3H				★	
SiC_4				★	
HCCCHO	★	★			
$c\text{-}C_3H_2O$	★	★			
H_2CCCO	★	★		•	
C_5H				★	
HC_4N	★		★	★	★
C_6				★	
C_5N	★			★	
C_5O	★	★		★	
C_2H_5				★	
CH_3NH_2				★	
CH_2CCH_2	★	★	★	★	★

Species	H^+	He^+	H_3^+	H_3O^+	HCO^+
CH_2CHCN				★	
C_6H				★	
HC_5N				★	
C_7				★	
C_6N	★	★		★	
C_2H_6				★	
C_3H_5	★	★	★	★	★
CH_3OCH_2	★		★		
CH_2CHC_2H	★	★	★	★	★
CH_3C_3N				★	
C_7H				★	
HC_6N	★		★	★	★
C_8				★	
C_7N	★			★	
C_7O	★	★		★	
CH_3CHCH_2	★	★	★	★	★
CH_3C_4H				★	
C_8H				★	
HC_7N				★	
C_9				★	
C_8N	★	★		★	
C_3H_7	★	★	★	★	★
$CH_2CHCHCH_2$	★	★	★	★	★
CH_3C_5N				★	
C_9H				★	
HC_8N	★		★	★	★
C_{10}				★	
C_9N	★			★	
C_9O	★	★		★	
C_3H_8	★	★	★	★	★
CH_3C_6H				★	
$C_{10}H$				★	
HC_9N				★	
C_{11}	★		★	★	★
$C_{10}N$	★	★		★	
$c\text{-}C_6H_6$	★				
CH_3C_7N				★	

Note. With the filled circles and filled stars we identify the neutral species reaction with the corresponding cation that was present in our network but was deleted in the cleaning process and the reaction that was originally absent, respectively.

Table 6
Species with Reported Observations in the ISM (McGuire 2022) That Are Not Present in the Final GRETOBAPE Network

Carbon-, Nitrogen-, and Oxygen-bearing Species					
4 atoms					
HCCO	HONO	HNCN	CNCN	HC_2N	
5 atoms					
HNCNH	NH_2OH	H_2NCO^+	$NCCNH^+$		
6 atoms					
HNCHCN	CH_3NC	C_5N^-			
7 atoms					
$c\text{-}C_2H_4O$	$c\text{-}C_3HCCH$	CH_2CHOH	HC_4NC	CH_3NCO	$HOCH_2CN$	HC_5O
8 atoms					
CH_2CHCHO	CH_3CHNH	NH_2CH_2CN	NH_2CONH_2		
9 atoms					
HC_7O	$H_2CCCHCCH$	$HCCCHCHCN$	H_2CCHC_3N	CH_3NHCHO	CH_3CONH_2
10 atoms					
$HOCH_2CH_2OH$	CH_3CH_2CHO	CH_3CHCH_2O	CH_3OCH_2OH		
11 atoms					
C_2H_5OCHO	CH_3COOCH_3	CH_3COCH_2OH	$NH_2CH_2CH_2OH$		
12 atoms					
$n\text{-}C_3H_7CN$	$i\text{-}C_3H_7CN$	$1\text{-}C_5H_5CN$	$2\text{-}C_5H_5CN$		
>13 atoms					
$HC_{11}N$	$c\text{-}C_6H_5CN$	C_9H_8	$1\text{-}C_{10}H_7CN$	$2\text{-}C_{10}H_7CN$	
Sulfur-bearing Species					
HSC	NCS	HNCS	HCCS	CHOSH	HCSCN	HCSCCH
H_2CCS	H_2CCCS	CH_3SH	C_5S	CH_3CH_2SH	
Silicon-bearing Species					
SiCN	SiCSi	SiC_3	SiH_3CN	CH_3SiH_3	

Note. Please note that different isomers of the listed species may be present in KIDA and UDfA.

1. The first group (top panels of Figure 5) contains species where the post- over pre-cleaning ratio is larger than 10^3 or lower than 10^{-3} during the evolution: HCSi, SiO^+, $CHSi^+$, and HCCSi. None of these species have been detected in space. Note that five of these six species contain Si and all of them have an abundance lower than 2×10^{-11}. The most abundant one is SiO^+, with a predicted steady-state abundance of 8×10^{-12}, 3×10^3 larger with respect to the original reaction network. The species whose ratio is the largest at steady state is $CHSi^+$, when its abundances is lower than 10^{-13}. HCCSi presents a ratio of 0.3 and an abundance of 5×10^{-13} at steady state, while at $\sim 2 \times 10^5$ yr the minimum ratio value is 0.008. We will discuss in Section 4.1.3 the Si chemistry and why it is so impacted by the reaction network cleaning process.

2. The second group (middle panels of Figure 5) contains species where the ratio is within 10 and 10^3 or 10^{-3} and 0.1 during the evolution: SiH, HNSi, $HSiNH^+$, Si^+, SiF^+, SiH_2^+, and H_2CS. H_2CS and Si^+ are the only species detected in space belonging to this group. H_2CS has a ratio ranging from 0.1 to 0.03 during the evolution, namely, its abundance is lower in the cleaned network simulation by about a factor 10 at steady state (3.4×10^{-12} instead of 4×10^{-13}). The cleaning process affects identically, with the same ratio profile and increasing the abundances by one order of magnitude at steady state, the following Si-bearing species: SiH, Si^+, SiF^+, and SiH_2^+. HNSi and $HSiNH^+$ share the same behavior, but with a steady-state reduced abundance by

about a factor 100. The predicted SO_2^+ abundance is a factor ~ 10 larger with the new network.

3. The third group (bottom panels of Figure 5) contains species where the ratio is within 3 and 10 or 0.1 and 0.3: SiC, SiH^+, $HSiO^+$, and SiN. Three species of this group have been detected in the ISM: SiC, SiN, and SiNC. At steady state, the abundance of SiC is about a factor 10 larger with the new reaction network, while that of SiN is a factor ~ 3 larger.

Overall, 5 out of 26 species affected by the cleaning process have been detected in the ISM: Si^+, SiC, SiN, SiNC, and H_2CS. Of those, only thioformaldehyde (H_2CS) has been detected so far in cold molecular clouds (Irvine et al. 1989; Minowa et al. 1997; Marcelino et al. 2005). We will discuss the case of H_2CS in Section 4.1.3. Finally, 18 out of 20 species affected by the cleaning process contain Si. We will discuss in some detail this case in Section 4.1.3.

4.1.2. Warm Molecular Outflow Shock

Young protostars are known to have spectacular ejections of matter that cause shocks when they hit the surrounding quiet environment. From the chemical point of view, the passage of the shock has two major effects: (i) it heats and compresses the gas, and (ii) species frozen on the icy grain mantles are sputtered and injected into the gas phase. As a result, molecular shocks have gas warm ($\sim 80\text{--}100$ K) and dense ($\geqslant 10^5$ cm^{-3}) regions with much larger abundances of some species with respect to cold molecular clouds.

In order to simulate the chemical composition of the gas after the passage of a shock, we adopt a two-step procedure where
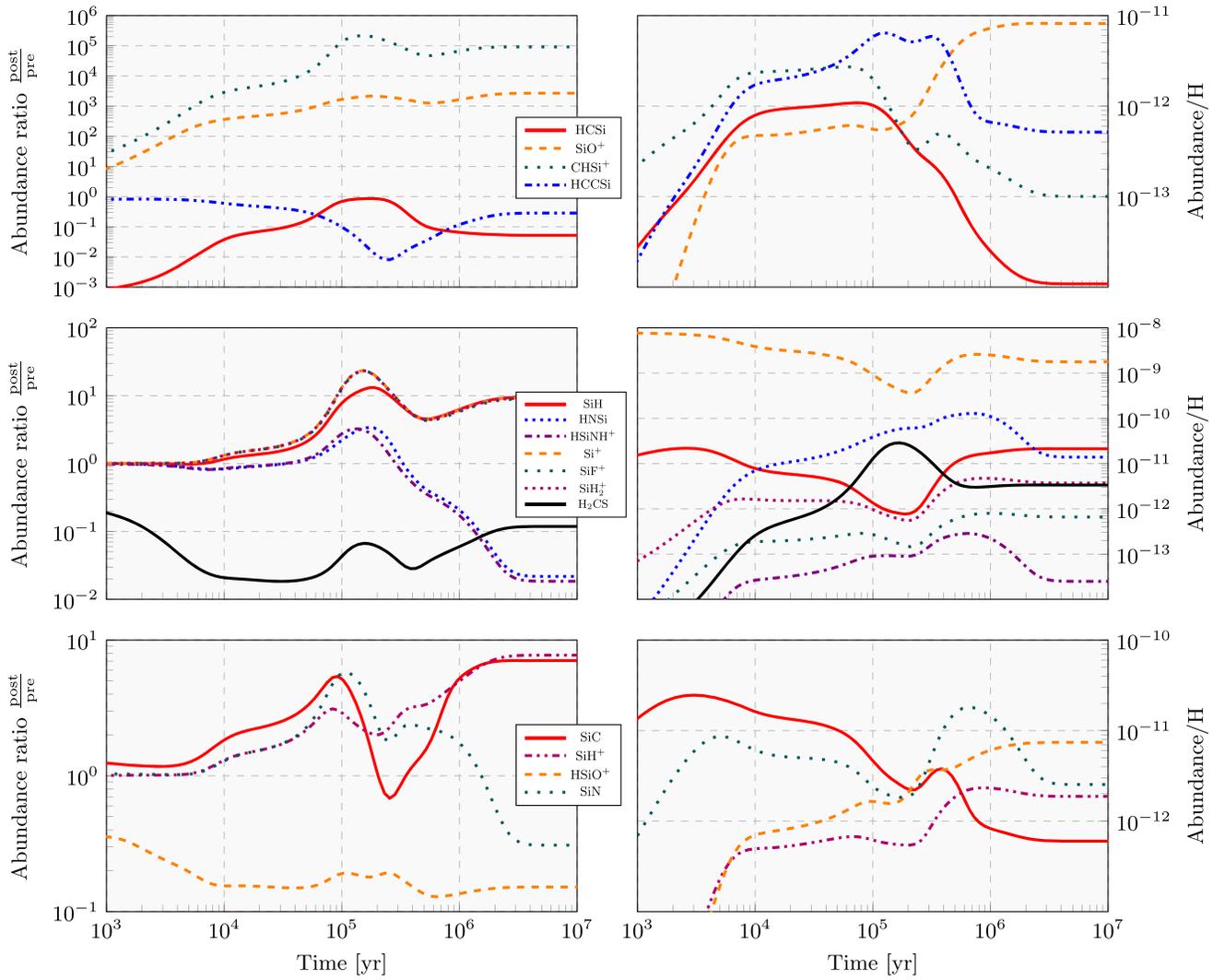
Figure 5. Comparison of the network postcleaning (GRETOBAPE) over precleaning predicted abundance ratios (left panels) and postcleaning abundance (right panels) as a function of time for the cold molecular cloud model (Section 4.1.1). The three rows show three groups of species where the abundance ratios during the evolution are $\geqslant 10^3$ or $\leqslant 10^{-3}$ (top panels), 10–10^3 or 0.1–10^{-3} (middle panels), and 3–10 or 0.1–0.3 (bottom panels). Only species with abundances larger than 10^{-13} at steady state are considered.

we first compute the chemical abundances of the cold molecular cloud (previous subsection), and then we suddenly (i) increase the gas temperature and the H nuclei number density and (ii) inject into the gas large abundances of species that are known to be present on the grain icy mantles. Although the physics is slightly different, a similar modeling can also be considered a fair enough simulation of what happens in hot cores and hot corinos, which are regions heated by the central forming star and where the grain icy mantles sublimate (see, e.g., the recent review by Ceccarelli et al. 2022). In addition, we chose to assume the preshock abundances from a steady-state cold cloud because our goal is not to reproduce a specific case but just to approximately understand the impact of the endothermicity correction in a "warm" case. In this respect, the steady-state choice has two advantages: it does not depend on the choice of the time (i.e., choice of initial abundances), and it only depends on the change in the reaction network.

In practice, our model has two steps:

Step 1: We first compute the steady-state chemical composition of a cold molecular cloud at 10 K with a H

nuclei number density of 2×10^4 cm^{-3} and with the parameters described in Section 4.1.1.

Step 2: We increase the gas temperature to 90 K, the H nuclei number density to 8×10^5 cm^{-3}, and the gaseous abundance of grain-mantle species as follows: $CO_2/H = 3 \times 10^{-5}$, $H_2O/H = 2 \times 10^{-4}$, $NH_3/H = 2 \times 10^{-5}$, $CH_3OH/H = 6.5 \times 10^{-6}$, $CH_3CHO/H = 3.8 \times 10^{-8}$, $C_2H_5/H = 8 \times 10^{-8}$, $SiO/H = 1 \times 10^{-6}$, $Si/H = 1 \times 10^{-6}$, and $OCS/H = 2 \times 10^{-6}$. We then let the chemical composition evolve for 10^4 yr.

The values adopted for the physical parameters and abundances of the injected species are based on the astronomical observations of the prototypical molecular out-flow shock L1157-B1, for which our group carried the same model described here (e.g., Codella et al. 2017; Podio et al. 2017; Codella et al. 2020).

The results of the modeling are reported in Table 8, which reports the ratio of the predicted abundances obtained with the cleaned GRETOBAPE and original reaction networks in the time intervals 1×10^2–1×10^3 yr and 1×10^3–5×10^3 yr for

Table 7
Results of the Modeling of the Cold Molecular Cloud (Section 4.1.1)

Interval Species	1×10^4–1×10^5 yr		1×10^5–1×10^6 yr		SS		max(Abd/H)	
	min(ratio)	max(ratio)	min(ratio)	max(ratio)	Ratio	Abd/H	Abd/H	time[yr]
HCSi	4.2×10^{-2}	6.2×10^{-1}	6.7×10^{-2}	8.6×10^{-1}	5.2×10^{-2}	1.1×10^{-14}	1.1×10^{-12}	7.5×10^4
SiO$^+$	3.8×10^2	1.6×10^3	1.3×10^3	2.1×10^3	2.7×10^3	8.2×10^{-12}	8.2×10^{-12}	2.4×10^6
CHSi$^+$	3.1×10^3	1.0×10^5	4.7×10^4	2.1×10^5	9.2×10^4	1.0×10^{-13}	2.7×10^{-12}	6.1×10^4
HCCSi	1.2×10^{-1}	5.8×10^{-1}	8.1×10^{-3}	1.1×10^{-1}	2.9×10^{-1}	5.2×10^{-13}	6.5×10^{-12}	1.3×10^5
SiH	1.2	6.7	4.5	1.3×10^1	9.5	2.1×10^{-11}	2.2×10^{-11}	2.6×10^3
HNSi	8.4×10^{-1}	2.0	2.2×10^{-1}	3.4	2.2×10^{-2}	1.4×10^{-11}	1.3×10^{-10}	7.1×10^5
HSiNH$^+$	8.6×10^{-1}	2.5	1.8×10^{-1}	3.2	1.8×10^{-2}	2.5×10^{-14}	2.9×10^{-13}	6.4×10^5
Si$^+$	1.4	1.0×10^1	4.5	2.4×10^1	9.4	1.8×10^{-9}	7.8×10^{-09}	1.0×10^2
SiF$^+$	1.4	1.0×10^1	4.3	2.3×10^1	8.9	6.6×10^{-13}	8.0×10^{-13}	9.6×10^5
SiH$_2^+$	1.4	1.0×10^1	4.4	2.3×10^1	9.2	3.7×10^{-12}	4.7×10^{-12}	8.7×10^5
H$_2$CS	1.8×10^{-2}	3.7×10^{-2}	2.8×10^{-2}	6.7×10^{-2}	1.2×10^{-1}	3.4×10^{-12}	2.9×10^{-11}	1.7×10^5
SiC	1.9	5.4	6.8×10^{-1}	5.1	7.1	6.0×10^{-13}	2.5×10^{-11}	2.9×10^3
SiH$^+$	1.3	3.1	2.0	4.9	7.7	1.9×10^{-12}	2.3×10^{-12}	8.7×10^5
HSiO$^+$	1.5×10^{-1}	1.9×10^{-1}	1.3×10^{-1}	1.9×10^{-1}	1.5×10^{-1}	7.4×10^{-12}	7.4×10^{-12}	1.0×10^7
SiN	1.3	5.4	1.8	5.8	3.1	2.5×10^{-12}	1.8×10^{-11}	6.4×10^5
SiNC	8.8×10^{-1}	2.8	1.6	5.8	3.4	2.5×10^{-15}	1.2×10^{-13}	6.8×10^4
l-SiC$_3$	1.1	5.5	3.6	7.7	6.5	1.0×10^{-14}	9.0×10^{-12}	3.5×10^5
SiC$_3$H	5.8×10^{-1}	8.8×10^{-1}	5.8×10^{-1}	1.1	4.5	1.8×10^{-15}	1.5×10^{-11}	2.1×10^5
SiCH$_3$	8.8×10^{-1}	9.6×10^{-1}	9.9×10^{-1}	3.3	6.4	5.9×10^{-14}	7.6×10^{-13}	1.5×10^5
l-C$_3$H$_2^+$	5.3×10^{-1}	9.0×10^{-1}	1.0×10^{-1}	4.7×10^{-1}	2.7×10^{-1}	2.5×10^{-15}	3.7×10^{-11}	1.0×10^5

Note. For each species, listed in the first column, columns (2)–(5) list the minimum (min) and maximum (max) values of the network postcleaning (GRETOBAPE) over precleaning predicted abundance ratios in the time range 1×10^4–1×10^5 yr and 1×10^5–1×10^6 yr, respectively. Columns (6)–(7) report the ratio and the abundance at steady state (SS). The last two columns list the largest reached abundance for each species and at what time. The species are ordered following the maximum ratio variation during the entire simulation. Finally, boldfaced species have been detected in the ISM. The detected molecules are boldfaced (McGuire 2022). The Si$^+$ atom was first detected by Haas et al. (1986).

species with a ratio larger than 3 or less than 0.3 and abundances larger than 10^{-13}.

Overall, 22 species are significantly affected by the reaction network cleaning process, namely, their abundances change by more than a factor 3 during the evolution from the passage of the shock to 5×10^3 yr. The species impacted by the reaction network cleaning process can be divided into three groups with respect to their ratio at 5×10^3 yr (e.g., the approximate L1157-B1 age):

1. *The first group contains species whose ratio is higher than 10^3 or lower than 10^{-3}*: SiS$^+$, CHSi$^+$, SiO$^+$, and SiN. SiN is the only species in this group that has been detected in outflows and will be discussed in detail in a dedicated paragraph of Section 4.1.3. The species with the largest variation due to the cleaning process is SiS$^+$, with an almost constant ratio of 4.2×10^7 and reaching an abundance of $\sim 4 \times 10^{-11}$ at 5×10^3 yr.
2. The second group contains species where the ratio is within 10^2 and 10^3 or 10^{-3} and 10^{-2}: HSiO$^+$, CH$_3$CHCH$_2$, SiH$_2^+$, Si$^+$, SiH, HNSi, SiNC$^+$. Two species of this group have been detected in the ISM: CH$_3$CHCH$_2$ and Si$^+$. Only the latter has been detected in outflows, and we discuss its case in Section 4.1.3. CH$_3$CHCH$_2$ has an almost constant ratio of $\sim 5 \times 10^{-2}$ and a maximum abundance of $\sim 5 \times 10^{-13}$ at the beginning of the simulation. On the other hand, Si$^+$ has a ratio of almost 4.5×10^2 that increases the abundance in the cleaned simulation with respect to the original.
3. *The third group contains species where the ratio is within 3 and 10^2 or 10^{-2} and 0.3 during the evolution*: HSS, c-HCCHSi, c-SiC$_2$, SiS, SO$_2^+$, HCCSi, Si, HSSH, SiH$^+$,

SiC, and C$_3$O. Five species belonging to this group have been detected in the ISM: HSS, c-SiC$_2$, SiS, SiC, and C$_3$O. Only SiS has been detected in outflows, and we discuss its case in Section 4.1.3. HSS has an abundance ratio between 0.2 and 0.9 in the 1×10^3–5×10^3 yr interval and a maximum abundance of 4.7×10^{-13} at the end of the simulation. c-SiC$_2$ at the L1157-B1 age has a ratio of 2×10^{-1} and an abundance of 7.2×10^{-13}. SiS has an almost constant ratio of 0.15; meanwhile, the C$_3$O ratio oscillates between ~ 20 and 2. SiC is the less affected species by the cleaning process with a factor 3 of difference.

As in the cold molecular cloud simulation (Section 4.1.1), the cleaning process mostly affected the Si reaction network. Indeed, 17 of the 22 species present in Table 8 are Si-bearing species. Three of them, SiN (Schilke et al. 2003), SiS (Podio et al. 2017), and Si$^+$ (Haas et al. 1986), are detected in outflows. They will be discussed in some detail in Section 4.1.3.

4.1.3. Discussion of Specific Cases

In this section, we discuss in some detail the species that are more significantly affected by the reaction network cleaning process.

Thioformaldehyde. In cold molecular gas (Section 4.1.1), the thioformaldehyde (H$_2$CS) abundance predicted after the cleaning process at steady state is about 10 times smaller than in precleaning. Here we first summarize the observations and then comment on what causes the difference in the predicted abundances.

Table 8
Results of the Modeling of the Warm Molecular Outflow Shock (Section 4.1.2)

Interval Species	1×10^2–1×10^3 yr		1×10^3–5×10^3 yr		5×10^3 yr		max(Abd/H)	
	min(ratio)	max(ratio)	min(ratio)	max(ratio)	Ratio	Abd/H	Abd/H	time[yr]
SiS$^+$	3.9×10^7	4.2×10^7	4.2×10^7	4.5×10^7	4.2×10^7	3.9×10^{-11}	6.3×10^{-11}	1.0×10^2
CHSi$^+$	1.5×10^6	1.6×10^6	1.4×10^6	1.9×10^6	1.4×10^6	4.3×10^{-13}	4.3×10^{-13}	4.9×10^3
SiO$^+$	1.3×10^3	1.5×10^3	1.6×10^3	7.9×10^3	7.9×10^3	1.1×10^{-11}	1.1×10^{-11}	5×10^3
SiN	1.2×10^{-2}	1.9×10^{-2}	2.8×10^{-3}	7.7×10^{-3}	5.5×10^{-3}	3.6×10^{-13}	3.6×10^{-13}	5×10^3
HSiO$^+$	1.1×10^{-1}	1.1×10^{-1}	8.3×10^{-2}	1.0×10^{-1}	8.2×10^{-2}	3.5×10^{-10}	3.5×10^{-10}	3.9×10^3
CH$_3$CHCH$_2$	5.0×10^{-2}	5.2×10^{-2}	5.3×10^{-2}	6.5×10^{-2}	6.5×10^{-2}	3.6×10^{-13}	4.2×10^{-13}	1.0×10^2
SiH$_2^+$	4.0×10^2	4.3×10^2	4.4×10^2	5.0×10^2	4.9×10^2	1.5×10^{-12}	1.5×10^{-12}	5×10^3
Si$^+$	4.0×10^2	4.3×10^2	4.4×10^2	4.9×10^2	4.8×10^2	1.5×10^{-10}	1.6×10^{-10}	3.9×10^3
SiH	7.1×10^1	1.4×10^2	1.6×10^2	3.7×10^2	3.7×10^2	3.9×10^{-12}	3.1×10^{-12}	4.3×10^3
HNSi	1.6×10^{-2}	1.9×10^{-2}	1.1×10^{-2}	1.5×10^{-2}	1.1×10^{-2}	1.3×10^{-11}	1.4×10^{-11}	8.0×10^2
SiNC$^+$	5.5×10^1	7.7×10^1	8.4×10^1	1.2×10^2	1.1×10^2	2.5×10^{-12}	2.4×10^{-12}	4.7×10^3
HSS	9.6×10^{-1}	1.0	2.3×10^{-1}	9.2×10^{-1}	2.3×10^{-1}	4.7×10^{-13}	4.7×10^{-13}	5×10^3
c-HCCHSi	4.3×10^{-1}	1.2	2.3×10^{-1}	3.7×10^{-1}	2.3×10^{-1}	6.2×10^{-12}	6.1×10^{-12}	4.9×10^3
c-SiC$_2$	4.3×10^{-2}	5.5×10^{-2}	6.3×10^{-2}	2.1×10^{-1}	2.1×10^{-1}	7.2×10^{-13}	7.3×10^{-13}	4.7×10^3
SiS	1.7×10^{-1}	2.8×10^{-1}	1.5×10^{-1}	1.7×10^{-1}	1.5×10^{-1}	8.4×10^{-12}	8.5×10^{-12}	4.9×10^3
SO$_2^+$	1.3×10^1	1.4×10^1	1.4×10^1	1.5×10^1	1.5×10^1	1.4×10^{-13}	1.4×10^{-13}	5×10^3
HCCSi	2.4×10^{-2}	3.8×10^{-2}	4.4×10^{-2}	1.3×10^{-1}	1.3×10^{-1}	4.8×10^{-12}	4.7×10^{-12}	4.9×10^3
Si	1.6×10^{-1}	1.7×10^{-1}	1.2×10^{-1}	1.6×10^{-1}	1.2×10^{-1}	3.3×10^{-10}	2.1×10^{-09}	1.0×10^2
HSSH	8.0×10^{-1}	1.0	1.0×10^{-1}	6.7×10^{-1}	1.0×10^{-1}	1.6×10^{-13}	1.6×10^{-13}	5×10^3
SiH$^+$	2.4	3.3	3.7	5.7	5.1	5.3×10^{-13}	5.4×10^{-13}	4.9×10^3
SiC	1.2	1.2	1.2	3.0	3.0	2.3×10^{-11}	2.3×10^{-11}	5×10^3
C$_3$O	9.8	2.1×10^1	2.0	1.6×10^1	2.0	2.4×10^{-12}	7.9×10^{-12}	1.0×10^2

Note. For each species, listed in the first column, columns (2)–(5) list the minimum (min) and maximum (max) values of the network postcleaning (GRETOBAPE) over precleaning predicted abundance ratios in the time range 1×10^2–1×10^3 yr and 1×10^3–5×10^3 yr, respectively. Columns (6)–(7) report the ratio and the abundance at 5×10^3 yr. The last two columns list the largest reached abundance for each species and at what time. The species are ordered following the value of the ratio at 5×10^3 yr, which is believed to approximately be the L1157-B1 age (Section 4.1.2). Finally, boldfaced species have been detected in the ISM. The detected species are boldfaced (McGuire 2022). The Si$^+$ atom was first detected by Haas et al. (1986).

Thioformaldehyde has been detected in cold molecular clouds, notably TMC-1 and L134, with an abundance (with respect to H nuclei) of about $\sim10^{-9}$ (e.g., Ohishi & Kaifu 1998). Vastel et al. (2018) and Spezzano et al. (2022) carried out a detailed study of thioformaldehyde toward the prototypical cold prestellar core L1544, reporting a column density of $\sim6 \times 10^{12}$ cm^{-2}, corresponding to an average abundance of $\leqslant10^{-10}$, depending on where the detected H$_2$CS emission comes from. These authors attribute the relatively low abundance of all the S-bearing species observed in these objects to a general depletion of the gaseous S elemental abundance. Likewise, Esplugues et al. (2022) have carried out a detailed study of H$_2$CS toward several cold starless cores in the Taurus, Perseus, and Orion regions and found H$_2$CS abundances ranging from 0.8 to 14×10^{-11}.

As discussed in detail in Esplugues et al. (2022), in cold gas, H$_2$CS is mainly formed by two reactions: S + CH$_3$ and H$_3$CS$^+$ + e$^-$. While H$_2$CS is not directly involved in the cleaning process (Table 2), the H$_3$CS$^+$ postcleaning abundance at steady state is lower by a factor ~18 than the precleaned one because one dominant reaction in the original network forming H$_3$CS$^+$, CH4 + S$^+$ $\rightarrow$ H + H$_3$CS$^+$, is endothermic by 48 kJ mol^{-1}. Please note that the steady-state postcleaning abundance of H$_3$CS$^+$ is $\leqslant10^{-13}$, so that this species does not appear in Figure 5 and Table 7. However, the low abundance (6×10^{-15}) is counterbalanced by the large recombination rate of H$_3$CS$^+$, whose product is H$_2$CS.

Si chemistry and Si$^+$. As already anticipated, the reaction network cleaning process significantly affects the Si chemistry, with the removal of 48 reactions and 6 species that represent

$\sim9\%$ and $\sim10\%$ of the total Si-bearing reactions and species, respectively. In addition, 18 out of 20 most affected species of the cold molecular cloud simulation (Section 4.1.1) contain Si, of which the vast majority (12) have less than four atoms. Similarly, 17 of the 22 most affected species in the the warm molecular outflow shock simulation (Section 4.1.2) are Si-bearing species.

The six removed Si-bearing species (Table 2) are (neutral) chains and are removed because their formation or destruction routes in the original network are endothermic (Section 3.2.2). Their removal very weakly affects two other Si-bearing species, c-HCCHSi and SC$_3$H, respectively (Figures 3 and 4), and consequently the rest of the Si-bearing species.

Therefore, the large impact of the cleaning process on the Si chemical network is due to the removal of the 48 reactions involving Si-bearing species. Figure 4 shows that those reactions connect 36 Si-bearing species out of 61, namely more than half of them. It is, therefore, not surprising that 17 Si-bearing species are among the most affected ones in the warm molecular outflow shock model. For example, SiS$^+$, the most affected species, has three removed reactions connecting it to Si$^+$, SiC$_3$H$_5$, and SiS, respectively. The first two are destruction routes of SiS$^+$; their removal is, therefore, responsible for the augmented predicted SiS$^+$ abundance. A similar argument applies to CHSi$^+$, the second most affected species in the modeling: a reaction of destruction is removed, causing the increase in the predicted abundance.

Probably most important, Si$^+$ has 10 removed reactions linking it to other Si-bearing species. Because of that, the predicted abundance of Si$^+$ is more than 400 times the

predicted abundance before the cleaning process. This causes a domino effect, propagating as a wave: 11 of the other 16 impacted Si-bearing species are one step away from Si^+ (i.e., one node distance in Figure 3), while the other 5 are two steps away. The species one step away are also those whose predicted abundances are the most impacted.

Silicon nitride. Silicon nitride, SiN, is the detected Si-bearing species most affected by the cleaning process (Table 8), so we will analyze its case in some detail here. SiN was first detected in the circumstellar envelope of the evolved star IRC+10216 (Turner 1992) and later toward the Galactic center cloud SgrB2 (M) (Schilke et al. 2003), probably in the warm shocked gas of the region. IRC+10216 and SgrB2(M) remain the only two objects where SiN has so far been detected, despite sensitive spectral surveys toward the high-mass star-forming region Orion (Tercero et al. 2011) and several low-mass star-forming regions, including the prototype of the warm molecular shocks L1157-B1 (Ceccarelli et al. 2017).

The predicted SiN abundance is more than 100 times lower when using the postcleaning reaction network than the precleaning one. Carefully looking at Figures 3 and 4, one can notice that the SiN major formation route in the precleaning network is the reaction involving $HSiNH^+$, followed by those involving SiC and $HNSi^+$ at a lesser extent. Although this reaction is not affected by the cleaning process, two reactions forming $HSiNH^+$ are removed after it (from Si^+ and HNSi, respectively). Indeed, the abundance of $HSiNH^+$ changes by three orders of magnitude (but it is not listed in Table 8 because of its low abundance, $\leqslant 10^{-13}$). About these reactions, as discussed for the $HSiO^+$ case (see Section 3.2.4), it is probable that KIDA erroneously considered the product $HSiNH^+$ instead of its more stable isomer $SiNH_2^+$ (Glosik et al. 1995; Parisel et al. 1996).

As said, the new reaction network predicts a much lower SiN abundance than before, which may explain the paucity of the SiN detections. For example, Tercero et al. (2011) found a lower limit of [SiO]/[SiN] $\geqslant 121$. In our postcleaning modeling, the SiO abundance at 5×10^3 yr is $\sim 2 \times 10^{-6}$, implying a predicted [SiO]/[SiN] $\sim 5 \times 10^{-6}$, therefore largely compatible with the nondetection of Tercero et al. (2011).

SiS. Silicon monosulfide, SiS, is another detected Si-bearing molecule affected by the cleaning process. Specifically, SiS has been so far detected toward two warm molecular shocks, Orion KL (Tercero et al. 2011) and L1157-B1 (Podio et al. 2017). The factor between the pre- and post-cleaning predicted abundances is ~ 0.15, again probably explaining the paucity of SiS detections. Peering into Figures 3 and 4, in the pre- and post-cleaning networks, SiS is mainly formed from $HSiS^+$, whose abundance ($\leqslant 10^{-13}$, so the species not present in Table 8) is affected by the cleaning process, with a decrease factor of ~ 0.06.

4.2. Predicted Abundances with the Reaction Network GRETOBAPE-RED versus GRETOBAPE

To explore the reliability of our reduced network, GRETO-BAPE-RED, we performed an exhaustive parameter space exploration and compared the steady-state abundances of all species with those obtained using the complete network, GRETOBAPE.

To this end, we initialized different combinations of parameters that can affect the resulting species abundances in the cold and warm conditions present during the star formation

process. Namely, the varied parameters are the cosmic-ray ionization rate for H_2, with values of 3×10^{-16} s^{-1}, 3×10^{-17} s^{-1}, and 3×10^{-18} s^{-1}; the gas temperature, with values of 10, 50, 100, and 500 K; and the H nuclei number density, with values of 10^2, 10^4, 10^6, and 10^8 cm^{-3}. All possible combinations of these parameters yield 48 different runs, the results of which we will explain in what follows. We used the KROME chemistry package (Grassi et al. 2014) to perform the parameter space exploration, considering only gas-phase chemistry. We keep the rest of the input parameters (visual extinction, dust-to-gas mass ratio, grain radius, and grain density) equal to those used to perform the run described in Section 4.1.1, and we let the chemistry evolve up to 1×10^7 yr.

In all runs, the reduced network GRETOBAPE-RED reproduces almost perfectly the steady-state abundances of GRETO-BAPE, for all species with abundances $>10^{-10}$, with differences smaller than a factor 3 in general. However, the differences progressively worsen for trace species. A few species with abundances $\leqslant 10^{-10}$ have differences larger than a factor 3 when comparing the results of GRETOBAPE-RED with those of GRETOBAPE. The main species affected are the carbon chains C_6, C_3, H_2CCCO, c-C_3H_2O, C_3O, H_2CCN, and H_2CCO, in addition to CH_2NH and O_2H. Larger differences might be present for other trace species with abundances $<10^{-15}$, but, as these abundances are already very low, these molecules may not be of interest in the astrochemical simulations where the reduced network GRETOBAPE-RED is suitable. As expected, the effect on carbon chains can be attributed to the removal of larger species that can be part of either their destruction or formation routes in the case of dissociative recombination of large cations that could produce the species listed here.

These results overall indicate that the reduced network GRETOBAPE-RED is reliable to be used when following the chemistry of species with fractional abundances larger than $\sim 10^{-10}$, while the complete network GRETOBAPE is needed when exploring the chemistry of trace species.

5. Conclusions

Gas-phase reaction networks are a crucial element of any astrochemical modeling. Present-day publicly available networks, e.g., KIDA and UMIST, are made up of more than 7000 reactions, of which only a tiny fraction (10%–20%) have been studied in laboratory or theoretical works.

In this work, we present a new gas-phase reaction network, GRETOBAPE, built from the publicly available KIDA2014 (Wakelam et al. 2015) with the addition of several new reactions from more recent studies. The most important novelty is that GRETOBAPE is cleaned by the most obvious source of error, the presence of endothermic reactions not recognized as such. We also present a reduced network, GRETOBAPE-RED, to be used in problems that do not require following the chemical evolution of trace species.

To this end, we performed an extended and systematic theoretical characterization of more than 500 species at CCSD (T)/aug-cc-pVTZ//M06-2X/cc-pVTZ computational level, providing for each of them: electronic state, electronic spin multiplicity, geometry, harmonic frequency, absolute electronic energy, and dipole moment.

We then computed the enthalpy at 0 K of each neutral–neutral, neutral–ion, and cation–anion reaction in the network. Finally, we identified the endothermic reactions not recognized as such, i.e., originally reported as barrierless or having an

Table 9
Summary of the Files Containing the Reaction Networks and Species Chemical Data Provided in the Supporting Material and a Brief Description of Their Content

File Name	Description
ReadMe.txt	File with the full description of all the files in the supporting material.
GRETOBAPE-pre.dat	Original reaction network, before the cleaning process, described in Section 3.1.
GRETOBAPE.dat	New reaction network, after the cleaning process, described in Section 3.2.
GRETOBAPE-red.dat	Reduced network from GRETOBAPE.dat, described in Section 3.3.
GRETOBAPE-endo.dat	List of reactions deleted because of the endothermicity criteria, described in Section 2.3.1.
GRETOBAPE-endo-info.csv	List of reactions in GRETOBAPE-endo.dat file with information on the endothermicity.
GRETOBAPE-endo-0-10kJmol.dat	List of reactions with an endothermicity lower than 10 kJ mol^{-1}: they are present in GRETOBAPE.dat.
GRETOBAPE-domino.dat	List of reactions deleted because of the domino effect, described in Section 2.3.2.
GRETOBAPE-domino-endo-Si.dat	List of reactions involving Si-bearing species and deleted because of the endothermicity and domino effect criteria, discussed in Section 3.2.2.
Database_Molecules.csv	Extracted and organized chemical information of all species in GRETOBAPE.dat, described in Section 2.1.
species	Directory containing the .XYZ files of all species studied in this work, described in Section 2.1.
codes	Directory containing all the Python scripts used in this work.

Note. The file structure of all the reaction networks follows the KIDA format.

exponential factor γ not taking into account the level of endothermicity.

Following these two criteria, we deleted about 5% of the studied reactions in the original precleaned network, leading to the removal of 11 species. The final cleaned network GRETOBAPE consists of 6911 reactions and 488 species, while the reduced network GRETOBAPE-RED contains 2810 reaction and 204 species.

We also reported a list of probably missing reactions, namely, (i) neutral species lacking reactions with the most abundant interstellar cations (H_3^+, HCO^+, H_3O^+, He^+, and H^+), and (ii) detected species absent in the network.

Using astrochemical model simulations of a typical cold molecular cloud and a warm molecular outflow shock, we measured the impact of the new cleaned GRETOBAPE on the predicted abundances. Overall, only about 5% of the species abundances are affected by more than a factor 3 with respect to the original precleaned network predictions. Thioformaldehyde (H_2CS), Si^+, SiN, and SiS are the detected species whose predicted abundances are affected by the cleaning process. Of the 5% affected abundances, the immense majority concerns Si-bearing species. We discuss in detail the origin of this large impact on the Si chemistry and conclude that a systematic review of this chemistry is needed.

We also verified that the abundances predicted by GRETO-BAPE-RED of the species with abundances larger than 10^{-10} are not affected by the reduction of the GRETOBAPE network, on a large parameter space. Therefore, GRETOBAPE-RED can be used when the goal is to reliably reproduce only the most abundant species and make the simulations less computationally demanding.

In conclusion, we carried out the first ever systematic study of the exo/endothermicity of the reactions included in an astrochemical network. The process removing the endothermic reactions not recognized as such and the sink/source species led to the new network GRETOBAPE and its reduced version GRETOBAPE-RED. The control of the exo/endothermicity and sink/source species is only the first and obvious step to make the astrochemical gas-phase reaction networks reliable. Inclusion of the reactions of neutrals with abundant interstellar cations is a second obvious but still not complete step. The next step would be to carefully review all the reactions, rate constants, and products, which is obviously unfeasible given the too large number of reactions present in the networks. A

compromise could be to start with the control/verification of subnetworks that form commonly detected species, like the work done by Vazart et al. (2020) on acetaldehyde. In addition, the lack of gas-phase reactions forming interstellar detected species does not necessarily mean that they do not exist, and a lot of work here is still necessary (e.g., Balucani et al. 2015; Skouteris et al. 2018; Sleiman et al. 2018; Cernicharo et al. 2022, just to mention some recent works). Last but not least, in addition to devoted studies on the formation routes, attention should also be dedicated to the destruction routes. For example, the study of the destruction of methyl formate and dimethyl ether by He^+ has shown that the rate constants included in the KIDA and UMIST networks are wrong by at least a factor 10 (Ascenzi et al. 2019).

Finally, the database created in this work and available online (Section 6) can be used to rapidly test or search existing or new possible formation and destruction reaction pathways and further refined via experiments or theoretical studies to computed rate constants.

6. Supporting Material

All the reaction networks used and produced in this work and cited in the main body are reported in the supporting material available under a Creative Commons Attribution license on Zenodo: doi:10.5281/zenodo.7799421. We provide all the .XYZ files of each studied species, as well as all the relative information in the Database_Molecules.csv file, following Tinacci et al. (2021). The codes used to encode and characterize the network properties are provided (and are also available in GitHub[18]). Table 9 summarizes the list of files with the various reaction networks along with a brief description of their content. The networks and the database of all molecular species properties are available on the ACO (AstroChemical Origins) Database.[19]

---

[18] https://github.com/TinacciL/GreToBaPe_Cleaning
[19] https://aco-itn.oapd.inaf.it/aco-public-datasets/theoretical-chemistry-calculations

ORCID iDs

Lorenzo Tinacci ● https://orcid.org/0000-0001-9909-9570
Simón Ferrada-Chamorro ● https://orcid.org/0000-0002-4874-838X
Cecilia Ceccarelli ● https://orcid.org/0000-0001-9664-6292
Stefano Pantaleone ● https://orcid.org/0000-0002-2457-1065
Daniela Ascenzi ● https://orcid.org/0000-0001-5393-9554
Andrea Maranzana ● https://orcid.org/0000-0002-5524-8068
Nadia Balucani ● https://orcid.org/0000-0001-5121-5683
Piero Ugliengo ● https://orcid.org/0000-0001-8886-9832

References

Anicich, V. G. 2003, An index of the literature for bimolecular gas phase cation-molecule reaction kinetics, v1, JPL Open Repository, https://hdl.handle.net/2014/7981
Ascenzi, D., Cernuto, A., Balucani, N., et al. 2019, A&A, 625, A72
Ayouz, M. A., Yuen, C. H., Balucani, N., et al. 2019, MNRAS, 490, 1325
Badnell, N. 2006, ApJS, 167, 334
Balucani, N. 2020, PhLRv, 34, 136
Balucani, N., Ceccarelli, C., & Taquet, V. 2015, MNRAS, 449, L16
Balucani, N., Skouteris, D., Ceccarelli, C., et al. 2018, MolAs, 13, 30
Barabási, A.-L. 2013, RSPTA, 371, 20120375
Blázquez, S., González, D., Neeman, E. M., et al. 2020, PCCP, 22, 20562
Canosa, A., Le Picard, S., Gougeon, S., et al. 2001, JChPh, 115, 6495
Cavallotti, C. 2022, Proc. Combust. Inst., in press
Ceccarelli, C., Caselli, P., Fontani, F., et al. 2017, ApJ, 850, 176
Ceccarelli, C., Codella, C., Balucani, N., et al. 2022, arXiv:2206.13270
Cernicharo, J., Fuentetaja, R., Agúndez, M., et al. 2022, A&A, 663, L9
Cheikh, S. E. 2012, PhD thesis, Université de Rennes
Codella, C., Ceccarelli, C., Bianchi, E., et al. 2020, A&A, 635, A17
Codella, C., Ceccarelli, C., Caselli, P., et al. 2017, A&A, 605, L3
Cramer, C. J. 2013, Essentials of Computational Chemistry: Theories and Models (Hoboken, NJ: John Wiley & Sons)
DeFrees, D., McLean, A., & Herbst, E. 1985, ApJ, 293, 236
Dóbé, S., Bérces, T., & Szilágyi, I. 1991, FaTr, 87, 2331
Dunham, T., Jr 1937, PASP, 49, 26
Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J., & Müller, H. S. 2016, JMoSp, 327, 95
Esplugues, G., Fuente, A., Navarro-Almaida, D., et al. 2022, A&A, 662, A52
Fontani, F., Ceccarelli, C., Favre, C., et al. 2017, A&A, 605, A57
Fournier, M. 2014, PhD thesis, Université de Rennes
Freedman, D., & Diaconis, P. 1981, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57, 453
Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al. 2016, Gaussian 16 Revision B.01 (Wallingford, CT: Gaussian Inc.)
Gao, L. G., Zheng, J., Fernández-Ramos, A., Truhlar, D. G., & Xu, X. 2018, JAChS, 140, 2906
Glosik, J., Zakouřil, P., & Lindinger, W. 1995, JChPh, 103, 6490
Glover, S. C., Federrath, C., Mac Low, M.-M., & Klessen, R. S. 2010, MNRAS, 404, 2
Grassi, T., Bovino, S., Schleicher, D., & Gianturco, F. 2013, MNRAS, 431, 1659
Grassi, T., Bovino, S., Schleicher, D. R. G., et al. 2014, MNRAS, 439, 2386
Grassi, T., Nauman, F., Ramsey, J., et al. 2022, A&A, 668, A139
Haas, M., Hollenbach, D., & Erickson, E. 1986, ApJL, 301, L57
Hagberg, A. A., Schult, D. A., & Swart, P. J. 2008, Proc. 7th Python in Science Conf., ed. G. Varoquaux, T. Vaught, & J. Millman, 11, https://conference.scipy.org/proceedings/SciPy2008/paper_2/
Hamberg, M., Zhaunerchyk, V., Vigren, E., et al. 2010, A&A, 522, A90
Harada, N., & Herbst, E. 2008, ApJ, 685, 272
Herbst, E. 2006, Gas Phase Reactions (New York: Springer), 561
Herbst, E., & Klemperer, W. 1973, ApJ, 185, 505
Hoyermann, K., Sievert, R., & Wagner, H. G. 1981, Berichte der Bunsengesellschaft für Physikalische Chemie, 85, 149
Irvine, W. M., Friberg, P., Kaifu, N., et al. 1989, ApJ, 342, 871
Jenkins, E. B. 2009, ApJ, 700, 1299
Kaiser, R., & Gu, X. 2009, JChPh, 131, 104311
Kendall, T., Jr 1992, JChPh, 96, 6796
Knowles, P. J., Hampel, C., & Werner, H.-J. 1993, JChPh, 99, 5219
Kooij, D. 1893, ZPC, 12, 155
Laidler, K. J. 1996, Pure Appl. Chem., 68, 149
Lara-Moreno, M., Stoecklin, T., & Halvick, P. 2019a, ESC, 3, 1556
Lara-Moreno, M., Stoecklin, T., Halvick, P., & Loison, J.-C. 2019b, PhRvA, 99, 033412
Lawson, P. A., Osborne, D. S., Jr, & Adams, N. G. 2012, JPCA, 116, 2880
Loison, J.-C., Agúndez, M., Marcelino, N., et al. 2016, MNRAS, 456, 4101
Loison, J.-C., Wakelam, V., & Hickson, K. M. 2014, MNRAS, 443, 398
Loison, J.-C., Wakelam, V., Hickson, K. M., Bergeat, A., & Mereau, R. 2013, MNRAS, 437, 930
Mancini, L., & Aragão, E. V. F. D. 2021, Computational Science and Its Applications – ICCSA 2021 (Berlin: Springer), 339
Marcelino, N., Cernicharo, J., Roueff, E., Gerin, M., & Mauersberger, R. 2005, ApJ, 620, 308
McElroy, D., Walsh, C., Markwick, A., et al. 2013, A&A, 550, A36
McEwan, M., Anicich, V., Huntress, W., Jr, Kemper, P., & Bowers, M. 1983, IJMSI, 50, 179
McGuire, B. A. 2022, ApJS, 259, 30
McKellar, A. 1940, PASP, 52, 187
Meot-Ner, M., Karpas, Z., & Deakyne, C. A. 1986, JAChS, 108, 3913
Minowa, H., Satake, M., Hirota, T., et al. 1997, ApJL, 491, L63
Mota, V., Varandas, A., Mendoza, E., Wakelam, V., & Galvão, B. 2021, ApJ, 920, 37
Nelson, R. P., & Langer, W. D. 1999, ApJ, 524, 923
Neufeld, D. A., Godard, B., Gerin, M., et al. 2015, A&A, 577, A49
Newman, M. 2018, Networks (Oxford: Oxford Univ. Press)
Ohishi, M., & Kaifu, N. 1998, FaDi, 109, 205
Oppenheimer, M., & Dalgarno, A. 1974, ApJ, 192, 29
Parisel, O., Hanus, M., & Ellinger, Y. 1996, JPhCh, 100, 2926
Penteado, E., Walsh, C., & Cuppen, H. 2017, ApJ, 844, 71
Podio, L., Codella, C., Lefloch, B., et al. 2017, MNRAS, 470, L16
Prasad, S., & Huntress, W., Jr 1980, ApJS, 43, 1
Raksit, A., Schiff, H., & Bohme, D. 1984, IJMSI, 56, 321
Rosi, M., Mancini, L., Skouteris, D., et al. 2018, CPL, 695, 87
Schilke, P., Leurini, S., Menten, K., & Alcolea, J. 2003, A&A, 412, L15
Skouteris, D., Balucani, N., Ceccarelli, C., et al. 2018, ApJ, 854, 135
Skouteris, D., Balucani, N., Ceccarelli, C., et al. 2019, MNRAS, 482, 3567
Skouteris, D., Vazart, F., Ceccarelli, C., et al. 2017, MNRAS, 468, L1
Sleiman, C., El Dib, G., Rosi, M., et al. 2018, PCCP, 20, 5478
Sleiman, C., El Dib, G., Talbi, D., & Canosa, A. 2018, ESC, 2, 1047
Smith, D., Adams, N. G., & Lindinger, W. 1981, JChPh, 75, 3365
Smith, I. W. 2006, AngCh, 45, 2842
Smith, I. W. 2011, ARA&A, 49, 29
Spezzano, S., Sipilä, O., Caselli, P., et al. 2022, A&A, 661, A111
Stancil, P., Schultz, D., Kimura, M., et al. 1999, A&AS, 140, 225
Su, T., & Chesnavich, W. J. 1982, JChPh, 76, 5183
Swings, P., & Rosenfeld, L. 1937, ApJ, 86, 483
Tercero, B., Vincent, L., Cernicharo, J., Viti, S., & Marcelino, N. 2011, A&A, 528, A26
Tinacci, L., Pantaleone, S., Maranzana, A., et al. 2021, ApJS, 256, 35
Turner, B. 1992, ApJL, 388, L35
Urso, R. G., Palumbo, M. E., Ceccarelli, C., et al. 2019, A&A, 628, A72
Urso, R. G., Scirè, C., Baratta, G. A., Compagnini, G., & Palumbo, M. E. 2016, A&A, 594, A80
Vastel, C., Quénard, D., Le Gal, R., et al. 2018, MNRAS, 478, 5514
Vazart, F., Calderini, D., Puzzarini, C., Skouteris, D., & Barone, V. 2016, J. Chem. Theory Comput., 12, 5385
Vazart, F., Ceccarelli, C., Balucani, N., Bianchi, E., & Skouteris, D. 2020, MNRAS, 499, 5547
Wakelam, V., Herbst, E., Loison, J.-C., et al. 2012, ApJS, 199, 21
Wakelam, V., Herbst, E., Selsis, F., & Massacrier, G. 2006, A&A, 459, 813

Wakelam, V., Loison, J.-C., Herbst, E., et al. 2015, ApJS, 217, 20
Watts, J. D., Gauss, J., & Bartlett, R. J. 1993, JChPh, 98, 8718
Wheeler, S. E., Houk, K. N., Schleyer, P. v. R., & Allen, W. D. 2009, JAChS, 131, 2547

Woon, D. E., & Dunning, T. H., Jr 1993, JChPh, 98, 1358
Woon, D. E., & Herbst, E. 2009, ApJS, 185, 273
Yu, M., Ruan, J., Qian, C., et al. 2020, ChemistrySelect, 5, 12764
Zhao, Y., & Truhlar, D. G. 2008, Theor. Chem. Acc., 120, 215