



<b>Publication Year</b>	2021
<b>Acceptance in OA</b>	2025-01-09T16:03:47Z
<b>Title</b>	Establishing the evolutionary timescales of the massive star formation process through chemistry
<b>Authors</b>	SABATINI, GIOVANNI, Bovino, S., GIANNETTI, Andrea, Grassi, T., BRAND, JAN, SCHISANO, EUGENIO, Wyrowski, F., Leurini, Silvia, Menten, K. M.
<b>Publisher's version (DOI)</b>	10.1051/0004-6361/202140469
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/35605">http://hdl.handle.net/20.500.12386/35605</a>
<b>Journal</b>	ASTRONOMY & ASTROPHYSICS
<b>Volume</b>	652

# Establishing the evolutionary timescales of the massive star formation process through chemistry

G. Sabatini<sup>1,2,3,\*</sup>, S. Bovino<sup>3</sup>, A. Giannetti<sup>2</sup>, T. Grassi<sup>4</sup>, J. Brand<sup>2</sup>, E. Schisano<sup>5</sup>, F. Wyrowski<sup>6</sup>,  
S. Leurini<sup>7</sup>, and K. M. Menten<sup>6</sup>

<sup>1</sup> Dipartimento di Fisica e Astronomia “Augusto Righi”, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy  
e-mail: giovanni.sabatini4@unibo.it

<sup>2</sup> INAF – Istituto di Radioastronomia - Italian node of the European ALMA Regional Centre (It-ARC), Via Gobetti 101, 40129 Bologna, Italy

<sup>3</sup> Departamento de Astronomía, Facultad Ciencias Físicas y Matemáticas, Universidad de Concepción, Av. Esteban Iturra s/n Barrio Universitario, Casilla 160, Concepción, Chile

<sup>4</sup> Universitäts-Sternwarte München, Scheinerstr. 1, 81679 München, Germany

<sup>5</sup> INAF – Istituto di Astrofisica e Planetologia Spaziali (IAPS), via Fosso del Cavaliere 100, 00133 Roma, Italy

<sup>6</sup> Max-Planck-Institut für Radioastronomie, Auf dem Hügel, 69, 53121 Bonn, Germany

<sup>7</sup> INAF – Osservatorio Astronomico di Cagliari, Via della Scienza 5, 09047 Selargius (CA), Italy

Received 1 February 2021 / Accepted 31 May 2021

## ABSTRACT

**Context.** Understanding the details of the formation process of massive (i.e.  $M \gtrsim 8\text{--}10 M_{\odot}$ ) stars is a long-standing problem in astrophysics. They form and evolve very quickly, and almost their entire formation process takes place deeply embedded in their parental clumps. Together with the fact that these objects are rare and at a relatively large distance, this makes observing them very challenging.

**Aims.** We present a method for deriving accurate timescales of the evolutionary phases of the high-mass star formation process.

**Methods.** We modelled a representative number of massive clumps of the ATLASGAL-TOP100 sample that cover all the evolutionary stages. The models describe an isothermal collapse and the subsequent warm-up phase, for which we followed the chemical evolution. The timescale of each phase was derived by comparing the results of the models with the properties of the sources of the ATLASGAL-TOP100 sample, taking into account the mass and luminosity of the clumps, and the column densities of methyl acetylene ( $\text{CH}_3\text{CCH}$ ), acetonitrile ( $\text{CH}_3\text{CN}$ ), formaldehyde ( $\text{H}_2\text{CO}$ ), and methanol ( $\text{CH}_3\text{OH}$ ).

**Results.** We find that the molecular tracers we chose are affected by the thermal evolution of the clumps, showing steep ice evaporation gradients from  $10^3$  to  $10^5$  AU during the warm-up phase. We succeed in reproducing the observed column densities of  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$ , but  $\text{H}_2\text{CO}$  and  $\text{CH}_3\text{OH}$  agree less with the observed values. The total (massive) star formation time is found to be  $\sim 5.2 \times 10^5$  yr, which is defined by the timescales of the individual evolutionary phases of the ATLASGAL-TOP100 sample:  $\sim 5 \times 10^4$  yr for 70- $\mu\text{m}$  weak,  $\sim 1.2 \times 10^5$  yr for mid-IR weak,  $\sim 2.4 \times 10^5$  yr for mid-IR bright, and  $\sim 1.1 \times 10^5$  yr for HII-region phases.

**Conclusions.** With an appropriate selection of molecular tracers that can act as chemical clocks, our model allows obtaining robust estimates of the duration of the individual phases of the high-mass star formation process. It also has the advantage of being capable of including additional tracers aimed at increasing the accuracy of the estimated timescales.

**Key words.** stars: formation – stars: massive – stars: evolution – astrochemistry – ISM: molecules – evolution

## 1. Introduction

Although high-mass stars (i.e.  $M \gtrsim 8\text{--}10 M_{\odot}$ ) are much rarer than their less massive counterparts, they have a critical effect on the physico-chemical characteristics of the interstellar medium (ISM). They also play an important role in the evolution of the host galaxies (e.g. Kennicutt 2005). Massive stars are responsible for the production of large amounts of the  $\alpha$ -elements involved in the formation of complex molecules (Woosley & Weaver 1995; Kobayashi et al. 2020) that are created during the final stages of their life-cycle through core-collapse supernovae (e.g. Smartt 2009). They also dominate the energy budget in their immediate surroundings by stirring, heating, and ionising the gas, and they affect the chemical evolution of the ISM as well as the star and planet formation process (e.g. Elmegreen 1998; Bally et al. 2005; Adams 2010). However, observing massive stars is challenging as they evolve quickly, and the initial

stages of their formation process take place when their progenitors are still embedded in the parental clump (Zinnecker & Yorke 2007; Motte et al. 2018, for reviews on this topic). In spite of the efforts made in recent years to understand the formation of massive stars, some fundamental questions still remain unanswered. One of these is related to the timescales of the various (evolutionary) phases of their formation process, which would provide crucial information to distinguish among competing star formation theories (e.g. McKee & Tan 2002; Mouschovias et al. 2006; Hartmann et al. 2012; Tigé et al. 2017; Padoan et al. 2020) and to reach a comprehensive view of the chemical evolution of the gas that resides in the parental clumps of massive stars.

Timescales are usually derived by applying statistical approaches (e.g. Wood & Churchwell 1989; Davies et al. 2011; Mottram et al. 2011; Battersby et al. 2017; Tigé et al. 2017; see also Motte et al. 2018). Typical values for the total massive star formation timescale range from  $\sim 1$  to 5 times  $10^5$  yr. The statistical lifetime of a phase is consequently obtained as the fraction of

\* Marco Polo fellowship of the University of Bologna.

the total time equal to the number of objects in that phase with respect to the total number of objects. These methods depend on the assumption of a total time for the star formation process, however, that is constrained by simulations (e.g. [Davies et al. 2011](#)) or derived with respect to the known age of OB stars (e.g. [Wood & Churchwell 1989](#); [Motte et al. 2007](#); [Russeil et al. 2010](#); [Mottram et al. 2011](#); [Tigé et al. 2017](#)).

An alternative way to quantify the timescales of the various steps is to study the effects that massive young stellar objects (YSOs) have on their immediate surroundings. The chemical evolution is affected by changes in density and temperature induced by the forming stars. During the so-called warm-up phase (e.g. [Viti & Williams 1999](#); [Viti et al. 2001](#); [Garrod et al. 2008](#)), molecules formed on the surface of the dust grains are rapidly released into the gas phase, increasing their observed abundance by several orders of magnitude ([Viti et al. 2004](#); [Garrod & Widicus Weaver 2013](#); [Choudhury et al. 2015](#)). Thus, the molecular composition of the material around YSOs contains information about the physical conditions of their past ([van der Tak 2005](#)). A well-known example are the rich molecular spectral features produced by hot molecular cores (HMCs), which are compact regions around YSOs ( $\lesssim 0.1$  pc radius; [Kurtz et al. 2000](#); [Cesaroni 2005](#)).

Chemical tracers that show a relation between their observed abundances and the different phases of the star formation process are commonly called chemical clocks (e.g. [Fontani et al. 2007](#); [Beuther et al. 2009](#); [Hoq et al. 2013](#); [Giannetti et al. 2019](#); [Urquhart et al. 2019](#); [Sabatini et al. 2020](#), as some recent examples; see also [van Dishoeck & Blake 1998](#) for a review). Through the comparison of their observed abundances obtained from large samples of massive clumps at different evolutionary stages and those predicted by accurate time-dependent chemical models, it is possible to quantify the timescales of each stage. An attempt to adopt this type of analysis was carried out by [Gerner et al. \(2014\)](#) based on the chemical properties of a sample of 59 massive clumps in different evolutionary stages, and modelling the warm-up phase alone. The radial profiles of temperature and density were assumed to be static with time and were described as power laws with the average properties of each evolutionary class (see their Sect. 5.1.1 for more details). The chemistry was then evolved in time until they found the best match with the average observed abundances. They interpreted this time as the typical age of each class. Based on these results, [Gieser et al. \(2019\)](#) studied the chemical history of the HMC VLA 3 in the high-mass star-forming region AFGL 2591 to assess the effects of different initial chemical conditions on the model results and on the derived evolutionary timescales. They showed that the latter are sensitive to the model assumptions. More recently, [Gieser et al. \(2021\)](#) used the same method to investigate 22 cores identified in 18 high-mass star-forming regions with ALMA.

An alternative to these approaches is to adopt a time-dependent analytical function of the temperature to simulate the gradual warming-up of the clumps in a range of typical observed values (i.e.  $\sim 10$ – $300$  K; e.g. [Viti & Williams 1999](#); [Garrod & Herbst 2006](#)). [Awad et al. \(2010\)](#) proposed an improved model of the warm-up phase applied to the low-mass regime, where the temperature evolution was based on the radiative transfer (RT) calculation of [Nomura & Millar \(2004\)](#). More recently, [Bonfand et al. \(2019\)](#) described the evolution of the physical parameters of the high-mass star-forming region Sgr B2 with 3D-RT simulations. They followed the trajectory of a parcel of gas under free-fall collapse and modelled its chemistry by updating the temperature and density of the gas. In this paper, we have developed a model similar to [Bonfand et al. \(2019\)](#) by employing a

time-dependent description of the thermal evolution of massive clumps during the warm-up phase, with the aim to propose a new method for deriving the evolutionary timescales of the massive star formation process.

This paper has the following structure: in Sect. 2 we present the reference sample of massive clumps, the evolutionary sequence, and the chemical tracers we employed to derive the duration of each phase. The model and chemical network are described in Sect. 3. In Sect. 4 we present our results, the post-processing procedure, and the derivation of the durations of each evolutionary phase. In Sect. 5 we discuss our estimates and how the selected tracers are reproduced by the models, providing information on the reliability of the chemical clocks. Finally, in Sect. 6 we summarise our conclusions.

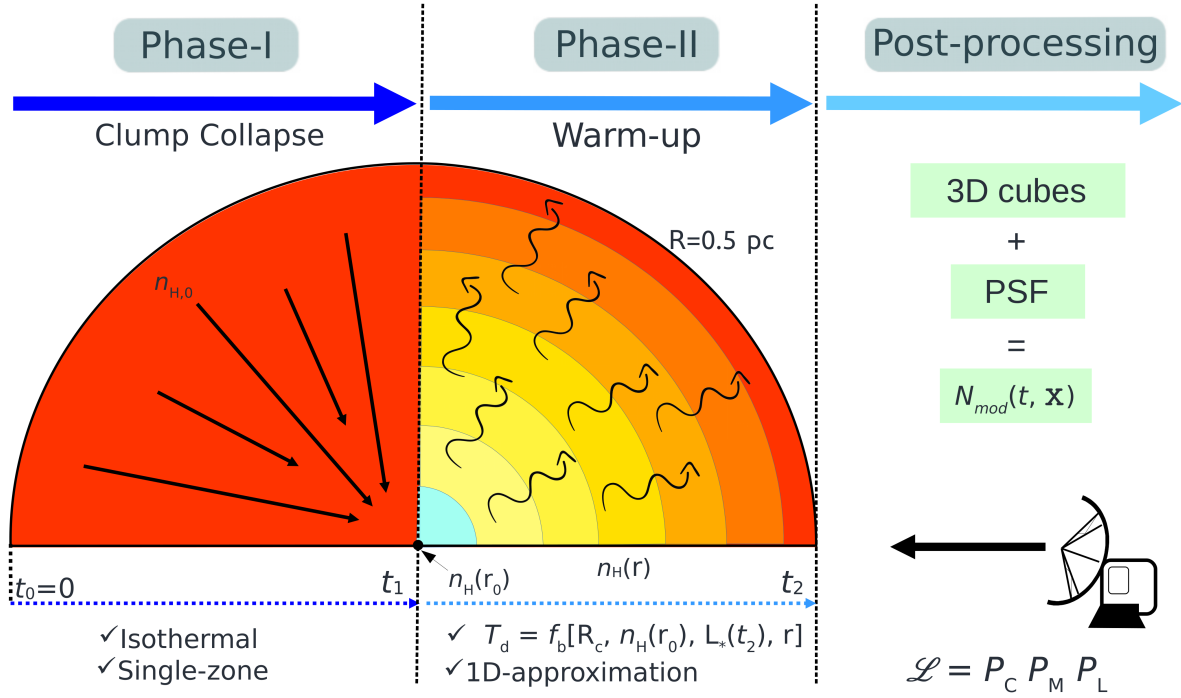
## 2. Reference sample and selected tracers

The survey we used to define the reference sample and the evolutionary stages of the massive star formation process is the APEX<sup>1</sup> Telescope Large Area Survey of the Galaxy (ATLASGAL; [Schuller et al. 2009](#)). ATLASGAL offers a complete view of the high-mass star-forming regions in the inner Galaxy at  $870 \mu\text{m}$  and provides the ideal basis for detailed studies of a large number of massive clumps in different evolutionary stages, with the unprecedented angular resolution of  $\sim 19''$  ([Contreras et al. 2013](#); [Csengeri et al. 2014](#); [Urquhart et al. 2013](#)). Estimates of luminosities, masses, kinematic distances, and dust temperature ([Urquhart et al. 2014, 2018](#) and [Wienen et al. 2015](#)) have been derived for more than  $\sim 10^4$  dust clumps.

Using the unbiased nature of ATLASGAL, the ATLASGAL-TOP100 sample (hereafter TOP100) has been defined as a flux-limited sample of high-column density clumps, selected with additional infrared (IR) criteria to include sources that potentially cover the whole spectrum of ages (see [Giannetti et al. 2014](#)). The number of sources included in the TOP100 was slightly refined by [König et al. \(2017\)](#), and to date, it includes 111 clumps. Combining the IR properties of the massive clumps in the TOP100 sample with radio-continuum measurements at 5–9 GHz, the objects were divided into the following four evolutionary classes: (1) 70  $\mu\text{m}$  weak stage (70w; i.e. quiescent), constituted by sources undetected at  $24 \mu\text{m}$  and showing no clear compact emission at  $70 \mu\text{m}$  (or are seen in absorption at this wavelength). This stage represents the earliest phase of massive-star formation and potentially includes starless or prestellar cores. (2) Mid-IR weak stage (IRw; protostellar), composed of compact and visible sources at  $70 \mu\text{m}$  that are still undetected at  $24 \mu\text{m}$  or are associated with a weak IR source ( $< 2.6$  Jy fluxes). The clumps in this evolutionary stage are young and likely dominated by cold gas. (3) Mid-IR bright stage (IRb; high-mass YSOs), made by the sources that show a strong mid-IR emission at 8 and  $24 \mu\text{m}$ , which was interpreted to be caused by the progressive dust heating induced by the forming (proto-)star(s). (4) HII regions (HII), where the sources are bright at 70 and  $24 \mu\text{m}$ , and detected in radio continuum at 5–9 GHz.

In the sample, the clump masses range from  $\sim 18$  to  $\sim 5 \times 10^4 M_{\odot}$ . The bolometric luminosities are between  $\sim 60$  and  $\sim 4 \times 10^6 L_{\odot}$  and correspond to a range in luminosity-to-mass ratio ( $L/M$ ) of  $\sim 0.2$ – $350 L_{\odot}/M_{\odot}$ . All the TOP100 clumps have the potential of forming high-mass stars and show no bias in terms of

<sup>1</sup> Atacama Pathfinder Experiment 12 meter submillimeter telescope ([Güsten et al. 2006](#)).



**Fig. 1.** Sketch of the physical model employed in this work. *Left panel:* collapse phase (see Sect. 3.1.1) solved in a single-zone approximation. The chemical output of phase I is used as initial condition of phase II. *Central panel:* warm-up phase solved in a 1D approximation from 1 to  $10^5$  AU (see Sect. 3.1.2). *Right panel:* post-processing applied to compare the final chemical outputs of phase II with the column densities observed in the TOP100 (see Sects. 4.1 and 4.2). The physical parameters assumed in phases I and II are summarised in Table 1. The time evolution of the model is indicated by the dotted blue arrows.

distance or mass between the evolutionary classes (König et al. 2017).

Four molecular tracers were studied with the APEX telescope in the TOP100 sample, revealing their potential as chemical clocks: formaldehyde ( $\text{H}_2\text{CO}$ ; Tang et al. 2018), methyl acetylene, acetonitrile, and methanol ( $\text{CH}_3\text{CCH}$ ,  $\text{CH}_3\text{CN}$  and  $\text{CH}_3\text{OH}$ , respectively; Giannetti et al. 2017). The detection rates, abundances, and excitation temperatures derived from each molecular species increase with the  $L/M$  of the clumps, which is known to trace the evolution of the star formation process (Saraceno et al. 1996; Molinari et al. 2008; Urquhart et al. 2018). With respect to the line-of-sight (LOS) and beam-averaged column densities reported by Giannetti et al. (2017) and Tang et al. (2018), we calibrated our method to estimate the duration of the four evolutionary phases defined in the TOP100 sample. We propose a general pipeline of comparisons of models and observations that can be expanded with additional tracers in future follow-ups.

### 3. Methods

In this section we present the model we developed to describe the time evolution of the abundances of the selected molecular tracers in the TOP100 survey (see Sect. 2). We describe the physical and the chemical model separately in Sects. 3.1 and 3.2, respectively.

#### 3.1. Physical model

Our physical model comprises two distinct stages that are sketched in Fig. 1. The first, namely the collapse (left panel of Fig. 1), describes the density evolution of an isothermal clump

with a single-zone approximation, exploring different conditions as in Viti & Williams (1999) and Garrod & Herbst (2006), for example. In the second phase, the warm-up (central panel of Fig. 1), the gas and dust temperatures (assumed to be in equilibrium) evolve as a function of time, driven by the luminosity of the central forming protostar. The temperature profiles are computed with the radiative transfer code MOCASSIN (see Ercolano et al. 2003, 2005), and the mass distribution of the clump is described by a static gas radial density profile (e.g. Tafalla et al. 2004).

In the following sections we describe the details of each phase. The specific details of the RT calculations will be presented in a forthcoming paper (Grassi et al., in prep.).

##### 3.1.1. Phase I: Isothermal collapse

The first phase of the model simulates a semi-analytical single-zone isothermal collapse (see Spitzer 1978; Brown et al. 1988; Viti & Williams 1999). The gas number density,  $n_{\text{H}}$ , at the centre of the clump, evolves with time as

$$\frac{dn_{\text{H}}(t)}{dt} = b \cdot \left( \frac{n_{\text{H}}(t)^4}{n_{\text{H},0}} \right)^{1/3} \left\{ C n_{\text{H},0} \left[ \left( \frac{n_{\text{H}}(t)}{n_{\text{H},0}} \right)^{1/3} - 1 \right] \right\}^{1/2}, \quad (1)$$

where  $n_{\text{H},0}$  is the initial central gas number density,  $b$  is a factor which aims to mimic a slower collapse compared to the ideal free-fall time (i.e.  $b = 1$ ),  $C = 24\pi G m_{\text{H}}$ ,  $G$  is the gravitational constant, and  $m_{\text{H}}$  is the hydrogen mass. Equation (1) is obtained assuming the conservation of mass during an isothermal contraction of a spherical clump (Spitzer 1978).

We set the initial gas number density to the typical values of a clump,  $n_{\text{H},0} = n(\text{H}) + 2n(\text{H}_2) = 10^4 \text{ cm}^{-3}$  (Bergin & Tafalla 2007;

**Table 1.** Parameter space explored in our models; i.e. phase I (Sect. 3.1.1) and phase II (Sect. 3.1.2).

	Parameter	Values	Unit
(1)	$b$	1; 0.5; 0.1	–
(2)	$n_{\text{H}}(r_0)$	$10^5$ ; $10^6$ ; $10^7$ ; $10^8$	$\text{cm}^{-3}$
(3)	$R_{\text{c}}$	$10^4$ ; $10^{4.5}$ ; $10^5$	AU
(4)	$\dot{M}$	$10^{-5}$ ; $10^{-3}$	$M_{\odot} \text{ yr}^{-1}$

Gerner et al. 2014). The collapse assumes a constant temperature of 15 K (for gas and dust temperatures; e.g. König et al. 2017) and a visual extinction  $A_{\text{v}} = 10$  mag (e.g. Semenov et al. 2010; Reboussin et al. 2014). The cosmic-ray ionisation rate of hydrogen molecules was set to  $\zeta_2 = 5 \times 10^{-17} \text{ s}^{-1}$ , as observationally constrained by van der Tak & van Dishoeck (2000) in high-mass star-forming regions, and in agreement with the results reported by Sabatini et al. (2020). The specific density of the dust grains was  $\rho_0 = 3 \text{ g cm}^{-3}$ , typical of silicates (e.g. Draine & Lee 1984), the dust-to-gas ratio  $\mathcal{D} = 10^{-2}$ , and we used a constant grain size  $\langle a \rangle = 0.1 \mu\text{m}$ . The gas mean molecular weight was  $\mu = 2.4$ .

To simulate various physical conditions for the environment in which the seeds of massive protostars are formed, the collapse was stopped at different final densities  $n_{\text{H}}(r_0)$  (see Fig. 1) that were observationally constrained (Table 1; e.g. Mueller et al. 2002; Sabatini et al. 2019) and were then used as central densities during the warm-up phase.

### 3.1.2. Phase II: Warm-up

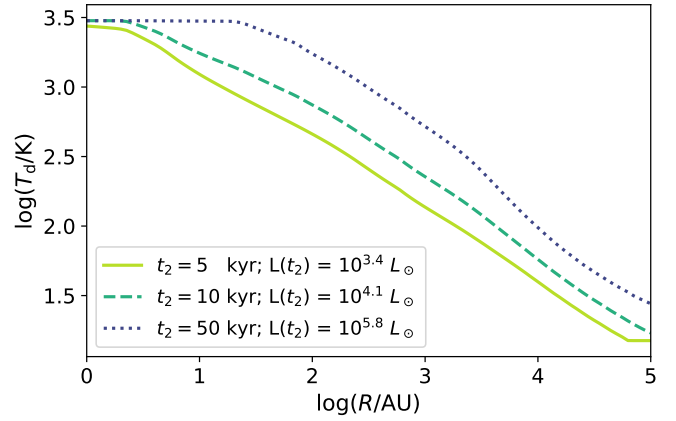
The second phase of the model, sketched in the central panel of Fig. 1, simulates the warm-up induced by a protostar at the centre of a spherical clump by using a 1D approximation of 100 logarithmic radial steps from 1 to  $10^5$  AU (i.e. up to a typical clump size of  $\sim 0.5$  pc; e.g. Motte et al. 2018).

The mass distribution of the core is described by the gas radial density profile (e.g. Tafalla et al. 2002, 2004),

$$n_{\text{H}}(r) = n_{\text{H}}(r_0) \frac{R_{\text{c}}^{5/2}}{R_{\text{c}}^{5/2} + r^{5/2}}, \quad (2)$$

where  $n_{\text{H}}(r_0)$  is the central number density at  $r_0$ , at which phase I is stopped (see second row in Table 1). This ensures continuity between the two physical phases of the model. In Eq. (2),  $R_{\text{c}}$  is the core radius, that is, the radius at which  $n_{\text{H}}(R_{\text{c}}) = 0.5 n_{\text{H}}(r_0)$ . The slope 5/2 agrees with the observations in massive star-forming regions (e.g. Mueller et al. 2002; Schneider et al. 2015), while similar profiles were successfully applied to model the  $\text{H}_2$  distribution in clumps and filamentary structures (e.g. Beuther et al. 2002a; Arzoumanian et al. 2011; André et al. 2016; Sabatini et al. 2019). Alternatively, power-law profiles can be assumed to model the  $\text{H}_2$  distribution especially on the clumps scale (e.g. Gieser et al. 2021). This provides minor differences in the final clump mass (a factor of  $\lesssim 2$  smaller for a slope of 5/2).

The physical parameters of phase II (i.e.  $\zeta_2$ ,  $\mathcal{D}$ ,  $\mu$ , and  $\langle a \rangle$ ) were assumed to be the same as those of the collapse phase, while the visual extinction at each radius  $r$ , was  $A_{\text{v}}(r) = N(\text{H}, r)/(2 \times 10^{21} \text{ cm}^{-2})$ , where  $N(\text{H}, r)$  is the column density of the hydrogen nuclei, obtained by integrating Eq. (2) from the edge of the clump (e.g. Tielens 2010; Zhu et al. 2017). We took the final abundances of phase I as initial chemical conditions for each radial grid point of phase II, rescaled by the



**Fig. 2.** Dust temperature radial profile at different times for an arbitrary model with the gas number density profile described by Eq. (2) and  $R_{\text{c}} = 10^{4.5}$  AU and  $n_{\text{H}}(r_0) = 10^7 \text{ cm}^{-3}$ . With time,  $T_{\text{d}}$  increases, driven by the protostar luminosity. The flat inner part represents the sublimation of the dust grains. The minimum temperature is set to 15 K to ensure continuity with the collapse phase.

density profile. The chemistry was evolved in time, assuming that the individual cells are independent throughout the entire time evolution.

To model the thermal evolution of the clump during the warm-up phase, we generated a set of models with different density distributions and protostar luminosities by employing the Monte Carlo radiative transfer code MOCASSIN. The models provide an interpolable look-up table  $T_{\text{d}}(r, t_2) = f_{\text{b}}[R_{\text{c}}, n_{\text{H}}(r_0), L_{\ast}(t_2), r]$ , where  $L_{\ast}(t_2)$  is the luminosity of the protostar and  $t_2$  the time spent since the start of the warm-up phase. The function that describes how  $L_{\ast}(t_2)$  evolves in time was derived by Hosokawa & Omukai (2009) and depends on the mass accretion rate,  $\dot{M}$ , assumed for the protostar. Based on the very few observational estimates (e.g. Beuther et al. 2002b; Herpin et al. 2012; Wang et al. 2013), we explored two typical cases:  $\dot{M} = 10^{-5} M_{\odot} \text{ yr}^{-1}$  and  $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$  (see Figs. 4 and 12 in Hosokawa & Omukai 2009). These values also agree with the results of Peters et al. (2011), who simulated the collapse of a magnetised rotating molecular cloud of  $1000 M_{\odot}$ . During phase II, we assumed that the gas radial density profile of the clump and its dust content remained constant over time so that the mass accretion rate of the protostar was completely balanced by the average infall rate observed in a massive clump (e.g. Schneider et al. 2010; Peretto et al. 2013; Wyrowski et al. 2016; Liu et al. 2018).

Analogously to phase I, we assumed that the gas and dust temperatures were in equilibrium; this assumption is less accurate at small radii (i.e.  $\lesssim 50$  AU), where the temperatures are close to the dust sublimation limit, and at large radii (i.e.  $\sim 10^5$  AU), where the gas-dust collision term is subdominant with respect to the radiation coupling (Draine 2011). However, neither region is relevant in our analysis and does not affect our findings. An example of how the  $T_{\text{d}}$  radial profile evolves is shown in Fig. 2 for three (typical) times,  $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$ , and for the following parameters  $R_{\text{c}} = 10^{4.5}$  AU and  $n_{\text{H}}(r_0) = 10^7 \text{ cm}^{-3}$ . For a qualitative comparison of the  $T_{\text{d}}$  radial profiles shown in Fig. 2, we refer to the recent results of Gieser et al. (2021), who derived the temperature structure of 18 massive star-forming regions based on two of the four tracers considered in this work (i.e.  $\text{H}_2\text{CO}$  and  $\text{CH}_3\text{CN}$ ; Sect. 2). In this case, the authors assumed a power law (i.e.  $T(r) \propto r^{-q}$ ; see their Eq. (1)) to fit the radial temperature in each source, providing an average slope of  $|q| = 0.4 \pm 0.1$ . Our

**Table 2.** Summary of the total masses of each model.

$n_{\text{H}}(r_0)$ [cm <sup>-3</sup> ]	$R_c/\text{AU}$		
	10 <sup>4</sup>	10 <sup>4.5</sup>	10 <sup>5</sup>
10 <sup>5</sup>	<b>0.4</b> $M_{\odot}$	<b>3.8</b> $M_{\odot}$	18.5 $M_{\odot}$
10 <sup>6</sup>	<b>3.5</b> $M_{\odot}$	38.4 $M_{\odot}$	184.6 $M_{\odot}$
10 <sup>7</sup>	34.7 $M_{\odot}$	383.7 $M_{\odot}$	1846.8 $M_{\odot}$
10 <sup>8</sup>	346.6 $M_{\odot}$	3836.7 $M_{\odot}$	18 468.2 $M_{\odot}$

**Notes.** The models with masses lower than the lowest mass associated with the TOP100 sources (König et al. 2017) are marked in boldface and were excluded.

profiles in Fig. 2 show a similar slope ( $|q| \sim 0.5$ ) at distances larger than the dust sublimation radius (i.e. the flat region in the same figure) and comparable temperatures at their fiducial radius (see Table 3 in Gieser et al. 2021). A more detailed description of the effect of  $L_*(t_2)$  and  $n_{\text{H}}(r)$  on  $T_{\text{d}}(r, t_2)$  will be discussed in a forthcoming paper (Grassi et al., in prep.).

To compare our results with the observed properties of the TOP100 clumps (Giannetti et al. 2014; König et al. 2017), we evolved one model for each mass in Table 2 (derived from  $n_{\text{H}}(r_0)$  and  $R_c$  in Table 1), and for each value of  $b$  and  $\dot{M}$  in Table 1. We considered only clumps with a total mass higher than 15  $M_{\odot}$ , which is the lowest mass associated with a TOP100 source, for a total of 54 models.

### 3.2. Chemical model

The chemical evolution of a massive clump (i.e. the collapse and the warm-up phases) is described by employing the publicly available time-dependent code KROME<sup>2</sup> (Grassi et al. 2014).

Adsorption and desorption<sup>3</sup> processes were included as in Hasegawa et al. (1992) and Hasegawa & Herbst (1993). The surface reactions between two species  $i$  and  $j$  follow Semenov et al. (2010), where the rate coefficient, in units of cm<sup>3</sup> s<sup>-1</sup>, is

$$k_{ij}^s = \frac{P_{ij}(k_{\text{diff}}^i + k_{\text{diff}}^j)}{N_{\text{sites}}n_{\text{d}}}, \quad (3)$$

with  $k_{\text{diff}}^i = v_0^i \exp(-T_{\text{diff}}^i/T_{\text{d}})$  the diffusion through thermal hopping ( $T_{\text{diff}}^i = 0.77T_{\text{b}}^i$ ; Semenov et al. 2010),  $T_{\text{d}}$  is the dust temperature,  $v_0^i = (2n_{\text{S}}E_{\text{b}}^i/\pi^2m_i)^{1/2}$  is the characteristic Debye vibration frequency for the adsorbed species,  $n_{\text{S}} = 1.5 \times 10^{15} \text{cm}^{-2}$  is the surface density of binding sites,  $m_i$  is the mass of the species, and  $E_{\text{b}}^i = k_{\text{B}}T_{\text{b}}^i$  is the binding energy of the  $i$ th species on the binding site. In Eq. (3),  $n_{\text{d}} = \mathcal{D}n_{\text{H}}m_{\text{H}}\mu/M_{\text{d}}$  is the dust number density, where  $M_{\text{d}} = 4/3\pi\rho_0 \langle a \rangle^3$  is the dust mass, and the total number of binding sites of a grain  $N_{\text{sites}} = 4\pi \langle a \rangle^2 / a_{\text{pp}}^2$  assumes an average distance between two contiguous sites of  $a_{\text{pp}} = 3 \text{ \AA}$  (Hocuk & Cazaux 2015).

The probability for a reaction to occur is  $P_{ij} = \alpha_{ij} \exp(-E_{\text{a}}/k_{\text{B}}T_{\text{d}})$ , where  $E_{\text{a}}$  is the activation energy of the reaction and  $\alpha_{ij}$  is a parameter that depends on the number and on the type of species in the reaction. In the case of exothermic reactions (i.e.  $E_{\text{a}} = 0$ ), (I) if  $i \neq j$ ,  $P_{ij} = 1$ ; (II) if  $i = j$ ,  $P_{ij} = \alpha_{ij} = 1/2$ . For endothermic reactions (i.e.  $E_{\text{a}} \neq 0$ ),  $\alpha_{ij}$  is the inverse of the number of paths in the branching ratios.

<sup>2</sup> <https://bitbucket.org/tgrassi/krome/wiki/Home>

<sup>3</sup> We include thermal and cosmic-ray induced desorption.

**Table 3.** Summary of fiducial initial elemental abundances,  $n_i$ , with respect to the abundance of H-nuclei,  $n_{\text{H}}$ .

Species	$(n_i/n_{\text{H}})_{t=0}$	Species	$(n_i/n_{\text{H}})_{t=0}$
H <sub>2</sub>	5.00(-1)	Si	1.95(-6)
He	1.00(-1)	S	1.50(-6)
O	1.36(-4)	Fe	7.40(-7)
CO	1.20(-4)	P	2.30(-8)
N	1.05(-5)	Na	2.00(-8)
N <sub>2</sub>	5.25(-6)	Cl	1.40(-8)
Mg	2.55(-6)	H <sub>3</sub> <sup>+</sup>	3.18(-9)

**Notes.** The “A(-B)” notation assumed in the table means “A × 10<sup>-B</sup>”.

The final chemical network was derived from Semenov et al. (2010)<sup>4</sup> and contains 654 chemical species and 5869 gas-phase, gas-grain, and grain-surface reactions. The photochemical rate coefficients follow the  $A_{\text{v}}$  formalism as in Draine (1978) (see KIDA<sup>5</sup>). In order to ensure that each depleted species was released into the gas phase, we modified the original network by adding 70 missing desorption processes, with the binding energy values updated to the most recent estimates in KIDA (see Table A.1 and Table A.2 for the reactions, and Appendix B for the network benchmark).

The abundances of chemical species evolve with time, starting from the assumed initial conditions following the recent large-scale simulations of molecular clouds (e.g. Hocuk et al. 2016; Clark et al. 2019), showing that CO is already formed at densities of few × 10<sup>3</sup> cm<sup>-3</sup>. Following these findings, we assumed H, C, and O to be in molecular form. In particular, the abundances of H<sub>2</sub>, H<sub>3</sub><sup>+</sup>, He, N, O, CO, and N<sub>2</sub> were taken from Bovino et al. (2019), while for the other elements, we refer to Garrod & Herbst (2006), as reported in Table 3.

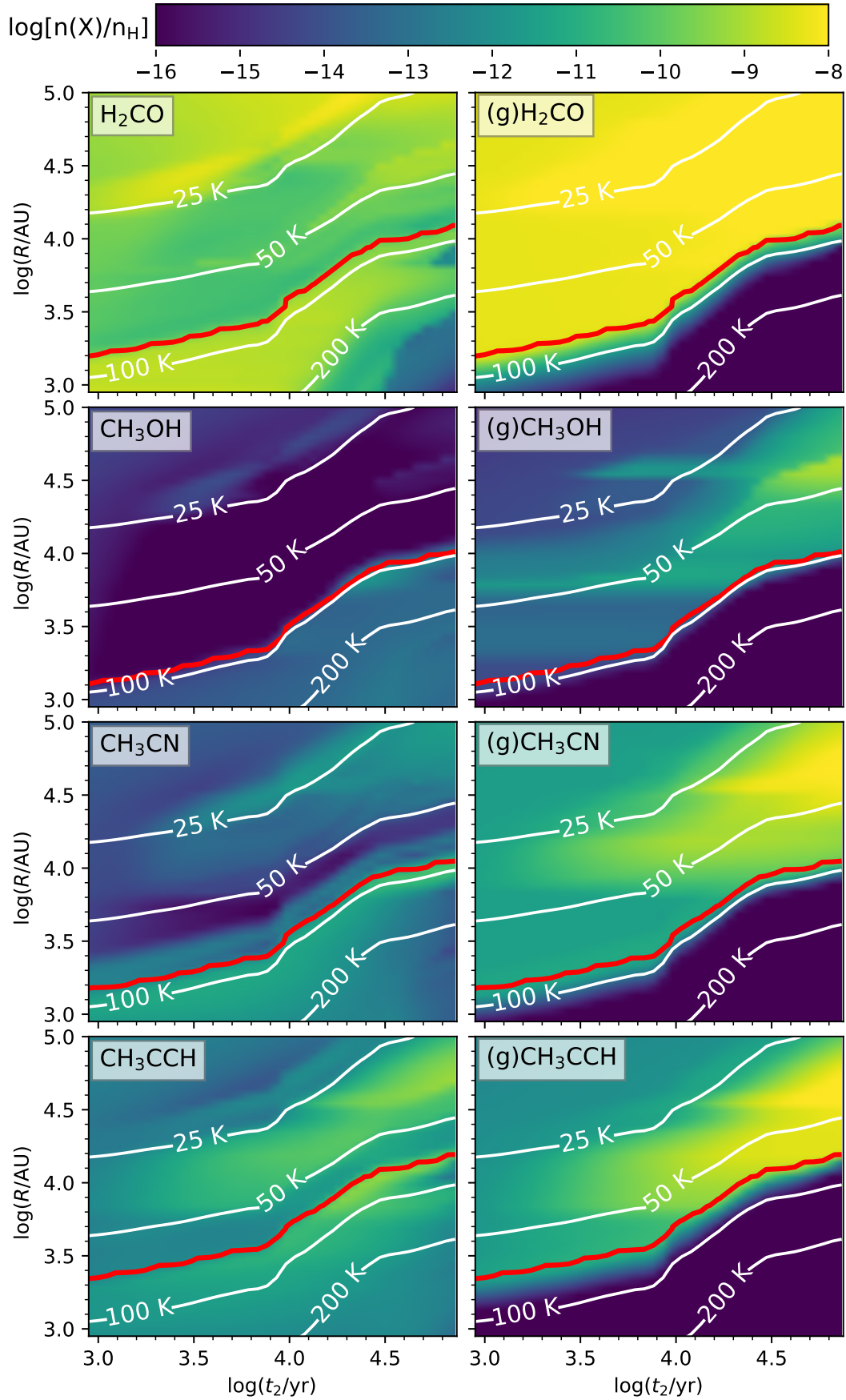
## 4. Results

An example of the chemical evolution of some observed species during phase I is reported in Appendix C, while here we focus on phase II and the comparison with observations. However, because the outcome of phase I represents the initial chemical conditions for phase II, a few considerations for phase I should be made. The fractional abundances are considerably affected by the dynamics of the collapse, showing a difference up to ~3 orders of magnitudes between the fastest ( $b = 1$ ) and the slowest ( $b = 0.1$ ) collapse, even if the global trend remains generally unaltered (more details in Appendix C). Based on the continuity between phase I and phase II, the final effect that  $b$  has on the models is to provide different inputs for the warm-up phase. For this reason, we ran phase II for each of the 54 models defined by the parameter space, taking the effect of parameter  $b$  on the second phase of the model into account (see also Sect. 3.1.2).

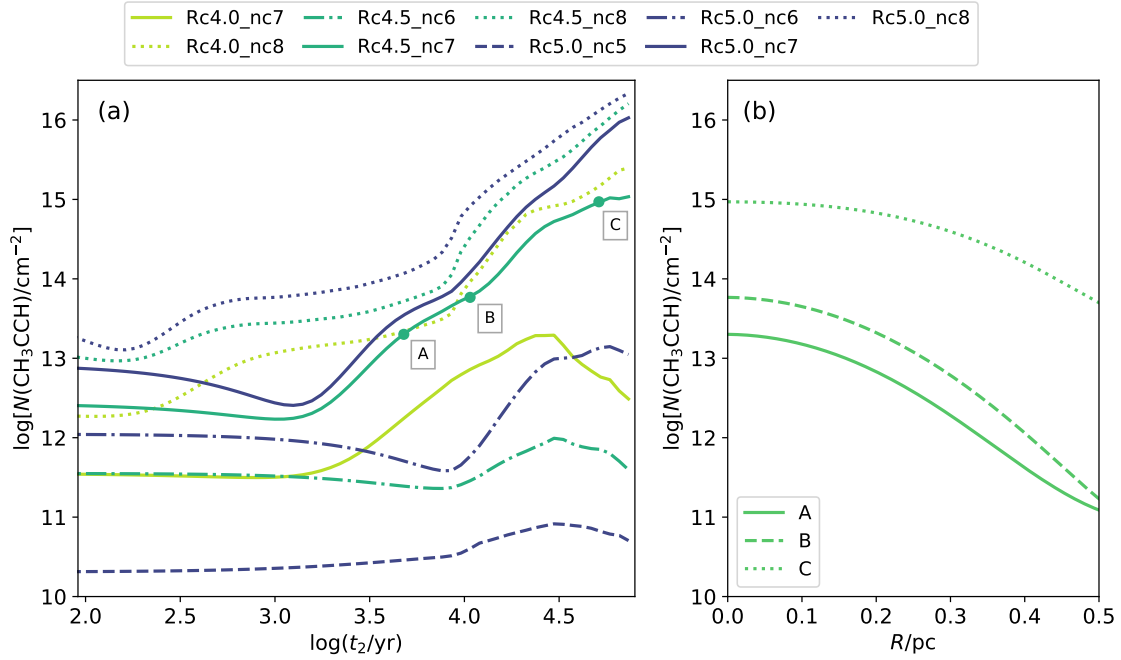
Figure 3 shows the abundances of H<sub>2</sub>CO, CH<sub>3</sub>CCH, CH<sub>3</sub>CN, and CH<sub>3</sub>OH as a function of time and distance in the gas phase (left panels) and on grains (right panels; labelled ‘(g)’) as a result of phase II. The white contours show the temperature, and the red curves the evaporation fronts, namely the distance from the protostar at which the desorption timescale of a given

<sup>4</sup> Ohio StateUniversity (OSU) chemical network, version March 2008 ([http://www.mpia.de/homes/semenov/Chemistry\\_benchmark/model.html](http://www.mpia.de/homes/semenov/Chemistry_benchmark/model.html))

<sup>5</sup> <http://kida.astrophy.u-bordeaux.fr/>



**Fig. 3.** Temporal and radial evolution of the abundances of the tracers observed in the TOP100 (Giannetti et al. 2017; Tang et al. 2018) during phase II in the gas phase (*left panels*) and on dust (*right panels*) for the same reference model as in Fig. 2. White contours indicate the temperature computed with MOCASSIN. Red curves corresponds to the evaporation front of the given tracer.



**Fig. 4.** *Panel a:* example of the LOS gas-phase column density profiles obtained at the end of the post-processing described in Sect. 4.1 for models with  $b = 1$  and  $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$  and for an arbitrarily chosen source at a distance of 3.5 kpc. For the U\_W items in the legend, U and W are  $R_c$  (colours) and  $n_{\text{H}}(r_0)$  (styles), respectively. *Panel b:* radial distribution of the gas-phase column densities extracted at the positions A, B, and C in panel a, i.e. the same times as in Fig. 2.

species (defined as the inverse of the desorption rate, involving the thermal and the cosmic-ray induced desorption; Semenov et al. 2010) is shorter than the dynamical timescale of the model. As a consequence, at distances larger than the evaporation front, the abundances of the tracers on dust grains (Fig. 3) increase rapidly. At the same time, the temperatures at these distances are low enough to enhance two-body reactions on the surface of dust grains. Hydrogenation chains play a major role in the formation of CH<sub>3</sub>CCH, CH<sub>3</sub>CN, and H<sub>2</sub>CO on dust grains, and methanol is also formed through the reaction (g)CH<sub>3</sub> + (g)OH → (g)CH<sub>3</sub>OH.

Each evaporation front grows as a function of time following the temperature evolution and ranges between  $\sim 10^3$  and  $\sim 10^5$  AU. Our findings agree with those of Choudhury et al. (2015), where the efficiency of the evaporation as a function of the thermal evolution of the clump is responsible for evaporation scales similar to ours (see their Fig. B1).

#### 4.1. Post-processing of the model outputs

The chemical structure of each clump was reconstructed assuming that each radial profile from phase II (i.e. each column in the relative abundance maps in Fig. 3) represents a radius of a spherically symmetric clump. For each profile we generated a data cube of  $150 \times 150 \times 150$  pixels at a resolution of  $\sim 10^3$  AU pixel<sup>-1</sup>, which corresponds to one-half of the size of the smallest evaporation front radius in Fig. 3, in order to always sample the region at which the gas-phase abundances are enhanced by the desorption of the products of dust-phase chemistry. The column density maps of each tracer were then obtained by integrating the abundances along the LOS, applying for each tracer a convolution to the APEX resolution of the observed transitions<sup>6</sup> (see Appendix D), and taking the different distances

<sup>6</sup> This step was done using the `fft2` function of `numpy.fft`: <https://numpy.org/doc/stable/reference/generated/numpy.fft.fft2.html>

of the TOP100 objects into account. The final column density profile of a given tracer X,  $N_{\text{mod}}(\text{X})$ , was calculated along the LOS at the centre of the clump.

With respect to this procedure, CH<sub>3</sub>OH and CH<sub>3</sub>CN required an additional constraint to reproduce the observed higher-K lines. In their spectral fit, Giannetti et al. (2017) distinguished the hot and cold components, assuming two domains separated at 100 K and providing the individual column densities. Following the same approach provides additional constraints to reduce the uncertainties and to test the reliability of the temperature of the background model, especially in the innermost part of the clump. After reconstructing the spherical distribution of CH<sub>3</sub>CN and CH<sub>3</sub>OH, we applied a 100 K temperature-mask, taking only regions into account in which  $T > 100$  K, to generate the averaged column densities profiles of the hot components alone.

As an example, we report the final gas-phase CH<sub>3</sub>CCH profiles in Fig. 4a, obtained assuming a distance of 3.5 kpc (arbitrarily chosen) to apply the post-processing. We show the results for  $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$  and for the standard free-fall collapse model (i.e.  $b = 1$ ), and we note that  $n_{\text{H}}(r_0)$  is the parameter that mostly affects the evolution of  $N_{\text{mod}}(\text{CH}_3\text{CCH})$ .

This behaviour is the consequence of two combined effects. First, the initial abundances of the warm-up phase are scaled by the density profile of each model, and thus, models with higher values of  $n_{\text{H}}(r_0)$  have higher initial column densities (see Fig. B.2). Second, the different thermal structures of the clumps have a strong effect on the chemical evolution. At a given time, the average temperature along the LOS is lower in clumps with larger  $n_{\text{H}}(r_0)$  because the radiation of the protostar is more attenuated. Because the luminosity of the protostar increases with time and affects the temperature of the surrounding gas, the evaporation front grows and so does the number of chemical species released into the gas. In Fig. 4b we show the radial profile of gas-phase  $N(\text{CH}_3\text{CCH})$  at three times (A, B, and C in panel a).

The column density of this species increases by about one order of magnitude over time, caused by the expanding evaporation front and by the decrease in  $\text{CH}_3\text{CCH}$  adsorption rate in the outer parts of the model. The adsorption process is directly proportional to the density, which drops at  $r > R_c$  (Fig. 3). This also provides an explanation for the observed increasing trend in the gas-phase  $\text{CH}_3\text{CCH}$  abundance with the evolutionary stage of massive clumps (e.g. Molinari et al. 2016; Giannetti et al. 2017).

#### 4.2. Comparison of the modelled column densities with sources from the dTOP100 sample

To infer the ages of the clumps and the relative duration of the evolutionary phases in the TOP100 sample, we compare the column densities derived from the observations (Appendix D and Table D.1) with those from our post-processed models. For a given clump the dataset,  $D = \{L, M, N(\mathbf{X})\}$ , comprises eight (two+six) measurements, that is, the luminosity  $L$ , the mass  $M$ , and the column densities of the six observed chemical tracers,  $N(\mathbf{X})$ . At each time step of the warm-up phase, with the free parameters  $\theta = \{b, \dot{M}, R_c, n_{\text{H}}(r_0)\}$  given the data  $D$ , the likelihood is

$$\mathcal{L}(\theta|D) = P_C(\theta|D) P_M(\theta|D) P_L(\theta|D), \quad (4)$$

where  $P_C(\theta|D)$ ,  $P_M(\theta|D)$ , and  $P_L(\theta|D)$  are probability density distributions of the chemical abundances, the mass, and the bolometric luminosity of each clump, respectively.

The first factor in Eq. (4) is

$$P_C = \prod_X P_{C,X}, \quad (5)$$

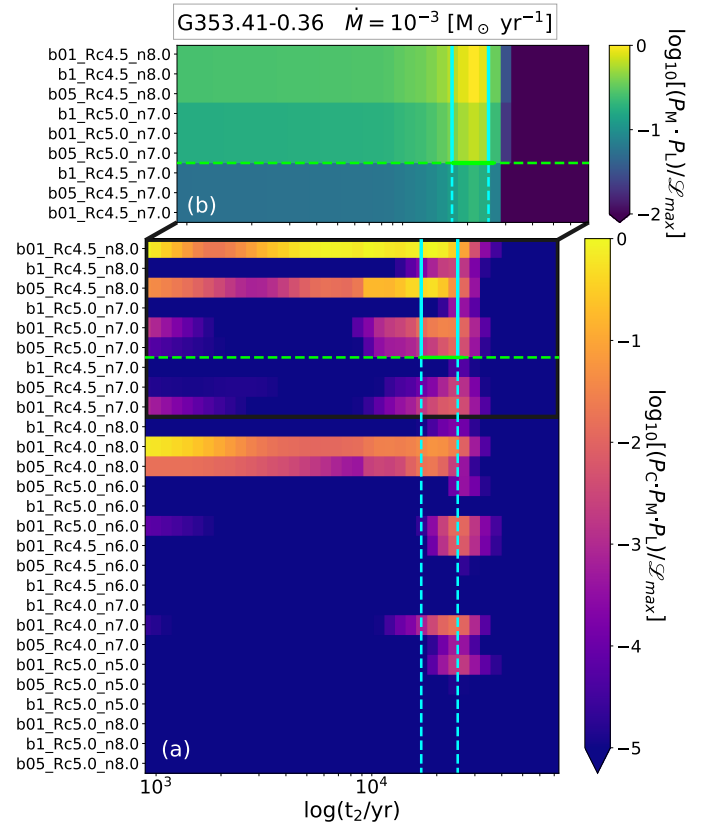
where for the sake of simplicity, we omitted the arguments of  $P_C$ , and where (Garrod et al. 2007; Gerner et al. 2014)

$$P_{C,X} = \text{erfc} \left\{ \frac{|\log_{10}[N_{\text{obs}}(\mathbf{X})] - \log_{10}[N_{\text{mod}}(\mathbf{X})]|}{\sqrt{2}\sigma} \right\}, \quad (6)$$

with ‘erfc’ the complementary error function. In Eq. (6),  $N_{\text{obs}}(\mathbf{X})$  and  $N_{\text{mod}}(\mathbf{X})$  are the observed and modelled column densities, respectively. As in Garrod et al. (2007), we set  $\sigma = 1$ , which corresponds to a difference of one order of magnitude between the observed and the modelled column densities. This assumption takes into account the uncertainties on the observed column densities, on the initial chemical conditions assumed for the phase I, and on the shape of the density profile assumed in phase II (see Sect. 3.1.2). For sources without a detection, we assumed a detection limit (see Appendix D) and set  $P_{C,X} = 1$  when  $N_{\text{mod}}(\mathbf{X}) < N_{\text{obs}}(\mathbf{X})$ , while we used Eq. (6) when  $N_{\text{mod}}(\mathbf{X}) > N_{\text{obs}}(\mathbf{X})$ .

For each clump,  $P_L$  and  $P_M$  are Gaussian functions with mean equal to the  $L$  and  $M$  values derived by König et al. (2017). The standard deviations were computed as follows: We added in quadrature a 50% uncertainty associated with the models based on the approximation of a single clump size of 0.5 pc and a constant accretion rate over time (Sect. 3.1.2), in addition to the errors of 50% on  $L$  and 20% on  $M$  proposed by Urquhart et al. (2018). This gives a standard deviation for  $P_L$  of about 60% of  $L$ , and for  $P_M$  of about 50% of  $M$ . We note that an additional variation of 10% on these uncertainties does not produce significant variations in our final results.

Figure 5a shows an example of how  $\mathcal{L}$  varies as a function of time when models with the same mass accretion rate ( $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$ ) are considered. The models are sorted in



**Fig. 5.** Panel a: example of the  $\mathcal{L}$  matrix calculated for G353.41-0.36 in all the models with  $\dot{M} = 10^{-3} M_{\odot} \text{ yr}^{-1}$ . Models on the y-axis are sorted from the top to the bottom by the difference between the mass of the models (see Table 2) and the value associated in the TOP100 (see Table D.1). For the sake of clarity, the colour bar is limited to five orders of magnitude and is normalised to maximum likelihood ( $\mathcal{L}_{\text{max}}$ ). Panel b: zoom on the age limits identified by  $P_L$  and  $P_M$  showing how the likelihood would appear if  $P_C$  in Eq. (4) were not taken into account. Cyan vertical lines indicate the range of time that corresponds to the uncertainties on the observed luminosity in the TOP100, and the horizontal green line shows the same limit associated with the clump mass. Times and  $\mathcal{L}$  are shown in log-scale.

ascending order by the difference in mass between the values reported in König et al. (2017) and those of the models (i.e. Table 2), while  $t_2$  is the time spent in the warm-up phase defined in Sect. 3.1.2. In both panels,  $P_L$  defines the range of time spent by the protostar in the warm-up phase (i.e. the range of time in between the vertical cyan lines), which only depends on the uncertainties associated with the bolometric luminosity of each source. Analogously,  $P_M$  is constrained by the mass range found for a given source (e.g. see discussions in König et al. 2017 and Urquhart et al. 2018), and it is therefore possible to define a lower mass limit that is indicated by the horizontal green line. The combined information given by  $P_L$  and  $P_M$  limits the age range of each source. In this area, the models with the same density profile defined by Eq. (2), but different values of  $b$ , are degenerate, showing the same  $\mathcal{L}$  as a function of  $t_2$  (see Fig. 5b).  $P_C$  is fundamental to break this degeneracy (see Fig. 5a), providing further information on the dynamics of the collapse that led to the observed chemical properties of each clump, and placing additional constraints on determining the relative duration of the evolutionary classes in the TOP100.

The absolute time is hence the sum of the collapse timescales,  $t_1$ , and the time spent during the warm-up phase,  $t_2$ . The collapse time depends on the setup of phase I, and it ranges

between  $(0.4\text{--}5) \times 10^6$  yr depending on  $b$  and  $n_{\text{H}}(r_0)$ , while the warm-up time depends on the mass accretion rate of phase II (see Sect. 3.1.2).

The absolute age for the  $\ell$ th source is the weighted mean,

$$t_{\text{A},\ell} = \left[ \sum_{i,j} \frac{\log_{10}[t_{1,i,\ell} + t_{2,j,\ell}]}{(\log_{10}\mathcal{L}_{i,j,\ell})^2} \right] \left[ \sum_{i,j} \frac{1}{(\log_{10}\mathcal{L}_{i,j,\ell})^2} \right]^{-1}, \quad (7)$$

where  $i$  ranges over the 54 models,  $j$  over the 100 times used to sample phase II, and we used the log-likelihood as weight.

We performed the comparison described above only for the sources in the TOP100 sample that agreed within the uncertainties with the masses of the models in Table 2. This reduced the sample to 48 objects, that is, 6 sources in the 70w stage, 16 IRw, 15 IRb, and 11 HII regions. Nevertheless, this limitation does not affect our approach because our model employs a subset of objects that are representative of the whole population, preserving their global features (e.g. Csengeri et al. 2016; Giannetti et al. 2017; König et al. 2017). We summarise the observed properties of the clumps in this subsample in Table D.1. It is worth noting that while the TOP100 sample does not show bias in terms of the source distances for each evolutionary phase (e.g. König et al. 2017), in Table D.1 only the IRw class extends beyond  $\sim 10$  kpc. To verify whether the distance has an effect on our results, we performed a test considering sources up to a maximum distance of 6 kpc. We did not find significant changes in the final durations reported in Sect. 5.

## 5. Discussion

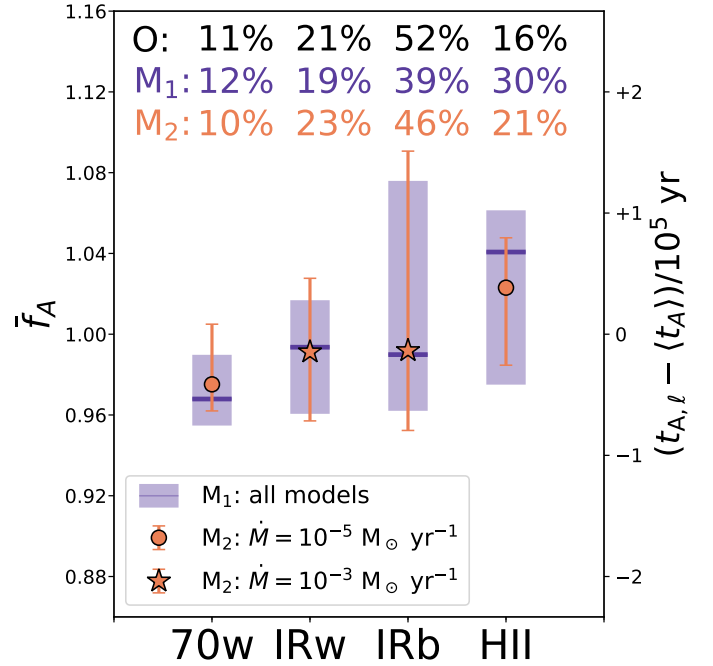
In this section, we estimate the relative duration of the evolutionary phases identified in the TOP100, following a different approach compared to the statistical lifetimes. Our findings are then compared with the relative number of objects observed in ATLASGAL (Urquhart et al. 2018), and we discuss whether the synthetic column densities of the selected chemical tracers match the observed densities.

### 5.1. Estimates of the duration of evolutionary phases

To compare the ages of the evolutionary stages defined in the TOP100 with the average absolute time estimated in a sample of  $N$  objects,  $\langle t_{\text{A}} \rangle = N^{-1} \sum_{n=1}^N t_{\text{A},n}$ , we defined the age factor of the  $\ell$ th source as

$$f_{\text{A},\ell} = \frac{t_{\text{A},\ell}}{\langle t_{\text{A}} \rangle}. \quad (8)$$

In this context, the models provide  $\langle t_{\text{A}} \rangle \sim 10^6$  yr starting from  $t_1 = 0$  (i.e. the beginning of phase I; see also Fig. 1). The average  $f_{\text{A}}$  were estimated as the median of the age-factor distributions in each evolutionary class, computed with Eq. (8) and reported as purple markers in Fig. 6. For the sake of clarity, Fig. 6 shows the corresponding value of  $t_{\text{A},\ell} - \langle t_{\text{A}} \rangle$  yr that is associated with each  $f_{\text{A}}$ . We quantified the duration of an evolutionary phase,  $\Delta t_{\text{phase}}$ , as the time between the minimum and maximum value of  $t_{\text{A},\ell} - \langle t_{\text{A}} \rangle$  (i.e. the lower and upper limits of the purple shaded areas in Fig. 6). These values were computed as the 5th and 95th percentiles of the  $f_{\text{A}}$  distributions in each evolutionary phase. The sum of the four  $\Delta t_{\text{phase}}$  is the total time of the high-mass star formation process  $t_{\text{MSF}}$ . The purple percent values in Fig. 6 ( $M_1$ ) indicate the contribution of each phase to the total time  $t_{\text{MSF}}$ .



**Fig. 6.** Summary of the final durations estimated in each evolutionary class defined for the TOP100. The relative number of massive clumps in each phase and observed in ATLASGAL are shown in black (“O”; Urquhart et al. 2018). Purple shaded areas indicate the 5th and 95th percentiles of the age factor,  $f_{\text{A}}$ , and the corresponding value in yr in our sample (model “ $M_1$ ”), and purple markers (“-”) represent their medians. The numbers associated with the model “ $M_2$ ” (orange), indicate the same quantities obtained by separating the models with respect to  $\dot{M}$  before calculating the average age of each source (see text). The different markers in  $M_2$  show the combination of  $\dot{M}$  that best match the observed relative number of object in each phase (legend).

We find that following the classification of Giannetti et al. (2014) and König et al. (2017), 12% of the star formation time is spent in the early phase (70w), and the IRw stage is associated with 19% of  $t_{\text{MSF}}$ . This suggests a fast evolution during the early stage of the massive star formation process, which in total correspond to the  $\sim 30\%$  of  $t_{\text{MSF}}$ . Advanced stages (i.e. IRb sources) have the longest duration (39%), while the remaining 30% of the  $t_{\text{MSF}}$  is spent in the final stage (HII).

Assuming that the number of objects in an evolutionary stage is also representative of its duration, we compared our findings with the relative number of objects per evolutionary class in ATLASGAL. This is represented as black percentages in Fig. 6 (“O”; from Urquhart et al. 2018). The duration found for the early stages agrees with the observed classification, and the later stages are probably biased by the different definition of IRb and HII in the TOP100 and ATLASGAL. In particular, when radio continuum emission is found at either 4 or 8 GHz within  $10''$  of the ATLASGAL peak, the source is classified as HII in the TOP100 (e.g. König et al. 2017). Different criteria have been applied in Urquhart et al. (2018) to classify the advanced stages that contain radio bright HII regions, massive young stellar objects, and sources associated with methanol masers. These surveys have a different coverage than the complete ATLASGAL and also have different sensitivities (see Urquhart et al. 2014 for more details), so that the final number of objects in the advanced stages might be underestimated. Because this limit affects the separation between IRb and HII, we note that if we consider a single phase to describe the more evolved sources (i.e.

IRb + HII), the total duration associated with this phase would agree better.

An additional source of uncertainty is a different value of  $\dot{M}$  throughout the evolutionary sequence. Evidence of an increasing mass accretion rate in the intermediate stages of the massive star formation process is discussed in [Beuther et al. \(2002b\)](#), but when the protostars are close to the main sequence, their radiation pressure might slow down or quench the mass accretion (e.g. [Nakano et al. 1995](#); [Stahler & Palla 2005](#); [Klassen et al. 2012](#)). To quantify the effects of  $\dot{M}$ , we repeated the calculation of the duration of each phase by separating the models by accretion rate and mixing the results of the different phases in all their possible combinations. We find that different  $\dot{M}$  produces different durations and  $t_{\text{MSF}}$ . The most accurate solution to interpret the observations in ATLASGAL shows an accretion rate that is initially slow and increases during the two intermediate classes, to decrease again in the more evolved phases (see the orange symbols and percentages in Fig. 6;  $M_2$ ). The latter result allows us to verify how reliable the assumption of a constant  $\text{H}_2$  radial density profile is that we made for the clumps during phase II (see Sect. 3.1.2). In the most variable scenario where  $\dot{M} = 10^{-3} M_{\odot} \text{yr}^{-1}$  over the  $t_2$  identified in our models (and assuming no mass replenishment), the majority of the clumps shows a mass loss of  $\lesssim 20\%$ . Only two clumps of  $\sim 20 M_{\odot}$  would suffer a relevant mass loss (i.e. G353.07+0.45 and G316.64-0.09 in Table D.1). However, these two sources belong to the IRb and IRw phases, which contain the majority of the clumps, and can be considered outliers in these classes. Conversely, for sources in the 70w and HII evolutionary phases, where  $\dot{M} = 10^{-5} M_{\odot} \text{yr}^{-1}$ , the mass loss is negligible. Alongside the few observational estimates available for  $\dot{M}$  and for the infall rate (e.g. from the parent filament; see Sect. 3.1.2), these results mean that the assumption of a constant-density  $\text{H}_2$  profile is reasonable on a clump scale.

The individual durations estimated in  $M_2$  lead to  $\sim 5 \times 10^4$  yr for younger objects (i.e. 70w),  $\sim 1.2 \times 10^5$  yr for the IR-weak,  $\sim 2.4 \times 10^5$  yr for IR-bright sources, and  $\sim 1.1 \times 10^5$  yr for HII regions. Additionally, the average  $f_A$  of each class increases with the evolution of the sources (circles and stars in Fig. 6), so that the typical object of each class appears statistically older than the average from the previous evolutionary phase. We therefore assume that  $M_2$  is the most representative model, with a total massive star formation time  $t_{\text{MSF},M_2} \sim 5.2 \times 10^5$  yr ( $t_{\text{MSF},M_1} \sim 1.1 t_{\text{MSF},M_2}$ ).

In this context, the chemistry plays a crucial role in breaking the model degeneracy when the likelihood is defined by  $P_M$  and  $P_L$  alone and in matching the results from ATLASGAL (see Fig. 6). If  $P_C$  is not included in  $\mathcal{L}$  (Eq. (4)), the duration of the advanced stages shows a difference up to a factor of  $\sim 8$  when compared to those predicted by  $M_2$ . This corresponds to a total time,  $t_{\text{MSF}}$ , 4.5 times shorter, and the final relative durations differ from the observed relative number of objects in each class by up to a factor of 5.

## 5.2. Selected tracers and chemical clocks

Four molecules were selected in this work: formaldehyde, methyl acetylene, acetonitrile, and methanol (see Sect. 2). They all manifest an observed upward trend in their abundances with increasing evolutionary stage, which is indicated by the luminosity-to-mass ratio of the clumps ([Giannetti et al. 2017](#); [Tang et al. 2018](#)). Two thermal components were also required to reproduce the observed emission of the higher K-ladders of methanol and acetonitrile lines in [Giannetti et al. \(2017\)](#).

To evaluate to which extent the models are able to reproduce the observed column densities, the same procedure as we used to calculate  $\Delta t_{\text{phase}}$  was employed (Eqs. (7) and (8)). Figure 7 summarises the comparison of the modelled and observed column densities, grouped by tracer type (colours) and evolutionary class. Shaded areas represent the range of the observed beam-averaged column densities (Sect. 2 and Appendix D), with an order of magnitude uncertainty (Eq. (6)), not shown in the figure. The median values of these observed column densities are shown as a solid line. The median column densities predicted by our models are plotted as coloured circles and are shown with an uncertainty of one order of magnitude, the same as assumed for the observations.

### 5.2.1. Methyl acetylene and acetonitrile

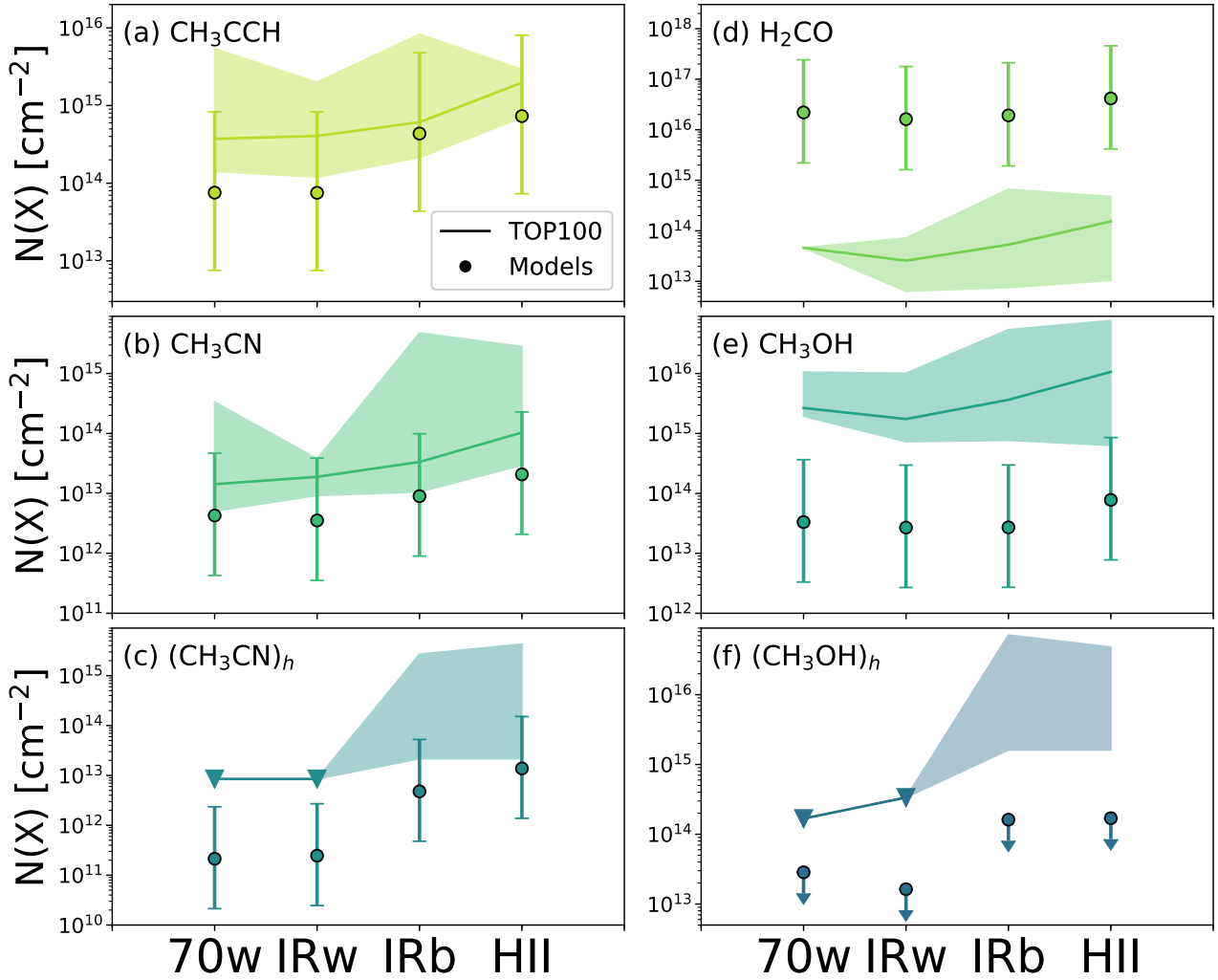
The reliability of  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$  as thermometers has been widely discussed in the past (e.g. [Zhang et al. 1998](#); [Molinari et al. 2016](#); [Giannetti et al. 2017](#)). In particular, [Giannetti et al. \(2017\)](#) detected an increasing trend of the temperature and column density of  $\text{CH}_3\text{CCH}$  from the less to the more evolved sources in the TOP100 (i.e. shaded area in panel a of Fig. 7). The same behaviour is found for  $\text{CH}_3\text{CN}$  if the contributions of a hot and cold component are taken into account.

In Fig. 7 the observed column densities of methyl acetylene (panel a) and acetonitrile (panels b, and c) derived from the observations increase by a factor of 10 between the 70w and the HII stages. The column densities provided by our models reflect the same trends and reproduce the observed values. The discrepancy between models and observations is within the uncertainties.

The robustness of  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$  as chemical clocks was finally tested by removing these two tracers from  $P_C$ . This had the same effect as removing  $P_C$  from  $\mathcal{L}$  altogether (see Sect. 5.1), and we conclude that methyl acetylene and acetonitrile are effective chemical clocks for characterising the evolution of massive star-forming clumps and that they contribute to constraining our findings.

### 5.2.2. Formaldehyde and methanol

Unlike  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$ , the model column densities of methanol and formaldehyde (panels d, e, and f in Fig. 7) agree less well with the observed densities. The slightly increasing trend in the observed column densities with evolution is reproduced by the models, but the modelled column densities under- or over-estimate those observed for the TOP100 by one order of magnitude at least, similarly to [Gerner et al. \(2014\)](#). This might be related to the uncertainties in the chemical pathways that determine the formation of methanol and its precursors. For temperatures below 20 K, hydrogenation chains are usually invoked to convert CO into  $\text{H}_2\text{CO}$  and  $\text{CH}_3\text{OH}$  on grains because hydrogenation is made very efficient by the strong CO depletion that occurs at low temperatures (e.g. [Caselli et al. 2008](#), [Giannetti et al. 2014](#) and [Sabatini et al. 2019](#)). However, at the same temperatures, thermal desorption is inhibited and methanol is not efficiently released from the dust into the gas phase, suggesting that the surface chemistry alone is not capable of explaining the detection of gas-phase methanol in the cold phases of massive star-forming regions as well (e.g. [Cosentino et al. 2018](#)). Alternative formation paths and mechanisms have been proposed to solve this issue (e.g. [Viti & Williams 1999](#); [Garrod et al. 2007](#); [Vasyunin & Herbst 2013](#)), but this was found to be relatively inefficient in reproducing the observed abundances



**Fig. 7.** Comparison of the median column densities observed in the TOP100 (solid lines) with those predicted by our models (circles) that were obtained by applying the same procedure as was used to quantify  $\Delta t_{\text{phase}}$  (see Sect. 4.2). Panels are separated by tracers (colours). The shape of the shaded areas indicates the minimum and maximum observed column density, and the error bars associated with each circle incorporate the uncertainty of one order of magnitude assumed in Eq. (6) for the comparison. The modelled column densities of the hot component of methanol (circles in panel *f*) are multiplied by a factor 1000. Triangles indicate the upper limits. The column densities are shown in log-scale.

of methanol (Geppert et al. 2006), and to no longer accurately predict the abundances of other chemical species (Garrod et al. 2007). The same issue also concerns  $\text{H}_2\text{CO}$ , a molecular precursor of methanol. In the advanced evolutionary stages, the amount of methanol is underestimated (although closer to the observed values). The reason might be that the thermal evaporation of methanol in the warmer environments is not well determined, or that further formation pathways are missing in the colder phases. We also note that the absence of the quantum tunneling diffusion in our models might affect the final abundance of methanol, enhancing the efficiency of CO hydrogenation in the colder evolutionary phases, and thus influencing the final abundance of methanol in the warm phases (Vasyunin et al. 2017).

An additional uncertainty that we explored is the effect of different activation energies for reactions that drive the formation of molecules such as formaldehyde and methanol at low temperatures. Recently, a study of CO hydrogenation on water ice by Rimola et al. (2014) reviewed the activation energy of the hydrogenation reaction between  $\text{H}_2\text{CO}$  and H producing  $\text{CH}_3\text{O}$ , the precursor of methanol. The authors proposed a barrier of

1300 K, which would favour the formation of  $\text{CH}_3\text{OH}$  compared to the value commonly employed ( $\sim 2500$  K, as in this work and derived for the gas phase; Woon 2002). Even a lower activation energy of  $\sim 500$  K has been predicted by Fuchs et al. (2009). However, we have tested these two scenarios without finding significant changes with respect to the modelled column densities of the tracers in Fig. 7. The most substantial change concerns  $\text{CH}_3\text{OH}$  when the extreme activation energy proposed by Fuchs et al. (2009) is assumed. In this case, methanol column densities approach the values observed in the two most advanced evolutionary stages, while those of formaldehyde remain always largely overestimated. This suggests that the activation energy has secondary effects in reproducing the observed abundances of methanol and formaldehyde, and other processes or observational strategies, such as different or multiple transitions, should be taken into account.

We finally note that the discrepancies in the abundances of  $\text{H}_2\text{CO}$  and  $\text{CH}_3\text{OH}$  are not relevant for our results. When these two tracers are removed from  $P_C$  and  $\mathcal{L}$ , the duration of each phase and the final  $t_{\text{MSF}}$  remain unaffected, showing a variation of  $\sim 10\%$  with respect to the numbers reported in Fig. 6.

## 6. Summary and conclusions

We have presented a new and generalised method for deriving the evolutionary timescales of the massive star formation process. Compared to typical statistical approaches (e.g. [Russeil et al. 2010](#); [Mottram et al. 2011](#); [Tigé et al. 2017](#); [Motte et al. 2018](#); [Urquhart et al. 2018](#)), our method follows a different procedure in which the  $t_{\text{MSF}}$  is the final result that is derived as the sum of the individual  $\Delta t_{\text{phase}}$  (Sect. 5.1 and Fig. 6), and not an a priori assumption.

We developed a set of 54 models, built to represent the entire population of massive clumps of the ATLASGAL-TOP100 sample (Sect. 1). The models consist of two physical phases: an initial isothermal collapse followed by a warm-up phase induced by a massive protostar at the centre of a spherical clump (Sect. 3). In particular, we assumed a Plummer-like density profile with a slope of  $5/2$ , while the temperature profiles were derived from accurate 3D RT simulations. The recent empirical estimates of the temperature profiles obtained by [Gieser et al. \(2021\)](#) from the same tracers used in this work agree with our computed profiles (see Sect. 3.1.2). We then compared the column densities of formaldehyde ( $\text{H}_2\text{CO}$ ), methyl acetylene ( $\text{CH}_3\text{CCH}$ ), acetonitrile ( $\text{CH}_3\text{CN}$ ), and methanol ( $\text{CH}_3\text{OH}$ ) as observed in the ATLASGAL-TOP100 sources ([Giannetti et al. 2017](#); [Tang et al. 2018](#)) with the modelled densities by post-processing the outputs of the models at the same angular resolution as the observed data (see Appendix D).

The timescales of the evolutionary stages associated with the massive star formation process were derived by considering the physical properties of the clumps, that is, their mass and luminosity, and the observed abundances of each selected molecular tracer. Considering a mass accretion rate depending on the evolutionary phases, we find a total star formation time,  $t_{\text{MSF}} \sim 5.2 \times 10^5$  yr, which agrees well with that assumed by the statistical methods. This provides a new and significant validation of the statistical methods. The individual  $\Delta t_{\text{phase}}$  that define  $t_{\text{MSF}}$  are found to be  $\sim 5 \times 10^4$  yr for the 70w,  $\sim 1.2 \times 10^5$  yr for the IRw,  $\sim 2.4 \times 10^5$  yr for IRb sources, and  $\sim 1.1 \times 10^5$  yr for HII evolutionary stages of the TOP100 sample (Sect. 2). Knowing  $t_{\text{MSF}}$  and  $\Delta t_{\text{phase}}$ , we derived the relative duration of each phase and found an agreement with the relative number of objects in each phase observed in the ATLASGAL survey. This provides further confirmation of the reliability of the method we presented. The chemical constraint included in the likelihood to determine the duration of the different phases is necessary to achieve results that agree with the observed relative number of objects in the ATLASGAL survey. Without this additional constraint,  $\Delta t_{\text{phase}}$  becomes shorter by up to a factor of  $\sim 8$  with respect to those reported above, and the relative durations differ by up to a factor of 5 with the relative number of objects in each class.

Of the selected molecular tracers,  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$  are best reproduced by our models, and  $\text{H}_2\text{CO}$  and  $\text{CH}_3\text{OH}$  differ from the observed values, although they do follow the observed trends. Therefore the final  $\Delta t_{\text{phase}}$  reported in this paper are mainly constrained by  $\text{CH}_3\text{CCH}$  and  $\text{CH}_3\text{CN}$ . Because the chemical constraint is important to identify reliable timescales, we plan to extend the number and complexity of the selected molecular tracers with several chemical clocks proposed in the literature (e.g.  $\text{HC}_3\text{N}$  by [Taniguchi et al. 2018](#);  $\text{CH}_3\text{OCHO}$  and  $\text{CH}_3\text{OCH}_3$  by [Coletta et al. 2020](#); see also [Urquhart et al. 2019](#); [Belloche et al. 2020](#)). This might help to better assess the different phases of the massive star formation process and increase the number of constraints on  $\Delta t_{\text{phase}}$  reported here.

We also found that the evaporation fronts of the discussed molecular tracers vary between  $10^3$  and  $10^5$  AU during the warm-up phase (Sect. 4), which in our models marks the regions in which each tracer becomes abundant in the gas phase. This result suggests that we compare our findings with the observations provided by the new astronomical facilities. In particular, the Atacama Large Millimeter/submillimeter Array (ALMA; [Wootten & Thompson 2009](#)) offers the perfect opportunity of achieving the high angular resolutions (e.g. [Csengeri et al. 2018](#); [Maud et al. 2019](#); [Sanna et al. 2019](#); [Johnston et al. 2020](#)) needed to sample scales of about the physical sizes of the modelled evaporation fronts. Moreover, ALMA can also reach the sensitivities needed to detect a large number of components in the  $J_{K_a, K_c}$  band. This would allow the removal of possible opacity effects and the accurate definition of the thermal state of the clump that can be compared with the results of this work. The increasing complexity of chemical models and the progress made in the observational techniques make this paramount goal increasingly achievable in the near future.

To conclude, we have reported relevant and robust results on the high-mass star formation process together with reliable estimates of the duration of the different phases of this complex process. The present pipeline is based on 1D models, which cannot capture the effect of dynamical processes such as magnetic fields and turbulence on the density evolution of the collapsing clumps and the subsequent (proto-)stellar accretion process fully. On the other hand, low-dimensionality models allow distinguishing between the different chemical processes, and building a large number of models to infer statistical properties that can be compared with observations. Although some authors have recently included the evolution of large chemical networks in 3D models (see e.g. [Bovino et al. 2019](#)), in the presence of a (proto-)stellar object, modelling the chemistry and the thermal evolution of the gas in 3D still represents a challenge.

*Acknowledgements.* The authors wish to thank the anonymous referee, for her/his suggestions to improve the manuscript, D. Semenov for fruitful discussions and feedbacks about the chemical model benchmark, and J. Ramsey how developed the KROME python-interface employed in this work (see also [Gressel et al. 2020](#)). This paper is based on data acquired with the Atacama Pathfinder EXperiment (APEX). APEX is a collaboration between the Max Planck Institute for Radioastronomy, the European Southern Observatory, and the Onsala Space Observatory. This research made use of Astropy, a community-developed core Python package for Astronomy ([Astropy Collaboration 2013, 2018](#); see also <http://www.astropy.org>), of NASA's Astrophysics Data System Bibliographic Services (ADS) and of Matplotlib ([Hunter 2007](#)). G.S. acknowledges R. Pascale for useful suggestions and comments, and the University of Bologna which partially provided the funds (Marco Polo fellowship) for this project. S.B. acknowledges for funds through BASAL Centro de Astrofísica y Tecnologías Afines (CATA) AFB-17 002.

## References

- Adams, F. C. 2010, *ARA&A*, 48, 47  
 André, P., Revéret, V., Könyves, V., et al. 2016, *A&A*, 592, A54  
 Arzoumanian, D., André, P., Didelon, P., et al. 2011, *A&A*, 529, L6  
 Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33  
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123  
 Awad, Z., Viti, S., Collings, M. P., & Williams, D. A. 2010, *MNRAS*, 407, 2511  
 Bally, J., Moeckel, N., & Throop, H. 2005, *ASP Conf. Ser.*, 341, 81  
 Battersby, C., Bally, J., & Svoboda, B. 2017, *ApJ*, 835, 263  
 Belloche, A., Maury, A. J., Maret, S., et al. 2020, *A&A*, 635, A198  
 Bergin, E. A., & Tafalla, M. 2007, *ARA&A*, 45, 339  
 Beuther, H., Schilke, P., Menten, K. M., et al. 2002a, *ApJ*, 566, 945  
 Beuther, H., Schilke, P., Sridharan, T. K., et al. 2002b, *A&A*, 383, 892  
 Beuther, H., Zhang, Q., Bergin, E. A., & Sridharan, T. K. 2009, *AJ*, 137, 406  
 Bonfand, M., Belloche, A., Garrod, R. T., et al. 2019, *A&A*, 628, A27  
 Bovino, S., Ferrada-Chamorro, S., Lupi, A., et al. 2019, *ApJ*, 887, 224  
 Brown, P. D., Charnley, S. B., & Millar, T. J. 1988, *MNRAS*, 231, 409

- Caselli, P., Vastel, C., Ceccarelli, C., et al. 2008, *A&A*, 492, 703
- Cesaroni, R. 2005, *IAU Symp.*, 227, 59
- Choudhury, R., Schilke, P., Stéphan, G., et al. 2015, *A&A*, 575, A68
- Clark, P. C., Glover, S. C. O., Ragan, S. E., & Duarte-Cabral, A. 2019, *MNRAS*, 486, 4622
- Coletta, A., Fontani, F., Rivilla, V. M., et al. 2020, *A&A*, 641, A54
- Contreras, Y., Schuller, F., Urquhart, J. S., et al. 2013, *A&A*, 549, A45
- Cosentino, G., Jiménez-Serra, I., Henshaw, J. D., et al. 2018, *MNRAS*, 474, 3760
- Csengeri, T., Urquhart, J. S., Schuller, F., et al. 2014, *A&A*, 565, A75
- Csengeri, T., Leurini, S., Wyrowski, F., et al. 2016, *A&A*, 586, A149
- Csengeri, T., Bontemps, S., Wyrowski, F., et al. 2018, *A&A*, 617, A89
- Davies, B., Hoare, M. G., Lumsden, S. L., et al. 2011, *MNRAS*, 416, 972
- Draine, B. T. 1978, *ApJS*, 36, 595
- Draine, B. T. 2011, *Physics of the Interstellar and Intergalactic Medium*, ed. Bruce T. Draine (Princeton: Princeton University Press)
- Draine, B. T., & Lee, H. M. 1984, *ApJ*, 285, 89
- Elmegreen, B. G. 1998, *ASP Conf. Ser.*, 148, 150
- Ercolano, B., Barlow, M. J., Storey, P. J., & Liu, X. W. 2003, *MNRAS*, 340, 1136
- Ercolano, B., Barlow, M. J., & Storey, P. J. 2005, *MNRAS*, 362, 1038
- Fontani, F., Pascucci, I., Caselli, P., et al. 2007, *A&A*, 470, 639
- Fuchs, G. W., Cuppen, H. M., Ioppolo, S., et al. 2009, *A&A*, 505, 629
- Garrod, R. T., & Herbst, E. 2006, *A&A*, 457, 927
- Garrod, R. T., & Widicus Weaver, S. L. 2013, *Chem. Rev.*, 113, 8939
- Garrod, R. T., Wakelam, V., & Herbst, E. 2007, *A&A*, 467, 1103
- Garrod, R. T., Widicus Weaver, S. L., & Herbst, E. 2008, *ApJ*, 682, 283
- Geppert, W. D., Hamberg, M., Thomas, R. D., et al. 2006, *Faraday Discuss.*, 133, 177
- Gerner, T., Beuther, H., Semenov, D., et al. 2014, *A&A*, 563, A97
- Giannetti, A., Wyrowski, F., Brand, J., et al. 2014, *A&A*, 570, A65
- Giannetti, A., Leurini, S., Wyrowski, F., et al. 2017, *A&A*, 603, A33
- Giannetti, A., Bovino, S., Caselli, P., et al. 2019, *A&A*, 621, L7
- Gieser, C., Semenov, D., Beuther, H., et al. 2019, *A&A*, 631, A142
- Gieser, C., Beuther, H., Semenov, D., et al. 2021, *A&A*, 648, A66
- Glassgold, A. E., & Langer, W. D. 1974, *ApJ*, 193, 73
- Grassi, T., Bovino, S., Schleicher, D. R. G., et al. 2014, *MNRAS*, 439, 2386
- Gressel, O., Ramsey, J. P., Brinch, C., et al. 2020, *ApJ*, 896, 126
- Güsten, R., Nyman, L. Å., Schilke, P., et al. 2006, *A&A*, 454, L13
- Hartmann, L., Ballesteros-Paredes, J., & Heitsch, F. 2012, *MNRAS*, 420, 1457
- Hasegawa, T. I., & Herbst, E. 1993, *MNRAS*, 261, 83
- Hasegawa, T. I., Herbst, E., & Leung, C. M. 1992, *ApJS*, 82, 167
- Herpin, F., Chavarría, L., van der Tak, F., et al. 2012, *A&A*, 542, A76
- Hickson, K. M., Wakelam, V., & Loison, J.-C. 2016, *Mol. Astrophys.*, 3, 1
- Hocuk, S., & Cazaux, S. 2015, *A&A*, 576, A49
- Hocuk, S., Cazaux, S., Spaans, M., & Caselli, P. 2016, *MNRAS*, 456, 2586
- Hoq, S., Jackson, J. M., Foster, J. B., et al. 2013, *ApJ*, 777, 157
- Hosokawa, T., & Omukai, K. 2009, *ApJ*, 691, 823
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Johnston, K. G., Hoare, M. G., Beuther, H., et al. 2020, *A&A*, 634, L11
- Kennicutt, R. C. 2005, in *Massive Star Birth: A Crossroads of Astrophysics*, eds. R. Cesaroni, M. Felli, E. Churchwell, & M. Walmsley (Cambridge: Cambridge University Press), 227, 3
- Klassen, M., Peters, T., & Pudritz, R. E. 2012, *ApJ*, 758, 137
- Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, *ApJ*, 900, 179
- König, C., Urquhart, J. S., Csengeri, T., et al. 2017, *A&A*, 599, A139
- Kulterer, B. M., Drozdovskaya, M. N., Coutens, A., Manigand, S., & Stéphan, G. 2020, *MNRAS*, 498, 276
- Kurtz, S., Cesaroni, R., Churchwell, E., Hofner, P., & Walmsley, C. M. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S. S. Russell (Tucson: University of Arizona Press), 299
- Liu, X.-L., Xu, J.-L., Ning, C.-C., Zhang, C.-P., & Liu, X.-T. 2018, *Res. Astron. Astrophys.*, 18, 004
- Loison, J.-C., Wakelam, V., & Hickson, K. M. 2014, *MNRAS*, 443, 398
- Maret, S., Hily-Blant, P., Pety, J., Bardeau, S., & Reynier, E. 2011, *A&A*, 526, A47
- Maud, L. T., Cesaroni, R., Kumar, M. S. N., et al. 2019, *A&A*, 627, L6
- McKee, C. F., & Tan, J. C. 2002, *Nature*, 416, 59
- Molinari, S., Pezzuto, S., Cesaroni, R., et al. 2008, *A&A*, 481, 345
- Molinari, S., Merello, S., Elia, D., et al. 2016, *ApJ*, 826, L8
- Motte, F., Bontemps, S., Schilke, P., et al. 2007, *A&A*, 476, 1243
- Motte, F., Bontemps, S., & Louvet, F. 2018, *ARA&A*, 56, 41
- Mottram, J. C., Hoare, M. G., Davies, B., et al. 2011, *ApJ*, 730, L33
- Mouschovias, T. C., Tassis, K., & Kunz, M. W. 2006, *ApJ*, 646, 1043
- Mueller, K. E., Shirley, Y. L., Evans, Neal J., I., & Jacobson, H. R. 2002, *ApJS*, 143, 469
- Müller, H. S. P., Thorwirth, S., Roth, D. A., & Winnewisser, G. 2001, *A&A*, 370, L49
- Nakano, T., Hasegawa, T., & Norman, C. 1995, *ApJ*, 450, 183
- Nomura, H., & Millar, T. J. 2004, *A&A*, 414, 409
- Padoan, P., Pan, L., Juvella, M., Haugbølle, T., & Nordlund, Å. 2020, *ApJ*, 900, 82
- Penteado, E. M., Walsh, C., & Cuppen, H. M. 2017, *ApJ*, 844, 71
- Peretto, N., Fuller, G. A., Duarte-Cabral, A., et al. 2013, *A&A*, 555, A112
- Peters, T., Banerjee, R., Klessen, R. S., & Mac Low, M.-M. 2011, *ApJ*, 729, 72
- Reboussin, L., Wakelam, V., Guilloteau, S., & Hersant, F. 2014, *MNRAS*, 440, 3557
- Rimola, A., Taquet, V., Ugliengo, P., Balucani, N., & Ceccarelli, C. 2014, *A&A*, 572, A70
- Russeil, D., Zavagno, A., Motte, F., et al. 2010, *A&A*, 515, A55
- Sabatini, G., Giannetti, A., Bovino, S., et al. 2019, *MNRAS*, 490, 4489
- Sabatini, G., Bovino, S., Giannetti, A., et al. 2020, *A&A*, 644, A34
- Sanna, A., Kölligan, A., Moscadelli, L., et al. 2019, *A&A*, 623, A77
- Saraceno, P., Andre, P., Ceccarelli, C., Griffin, M., & Molinari, S. 1996, *A&A*, 309, 827
- Schneider, N., Csengeri, T., Bontemps, S., et al. 2010, *A&A*, 520, A49
- Schneider, N., Ossenkopf, V., Csengeri, T., et al. 2015, *A&A*, 575, A79
- Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, *A&A*, 504, 415
- Semenov, D., Hersant, F., Wakelam, V., et al. 2010, *A&A*, 522, A42
- Smartt, S. J. 2009, *ARA&A*, 47, 63
- Spitzer, L. 1978, *Physical Processes in the Interstellar* (New York: Wiley-Interscience)
- Spitzer, Lyman, J., & Tomasko, M. G. 1968, *ApJ*, 152, 971
- Stahler, S. W., & Palla, F. 2005, *The Formation of Stars* (Hoboken: John Wiley & Sons, Ltd), 865
- Tafalla, M., Myers, P. C., Caselli, P., Walmsley, C. M., & Comito, C. 2002, *ApJ*, 569, 815
- Tafalla, M., Myers, P. C., Caselli, P., & Walmsley, C. M. 2004, *Ap&SS*, 292, 347
- Tang, X. D., Henkel, C., Wyrowski, F., et al. 2018, *A&A*, 611, A6
- Taniguchi, K., Saito, M., Sridharan, T. K., & Minamidani, T. 2018, *ApJ*, 854, 133
- Tielens, A. G. G. M. 2010, *The Physics and Chemistry of the Interstellar Medium*, ed. A. G. G. M. Tielens, (Cambridge, UK: Cambridge University Press)
- Tigé, J., Motte, F., Russeil, D., et al. 2017, *A&A*, 602, A77
- Urquhart, J. S., Figura, C. C., Moore, T. J. T., et al. 2013, *MNRAS*, 437, 1791
- Urquhart, J. S., Moore, T. J. T., Csengeri, T., et al. 2014, *MNRAS*, 443, 1555
- Urquhart, J. S., König, C., Giannetti, A., et al. 2018, *MNRAS*, 473, 1059
- Urquhart, J. S., Figura, C., Wyrowski, F., et al. 2019, *MNRAS*, 484, 4444
- van der Tak, F. F. S. 2005, in *Massive Star Birth: A Crossroads of Astrophysics*, eds. R. Cesaroni, M. Felli, E. Churchwell, & M. Walmsley (Cambridge: Cambridge University Press), 227, 70
- van der Tak, F. F. S., & van Dishoeck, E. F. 2000, *A&A*, 358, L79
- van Dishoeck, E. F., & Blake, G. A. 1998, *ARA&A*, 36, 317
- Vasyunin, A. I., & Herbst, E. 2013, *ApJ*, 769, 34
- Vasyunin, A. I., Caselli, P., Dulieu, F., & Jiménez-Serra, I. 2017, *ApJ*, 842, 33
- Viti, S., & Williams, D. A. 1999, *MNRAS*, 310, 517
- Viti, S., Caselli, P., Hartquist, T. W., & Williams, D. A. 2001, *A&A*, 370, 1017
- Viti, S., Collings, M. P., Dever, J. W., McCoustra, M. R. S., & Williams, D. A. 2004, *MNRAS*, 354, 1141
- Wakelam, V., Loison, J. C., Mereau, R., & Ruaud, M. 2017, *Mol. Astrophys.*, 6, 22
- Wang, K. S., Bourke, T. L., Hogerheijde, M. R., et al. 2013, *A&A*, 558, A69
- Wienen, M., Wyrowski, F., Menten, K. M., et al. 2015, *A&A*, 579, A91
- Wood, D. O. S., & Churchwell, E. 1989, *ApJ*, 340, 265
- Woon, D. E. 2002, *ApJ*, 569, 541
- Woodsley, S. E., & Weaver, T. A. 1995, *ApJS*, 101, 181
- Wootten, A., & Thompson, A. R. 2009, *IEEE Proceedings*, 97, 1463
- Wyrowski, F., Güsten, R., Menten, K. M., et al. 2016, *A&A*, 585, A149
- Zhang, Q., Ho, P. T. P., & Ohashi, N. 1998, *ApJ*, 494, 636
- Zhu, H., Tian, W., Li, A., & Zhang, M. 2017, *MNRAS*, 471, 3494
- Zinnecker, H., & Yorke, H. W. 2007, *ARA&A*, 45, 481

## Appendix A: Details of the chemical network

Table A.1 shows the types of chemical reactions that we included in the network. Column (1) contains names of each type of reaction; Col. (2) an example of chemical reaction between two generic reactants A and B, and, Col. (3) summarises the total number of reactions included in the network.

In Table A.2 we provide the complete list of binding energies assumed in this work, updated to the most recent estimates found in KIDA. Where it was not possible to find a recent estimate, we used [Semenov et al. \(2010\)](#) (species in bold). The 35 chemical species in the bottom part of the table are those added to the original network due to the missing desorption processes (see Sect. 3.2).

**Table A.1.** Summary of the reactions in the chemical network.

Reaction	Example <sup>(a)</sup>	#included <sup>(b)</sup>
Recomb. on grain	$A^+ + \text{grain}^- \rightarrow A + \text{grain}$	13
CR ionisation	$AB + cr \rightarrow AB^+ + e^-$	220
CR photodissociation	$AB + cr \rightarrow A + B$	
Gas-phase reactions	(see the note below) <sup>7</sup>	4016
Photo-ionisation	$AB + \gamma_{UV} \rightarrow AB^+ + e^-$	153
Photo-dissociation	$AB + \gamma_{UV} \rightarrow A + B$	
Grain-surf. reactions	$A_{\text{dust}} + B_{\text{dust}} \rightarrow AB_{\text{dust}}$	266
Thermal desorption	$AB_{\text{dust}} + \text{heat} \rightarrow AB$	195
Desorp. induced by CR	$AB_{\text{dust}} + cr \rightarrow AB$	195
(CR photodiss.) <sub>dust</sub>	$AB_{\text{dust}} + cr \rightarrow A_{\text{dust}} + B_{\text{dust}}$	185
(CR photoion.) <sub>dust</sub>	$AB_{\text{dust}} + cr \rightarrow AB_{\text{dust}}^+ + e^-$	56
(FUV photodiss.) <sub>dust</sub>	$AB_{\text{dust}} + \gamma_{UV} \rightarrow A_{\text{dust}} + B_{\text{dust}}$	204
(FUV photoion.) <sub>dust</sub>	$AB_{\text{dust}} + \gamma_{UV} \rightarrow AB_{\text{dust}}^+ + e^-$	171
(Freeze-out) <sub>dust</sub>	$AB \rightarrow AB_{\text{dust}} + \gamma$	195

**Notes.** <sup>(a)</sup>A and B are two generic reactants. <sup>(b)</sup>Number of reactions.

**Table A.2.** Complete list of binding energies.

Chemical species - $T_b/K$							
C	10 000	C <sub>6</sub> H <sub>6</sub>	7590	CO	1300	HNC <sub>3</sub>	4580
C <sub>10</sub>	8000	C <sub>7</sub>	5600	CO <sub>2</sub>	2600	HNCO	4400
C <sub>2</sub>	10 000	C <sub>7</sub> H	6140	CS	3200	HNO	3000
C <sub>2</sub> H	3000	C <sub>7</sub> H <sub>2</sub>	6590	Fe	4200	HS	2700
C <sub>2</sub> H <sub>2</sub>	2590	C <sub>7</sub> H <sub>3</sub>	7040	FeH	4650	HSS	2650
C <sub>2</sub> H <sub>3</sub>	2800	CH <sub>3</sub> C <sub>6</sub> H	7490	H	650	Mg	5300
C <sub>2</sub> H <sub>4</sub>	2500	C <sub>7</sub> N	6400	H <sub>2</sub>	440	MgH	5750
C <sub>2</sub> H <sub>5</sub>	3100	C <sub>8</sub>	6400	<b>H<sub>2</sub>C<sub>3</sub>N</b>	<b>5030</b>	MgH <sub>2</sub>	6200
CH <sub>3</sub> CH <sub>2</sub> OH	5400	C <sub>8</sub> H	6940	<b>H<sub>2</sub>C<sub>3</sub>O</b>	<b>3650</b>	N	720
C <sub>2</sub> H <sub>6</sub>	1600	C <sub>8</sub> H <sub>2</sub>	7390	<b>H<sub>2</sub>C<sub>5</sub>N</b>	<b>6630</b>	N <sub>2</sub>	1100
C <sub>2</sub> N	2400	C <sub>8</sub> H <sub>3</sub>	7840	<b>H<sub>2</sub>C<sub>7</sub>N</b>	<b>8230</b>	<b>N<sub>2</sub>H<sub>2</sub></b>	<b>4760</b>
CCO	1950	C <sub>8</sub> H <sub>4</sub>	8290	<b>H<sub>2</sub>C<sub>9</sub>N</b>	<b>9830</b>	Na	11 800
C <sub>2</sub> S	2700	C <sub>9</sub>	7200	H <sub>2</sub> CN	2400	NaH	12 300
C <sub>3</sub>	2500	C <sub>9</sub> H	7740	H <sub>2</sub> CO	4500	NaOH	14 700
C <sub>3</sub> H	4000	C <sub>9</sub> H <sub>2</sub>	8190	H <sub>2</sub> CS	4400	NH	2600
C <sub>3</sub> H <sub>2</sub>	3390	C <sub>9</sub> H <sub>3</sub>	8640	H <sub>2</sub> O	5600	NH <sub>2</sub>	3200
C <sub>3</sub> H <sub>3</sub>	3300	C <sub>9</sub> H <sub>4</sub>	9090	H <sub>2</sub> O <sub>2</sub>	6000	NH <sub>2</sub> CHO	6300
C <sub>3</sub> H <sub>3</sub> N	5480	C <sub>9</sub> N	8000	H <sub>2</sub> S	2700	<b>NH<sub>2</sub>OH</b>	<b>6810</b>
CH <sub>3</sub> CCH	3800	CH	925	HSSH	3100	NH <sub>3</sub>	5500
C <sub>3</sub> N	3200	CH <sub>2</sub>	1400	<b>H<sub>3</sub>C<sub>5</sub>N</b>	<b>7080</b>	NO	1600
C <sub>3</sub> O	2750	H <sub>2</sub> CCN	4230	<b>H<sub>3</sub>C<sub>7</sub>N</b>	<b>8680</b>	NS	1900
C <sub>3</sub> S	3500	H <sub>2</sub> CCO	2800	<b>H<sub>3</sub>C<sub>9</sub>N</b>	<b>10 300</b>	O	1600
C <sub>4</sub>	3200	CH <sub>3</sub> N	5530	<b>H<sub>4</sub>C<sub>3</sub>N</b>	<b>5930</b>	O <sub>2</sub>	1200
C <sub>4</sub> H	3740	CH <sub>2</sub> NH <sub>2</sub>	5530	<b>H<sub>5</sub>C<sub>3</sub>N</b>	<b>6380</b>	O <sub>2</sub> H	5000
C <sub>4</sub> H <sub>2</sub>	4190	CH <sub>3</sub> O	4400	HCCNC	4580	O <sub>3</sub>	2100
C <sub>4</sub> H <sub>3</sub>	4640	CH <sub>3</sub>	1600	HC <sub>2</sub> O	2400	OCN	2400
C <sub>4</sub> H <sub>4</sub>	5090	CH <sub>3</sub> C <sub>3</sub> N	6480	HC <sub>3</sub> N	4580	OCS	2400
C <sub>4</sub> N	4000	CH <sub>3</sub> C <sub>4</sub> H	5890	<b>HC<sub>3</sub>O</b>	<b>3200</b>	OH	4600
C <sub>4</sub> S	4300	CH <sub>3</sub> C <sub>5</sub> N	7880	HC <sub>5</sub> N	6180	S	2600
C <sub>5</sub>	4000	CH <sub>3</sub> C <sub>6</sub> H	7490	HC <sub>7</sub> N	7780	S <sub>2</sub>	2200
C <sub>5</sub> H	4540	CH <sub>3</sub> C <sub>7</sub> N	9480	HC <sub>9</sub> N	9380	Si	11 600
C <sub>5</sub> H <sub>2</sub>	4990	CH <sub>3</sub> CHO	5400	<b>HCCN</b>	<b>3780</b>	SiC	3500
C <sub>5</sub> H <sub>3</sub>	5440	CH <sub>3</sub> CN	4680	HCN	3700	SiH	13 000
CH <sub>3</sub> C <sub>4</sub> H	5890	CH <sub>2</sub> NH <sub>2</sub>	5530	HNCCC	4580	SiH <sub>2</sub>	3600
C <sub>5</sub> N	4800	CH <sub>3</sub> OCH <sub>3</sub>	3150	HCO	2400	SiH <sub>3</sub>	4050
C <sub>6</sub>	4800	CH <sub>3</sub> OH	5000	HCOOCH <sub>3</sub>	6300	SiH <sub>4</sub>	4500
C <sub>6</sub> H	5340	CH <sub>4</sub>	9600	CH <sub>2</sub> O <sub>2</sub>	5570	SiO	3500
C <sub>6</sub> H <sub>2</sub>	5790	CH <sub>3</sub> NH <sub>2</sub>	6580	HCS	2900	SiS	3800
C <sub>6</sub> H <sub>3</sub>	6240	H <sub>2</sub> CN	2400	He	100	SO	2800
C <sub>6</sub> H <sub>4</sub>	6690	CN	2800	HNC	3800	SO <sub>2</sub>	3400
Added in this work							
CH <sub>3</sub> COCH <sub>3</sub>	3500	H <sub>2</sub> SiO	4050	NO <sub>2</sub>	2400	SiC <sub>3</sub>	5100
C <sub>3</sub> P	5900	HCCP	4750	P	1100	SiC <sub>3</sub> H	5550
C <sub>4</sub> P	7500	HCL	5170	PH	1550	SiC <sub>4</sub>	5900
CCl	1900	HCP	2350	PH <sub>2</sub>	2000	SiCH <sub>2</sub>	3750
CCP	4300	HCSi	3630	PN	1900	SiCH <sub>3</sub>	4200
CH <sub>2</sub> PH	2600	HNSi	5080	PO	1900	SiN	3500
Cl	3000	HPO	2350	SiC <sub>2</sub>	4300	SiNC	4300
ClO	1900	N <sub>2</sub> O	2400	SiC <sub>2</sub> H	4700	SiO <sub>2</sub>	4300
CP	1900	NH <sub>2</sub> CN	5560	SiC <sub>2</sub> H <sub>2</sub>	5200		

**Notes.** Binding energies were taken from [Wakelam et al. \(2017\)](#) and [Semenov et al. \(2010\)](#) (see text).

<sup>7</sup> The gas-phase processes included are ion-molecule, neutral-neutral, charge exchange, radiative association, radiative recombination and dissociative recombination.

## Appendix B: Benchmark of the chemical network

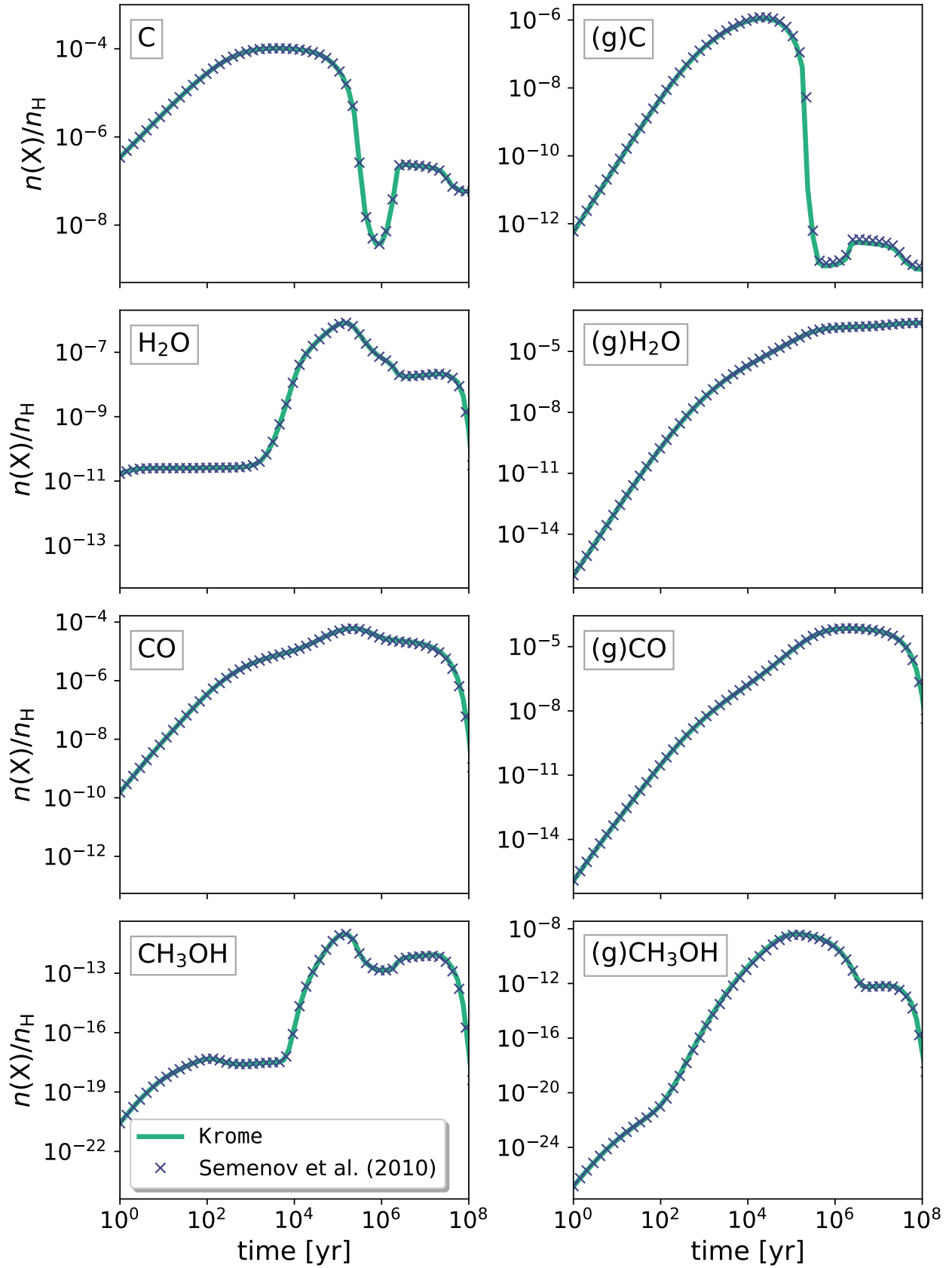
We benchmarked our network against [Semenov et al. \(2010\)](#). The initial conditions are summarised in [Table B.1](#). All the elements are initially atomic, with the exception of hydrogen, which was assumed to be completely in molecular form. Except for He, N, and O, all the elements are also ionised, and grains are initially neutral.

**TMC1-model:** we benchmarked the physical case called TMC1 model. The gas has a constant temperature of 10 K, a constant hydrogen nuclei density  $n_{\text{H}} = 2 \times 10^4 \text{ cm}^{-3}$ , and visual extinction  $A_{\text{v}} = 10 \text{ mag}$ . The sticking coefficient was  $S = 1$ , and the cosmic-ray ionisation rate of hydrogen molecules was set to  $\zeta_2 = 1.3 \times 10^{-17} \text{ s}^{-1}$  ([Spitzer & Tomasko 1968](#); [Glassgold & Langer 1974](#)). We have assumed an average grain size of  $\langle a \rangle = 0.1 \text{ }\mu\text{m}$  and a bulk density  $\rho_0 = 3 \text{ g cm}^{-3}$ , corresponding to silicate grains, as in [Sect. 3.1](#). [Figure B.1](#) shows the perfect agreement with the results reported by [Semenov et al. \(2010\)](#).

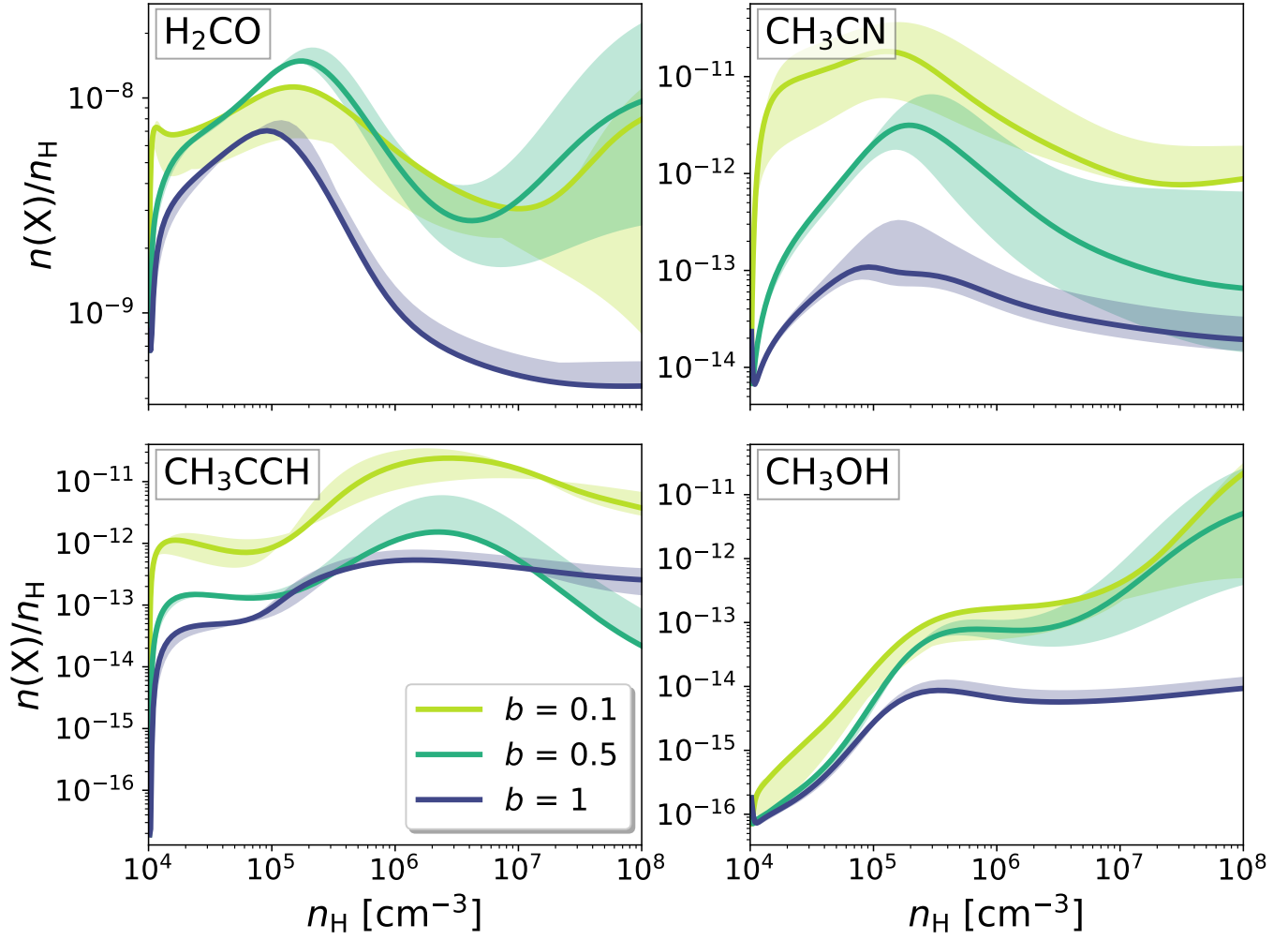
**Table B.1.** Summary of fiducial initial elemental abundances set for the TMC1 model.

Species	$(n_i/n_{\text{H}})_{t=0}$	Species	$(n_i/n_{\text{H}})_{t=0}$
He	9.00(-2)	Si <sup>+</sup>	8.00(-9)
H <sub>2</sub>	5.00(-1)	Na <sup>+</sup>	2.00(-9)
C <sup>+</sup>	1.20(-4)	Mg <sup>+</sup>	7.00(-9)
N	7.60(-5)	Fe <sup>+</sup>	3.00(-9)
O	2.56(-4)	P <sup>+</sup>	2.00(-10)
S <sup>+</sup>	8.00(-8)	Cl <sup>+</sup>	1.00(-9)

**Notes.** Same notation as in [Table 3](#).



**Fig. B.1.** Comparison of the time-dependent variation in the abundances of four arbitrary species (i.e. C, H<sub>2</sub>O, CO, and CH<sub>3</sub>OH) in gas phase (*left panels*) and on dust (*right panels*) in the TMC1 model. The green line shows the results obtained with the KROME-package, and the blue crosses are the results of Semenov et al. (2010) with ALCHEMIC.



**Fig. B.2.** Fractional abundance evolution for  $\text{H}_2\text{CO}$ ,  $\text{CH}_3\text{CN}$ ,  $\text{CH}_3\text{CCH}$ , and  $\text{CH}_3\text{OH}$  as a function of the central density of the clump during the collapse phase. Colours indicate the value of  $b$  assumed for the collapse described by Eq. (1). Solid lines show the results employing the canonical binding energies from KIDA, and shaded areas represent the solutions when  $T_b$  was varied by  $\pm 10\%$ .

## Appendix C: Chemical evolution and model uncertainties during phase I

Figure B.2 shows the fractional abundance evolution of the relevant tracers (see Sect. 1) as a function of the central density of the clump during the isothermal collapse. We first tested the effect of the collapse velocity (parameter  $b$  in Eq. (1)) on the chemical evolution of some of the observed species by delaying the collapse by 50 and the 90%, that is,  $b = 0.5$  and  $b = 0.1$ , respectively. Different colours in Fig. B.2 represent different values of  $b$ .

Despite the effects induced by varying  $b$ , the relative importance of the chemical reactions involving some of the main tracers remains unaltered. In particular,  $\text{H}_2\text{CO}$  is mainly formed through  $\text{CH}_3 + \text{O} \rightarrow \text{H}_2\text{CO} + \text{H}$ , and it is destroyed by ion-molecule reactions that involve  $\text{H}^+$ ,  $\text{C}^+$ , and  $\text{S}^+$ . For densities higher than  $\sim 10^5 \text{ cm}^{-3}$ , ion-molecule reactions become predominant and the abundance of  $\text{H}_2\text{CO}$  decreases. At higher densities, some relative abundance profiles again increase because the dominant reaction becomes  $\text{H}_3\text{CO}^+ + \text{e}^- \rightarrow \text{H}_2\text{CO} + \text{H}$ .

Methyl acetylene is formed by the dissociative recombination (i)  $\text{C}_3\text{H}_5^+ + \text{e}^- \rightarrow \text{CH}_3\text{CCH} + \text{H}$ , the main gas-phase source of  $\text{CH}_3\text{CCH}$  (see Hickson et al. 2016). For densities above  $\sim 10^6 \text{ cm}^{-3}$ , the relative abundance is almost constant because  $\text{CH}_3\text{CCH}$  reacts with  $\text{C}$  with a rate coefficient similar to (i).

Conversely, the abundance of acetonitrile is more sensitive to  $b$ . The main source of gas-phase  $\text{CH}_3\text{CN}$  in our model is (ii)  $\text{HCN} + \text{CH}_3^+ \rightarrow \text{CH}_3\text{CNH}^+$ , followed by the dissociative recombination (iii)  $\text{CH}_3\text{CNH}^+ + \text{e}^- \rightarrow \text{CH}_3\text{CN} + \text{H}$ . We found that at beginning of the collapse,  $\text{CH}_3\text{CN}$  is rapidly formed, in agreement with what was estimated by Loison et al. (2014), and is then slowed down by (iv)  $\text{CH}_3\text{CN} + \text{H}_3^+ \rightarrow \text{CH}_3\text{CNH}^+ + \text{H}_2$ . The combined effect of reactions (iii) and (iv) is also sensitive to  $b$  because for fast collapse ( $b = 1$ ), reactions (ii) and (iii) have less time to form  $\text{CH}_3\text{CN}$  before reaction (iv) becomes effective. For densities between  $10^5$  and  $10^6 \text{ cm}^{-3}$ ,  $\text{CH}_3\text{CN}$  starts to react efficiently with  $\text{H}_3^+ \rightarrow \text{CH}_3\text{CNH}^+ + \text{CO}$ . The effect of this reaction in combination with (iv) produces a slow decrease in the  $\text{CH}_3\text{CN}$  profiles at high densities.

The abundance of methanol increases with density, driven by reaction  $\text{CH}_3^+ + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{OH}_2^+$ , followed by  $\text{CH}_3\text{OH}_2^+ + \text{e}^- \rightarrow \text{CH}_3\text{OH} + \text{H}$  (see also Viti & Williams 1999).

To determine other sources of uncertainties, we analysed the variation in chemical evolution by changing the binding energies reported in Table A.2 by  $\pm 10\%$ , repeating the collapse with the same physical conditions. The effects on the fractional abundances are represented by the coloured areas in Fig. B.2. The discrepancy is more than one order of magnitude at the highest densities. We note that even larger discrepancies are reported by Penteado et al. (2017), when the binding energies are randomly selected from a normal distribution (see results reported in their Fig. 2).

These tests show that the uncertainties due to binding energies variation are not negligible (see Fig. B.2), but they are not greater than variations caused by variations in  $b$ . This suggests that the dynamical state of the clump may play an important role in regulating its chemical evolution (see also the recent results by Kulterer et al. 2020).

Finally, we tested our collapse-phase with a set-up similar to that of Garrod & Herbst (2006), but employing the chemical network and binding energies of Semenov et al. (2010) because both are publicly available and consequently easy to benchmark. The difference between the abundances found by Garrod & Herbst (2006) and ours is within an order of magnitude. It is caused by the adopted details of the chemical networks.

## Appendix D: Notes on the observed column densities

In Sect. 4.1 we discussed that we post-processed the outputs of the models to compare the synthetic column densities with a number of available observations of the TOP100 clumps (right panel of Fig. 1). In this appendix we report some notes about our treatment of the observed column densities before the comparison that we discussed in Sect. 4.2. The observed column densities we used in the comparison are summarised in Table D.1.

### D.1. Formaldehyde $\text{H}_2\text{CO}$

Formaldehyde data were taken from Tang et al. (2018). For simplicity, we assumed a single half power beam width (HPBW) size of  $29''$  to convolve the modelled column densities (i.e. the APEX-telescope resolution corresponding to the  $J = 3-2$  transition of the  $\text{H}_2\text{CO}$  ortho and the para forms at  $\sim 211$  and  $\sim 218$  GHz, respectively). For the sources in which both isomers are observed, the total column density,  $N_{\text{obs}}(\text{H}_2\text{CO})$ , is derived as  $N(o\text{-H}_2\text{CO}) + N(p\text{-H}_2\text{CO})$ . For sources not detected in  $o$ - and/or  $p$ - $\text{H}_2\text{CO}$ , we estimated the corresponding detection limit column density using WEEDS (Maret et al. 2011), with a  $3\sigma$  detection with respect to the average noise level (Tang et al. 2018; 0.06 K main-beam temperature scale) and assuming the  $\text{H}_2\text{CO}$  molecular line parameters provided by the CDMS<sup>8</sup> database (Müller et al. 2001). For each evolutionary class, we assumed the average values of  $T_{\text{kin}}$  (i.e. 52, 73, 81, and 110 K for 70w, IRw, IRb, and HII, respectively) and a full width at half maximum (FWHM) of  $5 \text{ km s}^{-1}$  for the line reported by Tang et al. (2018), and that the source size was equal to the APEX beam at the corresponding frequency.

### D.2. Methyl acetylene ( $\text{CH}_3\text{CCH}$ ), methanol ( $\text{CH}_3\text{OH}$ ), and acetonitrile ( $\text{CH}_3\text{CN}$ )

Column densities of methyl acetylene, methanol, and acetonitrile were taken from Giannetti et al. (2017), who observed the  $J = 20_{\text{K}} - 19_{\text{K}}$ ,  $J = 7_{\text{K}} - 6_{\text{K}}$ , and  $J = 19_{\text{K}} - 18_{\text{K}}$  bands for  $\text{CH}_3\text{CCH}$ ,  $\text{CH}_3\text{OH}$ , and  $\text{CH}_3\text{CN}$ , respectively, with APEX. Because these lines are very close to each other, we assumed the same angular resolution for APEX, that is,  $18''$ . The final column densities in Giannetti et al. (2017) were corrected for the APEX-beam dilution,  $\eta = \theta_{\text{S}}^2 / (\theta_{\text{S}}^2 + \theta_{\text{beam}}^2)$ , where  $\theta_{\text{S}}$  and  $\theta_{\text{beam}}$  are the source and beam HPBW sizes, respectively. We (re-)applied this correction to each source where the detections were not resolved in the APEX-beam angular size (i.e. when  $\theta_{\text{S}} < \theta_{\text{beam}}$ ) to obtain the beam-averaged column densities, whereas for sources with  $\theta_{\text{S}} > \theta_{\text{beam}}$ ,  $\eta = 1$ . Detection limits were estimated assuming the mean noise of the source spectra, the average line width, and the median excitation temperatures reported in Tables 6 and 7 of Giannetti et al. (2017) for each evolutionary class. We treated the observed hot components for these tracers (see Sect. 4.1) in the same way as the colder ones, considering the appropriate temperatures reported by Giannetti et al. (2017), when we calculated the detection limits.

<sup>8</sup> <https://cdms.astro.uni-koeln.de/cdms/portal/>

**Table D.1.** Summary of the observed properties of the selected TOP100 sources.

(1) Source-ID	(2) Class	(3) $d_{\odot}$ kpc	(4) $L_{\text{bol}}$ $10^2 L_{\odot}$	(5) $M_{\text{clump}}$ $10^2 M_{\odot}$	(6) $T_{\text{d}}$ K	(9) $\log_{10}[N_{\text{obs}}(\text{X}/\text{cm}^{-2})]$					
						(7) H <sub>2</sub> CO	(8) CH <sub>3</sub> CCH	(9) (CH <sub>3</sub> CN) <sub>cold</sub>	(10) (CH <sub>3</sub> CN) <sub>hot</sub>	(11) (CH <sub>3</sub> OH) <sub>cold</sub>	(12) (CH <sub>3</sub> OH) <sub>hot</sub>
G13.18+0.06	70w	2.40	83.18	3.68	24.24	13.66	14.62	13.41	<12.89	16.02	<14.22
G14.49-0.14	70w	3.87	7.50	19.16	12.39	<12.68	14.52	13.22	<12.89	15.85	<14.22
G30.89+0.14	70w	4.90	4.96	19.18	11.44	<12.68	14.16	13.08	<12.93	15.29	<14.17
G351.57+0.77	70w	1.34	4.33	1.64	17.04	<12.68	14.15	<14.53	<12.72	<15.42	<14.20
G353.42-0.08	70w	6.06	45.08	17.91	17.12	<12.68	<15.73	12.74	<12.74	<15.42	<14.20
G354.95-0.54	70w	1.91	4.84	1.47	19.13	<12.68	<15.73	12.71	<12.70	<15.42	<14.20
G08.68-0.37	IRw	4.78	275.00	14.81	24.18	13.72	15.11	13.58	<12.85	15.22	<14.41
G08.71-0.41	IRw	4.78	5.04	16.62	11.81	<12.80	14.46	12.99	<12.84	15.19	<14.36
G10.45-0.02	IRw	8.55	113.14	16.06	20.70	<12.80	14.36	13.34	<12.91	14.96	<14.53
G14.11-0.57	IRw	2.57	31.77	3.53	22.35	13.46	14.63	13.14	<12.87	15.37	<14.40
G14.63-0.58	IRw	1.83	27.78	2.54	22.54	13.54	14.67	13.22	<12.83	15.82	<14.41
G24.63+0.17	IRw	7.72	50.36	15.28	18.12	13.20	14.08	13.21	<12.89	15.03	<14.35
G28.56-0.24	IRw	5.45	17.67	54.14	11.73	<12.80	14.58	12.96	<12.84	14.97	<14.32
G305.19-0.01	IRw	3.80	125.29	5.18	26.07	13.52	<15.30	13.33	<12.85	<14.86	<14.40
G317.87-0.15	IRw	2.95	16.41	3.62	19.26	13.76	14.82	13.51	<12.84	15.88	<14.40
G318.78-0.14	IRw	2.78	64.00	3.59	24.92	13.26	14.22	13.28	<12.92	15.26	<14.39
G326.99-0.03	IRw	3.95	11.44	4.44	17.95	13.84	14.88	13.57	<12.89	15.65	<14.39
G331.71+0.60	IRw	10.53	373.43	51.12	21.00	13.86	14.74	13.48	<12.93	16.00	<14.42
G336.96-0.23	IRw	10.91	36.25	24.22	15.77	13.36	14.71	13.21	<12.89	15.54	<14.40
G337.26-0.10	IRw	11.00	300.11	31.87	21.73	13.35	14.40	13.27	<12.93	15.34	<14.43
G342.48+0.18	IRw	12.55	641.50	49.13	23.61	13.47	14.44	13.18	<12.81	15.15	<14.39
G353.07+0.45	IRw	0.86	0.57	0.18	17.79	<12.80	14.55	<13.48	<12.70	15.11	<14.46
G34.41+0.23	IRb	1.56	48.39	2.14	26.13	14.01	15.20	14.78	14.72	16.15	15.92
G35.20-0.74	IRb	2.19	235.80	4.63	29.54	14.03	14.94	14.26	<14.14	16.33	15.19
G59.78+0.07	IRb	2.16	97.71	2.55	28.19	<12.88	14.36	13.26	<13.11	15.21	<14.59
G305.56+0.01	IRb	3.80	517.18	4.06	33.41	<13.72	<15.00	13.02	<12.96	15.37	<14.55
G310.01+0.39	IRb	3.61	496.33	4.15	32.20	13.46	14.43	13.23	<12.94	14.91	<14.53
G313.58+0.32	IRb	3.78	94.14	1.83	29.17	<12.88	<15.00	<13.30	<12.78	14.88	<14.88
G316.64-0.09	IRb	1.19	9.94	0.18	30.63	<12.88	14.67	13.60	13.57	15.20	15.12
G333.31+0.11	IRb	3.60	107.33	4.27	25.85	13.73	14.73	13.52	<13.01	15.63	<14.58
G339.62-0.12	IRb	3.01	149.96	3.20	28.66	13.65	14.69	13.51	<12.94	14.94	<14.57
G340.75-1.00	IRb	2.76	76.82	2.13	27.07	13.64	14.35	13.19	<12.97	15.75	<14.49
G341.22-0.21	IRb	3.67	162.23	4.90	27.03	13.74	14.33	13.62	<12.91	15.56	<14.59
G345.51+0.35	IRb	2.25	432.67	4.24	32.70	14.04	14.85	13.66	<12.94	16.10	16.05
G345.72+0.82	IRb	1.56	18.73	2.01	22.05	13.28	14.79	13.09	<12.89	15.21	<14.53
G351.77-0.54	IRb	1.00	164.69	2.64	31.79	14.82	15.92	15.68	15.63	16.73	16.68
G353.41-0.36	IRb	3.44	1272.01	34.83	28.26	13.98	15.03	13.59	<12.91	15.88	<14.68
G10.62-0.38	HII	4.95	4227.57	37.87	34.45	14.68	15.43	13.98	13.34	15.95	15.21
G12.81-0.20	HII	2.40	2468.69	18.80	35.15	14.19	15.36	13.64	<13.23	15.32	<14.54
G31.41+0.31	HII	4.90	689.42	30.63	26.32	14.14	15.34	15.45	15.43	16.41	16.38
G34.40+0.23	HII	1.56	29.91	2.76	22.82	13.93	14.85	13.75	<13.09	16.04	<14.54
G301.14-0.23	HII	4.40	2137.40	19.41	34.56	<13.02	15.13	14.01	13.76	15.91	15.78
G330.88-0.37	HII	4.16	1545.88	15.87	33.44	14.26	15.19	14.61	14.49	16.11	16.01
G330.95-0.18	HII	9.32	13 064.99	173.06	33.04	14.45	15.42	14.58	14.44	16.21	16.07
G332.83-0.55	HII	3.60	2419.56	19.36	35.67	14.25	15.30	14.22	14.08	16.03	15.90
G333.28-0.39	HII	3.60	1291.69	20.77	30.37	13.80	14.95	13.47	<13.28	15.76	<14.56
G333.60-0.21	HII	3.60	12 416.71	34.61	41.12	14.23	15.47	13.75	<13.53	14.80	<14.59
G351.42+0.65	HII	1.34	398.57	4.63	33.38	<13.02	15.27	14.85	<14.72	16.88	16.85

**Notes.** Column 1: TOP100 sources considered in this work; Cols. from (2) to (6): physical properties of each source (König et al. 2017); Cols. from (7) to (12): final beam- and LOS-average column densities, obtained by applying the procedure described in Appendix D to the values reported in Giannetti et al. (2017) and Tang et al. (2018). Upper limits are marked with ‘<’.