ALMA MATER STUDIORUM - Università di Bologna

in cotutela con Universidad de Concepción

DOTTORATO DI RICERCA IN

ASTROFISICA

Ciclo XXXIII

ESTABLISHING A TIMELINE FOR THE HIGH-MASS STAR FORMATION PROCESS

Presentata da:
Giovanni Sabatini

Coordinatore di Dottorato:
Prof. **Francesco Rosario Ferraro**

Supervisore (UniBo):
Dr. **Jan Brand**

Supervisore (UdeC):
Prof. **Stefano Bovino**

Co-supervisore:
Dr. **Andrea Giannetti**

Esame finale anno 2021

Settore Concorsuale: 02/C1 – ASTRONOMIA, ASTROFISICA, FISICA DELLA TERRA E DEI PIANETI
Settore Scientifico Disciplinare: FIS/05 – ASTRONOMIA E ASTROFISICA

To Giorgia,
for being there.

“Home is behind, the world ahead,
And there are many paths to tread
Through shadows to the edge of night,
Until the stars are all alight.
Then world behind and home ahead,
We’ll wander back to home and bed.
Mist and twilight, cloud and shade,
Away shall fade! Away shall fade!
Fire and lamp, and meat and bread,
And then to bed! And then to bed!”

J. R. R. Tolkien - *“A Walking Song”; The Lord of the Rings*

The research presented in this Thesis
was carried out as part of the scientific activities of the
Italian node of the European ALMA Regional Centre,
of the Istituto di Radioastronomia
(Istituto Nazionale di Astrofisica - INAF, Bologna),
and in cotutelle with the Astronomy Department
of the University of Concepción, Chile

La ricerca presentata in questa Tesi
è stata svolta nell'ambito delle attività
del nodo Italiano del Centro Regionale ALMA Europeo,
dell'Istituto di Radioastronomia
(Istituto Nazionale di Astrofisica - INAF, Bologna),
ed in cotutela con il Dipartimento di Astronomia
dell'Università di Concepción, Cile.

Abstract

In this Thesis we aim to answer a long-standing astrophysical problem, quantifying the timescales of the evolutionary phases characterising the high-mass star formation process. Understanding the details of the formation of massive stars (i.e. M>8-10 $M_{\odot}$) is not trivial, since these objects are rare and at a relatively large distance. They also form and evolve very quickly and almost their entire formation takes place deeply embedded in their parental clumps. During the evolution, the chemical composition of massive clump can be heavily affected by the changes in density and temperature induced by the presence of massive young stellar objects. Chemical tracers that show a relation between their observed abundances and the different phases of the star formation process are commonly called chemical clocks. In this Thesis, through the comparison of observations of a large sample of massive clumps in different evolutionary stages, and accurate time-dependent chemical models, we estimate the timescales of the different phases over the entire star formation process. In addition, we provide relevant information on the reliability of crucial chemical clocks, both for the early and the late stages, confirming that the chemistry is a powerful tool to establish a timeline for the high-mass star formation process.

Contents

Contents **ix**

1 The star formation process **1**
1.1 The interstellar medium . . . 1
1.1.1 The hierarchy of molecular clouds . . . 1
1.2 Low-mass star formation . . . 5
1.3 High-mass star formation . . . 7
1.3.1 Massive star formation scenarios . . . 8
1.4 Mass-Luminosity diagram . . . 12
1.5 The ATLASGAL-survey: a census of high-mass star-forming regions . . . 14
1.5.1 The TOP100 sample . . . 15

2 Chemical evolution during the star formation process **21**
2.1 Molecules in the Interstellar Space . . . 21
2.2 Astrochemistry and star formation . . . 24
2.2.1 CO-depletion . . . 29
2.2.2 Deuterium fractionation . . . 32
2.2.3 Warm-up and molecular desorption . . . 34
2.3 Star formation chemical clocks . . . 37
2.4 Aims of this thesis . . . 44

3 The early stages of the high-mass star formation process (I) **47**
3.1 The IRDC G351.77-0.51 . . . 48
3.1.1 $C^{18}O$ map . . . 49
3.2 Derived quantities . . . 50
3.2.1 Temperature and H_2 column density maps . . . 50
3.2.2 Column densities of $C^{18}O$ . . . 52
3.3 The large-scale depletion map in G351.77-0.51 . . . 56
3.4 Depletion modeling . . . 59

3.4.1 Estimate of R_{dep} . . . 62
3.4.2 Comparison with 3D models . . . 65
3.4.3 The influence of a different abundance profile on R_{dep} . . . 65
3.5 Chapter conclusions . . . 68

4 The early stages of the high-mass star formation process (II) 71
4.1 Observed o-H_2D^+ sample in ATLASGAL . . . 72
4.2 Observations and data reduction . . . 74
4.2.1 o-H_2D^+ detections . . . 74
4.2.2 Additional tracers: $H^{13}CO^+$, DCO^+ and $C^{17}O$ . . . 78
4.3 Deriving the physical properties of the sample . . . 78
4.3.1 Column densities . . . 80
4.3.2 Estimates of dynamical quantities . . . 81
4.4 New correlations with $X(o\text{-}H_2D^+)$ . . . 83
4.4.1 Source parameters . . . 87
4.4.2 Evolutionary parameters . . . 88
4.4.3 Dynamical parameters . . . 89
4.5 Estimates of the CRIR . . . 90
4.6 o-H_2D^+ as an evolutionary tracer for massive clumps . . . 95
4.6.1 o-H_2D^+ in a broader scenario of star formation . . . 97
4.7 Chapter conclusions . . . 98

5 The late stages of the high-mass star formation process 101
5.1 Physical model . . . 102
5.1.1 Phase-I: isothermal collapse . . . 103
5.1.2 Phase-II: the warm-up phase . . . 104
5.2 Chemical model . . . 106
5.3 Results . . . 108
5.3.1 Post-processing of the models' outputs . . . 112
5.3.2 Notes on the column densities derived for the TOP100 sources . . . 113
5.3.3 Comparison of the modelled column densities with sources from TOP100 sample . . . 115
5.4 Estimates of the duration of evolutionary phases . . . 118
5.5 Selected tracers and chemical clocks . . . 121
5.5.1 Methyl acetylene and acetonitrile . . . 123
5.5.2 Formaldehyde and methanol . . . 123
5.6 Chapter conclusions . . . 126

6 Conclusions and future perspectives **129**
6.1 Final remarks 131

A Notes on the chemical network **133**
A.1 Reaction types and binding energies 133
A.2 Benchmark of the chemical network 135

Bibliography **143**

Chapter 1

The star formation process

1.1 The interstellar medium

The space between the stars is not empty. It is filled with a mixture of gas (~99%) and small particles of interstellar dust (~1%), that together form the interstellar medium (ISM). Depending on its chemistry, temperature and density, the ISM is classified in five main phases: hot ionised, warm ionised, warm neutral, cold neutral, and molecular gas. From the hot ionised to the molecular phase, the temperature of the ISM (T_g) ranges from $\gtrsim 5\times10^5$ to 10 K, more than four orders of magnitude, and the particle volume density (n_{tot}) lies between $\sim 3\times10^{-3}$ and >300 particles cm^{-3} (e.g. Stahler & Palla 2004; Draine 2011). Among the ISM phases, the molecular gas is the coldest and the densest one, and it is the principal ingredient from which to form new stars (first hypothesised by Bok 1948).

The ISM within our Galaxy can be easily seen by naked eye in the direction of the Galactic Centre (Fig. 1.1). Without doubt, this is one of the most spectacular views we can have if we are far from the light pollution of cities, made special by the peculiar over-density of stars that make up the Galactic disc, and the presence of clouds of gas and dust that obscure the starlight, called giant molecular clouds (GMCs). The GMCs contain approximately 80% of the total mass of the molecular gas of the Milky Way ($\sim 7\times10^9$ M$_\odot$; Draine 2011), and their characterisation is the first step towards a clear understanding of the star formation process.

1.1.1 The hierarchy of molecular clouds

In Fig. 1.1 molecular clouds are seen as dark patches against a stellar background because of extinction by dust, which is the combined effect of the scattering and the absorption of the optical starlight. The absorbed light is also re-emitted at larger wavelengths, mostly in the mid- and far-infrared bands (hereafter mid-IR and far-IR, respectively). The starlight not penetrating the clouds due to the extinction, reduces the heating effects from external radiation inside the clouds, and leads to internal temperatures of just a few Kelvin degrees.

Figure 1.1: View of the Galactic plane of the Milky Way, here seen looking toward its centre (yellowish and bright bulge in the sky). The dark lanes are clouds of interstellar gas and dust that lie in the disc of the Galaxy. The photo was taken from the Chajnantor plateau, in the North of Chile, Region of Antofagasta. In the foreground, there are four antennas of the Atacama Large Millimetre/submillimetre Array (ALMA) telescope. Credits: ESO/José Francisco Salgado. The image is taken from https://www.eso.org/public/images/alma-jfs-2010-10/.

Thanks to the advent of astronomical facilities such as the *InfraRed Astronomical Satellite* (IRAS; Neugebauer et al. 1984) that operated in the IR band, and the release of the first large scale molecular surveys in the millimetre and sub-millimetre spectral bands (see Heyer & Dame 2015 as a review) that opened up new windows on the Galaxy, it was possible to identify those dark patches as the sites of star formation. What in the optical band appeared dark and difficult to characterise, became bright in the IR and mm/sub-mm bands where dust and molecules radiate. Observations in these spectral bands revealed that the gas distribution within GMCs is not homogeneous but fragmented into a network of filamentary and/or clumpy hierarchical substructures (e.g. Pokhrel et al., 2018). The typical properties of molecular clouds and their substructures are reported in Tab. 1.1, while Fig. 1.2 shows an example of each observed category. On the large scale, GMCs are cold, with a mean T_g of 15 K, and a mean hydrogen volume density of $n(H_2)$ of 100 cm^{-3}, and typical sizes of several tens of pc (see Fig. 1.2a). The largest scale fragmentation is from clouds to clumps (Fig. 1.2b). Each cloud can contain several clumps of size of the order of a parsec, with $n(H_2)$~10^3-10^4 cm^{-3} and T_g ~10-20 K. Clumps, like clouds, are not homogeneous and can fragment in smaller entities called cores (Fig. 1.2c), with typical sizes of the order of 0.1 pc, $n(H_2)$~ $10^4 - 10^5$ cm^{-3} and T_g around 10 K. Not every core in a clump goes through collapse to form a new star. Those that will become gravitationally

Structure	$n(H_2)$ (cm^{-3})	size (pc)	T_g (K)	M ($M_\odot$)	linewidths (km s^{-1})	A_V (mag)
GMCs	~100	10-50	15	10^4-10^5	~7	2-5
Clumps	10^3-10^4	~1	10-20	50-500	0.3-3	7-10
Cores	10^4-10^5	~0.1	8-12	0.5-5	0.1-0.3	10-25
Envelopes	10^5-10^6	~0.01	12-30	0.01-0.5	0.1-0.2	>20

Table 1.1: Summary of physical properties of GMCs and their substructures. The table summarises the average H_2 number density, size, temperature, molecular mass, typical line widths and visual extinction in the optical, respectively. The table is adapted from Stahler & Palla (2004), Bergin & Tafalla (2007) and Pokhrel et al. (2018).

unstable are called prestellar cores and become sites of active star formation that can be isolated or grouped in intermediate structures named envelopes (Fig. 1.2d). The envelopes can reveal the presence of embedded young stellar objects (YSOs) visible as point sources at IR wavelengths (Fig. 1.2e).

The final mass of the stars that can be formed from the gravitational collapse of a core spans a large range, from ~0.08 $M_\odot$ to ~100 $M_\odot$; these are the minimum mass required for the ignition of hydrogen-burning in a protostellar nucleus and the approximate mass upper limit that allows the balance between radiation pressure and gravity, respectively (Salaris & Cassisi 2005). Within this range of masses, the stars are generally separated according to what happens at the end of their lifecycles (e.g. Smartt 2009; Crowther 2012). The stars that end their life in a core-collapse supernova (cc-SNe[1]) are those with initial masses >8-10 $M_\odot$, and are called high-mass stars. The remnants of high-mass stars are neutron stars or black holes, depending on their mass after the supernova explosion (*Tolman-Oppenheimer-Volkoff mass limit*; Kippenhahn, Weigert & Weiss 2012). Stars with initial masses less than 8-10 $M_\odot$, are generally defined as low-mass stars[2]. These go through the planetary nebula phase, ending their life as a white dwarf; under certain conditions, they could finish as thermonuclear supernovae (SNe Type Ia; Smith 2014). The extent of our knowledge about low- and high-mass stars, and in particular about their formation process, depends on many factors and it is better constrained for the low-mass regime. The main reason is the observational advantage since low-mass stars are much more numerous than massive stars and, consequently, they are also statistically closer to the Sun. For example, the Salpeter initial mass function (IMF) predicts ~60 times more

[1]For the sake of clarity, supernova are empirically classified based on their spectral features and not considering their physical origin. Contrary to SNe Type I, SNe Type II show H lines in their spectra and are cc-SNe. SNe Type I are divided into three classes depending on the presence or absence of Si and He spectral features: Type Ia (show Si lines), Type Ib (no Si but show He lines) and Type Ic (both Si and He are absent). Among SNe Type I, types Ib and Ic are cc-SNe, while the remaining SNe type Ia are thermonuclear.

[2]Often the range 2 $M_\odot < M <$ 8-10 $M_\odot$ is referred to as intermediate mass, reserving the term low-mass for stars with $M < 2\ M_\odot$.

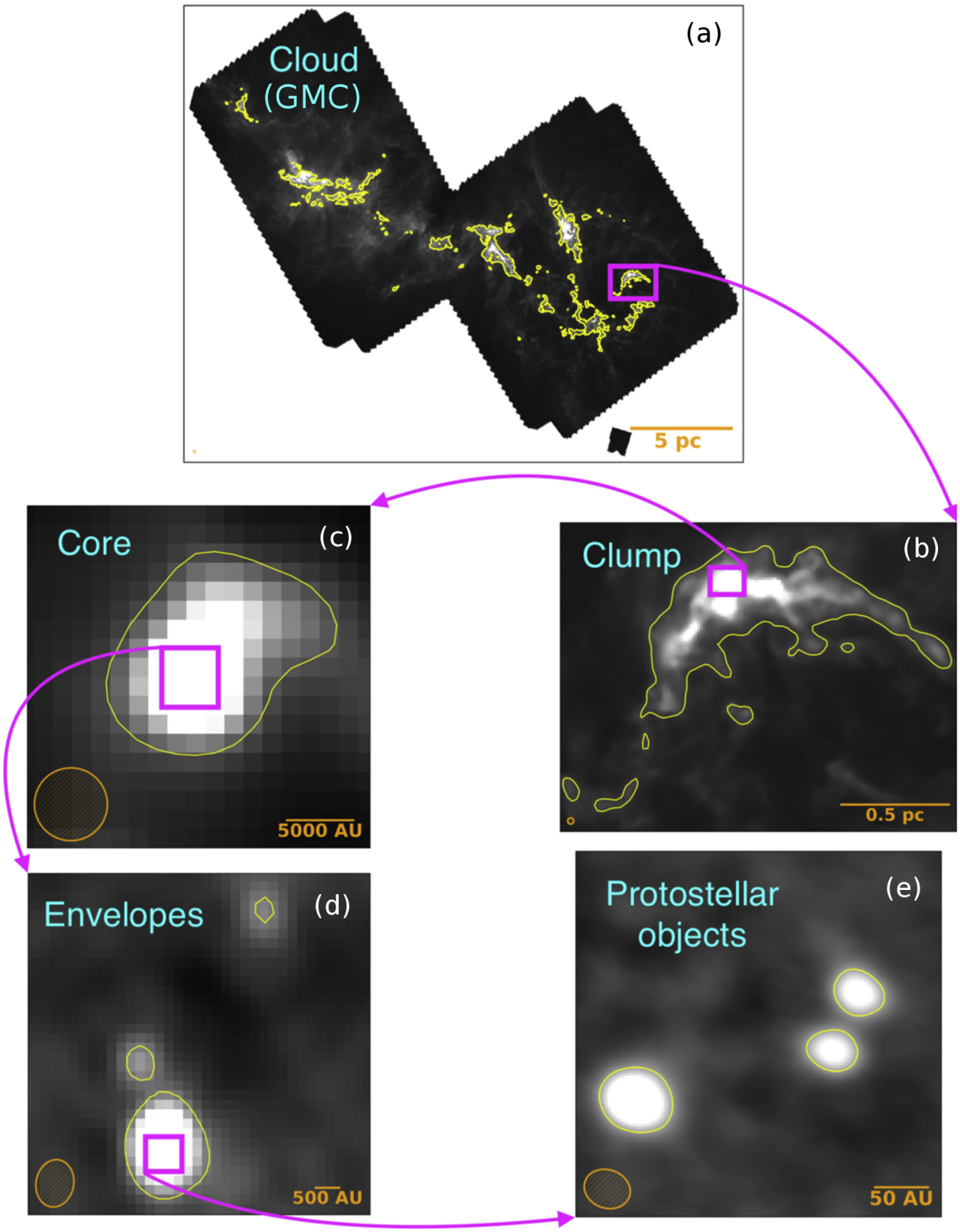

Figure 1.2: Hierarchical cloud structures in the Perseus molecular cloud. Panel (a) shows the whole GMC scale observed with Herschel at 350 μm, where seven clumps were detected (see also Sadavoy et al. 2014; Mercimek et al. 2017). Panel (b) is a zoom of one of those clumps (L1448; Kwon et al. 2006). Panel (c) reports the SCUBA 850 μm map of one of the cores that resides in L1448 (J032536.1+304514; see Di Francesco et al. 2008). Panel (d) reveals the three envelope-fragments observed with the Submillimetre Array (SMA) in J032536.1+304514, one of those harbouring three protostellar objects (Panel e observed with the Very Large Array; Tobin et al. 2016). This figure is adapted from Fig. 2 of Pokhrel et al. (2018).

stars like our Sun than stars with $M \sim 20$ $M_\odot$ (Salpeter 1955). This makes it possible to study the formation process of low mass-stars in more detail and with better statistics.

In the following, we describe the main phases that define the low-mass star formation process, and the differences with the high-mass regime.

1.2 Low-mass star formation

Solar-mass stars have formation timescales of ~10 Myr, after which they spend ~10 Gyr on the main sequence consuming their hydrogen (see Tab. 1.1 in Stahler & Palla, 2004). On the other hand, for stars ten times more massive, the formation timescale is less than a million years and these exhaust their supply of hydrogen in some tens of Myrs. Thus, compared to the high-masses, the low-mass stars are long-lived objects and this allows the detailed investigation of their more easily distinguished different evolutionary stages.

The first step in the star formation process is the formation of a core within the GMCs. When a core becomes unstable, it starts to collapse under the influence of its own gravity, and it contracts to form a proto-star, that continues to accrete new material and grow in mass. During this process, we distinguish four principal evolutionary stages based on observed properties. In the younger sources, large amounts of cool material will surround a protostar and the spectral energy distribution (SED) will show an 'infrared excess' at near-IR and mid-IR wavelengths (Lada 1987 and Andre, Ward-Thompson & Barsony 1993). This excess can be used to empirically arrange the objects on a relative young(er)-old(er) scale by evaluating the SED spectral index, $\alpha_{\rm IR} = d\log_{10}(\lambda F_\lambda)/d\log_{10}(\lambda)$, conventionally computed around 10 μm (Stahler & Palla 2004).

The four classes identified in this way are shown in Fig. 1.3: (*i*) the 'Class 0' includes young accreting protostars that are still deeply embedded in a dense envelope, which is heated by the material accreting onto the protostars. The SEDs of these objects are close to a single-temperature greybody, so no infrared excess is observed at this stage. In this phase, the envelope mass is still larger than that of the protostar ($M_{\rm env} > M_\star$). The angular momentum of the cloud is conserved during the collapse, so that the rotational velocity of the contracting gas increases as the radius shrinks, giving rise to a flattened structure that will become an accretion disk. At this stage molecular outflows and jets have been observed along the rotational axis (e.g. Gibb & Davis 1998 and Busch et al. 2020), implying that this phenomenon is directly linked to the mass accretion process and already activated during the earliest evolutionary phases of star formation. (*ii*) The 'Class I' includes evolved accreting protostars with SEDs broader than a single-temperature greybody. The material surrounding the embryonic star is now arranged in a circumstellar accretion disk, that continues accreting from the envelope, and the SED is dominated by the IR excess mainly due to the emission of the disk ($\alpha_{\rm IR} > 0$). In this phase, most of the mass of the system is concentrated in the protostar ($M_{\rm env} < M_\star$). (*iii*) In the 'Class II' most of the surrounding material has joined the circumstellar disk, and the central ob-

ject, from now on called a pre-main sequence star, becomes visible also in the optical band. Accretion onto the protostar is still continuing but it is less efficient. The SEDs are broader than a single-temperature greybody and still show a less pronounced IR excess ($-1.5 < \alpha_{IR} < 0$). Classical T-Tauri stars are in this evolutionary stage. (*iv*) The final stage is the 'Class III': the protostellar disk is slowly consumed, eventually leading to the formation of a protoplanetary system. No IR excess is shown in the SED that now returns close to a single-temperature greybody (with $\alpha_{IR} < -1.5$).

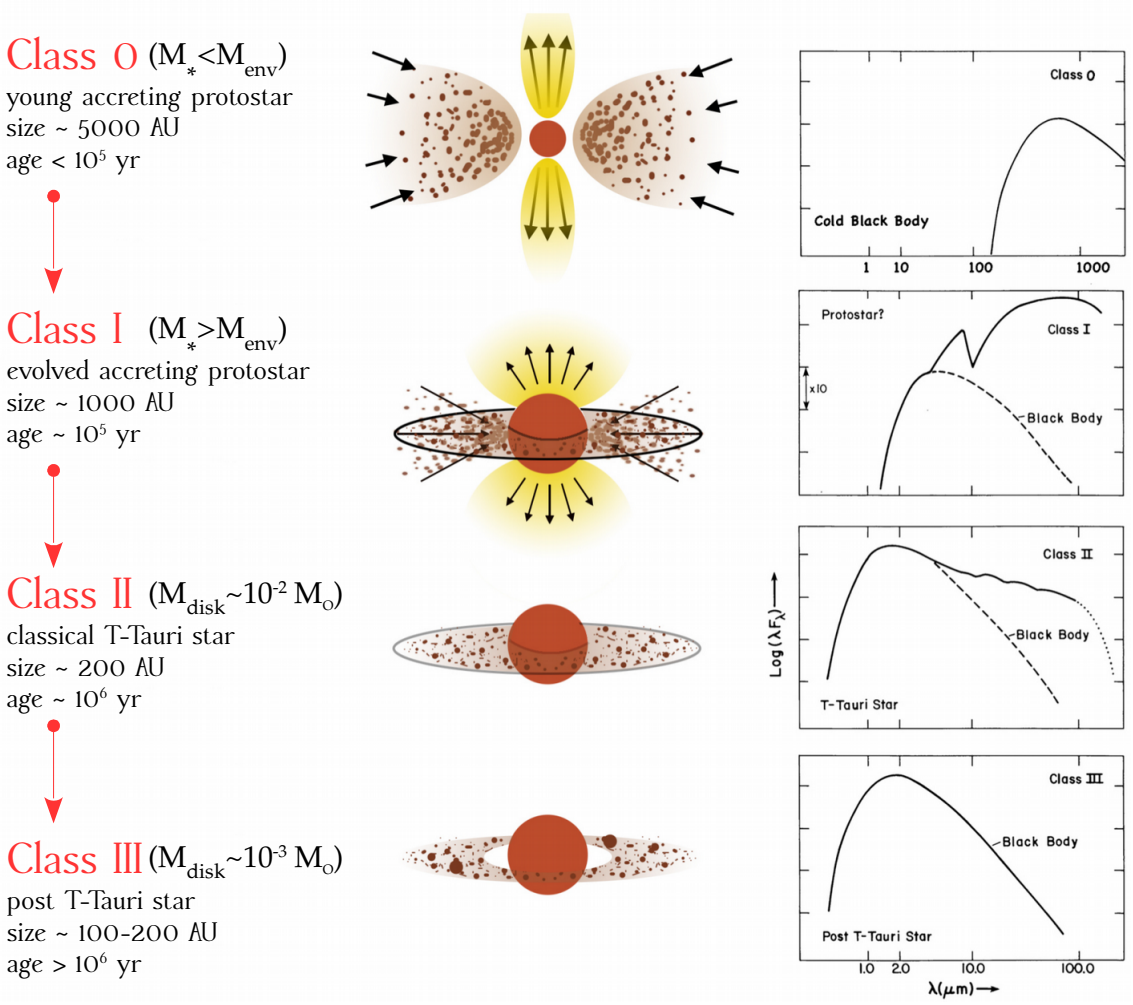

Figure 1.3: Schematic summary of the evolutionary stages defined in the low-mass regime (left panel). Central panels show the sketch of each phase (credit: Guenthner & Kauffmann 2009). Right panels report an example of the typical SED of each phase (Lada 1987). The time evolves from top to bottom and the age and sizes of each stage are reported on the left.

1.3 High-mass star formation

Although the number of high-mass stars is smaller than that of their less massive counterparts (as discussed in Sect. 1.1.1), their influence reaches, in some cases, scales comparable with those of the entire Galaxy (Kennicutt, 2005). For example, when thermonuclear reactions are ignited in the nucleus of a massive forming-star still embedded in its parental clump, the chemical and physical conditions of the clump changes rapidly. In the immediate proximity of the YSO, the ISM is exposed to an intense flux of UV photons with energies $h\nu$>13.6 eV that exceed the ionisation potential of atomic hydrogen (where h is the Planck constant and ν the frequency of the emitted photon). The gas is rapidly photodissociated around the star (*Strömgren sphere*; Strömgren 1939), up to a distance that depends on the rate at which the star emits the UV photons and the average $n(H_2)$ around the star (typically a few parsec; Cimatti, Fraternali & Nipoti 2019). An HII region is formed and the cocoon of gas surrounding the star is disrupted and pushed away, forming favourable conditions for the triggering of a second star formation phase (Elmegreen, 1998).

When a massive star reaches the main sequence (MS), it rapidly evolves. The MS phase of a 20 $M_\odot$ star is approximately a thousand times more rapid than a Sun-like star (i.e. $\sim 8.1\times10^6$ yr and 10^{10} yr, respectively; see Table 1.1 in Stahler & Palla 2004). However, during this time a star of 20 $M_\odot$ maintains a luminosity of about 10^5 times greater than a 1 $M_\odot$ star, with the consequence that the energy budget released by the most massive star during its MS phase is approximately that of 100 Sun-like stars. We have already seen in Sect. 1.1.1 that the Salpeter IMF predicts a ratio of 60 between the number of stars with 1 $M_\odot$ and 20 $M_\odot$, meaning that, in a galaxy, the energy budget released by a single massive stars may be dominant compared to the low-mass ones.

At the end of their life-cycle, massive stars die as cc-SNe, releasing large amount of α-elements[3] in the ISM (Woosley & Weaver, 1995; Kobayashi, Karakas & Lugaro, 2020), while not all the low-mass stars explode as SNe-Ia. Recent estimates by Adams et al. (2013) predict about 3.2 cc-SNe and 1.4 SNe-Ia per century in the Milky Way. Since for each supernova explosion, the kinetic energy released in its surrounding is roughly the same ($\sim10^{51}$ erg; e.g. Smartt 2009), massive star dominate the energy released in a galaxy also during the final phase of the stellar evolution.

In spite of the efforts made in the recent years to characterise the formation process of massive stars and the progress made in this field, there are still some fundamental questions that remain open. For instance, a relevant question concerns the timescales of the massive star formation process, which are strongly related to the galaxy evolution.

This is due to several reasons. First, as long as the gas of the clump accretes onto the quasi-hydrostatic core, the core grows in mass, radius, and luminosity ($M_\star$, $R_\star$ and

[3]Chemical element synthesised in massive stars by the α-particle capture (e.g. ${}^{12}_{6}C + {}^{4}_{2}He \rightarrow {}^{16}_{8}O + \gamma$, where helium is the α-particle), leading to iron. The stable α-elements are: ${}^{12}C$, ${}^{16}O$, ${}^{20}Ne$, ${}^{24}Mg$, ${}^{28}Si$, ${}^{32}S$, ${}^{36}Ar$, ${}^{40}Ca$, ${}^{44}Ti$, ${}^{48}Cr$, ${}^{52}Fe$ and ${}^{56}Ni$.

$L_\star$, respectively). Simultaneously, the core contracts on the thermal *Kelvin-Helmholtz* timescale

$$\tau_{\rm KH} = \frac{GM_\star^2}{R_\star L_\star} \ \ {\rm yr}\,, \tag{1.1}$$

toward densities and temperatures suitable for the activation of hydrogen-burning in the stellar core. For high-mass stars, the $\tau_{\rm KH}$ is shorter than the accretion timescale, and therefore they start hydrogen fusion while still accreting large amounts of material and a significant fraction of their final mass (Kahn, 1974). Therefore, the pre-MS evolution of a massive star takes entirely place deeply embedded in the parental cloud. For this reason observations at long-wavelengths (e.g. in the mid-, far-IR and (sub-)mm spectral regions) are necessary to directly investigate the massive star formation process. The discovery of infrared dark clouds (IRDCs), revealed in absorption against the mid-IR Galactic background almost twenty years ago (Perault et al. 1996 and Egan et al. 1998), started the quest to identify some of the most likely birthplaces of high-mass stars (e.g. Menten, Pillai & Wyrowski 2005, Rathborne, Jackson & Simon 2006, and Peretto et al. 2016). IRDCs are usually filamentary structures and host complexes of clumps with high H_2 column densities, $N(H_2)\sim10^{23}$ cm^{-2}, and masses that usually exceed 100 $M_\odot$.

Second, because of their rarity in the Galaxy, massive stars are typically found at larger distances from the Sun than low-mass stars. This complicates to achieve the same spatial resolution reached in the study of low-mass objects to describe the star formation process with the same accuracy.

In addition to the observational limitations, the problem of massive star formation is in general also not easy to characterise from a theoretical point of view, due to the complexity of the physical processes that must be taken into account, such as the influences of magnetic fields, the heating and the cooling of gas and dust, their chemistry, the accretion from the envelope onto the disk and from the disk onto the star or the mechanical feedbacks (McKee & Ostriker, 2007). Nevertheless, since the early 1990s, different scenarios have been proposed to describe the origin of massive stars, that depend on the initial conditions and the properties of their parent clumps.

1.3.1 Massive star formation scenarios

An extension of the classic paradigm of the low-mass star formation process (discussed in Sect. 1.2) to the high-mass regime is the *turbulent core accretion* (or *monolithic collapse*) scenario (McKee & Tan, 2002, 2003), describing the isolated formation of a single massive star in a clump. Following this scenario, a massive clump that becomes gravitationally unstable, collapses into a YSO, which continues to grow by accreting matter. The accretion rate and the mass of the final star may be influenced by negative feedback such as the radiative pressure of the proto-star or the stellar winds and outflows, produced to conserve the angular momentum during the contraction (e.g. Zinnecker, 2011). Outflows

should have a secondary effect on the accretion, since they are oriented along the polar direction of the protostellar disk and remove only a small quantity of the gas (Zinnecker & Yorke, 2007). When the collapse of the core is triggered, the total star formation timescale estimated by the turbulent core accretion scenario is between $\sim 10^4$ and $3{\times}10^5$ yr, depending on the final mass of the star and the accretion rate considered (McKee & Tan, 2003). From the observational point of view, there are two main consequences of this theory, considering that it is strongly based on the assumption that most of the mass of a clump is involved in the formation of a single massive star. On the one hand, the monolithic collapse theory requires the existence of pre-stellar cores of a few dozen solar masses to form a star with mass greater than 8-10 $M_\odot$. On the other hand, mid-/far-IR observations must confirm the presence of a single object within those cores. However, despite the progress of the observing techniques and the release of large-scale surveys spanning several spectral ranges with high spatial resolution (e.g. Schuller et al. 2009, Molinari et al. 2010 and Aguirre et al. 2011), observations have not been successful in confirming the existence of a significant population of cores massive enough to ensure the formation of stars with $M >$ 8-10 $M_\odot$ on their own. On the contrary, low- and intermediate-mass star formation sites have been discovered in many candidate high-mass star-forming regions, invoking different scenarios to explain the formation of massive stars (e.g. Kauffmann & Pillai 2010, Pillai et al. 2019, Könyves et al. 2020 and Retes-Romero et al. 2020).

Alternatively to the monolithic collapse, the *competitive accretion* scenario (Bonnell et al., 2001) suggests that several low-mass seeds are formed through the fragmentation of a massive clump. All seeds have similar mass, temperature and density when they are formed. When the seeds settle in the protocluster centre, it becomes more likely to form massive stars since that environment is gas-rich due to the inflow of new material from the outer parts of the clump. The more the core accretes gas, the more it grows. The more it grows the more it accretes gas, in a (positive) feedback process that eventually will form high mass stars. At the centre of the clump, where the density of seeds is considerable, their spheres of influence start to overlap so that different objects are forced to compete to accrete the gas still available (hence the name "competitive accretion"). In this case, an estimate for the star formation timescale is less trivial, because it would depend on the number and the distribution of the low-mass seeds, the amount of available material and the rate at which this is accreted from each seed. Since one of the major differences between the competitive- and the core-accretion scenarios is that the former predicts a much smaller accretion rate to the forming star (Tan et al. 2014), one can conclude that competitive accretion occurred relatively slowly and implies star formation timescales one order of magnitude larger (on average) than in a core-accretion scenario (e.g. Wang et al., 2010).

More recently, Tigé et al. (2017) and Motte, Bontemps & Louvet (2018) proposed a scenario where the low-mass cores, formed by the fragmentation of a massive clump, can increase their potential to form massive stars by being intermittently fed by a large-scale

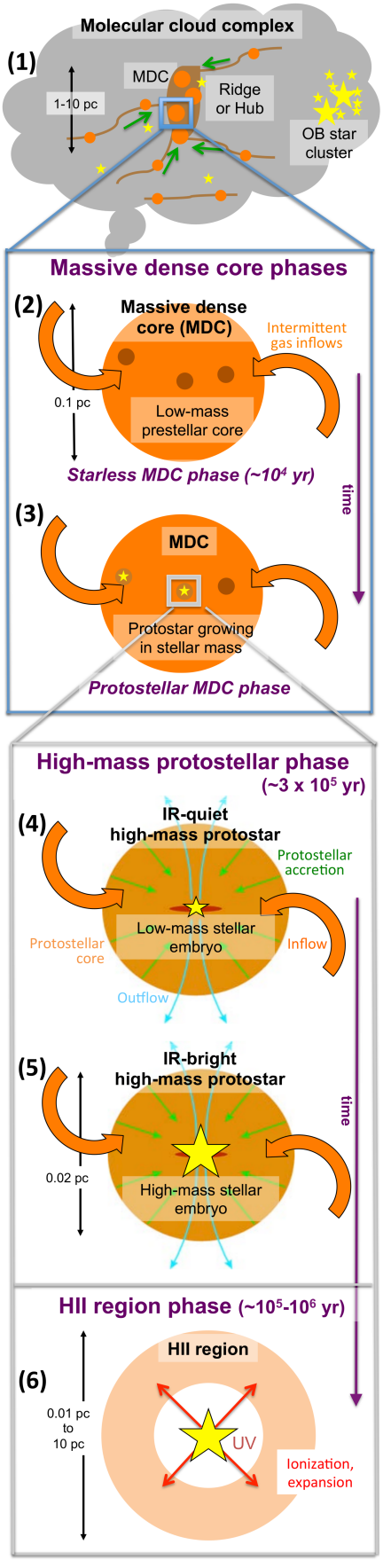

Figure 1.4: Schematic evolutionary diagram proposed by Tigé et al. (2017) for the formation of high-mass stars. (1) Massive accretion filaments and spherical clumps, respectively called ridges and hubs, host massive dense cores (MDCs, 0.1 pc). (2) During their starless phase, MDCs only harbor low-mass prestellar cores. (3) IR-quiet MDCs become protostellar when hosting a stellar embryo of low-mass. The local protostellar collapse is accompanied by the global, 0.1-1 pc, collapse of MDCs and ridges/hubs. (4) Protostellar envelopes feed from these gravitationally-driven inflows, leading to the formation of a high-mass YSO. The latter are IR-quiet as long as their stellar embryos remain low-mass. (5) High-mass YSOs become IR-bright for stellar embryos with mass larger than 8-10 $M_{\odot}$. (6) The main accretion phase terminates when the stellar UV field ionises the protostellar envelope and an H II region develops. This sketch is taken from Fig. 8 of Tigé et al. (2017).

flow of new material. Figure 1.4 summarises the entire process in six main phases: (1) the fragmentation of a GMC favours the formation of denser substructures of gas called ridges or hubs, which have a size of ~1-10 pc. On sub-parsec scale, hubs host denser structures called massive dense cores (MDC). (2) In this scenario, the MDC represent the IR-quiet starless phase of the massive star formation process and their mass can continue to grow due to intermittent gas inflows from the ridge/hub. (3) In the MDC interiors, low-mass protostellar objects are the first (and only) to be formed, while on the ridges/hubs scale, the collapse continues feeding the mass of the MDC with new material. (4) A low-mass stellar embryo is formed, and can continue to increase its mass from the surrounding gas. (5) When stellar embryos reach 8-10 $M_{\odot}$, they start H-burning, their luminosity sharply increases and YSOs become IR-bright. (6) An HII region is developed by the increasing flux of UV photons generated by the massive stellar embryos, and stellar winds and radiation pressure slow down or quench the accretion of additional mass.

A similar scenario has been recently discussed by Padoan et al. (2020), from high-resolution simulations of molecular cloud formation. In this *inertial-inflow* model, a massive clump is fragmented into one or more cores that become gravitationally unstable when their mass exceeds the critical limit of the Bonnor-Ebert mass (Ebert 1955 and Bonnor 1956). The cores collapse into low or intermediate-mass stars, and accrete the remaining mass of the clump to form massive stars through a circumstellar disk accretion. The disk is continuously supplied by a large-scale converging flow of matter. The efficiency and duration of the accretion is what determines the final mass of the star. Since most of the final stellar mass is channeled towards the accreting star by the random velocity field from the large scale, and is unaffected by the stellar gravity during most of the star formation process, this scenario has been named "inertial inflow". The accretion rate of the star and the timescale of the process are determined by the turbulence of the cloud. The former is controlled by the pc-scale inertial inflow that is insensitive to the gravity of the star (unlike in the competitive-accretion scenario), the latter depends on the final stellar mass on average, and it is approximately in the range of 1-6 Myr for stars above 10 $M_{\odot}$.

Moreover, dynamical interactions between the large-scale filamentary regions observed inside a GMC were also proposed as the possible triggering event to form massive stars in the *filaments-to-clusters* formation paradigm (Kumar et al. 2020). In this scenario, similarly to what proposed by Padoan et al. (2020), two or more filaments interact to form a hub where the star formation takes place. For massive stars, the formation timescale is short (~10^5 yr) in the hub, while the formation of lower-mass objects would begin in the individual filamentary structure even before assembling the hub and proceed slowly (~10^6 yr).

From recent observational evidence, and from the multiple formation scenarios proposed, it seems reasonably clear that an unequivocal position on what mechanism dominates the high mass star formation process cannot be assumed.

In the absence of overwhelming evidence favouring one scenario over others, the timescales over which the star formation process takes place still remain poorly known. The aforementioned scenarios can influence the evolution of galaxies in different ways, depending on their timescales, that range from 10^4 to $\gtrsim 10^6$ yr. Is there a process that dominates over the others? Or is a combination of some (or all) of them more credible to explain what is observed? Systematic surveys at mid-/far-IR wavelengths of the Galactic Plane can help to answer these questions, providing a statistical basis for an unbiased classification of massive star-forming regions within a broader evolutionary scenario.

1.4 Mass-Luminosity diagram

The evolutionary sequence described in Sect. 1.2 for the low mass regime (Class 0-III in Fig. 1.3) is based on the observational evidence that the SEDs of objects in different evolutionary classes are profoundly different. A system of molecular gas, initially at equilibrium, can form a low-mass star if it becomes gravitationally unstable, and an observational consequence of this process is the increase of flux at near-IR and mid-IR wavelengths. A further step forward in the characterisation of this process was made by Saraceno et al. (1996). Using a sample of Class 0-II objects, the authors constructed a diagnostic diagram comparing the circumstellar envelope mass and the bolometric luminosity of each source (grey symbols and curves in Fig. 1.5). Objects known to be in different stages according to their SEDs occupy different regions in the $M_{\rm env}$-L diagram, where the entire evolutionary path can be identified.

For a given mass of the envelope, the younger objects show lower luminosities. In Fig. 1.5, a low-mass system evolves along the grey solid curves starting from low-luminosity end. As the protostar is formed, the bolometric luminosity increases, while the mass of the envelope stays roughly constant. The evolutionary tracks in this phase show a vertical increase. The gas envelope surrounding the protostar is slowly consumed during the star formation process and therefore the evolutionary tracks curve towards lower masses. When the protostar is almost completely formed and the accretion becomes less efficient, its luminosity stays approximately constant, while the envelope is further consumed or disrupted. This behaviour shows that the evolution of a protostar in the low-mass regime is controlled mainly by its mass, that determines its bolometric luminosity, L, and depends on the amount of circumstellar material, $M_{\rm env}$. $M_{\rm env}$ and L can be directly derived from the observed fluxes, and their combination provides a straightforward method to determine the evolutionary phases of a low-mass protostar.

The $M_{\rm env}$-L diagram was extended to the high-mass regime by Molinari et al. (2008), finding that the method may also be used to infer the evolutionary stage of massive objects (black symbols and curves in Fig. 1.5). The high-mass sources in the earliest evolutionary stage are shown as black open circles, and in subsequent more advanced stages of evolution they are marked as circles, asterisks and pluses, respectively. The classification was

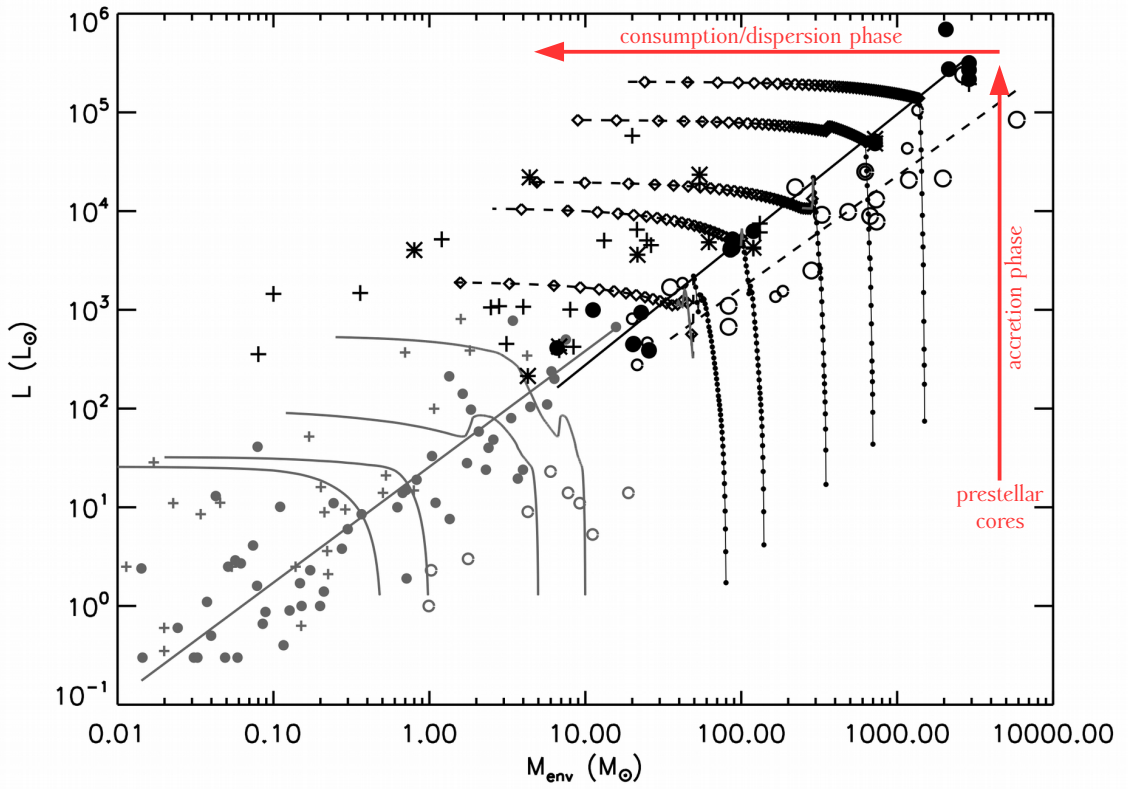

Figure 1.5: $M_{\rm env}$-L diagram for low- (grey) and high-mass (black) regimes. Lines and symbols in grey are from Saraceno et al. (1996). Class 0, I and II sources are represented respectively by open circles, filled circles and crosses, with the solid grey line representing the log-log linear fit to the Class I source distribution. The grey curves show the protostellar evolution in the low-mass regime, for different values of the final stellar mass. Lines and symbols in black represent the high-mass clumps in Molinari et al. (2008); open circles show sources in the earliest evolutionary stage, and progressively more evolved objects are indicated by filled circles, asterisks and pluses. The black curves represent how the clumps evolve in time starting with different initial envelope masses. The black solid line is the best fit for those clumps likely hosting a Zero Age Main Sequence (ZAMS) star. The black dashed line is the log-log fit to the less-evolved massive sources. In both regimes, time increases following the red arrows trough the accretion and dispersion phase of the clumps (Urquhart et al. 2018). This plot was modified starting from Fig. 9 in Molinari et al. (2008).

done evaluating the SEDs in the 8-1200 μm spectral range. The black curves represent the evolution of the massive clumps with different initial envelope masses, and evolve as in the low-mass case, following the red arrows in Fig. 1.5. The black solid line is the best fit to the sources likely hosting a Zero Age Main Sequence (ZAMS) star. In this case, L and $M_{\rm env}$ are typically those of the clump that hosts the high-mass star-forming regions, rather than a single (proto)star, since high mass stars are often formed in clusters (see sect. 1.3) where it is more difficult to resolve individual sources, as they are typically at relatively large distances. Because of this, also the shapes of the evolutionary tracks are slightly different compared to those of the low-mass regime. For the high-mass star-forming regions, the accretion phase shows a quasi-vertical rise, that is abruptly stopped by the disruption of the clump and causes the horizontal trend in the evolutionary tracks, instead of a slow consumption due to the accretion as observed for the less massive stars.

The evolution of the L/M ratio and the observed infrared excess are two of the most widely used methods to identify the phases in the star formation process. To ensure the robustness of these methods and to refine them, it is necessary to select a representative sample of high-mass star forming regions.

1.5 The ATLASGAL-survey: a census of high-mass star-forming regions

The APEX *Telescope Large Area Survey of the Galaxy* (ATLASGAL; Schuller et al. 2009) provides an ideal basis for detailed studies of large numbers of massive clumps in different evolutionary stages. The survey is the result of three years of observations with the Large APEX Bolometer Camera (LABOCA) on the *Atacama Pathfinder EXperiment 12-meter submillimetre telescope* (APEX; Güsten et al. 2006). The LABOCA array (beam size ~19.2 arcsec and field-of-view ~11.4 arcmin) operates at 870 μm (345 GHz) and traces the thermal emission of dust.

Located at an altitude of 5100 metres on the Chajnantor plateau in northern Chile, APEX is ideally placed to scan the inner Galactic plane. The ATLASGAL survey covers nearly 420 square degrees of sky with a typical noise level of 50-70 mJy/beam (see Fig. 1.6). The goal of the survey was to produce a large-scale, systematic database of massive pre- and proto-stellar clumps in the Galaxy, to understand how and under what conditions star formation takes place (Schuller et al., 2009).

The ATLASGAL *Compact Source Catalog* has delivered ~11000 of those clumps above a conservative 5σ threshold (~300 mJy; e.g. Contreras et al. 2013; Urquhart et al. 2013; Csengeri et al. 2014; Li et al. 2016), and distributed over a range of heliocentric distances between ~0.2-17 kpc (Wienen et al., 2015). The clump size distribution ranges between 0.01 and a few parsecs of radius (peaking at ~0.5-0.6 pc; Urquhart et al. 2018), while their masses cover a wide range of values, from as little as a few $\times$10 $M_\odot$ to 10^5 $M_\odot$, and peaks at ~(2-3)$\times10^3$ $M_\odot$. The dust temperatures, $T_{\rm d}$, are between 10 and 40 K with a peak at ~16 K. Almost all of the star forming clumps have a virial parameter $\alpha < 2$

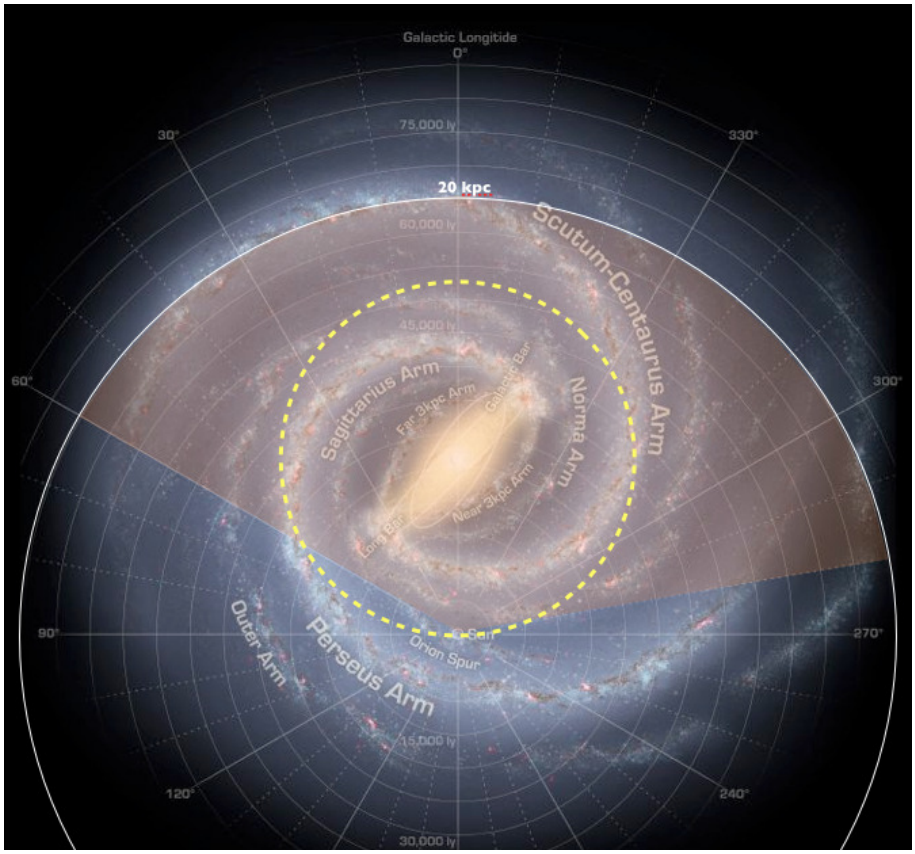

Figure 1.6: The region covered by the ATLASGAL survey is shown as the yellow shaded area, superimposed on an artist's conception of our Galaxy, incorporating much of what is known about its large-scale structure. The solar circle is shown by the yellow dashed circle; the Sun is at the lowest point thereof. Image taken from Fig. 1 by Urquhart et al. (2014a). Credits for the background image: NASA/JPL-Caltech/R. Hurt(SSC/Caltech).

(estimated from ammonia observations; Urquhart et al. 2013) and, unless supported by strong magnetic fields, appear gravitationally unstable against collapse. The molecular hydrogen column densities peak at values between $\sim 10^{21.5}$ and $\sim 10^{24.5}$ cm^{-2}. The clumps have bolometric luminosities between $\sim$0.01 and 10^6 L$_\odot$, and cover a range of luminosity-to-mass ratio of $\sim$0.01-100, reflecting the completeness of the ATLASGAL survey in identifying massive star forming clumps in different stages of evolution (Urquhart et al., 2019, see also Fig. 1.5 and Fig. 24 in Urquhart et al. 2018)

1.5.1 The TOP100 sample

Using the unbiased nature of the ATLASGAL survey, the TOP100 sample (hereafter TOP100; see Giannetti et al. 2014) was defined. It includes 111 sources, selected from ATLASGAL as a flux-limited sample, using additional infrared (IR) criteria in order to include sources in different evolutionary stages. Four stages have been defined by Giannetti et al. (2014), Csengeri et al. (2016) and König et al. (2017) based on the observed

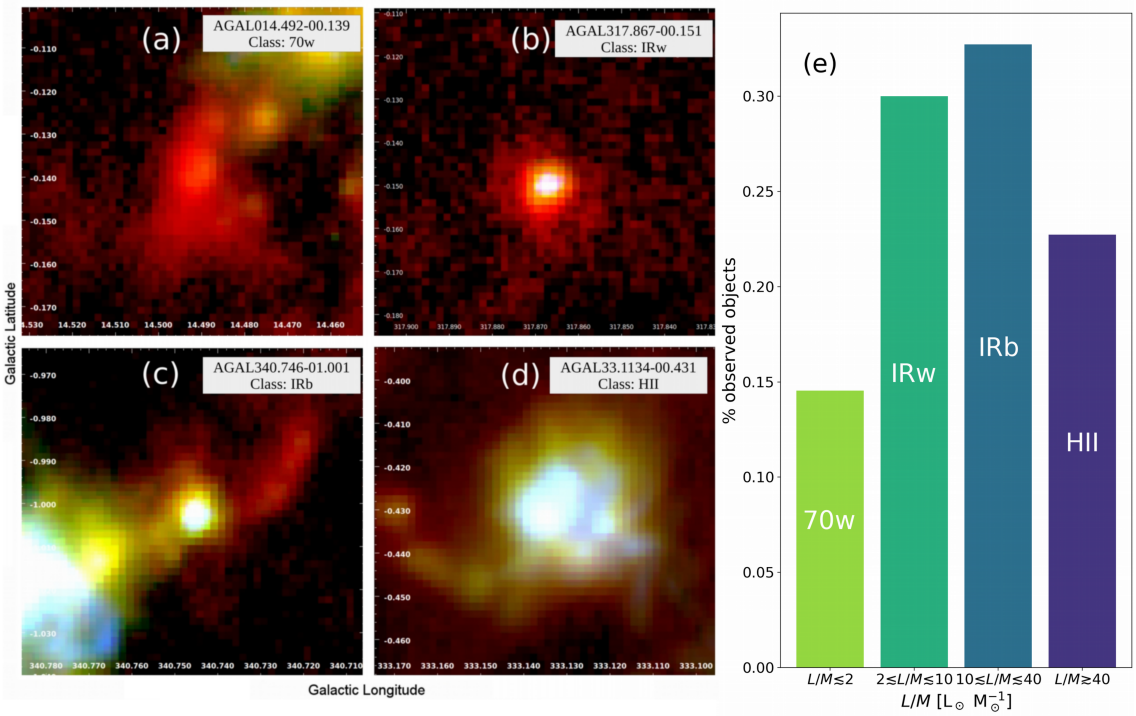

Figure 1.7: Panels (a-d): Illustrative three-colour images for each evolutionary class defined in the TOP100 sample: (a) 70w; (b) IRw; (c) IRb and (d) HII region. The size of each map is 5×5 arcmin; the colors are: the ATLASGAL 870 μm (red), the Hi-Gal PACS 160 μm (green) and 70 μm (blue). Readapted image from Fig. 2 of König et al. (2017). Panel (e): Bar-plot of the TOP100 sources divided by evolutionary class, showing the range of L/M and percentage of objects observed compared to the whole sample.

properties of clumps, and an example map of each phase is shown in Fig. 1.7:

1. 70 μm-weak sources (70w; see Fig. 1.7a): represent the earliest stage of massive-star formation and includes starless or clumps with very low level of ongoing star formation. This class of objects have $L/M \lesssim 2$ $L_\odot$ $M_\odot^{-1}$ and was identified from a visual inspection of the *Herschel*[4] Infrared Galactic Plane Survey (Hi-GAL, Molinari et al. 2010) images at 70 μm from PACS[5], either mostly devoid of any embedded point-like sources. The sources showing no compact emission at 70 μm within 10″of the submm emission peak, are considered 70w.

2. Infrared-weak sources (IRw; see Fig. 1.7b): detected at 70 μm, but not in the mid-IR band, and associated with sources in an early stage of star formation. The classification of the sources in this evolutionary phase was done inspecting the emission at 21 μm and 24 μm from the *Midcourse Space Experiment Survey* (MSX; Price et al. 2001) and the *Multiband Imaging Photometer Galactic Plane Survey* (MIPSGAL;

[4]Pilbratt et al. (2010)
[5]Poglitsch et al. (2010)

Carey et al. 2009) maps, respectively. A source is considered mid-infrared weak if there is no compact mid-infrared source associated with the submm emission peak and the flux reported in the compact source catalogues is below 2.6 Jy. This flux threshold indicates that the star formation is at an early (prestellar) stage and that the clumps are likely to be dominated by cold gas (Heyer et al., 2016). The L/M of the IRw sources varies between ~2 and ~10 $L_\odot$ $M_\odot^{-1}$.

3. Infrared-bright sources (IRb; see Fig. 1.7c): bright both at 70 and at 21-24 μm, but not detected in radio continuum at 4-8 GHz and associated with the high-mass protostellar stage, with L/M in the range of ~10-40 $L_\odot$ $M_\odot^{-1}$. These clumps are likely undergoing collapse in the absence of a strong magnetic field (Urquhart et al. 2013, 2014b), show signs of infall (Wyrowski et al., 2016) and are likely to be driving strong outflows (Navarete et al., 2019). With these properties, sources in the IRb stage are also likely to be significantly hotter than less evolved sources, giving rise to the bright emission at mid-IR wavelengths, and are one of the most active stages of massive-star formation.

4. Compact HII region sources (HII; see Fig. 1.7d): identified checking the radio continuum emission between 4 and 8 GHz using the *Coordinated Radio and Infrared Survey for High-mass star formation* survey (CORNISH; Hoare et al. 2012 and Purcell et al. 2013), the *The Red MSX Source* survey (RMS; Urquhart et al. 2007; Urquhart et al. 2013), or the targeted observations towards methanol masers reported by Walsh et al. (1998). When radio continuum emission is found at either 4 or 8 GHz within 10″of the ATLASGAL peak, the source is considered to be a compact HII region. In this stage, the objects show $L/M \gtrsim 40$ $L_\odot$ $M_\odot^{-1}$.

The range of L/M covered by each evolutionary phase is reported in Fig. 1.7e, together with the observed percentages of sources in each class. A summary of the evolutionary scenario for high-mass star-forming regions discussed above is sketched in the upper panel of Fig. 1.8. Molecular line surveys have been carried out on the TOP100 with APEX-12m, Mopra-22m, and the IRAM-30m single-dish telescopes, between 80 and 345 GHz, covering more than 120 GHz of bandwidth in three spectral windows[6] that contain a multitude of emission lines of both simple and complex molecules.

The survey carried out with APEX-PI230 (MPIfR) heterodyne receiver at 1.4 mm is the most extended among the aforementioned molecular surveys, ranging between 207 and 278 GHz. An example of spectra for each evolutionary class is shown in Fig. 1.8. A detailed analysis of a small sample of chemical species allowed to estimate their excitation parameters, and to derive accurate column densities (e.g. Giannetti et al. 2014, Csengeri et al. 2016, Giannetti et al. 2017b, and Tang et al. 2018). From the analysis of this dataset, Giannetti et al. (2017b) demonstrate that temperature, column densities of

[6]Not yet completely published.

several tracers, and H_2 volume density increase with time, as a function of evolutionary class and luminosity-to-mass ratio (L/M; see Fig. 1.8). Moreover, the dominant phases of the massive star formation process (compression, collapse, accretion and disruption; e.g. Zinnecker & Yorke 2007) have been identified in different regimes of L/M by Giannetti et al. (2017b). Therefore, the evolutionary sequence defined for the TOP100 is statistically valid, and the TOP100 sample can be considered to be representative of the Galactic protocluster population through all the evolutionary stages.

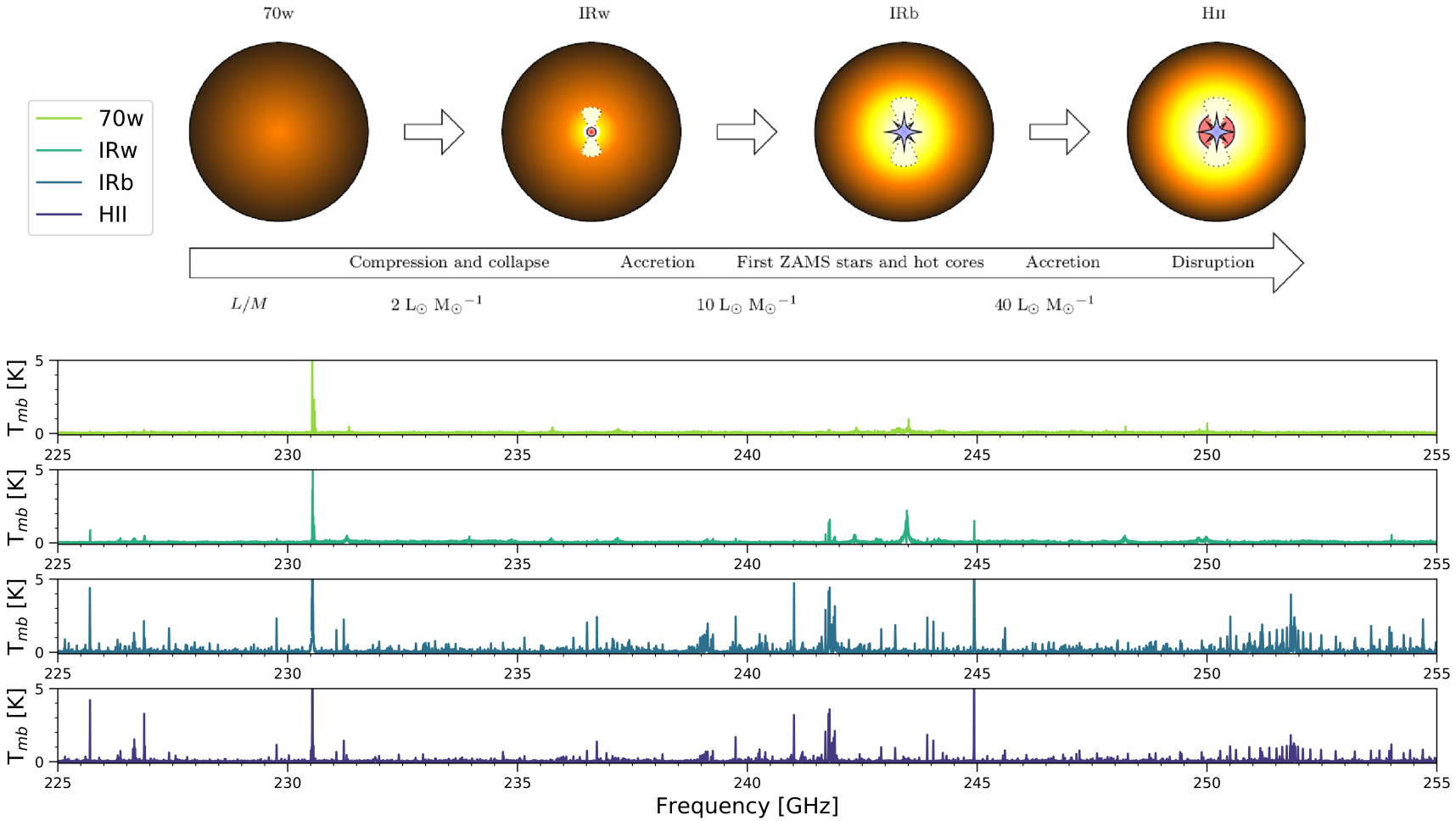

Figure 1.8: Top panel: Simplistic sketch of the different phases (on top of each clump) identified for the TOP100 (see also Fig. 1.7). The colour scale indicates the progressive warm-up of the clump around a young stellar object. The physical processes dominating in the corresponding interval of L/M are also shown at the bottom of the panel. This figure is taken from Fig 18 of Giannetti et al. (2017b). Other panels: example of a typical spectra in each evolutionary stages. The spectra are a 30 GHz window of the APEX-PI230 TOP100 survey.

Chapter 2

Chemical evolution during the star formation process

In this chapter, we propose an overview of the main chemical processes that take place during the early and late stages of the massive star formation process, as a basis for the following chapters of this Thesis. For a broader view of these topics, we refer to Kurtz et al. (2000), as a review on hot molecular cores (HMCs) and their chemistry; Caselli & Ceccarelli (2012) and Ceccarelli et al. (2014) in particular for a detailed description of depletion and deuteration processes; Ehrenfreund & Charnley (2000), Herbst & van Dishoeck (2009) and Jorgensen, Belloche & Garrod (2020), for a general discussion on the astrochemistry of dust, ice and gas and their interplay in the formation of complex molecules.

2.1 Molecules in the Interstellar Space

Astrochemistry, namely the study of the formation, destruction and excitation of atoms, molecules and ions in space, and how these processes affect the formation of galaxies, stars, and planets, is currently in its *golden age* thanks to the several new astronomical facilities operating at mm and sub-mm bands. Some of the most relevant examples are the IRAM *NOrthern Extended Millimetre Array* (NOEMA; Guilloteau et al. 1992), the *Atacama Pathfinder EXperiment 12-meter submillimetre telescope* (APEX; Güsten et al. 2006), the *Atacama Large Millimetre/submillimetre Array* (ALMA; Wootten & Thompson 2009), the *Stratospheric Observatory for Infrared Astronomy* (SOFIA; Young et al. 2012) and the *James Webb Space Telescope* (JWST; Gardner et al. 2006), to be launched in October 2021.

After almost a century from the first molecular detections in the interstellar space (CH by Swings & Rosenfeld 1937 and McKellar 1940; and CH^+ by Douglas & Herzberg 1941), more than 200 different molecules have been observed and confirmed in the inter-

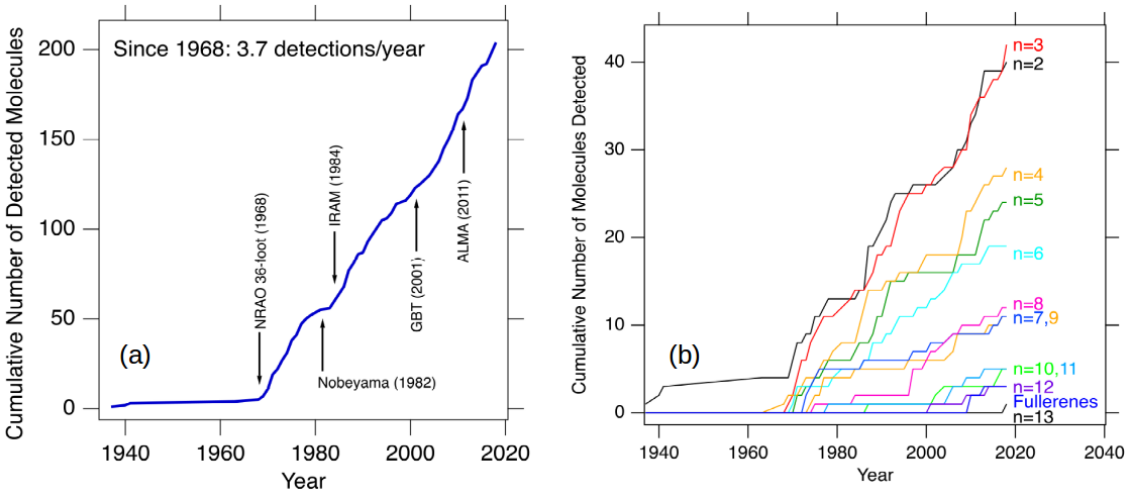

Figure 2.1: *Panel (a):* Cumulative number of confirmed detected interstellar molecules over time. The commissioning dates of several astronomical facilities are reported with arrows. After the birth of molecular radio astronomy in the 1960s, there have been on average 3.7 new molecular detections per year until 2005, and 5.4 new molecular yr^{-1} between 2005 and 2018; *Panel (b)*: The same as Panel (a), distinguished by the number of atoms, n, that compose the molecular structures. The term 'fullerenes', indicates the largest carbon-based molecules detected to date (i.e. C_{60}, C_{60}^{+} and C_{70}; see the text). The curves are colour-coded by n, and labelled on the right. Both plots are taken from McGuire (2018).

stellar medium or circumstellar shells[1].

Fig. 2.1a, taken from McGuire (2018), shows the cumulative number of detected molecules with time to date, as well as the commissioning dates of a number of key observational facilities. Fitting the data, McGuire (2018) reported that the number of new molecules detected per year, ~3.7 between 1960 and 2005, has increased to ~5.4 molecules yr^{-1} between 2005 and 2018, reflecting the significant progress made in recent decades.

Compared with the terrestrial standards, interstellar molecules are quite small, with sizes that range from 2 to 13 atoms, except for some rare cases (e.g. C_{60} - Werner et al. 2004 and Sellgren, Uchida & Werner 2007; C_{60}^{+} - Foing & Ehrenfreund 1994 and C_{70} - Cami et al. 2010). Molecules with multiple carbons and composed of at least 6 atoms, which represent among 35% of the molecules detected to date, are usually defined as *Complex Organic Molecules* (COMs; e.g. Herbst & van Dishoeck 2009) or *interstellar Complex Organic Molecules* (*i*COMs, as recently re-defined by Ceccarelli et al. 2017)[2].

A recent example of the new frontier unveiled by ALMA is well represented by the first publication of the GUAPOS project[3] results by Mininni et al. (2020); see Fig. 2.2. A

[1]For an updated list of the confirmed molecular detections see the Cologne Database for Molecular Spectroscopy (CDMS) at https://cdms.astro.uni-koeln.de/classic/molecules.

[2]In this Thesis, we choose to adopt the term COMs.

[3]G31.41+0.31 Unbiased ALMA sPectral Observational Survey (GUAPOS; visit the website)

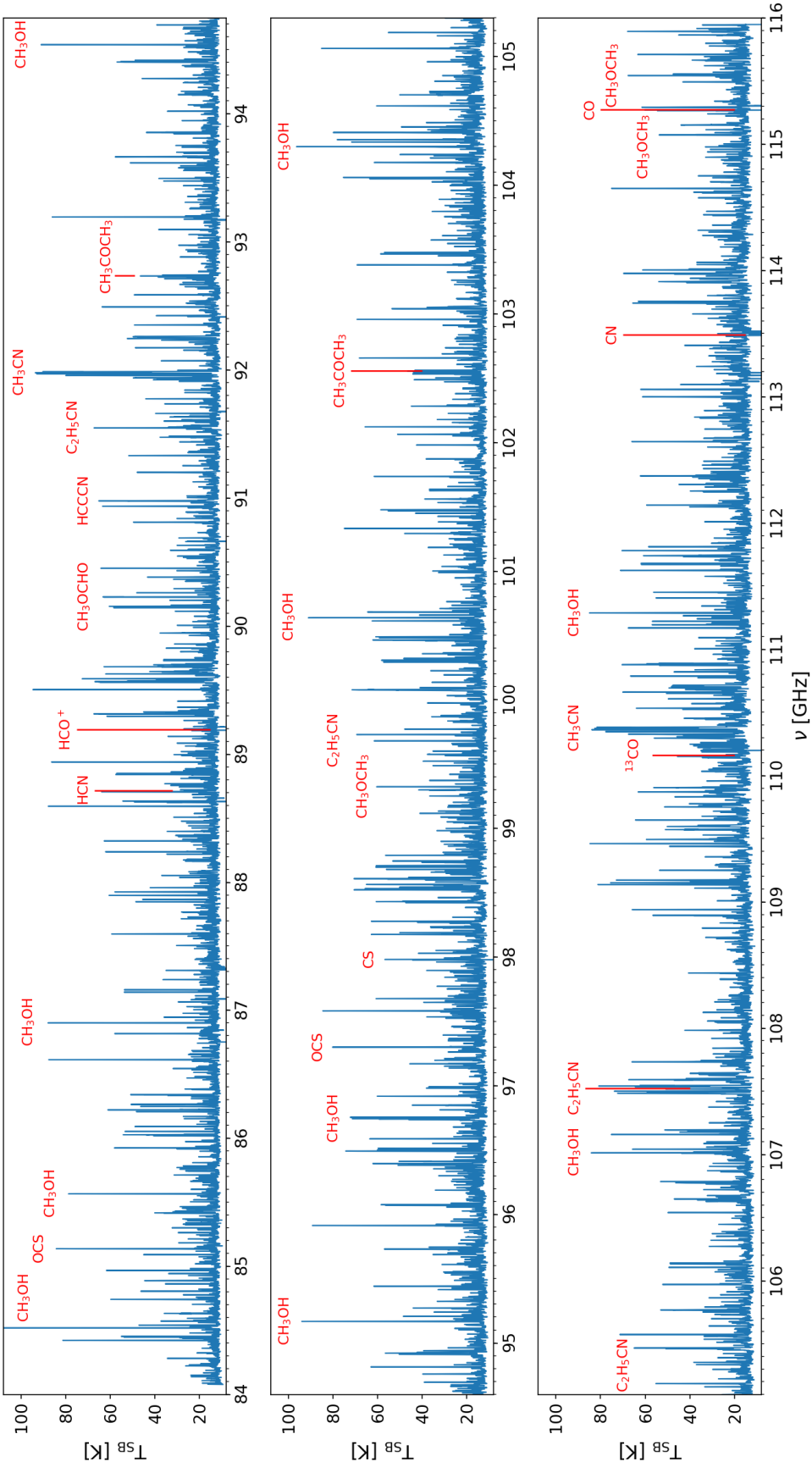

Figure 2.2: Full final GUAPOS spectrum from 84 to 116 GHz. In red are shown some of the most common molecular species. The temperatures are reported on the synthesised-beam brightness temperature, T_{SB}, scale in units of Kelvin. This figure is adapted from figure 4 of Mininni et al. (2020) (Mininni, private communication).

spectral survey covering the entire ALMA Band-3 (i.e. from ~84 to ~116 GHz) with a spectral resolution of ~0.488 MHz (i.e. ~1.3-1.7 km s^{-1}) was performed towards G31.41+0.31 (Immer et al. 2019 and Cesaroni 2019), one of the most chemically rich HMCs outside the Galactic Centre (GC). Previous works have pointed out the chemical peculiarity of G31.41+0.31, confirming the detection of several large COMs: e.g. ethyl cyanide, C_2H_5CN (Beltrán et al. 2005); dimethyl ether, CH_3OCH_3 (Fontani et al. 2007 and Coletta et al. 2020); glycolaldehyde, $CH_2(OH)CHO$ (Beltrán et al. 2009); acetone, CH_3COCH_3 (Isokoski, Bottinelli & van Dishoeck 2013); and methyl formate, CH_3OCHO (Rivilla et al. 2017). While the work of Mininni et al. (2020) is focused on the study of the isomeric forms[4] of $C_2H_4O_2$, as precursors of prebiotic molecules, the GUAPOS survey has the potential to confirm the detection of a larger number of COMs as well as possibly find new ones.

Fig. 2.2 shows the remarkable richness of emission lines in a molecular hot core, and also how much this spectrum is close to the so-called confusion limit. This means that longer integration times, with the same observational setup, would not bring any new information. While reaching the line confusion limit raises some practical problems for the data reduction and interpretation, for which we refer to Jorgensen, Belloche & Garrod (2020), the remarkable number of lines and hypothetical COMs that can be detected through these surveys leads us to wonder how this chemical wealth is produced and why the spectra of the star formation regions are particularly rich in molecular lines (e.g. Herbst & van Dishoeck 2009, Caselli & Ceccarelli 2012 and McGuire 2018).

In the following Sections, we discuss the main chemical processes that occur under the typical conditions of the early and late stages of the star formation process.

2.2 Astrochemistry and star formation

Due to the low densities of the interstellar medium (see Tab. 1.1) compared to terrestrial standards (n~10^{19} cm^{-3}; Fraser, McCoustra & Williams 2002), most of the gas-phase chemical reactions take place through two-body collisions. In this scenario, two generic reactants A and B react to form one or more generic products, i.e. $A + B \longrightarrow C + D$. The rate coefficient, k^{gas}, that governs the gas-phase chemistry can be expressed by the modified *Arrhenius* equation (McElroy et al. 2013), in unit of cm^3 s^{-1},

$$k^{gas}(T_g) = \alpha \left(\frac{T_g}{300}\right)^{\beta} \exp\left(-\frac{\gamma}{T_g}\right), \tag{2.1}$$

where α is the reaction rate, β characterises the temperature dependence of the rate, and γ usually represent the activation barrier (in Kelvin) for the reaction. Given the dependence of Eq. 2.1 on T_g and γ, one can note that at low temperatures exothermic reactions without activation barrier are favoured in gas-phase. Ion-neutral reactions, i.e. re-

[4] *isomers*: molecules with the same chemical composition, but with a different molecular structure.

actions involving positive ions and neutrals, often obey these conditions (van der Tak 2005). In dense regions, positive ions are mainly formed by cosmic-ray-induced ionisations, AB + c.r. $\longrightarrow$ $AB^+ + e^-$, with a rate-coefficient that depends on the cosmic-ray ionisation rate for molecular hydrogen, ζ_2. Starting from the pioneering work of Black, Hartquist & Dalgarno (1978), based on OH and HD observations, the ζ_2 has been probed through a mix of observations and chemical models both in low- and high-mass star forming regions (e.g. Caselli et al. 1998, McCall et al. 1999, van der Tak & van Dishoeck 2000, Ceccarelli et al. 2004, Vaupré et al. 2014, Indriolo et al. 2015, Ivlev et al. 2019 and Bovino et al. 2020). Overall, there is no consensus on the value of ζ_2 in dense regions and the various estimates differ by orders of magnitude, between $\sim 10^{-18}$ and 10^{-15} s^{-1}, and still represents a great uncertainty in many astrochemical models (see Sect. 4.5).

While gas-phase chemistry is always active throughout all evolutionary stages of high-mass star-forming regions introduced in Chapter 1, other chemical processes need particular conditions of temperatures and densities to take place. As the average density and temperature change in the cloud, different chemical processes are activated, giving birth to new chemical species and/or destroying the existing ones. This provides a variety of chemical tracers that can be used to probe the different regions (and scales) of the star-forming environments.

Fig. 2.3, taken from Wang et al. (2020), shows an example of the tomographic view of molecular filaments revealed by different tracers. What appears as a single structure of dozens of parsec in size, when observed in CO (black contours in each panel of Fig. 2.3), reveals smaller internal structures when mapped in other species (see the different white contours in each panel). This view becomes more complicated if we look on smaller scales. Fig. 2.4 is adapted from Oberg & Bergin (2020) and shows the prestellar core Barnard 68 (panel a). The thermal dust continuum emission at 850μm (panel b; Bianchi et al. 2003) has a different distribution than the molecular line emission traced by N_2H^+ and $C^{18}O$, shown in panels c and d (Bergin et al. 2002), respectively. While the N_2H^+ emission coincides with the more intense part of the dust emission at the centre of the cloud, a different structure is displayed by $C^{18}O$. The latter molecule outlines the shape of the cloud, but is less intense at the centre, revealing a ring-like structure around the peak of the dust emission, indicated by the black cross (Bergin & Tafalla 2007).

Why is this happening? Why is the spatial distribution of the various molecules in star-forming regions so different? This becomes more clear when looking at the different physical conditions of a clump and how these change when moving from the large to the small scales. Fig. 2.5 shows the internal density structure of a generic clump. In the upper panel are reported, in arbitrary scales, the density (violet), the temperature (orange) and the external UV flux (red) profiles from the outer to the inner regions of the clump. The middle panel of Fig. 2.5 is a sketch of the chemical structure of the clump, while the lower panel shows the principal chemical processes that are at work in the regions indicated. The density profile of the clump increases from the edge to the centre. If no YSOs are formed

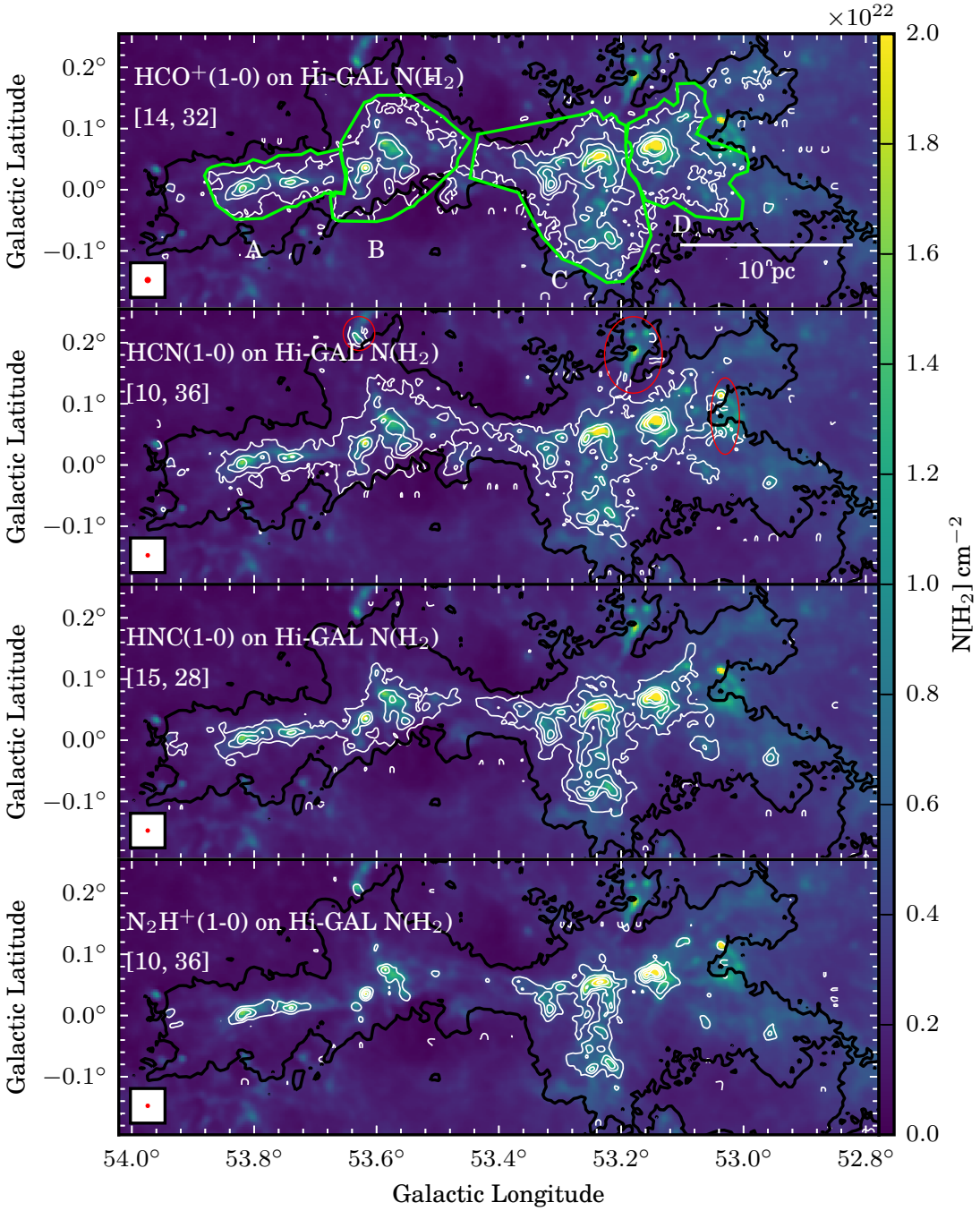

Figure 2.3: Tomographic view of a molecular filament revealed by different chemical tracers. The black external contour in each panel is associated to the CO (1-0) integrated emission in the range of 10< $v_{\rm lsr}$ <36 km s^{-1}, which reveals four sub-filamentary regions indicated in green (labelled A, B, C, and D). From the top to the bottom panel, white contours correspond to the integrated intensity of the $J = 1-0$ transition of HCO^+, HCN, HNC and N_2H^+. The number between square brackets in each panel indicate the velocity ranges of the integrations. The colour-map traces the H_2 column density derived from the SED-fitting including the Hi-GAL 160, 250, 350, and 500 μm continuum maps (Zucker, Battersby & Goodman 2018). The red ellipses mark the emissions that are at a different distance and not associated with the filament. These maps are taken from Fig. 2 of Wang et al. (2020).

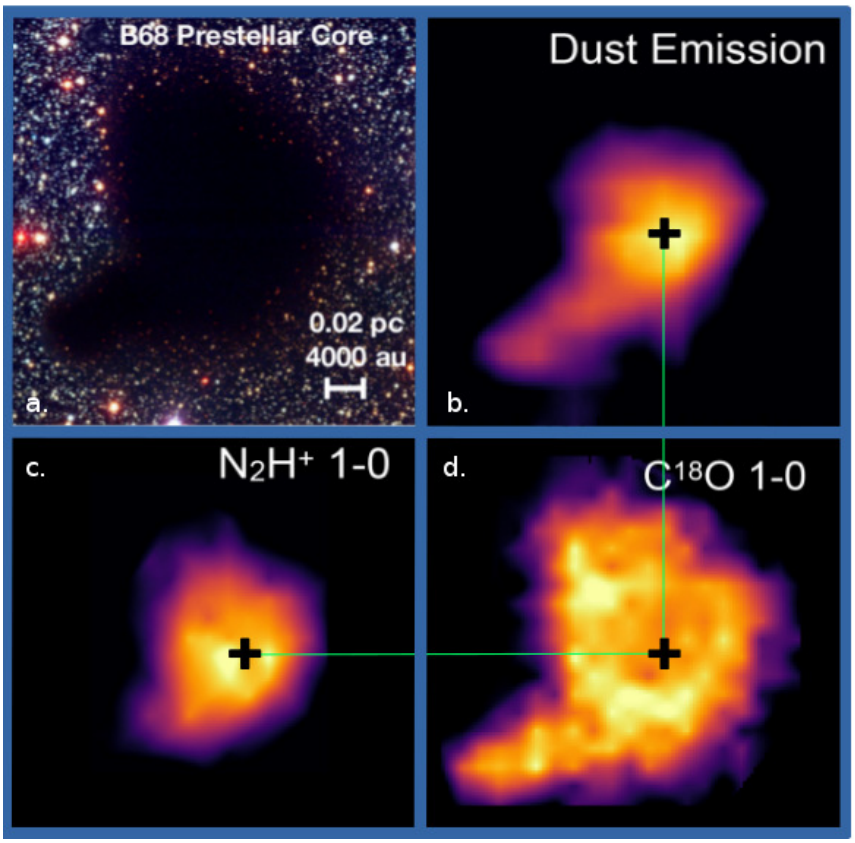

Figure 2.4: Multiwavelength image of the molecular core Barnard 68, taken from Oberg & Bergin (2020) and readapted from Bergin & Tafalla (2007). Panel a shows the obscuration region observed in the optical. Panel b shows the dust continuum emission at 850 μm. Panels c and d show the integrated molecular emission of N_2H^+ and $C^{18}O$(1-0), respectively. The different spatial distribution is clear, with N_2H^+ (1-0) closely following the emission profile of the dust, while $C^{18}O$ shows a ring-shaped emission around the dust sub-mm emission peak.

in the central region, the UV flux and the temperature have the opposite trend, as the external radiation field is progressively attenuated moving towards the innermost, denser regions of the cloud. When the volume density of molecular hydrogen, $n(H_2)$ becomes larger than $\sim 10^4$ cm^{-3} and the temperature of the cloud, T_g, is lower than ~20 K (see point 4 in the lower panel of Fig. 2.5), several C-, N- and O-bearing species, in particular carbon monoxide, start to be removed from the gas phase as they are trapped on the surface of the dust via the freeze-out, or depletion, process (e.g. Kramer et al. 1999, Bergin et al. 2002, Caselli et al. 2008, Fontani et al. 2012, Wiles et al. 2016 and Sabatini et al. 2019; see also Bergin & Tafalla 2007 for a review). This process has a characteristic timescale, τ_{dep}, that decreases with increasing H_2 volume density, e.g. Caselli et al. (1999),

$$\tau_{dep} = \frac{10^9}{\mathcal{S}\, n(H_2)} \quad \mathrm{yr}\,, \tag{2.2}$$

where $\mathcal{S}$ is the sticking coefficient and represents the probability that a given molecule

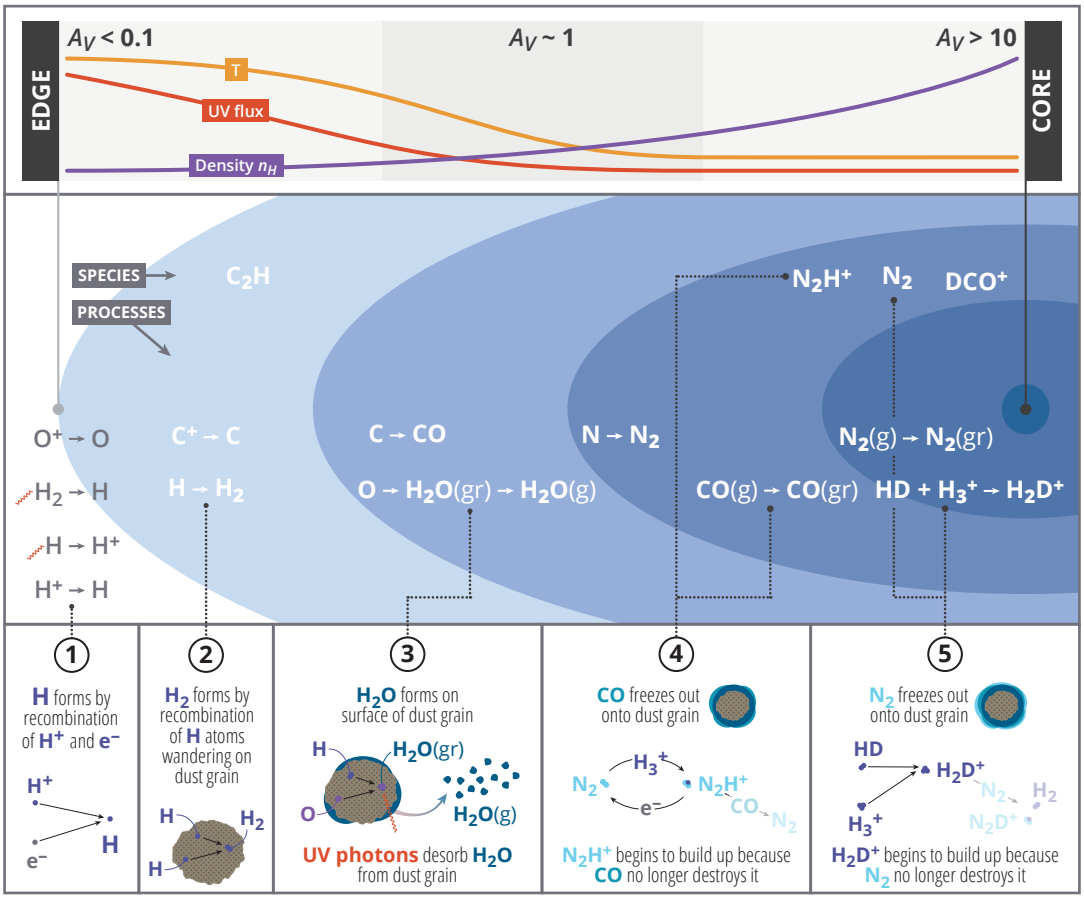

Figure 2.5: Sketch of the chemical structure of a molecular cloud from the UV exposed cloud edge (left) to the protected cloud core (right), where cloud depth is parameterised by extinction (measured in visual magnitudes A_V). The density, temperature and UV flux profiles as a function of the distance from the centre of the core are shown on the top of the figure. The bottom panels summarise the principal chemical processes that alter the internal composition of the cloud. This scheme is taken from Oberg & Bergin (2020).

is adsorbed on the surface of a dust grain after a two-body collision (e.g. Hollenbach & McKee 1979 and Burke & Hollenbach 1983). The observational effect of the freeze-out is a deficiency of CO inside the clouds, which produces the empty ring-like structure seen in Fig. 2.4d.

The CO adsorption onto the grains' surface also favours the progressive enrichment of deuterium atoms with respect to hydrogen in the molecules, through the so-called deuterium fractionation process (e.g. Ceccarelli et al. 2014). The main driver of deuteration, the trihydrogen cation H_3^+, leads to the enrichment of deuterated molecules like H_2D^+ and N_2D^+ in the gas phase (e.g. Pagani et al. 2009 and Vastel et al. 2012), mainly due to the quenching of its destruction path (Dalgarno & Lepp 1984)

$$H_3^+ + CO \rightarrow HCO^+ + H_2\,. \tag{2.3}$$

Diazenylium (N_2H^+) can be formed directly from H_3^+, e.g. Jørgensen, Schöier & van Dishoeck (2004), via

$$H_3^+ + N_2 \rightarrow N_2H^+ + H_2\,. \tag{2.4}$$

Reaction 2.4 depends on the degree of CO depletion, so that this explains why the scales mapped by N_2H^+ are smaller than those traced by CO (see Fig. 2.3) and also why their observed abundances are anti-correlated on the core scales (see Fig. 2.4c and d). Due to the low abundances of N_2, deuterium enrichment in the gas phase is enhanced by the formation of H_2D^+ through the reactions:

$$H_3^+ + HD \rightleftharpoons H_2D^+ + H_2 + \Delta E\,, \tag{2.5}$$

and its subsequent deuterations that lead the formation of D_2H^+ and D_3^+. In reaction 2.5, the released energy ΔE depends on the H_3^+ and H_2D^+ spin states involved in each reaction (e.g. Hugo et al. 2007). Reaction 2.5 is thus the first step that favours the deuterium fractionation and proceeds from left-to-right, unless there is a substantial fraction of ortho-H_2 (e.g. Gerlich, Herbst & Roueff 2002).

Deuteration and depletion are thus two strongly connected processes, and the latter is also thought to have drastic consequences on the formation routes of several COMs during the early stages of star formation (e.g. Caselli & Ceccarelli 2012 and Ceccarelli et al. 2014). In Chapters 3 and 4 of this Thesis, we present recent results connected to these processes and published in Sabatini et al. (2019) and Sabatini et al. (2020), respectively. The former presents large-scale CO-depletion maps of a typical high-mass star-forming region (G351.77-0.51; Leurini et al. 2019), while the latter discusses the reliability of *ortho*$-H_2D^+$ (hereafter *o*-H_2D^+) as the chemical signature of the evolutionary stage of a sample of massive clumps in different evolutionary stages. For this reason the following Sections are further focused on what is known about CO-depletion and *o*-H_2D^+ in high-mass star forming regions.

2.2.1 CO-depletion

In the last two decades, evidence of CO depletion has been widely found in numerous sources (e.g. Kramer et al. 1999; Caselli et al. 1999; Bergin et al. 2002; Fontani et al. 2012; Wiles et al. 2016; Sabatini et al. 2019), with important consequences for the chemistry, such as an increase in the fraction of deuterated molecules (Bacmann et al. 2003). Observationally, how much of the CO is depleted onto the surface of dust grains is usually characterised by the depletion factor (e.g. Caselli et al. 1999; Fontani et al. 2012), defined as the ratio between the "expected" CO abundance with respect to H_2 (χ^E_{CO}) and the observed one (χ^O_{CO}):

$$f_D = \frac{\chi^E_{\rm CO}}{\chi^O_{\rm CO}} \,. \tag{2.6}$$

The expected CO abundance is $\sim 10^{-4}$ in the solar neighbourhood (e.g. Lacy et al. 1994) and has a gradient with Galactocentric distance (Wouterloot et al. 2008); the observed abundance is directly obtainable by the ratio between the CO and H_2 column densities (Giannetti et al. 2014). The main CO isotopologue[5], $^{12}C^{16}O$, is virtually always optically thick, and thus its intensity is not proportional to the CO column density (Rigby et al. 2019). One of the less abundant CO isotopologues (such as $^{12}C^{18}O$, see Chapter 3) can be used to obtain a more accurate estimate of f_D.

High-mass star-forming regions are potentially more affected by large-scale depletion compared to their low-mass counterparts: the high volume densities of H_2 both in the clumps and in the surrounding inter-clump regions make these sources more prone to high levels of molecular depletion (e.g. Giannetti et al. 2014), since the depletion timescale is inversely proportional to the molecular hydrogen volume density (Eq. 2.2).

CO-depletion has been studied in different samples of young high-mass star-forming regions, with observed factors varying between one, in the case of complete absence of depletion, and a few tens (e.g. Thomas & Fuller 2008; Fontani et al. 2012; Feng et al. 2016). Large f_D are reported in Fontani et al. (2012) where it reaches values of up to 50-80, a factor of $\sim$10 larger than those observed in samples of low mass clouds (e.g. Bacmann et al. 2002, 2003; Ceccarelli et al. 2007; Christie et al. 2012). However, Giannetti et al. (2014) noted that the method used by Fontani et al. (2012) to calculate the H_2 column density, $N(H_2)$, yields values $\sim$2.7 times larger than what they themselves used to obtain this quantity, mainly due to the different dust absorption coefficient assumed. Rescaling the Fontani et al. (2012) depletion factors under the same assumptions, the typical values of f_D vary from a minimum of 1 to a maximum of $\sim$10, except for some peculiar cases. In fact, there are studies where high-resolution observations in massive clumps revealed even larger depletion factors: this is the case for the high-mass core in the IRDC G28.34+0.06 where, at a spatial resolution of $\sim 10^{-2}$ pc (using a source distance of 4.2 kpc; Urquhart et al. 2018), f_D reaches values of $10^2 - 10^3$, the highest values of f_D reported to date (Zhang et al. 2009).

Through the TOP100 sample, empirical correlations were found between f_D and few relevant physical quantities of the clumps, such as the excitation temperature ($T_{\rm ex}$) of CO, or the H_2 column density (Giannetti et al. 2014). Figures 2.6a shows that f_D decreases with increasing temperature, suggesting that CO progressively returns to the gas-phase in the most evolved and hot sources, as will be further discussed in Sect. 2.2.3. The f_D values obtained by different isotopologues (compare the two panels a_1 and a_2 in Figure 2.6) seem to suggest that $^{12}C^{17}O$ and $^{12}C^{18}O$ are mapping (on average) the same clump region.

[5] *isotopologue*: molecules that only differ in their isotopic composition, i.e. with at least one atom with a different number of neutrons than its parent.

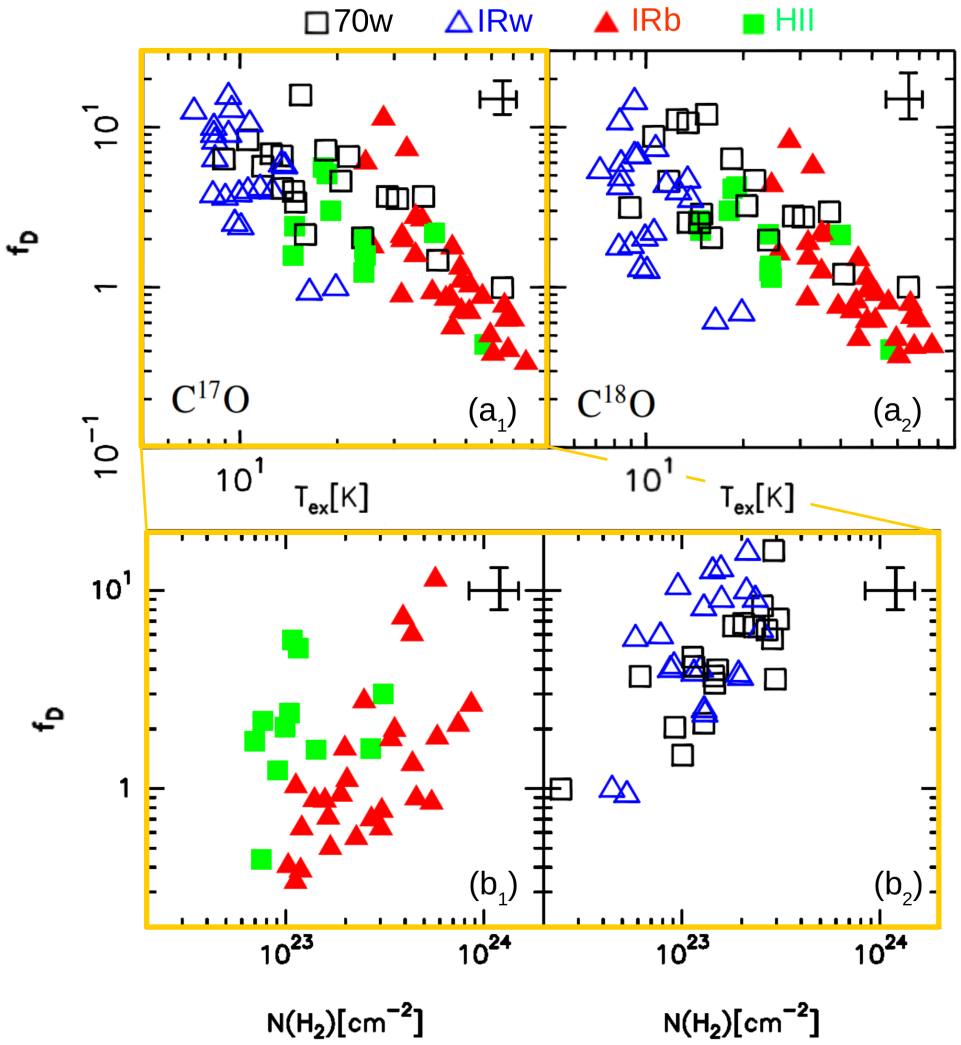

Figure 2.6: Panels (a) show CO depletion factors from $^{12}C^{17}O$ (on the left) and $^{12}C^{18}O$ (on the right) as a function of T_{ex} for the TOP100 sample. Panels (b) show the CO depletion factor (from $^{12}C^{17}O$) as a function of $N(H_2)$, for the TOP100 sample. A typical uncertainty is shown in the top right corner of each panel. Each evolutionary class is shown with a different symbol and colour, as indicated. This Figure is adapted from Figures 8 and 9 of Giannetti et al. (2014).

In both the panels of Figure 2.6a there are no evident trends between f_D and the four clump's evolutionary classes identified in the TOP100 sample (see Sect. 1.5.1), but a separation between the younger (i.e. 70w and IRw) and the more evolved (i.e. IRb and HII) sources can be observed. This separation is more evident in Figure 2.6b_1 and b_2, where the depletion factor derived by $^{12}C^{17}O$ is shown as a function of $N(H_2)$ and the evolutionary class. Younger sources (Figure 2.6b_2) have f_D that, on average, are larger by a factor of ~5 compared to those computed in the evolved sources (Figure 2.6b_1).

Observing depletion factors ranging from 1 to 10 means that along the line of sight there will be regions in which the CO is completely in gas phase (i.e. not depleted), and other regions in which the depletion factor reaches values larger than 10, as recently

demonstrated by Bovino et al. (2019). The size of the region where most of the CO is frozen onto dust grains is called the depletion radius, and provides information on the approximate spatial scales on which different chemical processes operate in high-mass star forming regions. At these scales, for example, it is reasonable to think that derivation of H_2 column densities from CO lines (see e.g. Bolatto, Wolfire & Leroy 2013 and references therein) and the studies of the gas-dynamics using lines of the carbon monoxide isotopologues might be strongly affected by the depletion process, as not all of the CO present is directly observable. Moreover, the mild correlations between f_D and the evolutionary stage of massive clumps observed in Figures 2.6b suggests that the size of the depletion radius might be an additional tool to study the evolution of massive star-forming regions, and to test the evolutionary sequence defined by L/M (Sections 1.4 and 1.5.1).

One way to verify the latter hypothesis, is by mapping high-mass star-forming regions on both large and small scales to determine the depletion factor fluctuations over a broad range of densities and temperatures (see Chapter 3).

2.2.2 Deuterium fractionation

As discussed in Sect. 2.2, the CO depletion onto the surface of dust grains also favours the progressive enrichment of deuterium atoms compared to hydrogen in the molecules, through the so-called deuterium fractionation process (e.g. Caselli & Ceccarelli 2012). Deuterium, D, was formed at the birth of the Universe with an abundance with respect to hydrogen atoms, D/H, estimated to be $\sim 1.6 \times 10^{-5}$, and subsequently destroyed in the interiors of the stars (Ceccarelli et al. 2014). Therefore, D/H may vary from place to place. For a generic molecular species, $\mathtt{X}$, the deuterium fraction is characterised by the ratio

$$D^{\mathtt{X}}_{\rm frac} = \frac{\mathtt{X(D)}}{\mathtt{X(H)}}, \tag{2.7}$$

that quantifies the fraction of the deuterated form of that species, $\mathtt{X}$(D), with respect to the non-deuterated one, $\mathtt{X}$(H). Without the deuterium fractionation process, $D^{\mathtt{X}}_{\rm frac}$ should be equal to the cosmic D/H for molecules with just one D (i.e. $D^{\rm HCN}_{\rm frac}$ = DCN/HCN = D/H), and reach smaller values for molecules with more than one deuterium (e.g. $D^{\rm H_2O}_{\rm frac}$ = D_2O/H_2O < D/H); the abundance of D-bearing molecules would be extremely low. But deuterium has favourable conditions for fractionation, making the value of $D^{\mathtt{X}}_{\rm frac}$ a powerful tool to study the origins of different astrophysical objects (e.g. star-forming regions, planets and comets) and to follow their evolution.

In the astrochemical context of star formation, the deuteration has been exploited to measure the age of low-mass (i.e. M $\lesssim 8-10$ $M_\odot$) cores via different tracers (e.g. via H_2D^+ by Brünken et al. 2014), while in the high-mass regime, age estimates are much more complex due, in particular, to the uncertainties in the dynamical and chemical ini-

tial conditions (see Sect. 1.3). A first attempt to use cold-gas chemistry to derive the age of high-mass star-forming regions was carried out by Kong et al. (2016) for cores of $\sim 15-60\,M_{\odot}$, by measuring the deuterium fraction via N_2H^+. However, it has been shown by Pillai et al. (2012) that this species may not be ideal to trace the first stages of the high-mass star formation process. Pillai et al. (2012) mapped part of the DR21 complex in o-H_2D^+ $J_{K_a,K_c} = 1_{10} - 1_{11}$ and N_2D^+ $J = 3 - 2$, observed with the *James Clerk Maxwell Telescope* (JCMT; Holland et al. 1999) and the *Submillimiter Array* (SMA; Ho, Moran & Lo 2004), respectively. The data reveal very extended o-H_2D^+ emission, in agreement with the results of Vastel et al. (2006) for the low-mass source L1544. The data also show that this species mainly traces gas that is not seen in dust continuum emission or in the interferometric N_2D^+ data. H_2D^+ may thus be sensitive to gas that eludes detection in the most commonly used tracers, and its presence might indicate an even earlier stage in the process of star formation.

More recently, Giannetti et al. (2019) have simultaneously detected for the first time o-H_2D^+ and N_2D^+ in three high-mass star-forming clumps with the APEX-telescope, opening the possibility of investigating their abundance variation in sources in different evolutionary stages (the G351.77-0.51 complex; e.g. Leurini et al. 2019, see also Sect. 3.1), from a clump that is still quiescent at 70 μm, to one which hosts luminous YSOs. Giannetti et al. (2019) observed that the abundance of N_2D^+ increases with evolution, showing a difference of a factor of ~2 between the least and the most evolved clump (see Fig. 2.7). Particularly relevant is also that the o-H_2D^+ abundance decreases in the two most evolved clumps (a factor of ~10 in Clump 2), likely due to the chemical conversion of H_2D^+ into D_2H^+ and D_3^+ (which would also boost the production of N_2D^+), or to the destruction by desorbed CO in the presence of a protostellar object. Similar results were reported by Kong et al. (2016), who obtained upper limits for the abundances of o-H_2D^+ of $\sim 10^{-11}$ in two high-mass star-forming regions associated with outflow activity (see Tan et al. 2016), and high abundances of N_2D^+, confirming the fact that N_2D^+ is forming at later times compared to o-H_2D^+ and under different physical conditions. Attempts to demonstrate this hypothesis have been recently reported by Miettinen (2020), using APEX observations in three pre-stellar and three protostellar cores in Orion-B9. Although observing a downward trend in o-H_2D^+ abundance of about a factor of 4 with evolution, the anti-correlation with N_2D^+ was not confirmed. However, as pointed out by the authors, their spatial offset between the N_2D^+ and the o-H_2D^+ observations ($\sim 10''$ on average) might be the cause of the observed behaviour, i.e. both molecules are sampling different regions.

The observations described above have been carried out on only a small samples of massive clumps, and to ascertain the reliability of o-H_2D^+ as an age indicator requires studying a larger sample of high-mass star-forming regions and containing sources at all evolutionary phases (see Chapter 4).

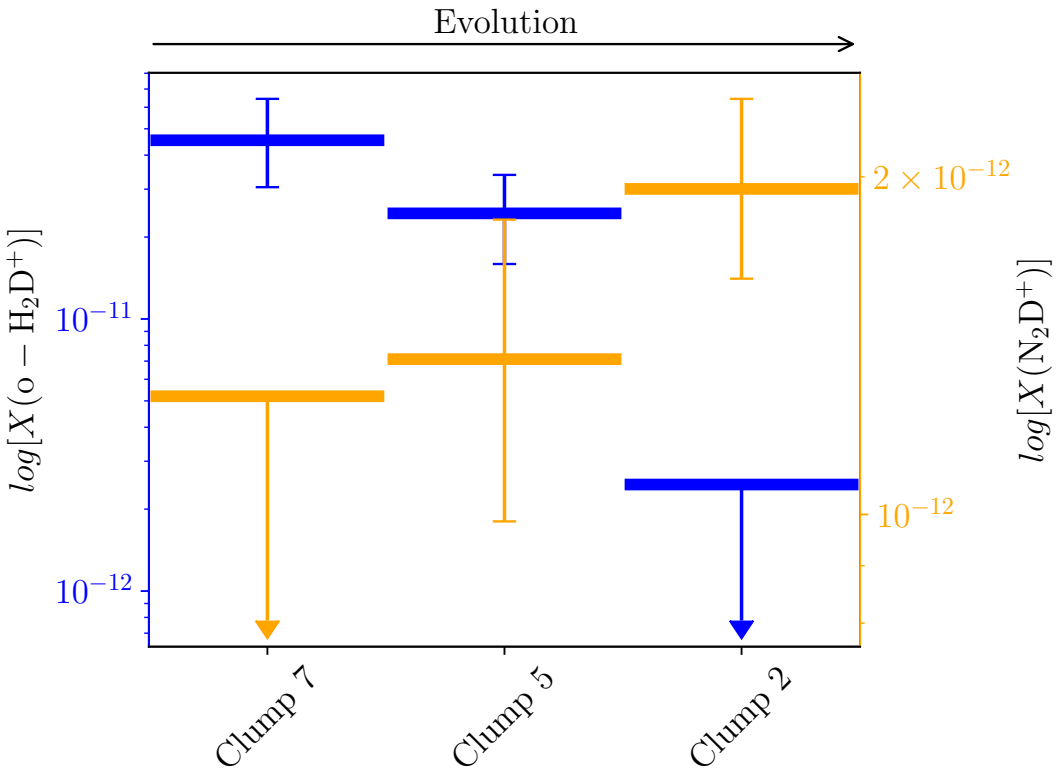

Figure 2.7: Observed anti-correlation between the abundances of o-H_2D^+ (in blue) and N_2H^+ (in orange) as a function of the evolutionary stage of three massive clumps located in the IRDC G351.77-0.51. The clumps are ordered by evolutionary stage, from the least- to the most evolved. The 95% credible interval is indicated as errorbar. This Figure is taken from Figure 3 of Giannetti et al. (2019).

2.2.3 Warm-up and molecular desorption

Figure 2.8, taken from Oberg & Bergin (2020), shows a schematic sketch of the rather complex and dynamical inner structure of a clump during its late stages of evolution.

Far from the YSO (Figure 2.8a), the temperatures and the densities of the dusty-gaseous envelope are low, i.e. T_g ~10 - 30 K and $n(H_2)$~10^2 cm^{-3}, and the chemistry can be considered, initially, as the continuation of the processes discussed during the earlier phases (see Fig. 2.5). As discussed in Sect. 2.1, at low gas temperatures many atoms and molecules can deplete (or accrete) onto dust particles (see points 3 to 5 in the lower panel of Fig. 2.5). The interactions between dust grains and gas-phase species can occur via chemical valence forces, known as *chemisorption*, or via weak van der Waals forces, known as *physisorption*. In cold molecular clouds, it is normally (but not universally) assumed that sticking occurs via physisorption, since often chemisorption has a potential barrier (see Ehrenfreund & Charnley 2000 and Herbst & van Dishoeck 2009).

The slow depletion of different chemical species on dust-grains leads to the production of icy layers around their carbonaceous and silicate cores (e.g. Hasegawa, Herbst

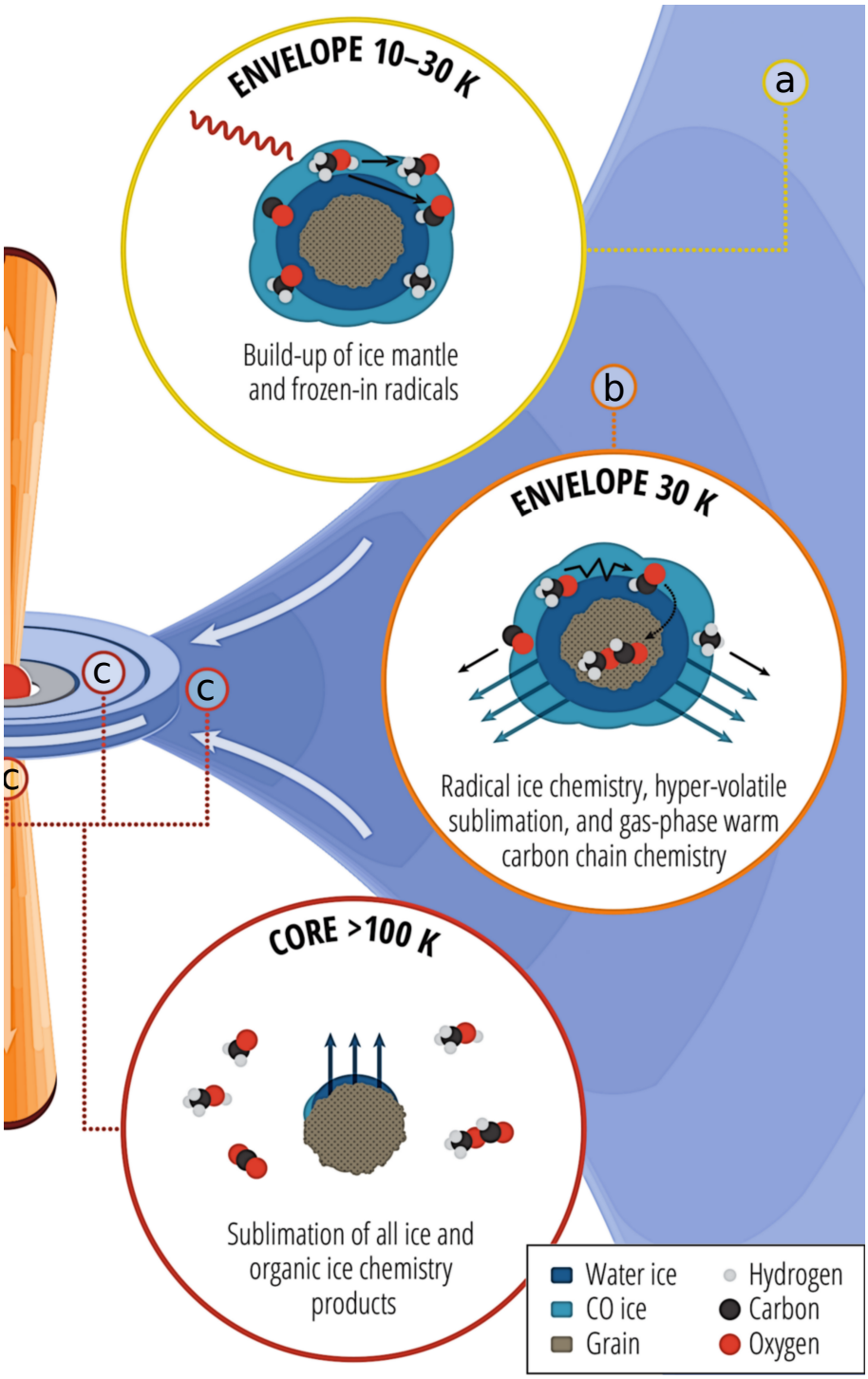

Figure 2.8: Sketch of the inner structure of a clump during the late stages of the star formation process. The YSO is on the central left side of the image. Around the YSO an accretion disk can be formed to preserve angular momentum during the accretion. Angular momentum is also removed from the system through the formation of protostellar outflows and jets. The chemistry of the cloud is sketched at three different places of the clump (points a, b and c). The blue colour scale of the clump follows its density structure. This image is taken from Oberg & Bergin (2020).

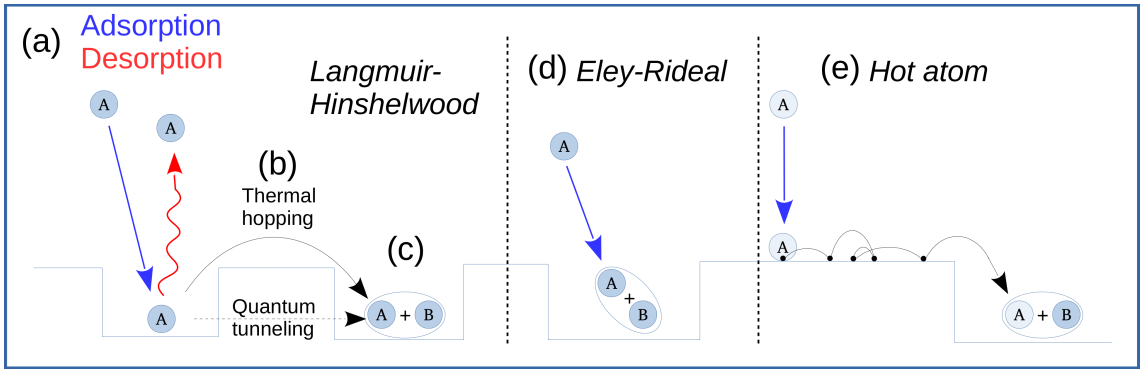

Figure 2.9: Sketch of the three major mechanisms for surface reactions on a regular grain surface (blue-rectangular surface). Panel (a): Adsorption (blue arrow) and desorption (red arrow) of a molecule on the dust-grain surface; Panel (b): thermal hopping and quantum tunnelling effects, two of the major diffusion mechanism through which molecules can scan the dust-grain surfaces; Panel (c): a new molecular species is formed from the generic reaction 'A+B'; Panels (a+b+c) Langmuir-Hinshelwood mechanism; Panel (d): Eley-Rideal (or 'stick-and-hit') mechanism; Panel (e): hot-atom mechanism. Figure adapted from Herbst & van Dishoeck (2009).

& Leung 1992; Garrod & Pauly 2011). On the outer grain surface, and also through the same layers of ice, chemical species can move and then react via grain-surface reactions to form COMs, which can be released into the gas phase when the temperature of the environment increases (when $T_{\rm d} \gtrsim$ 100-300 K as in the warm regions surrounding YSOs). Fig. 2.9 summarises the three major mechanisms for grain-surface reactions:

1. **Langmuir-Hinshelwood (or diffusive) mechanism:** Accretion of molecules occurs onto binding sites[6] (Fig. 2.9a). In the Langmuir-Hinshelwood mechanism, diffusion occurs by tunneling or thermal hopping mechanisms over/through the barrier between two adjacent binding sites (Fig. 2.9b). Once two reactants (A and B) lie on the same binding site, they can react to form a new species (Fig. 2.9c).

2. **Eley-Rideal mechanism:** A gas-phase molecule accretes, landing atop a (stationary) reactant in a binding site (Fig. 2.9d; e.g. Ruaud et al. 2015).

3. **Hot-atom mechanism:** The non-thermalised species landed on the grains have enough energy to travel some distance over the surface before finding an other reactant. Hot-atom mechanism is often considered less important then Langmuir-Hinshelwood and Eley-Rideal (Fig. 2.9e; e.g. Cuppen et al. 2017).

Among these three mechanisms, the Langmuir-Hinshelwood mechanism is the best studied, and it is often used to treat surface reactions in astrochemical models (Kolasinski 2012). Typically, the products of surface chemistry via the diffusion mechanism, are saturated molecules, because atomic hydrogen accretes from the gas-phase and then diffuses very rapidly. Examples of some chemical species that are thought to be formed in

[6] which are the points of the surface of the grains where chemisorption or physisorption occur.

this way are H_2O, H_2CO, CH_3O and CH_3OH (e.g. Charnley, Tielens & Rodgers 1997; Ioppolo et al. 2008; Herbst & van Dishoeck 2009; see also the discussion of Chapter 5).

The paradigm of the grain-surface chemistry has remained the main theory for the formation of COMs for almost a decade, until recent observations have revealed the presence of gas-phase COMs also in cold environments, where the desorptions are not efficient (e.g. Öberg et al. 2010, Bacmann et al. 2012, Vastel et al. 2014, Cernicharo et al. 2012 and Jiménez-Serra et al. 2016). These discoveries opened new scenarios to explain the formation route of COMs involving gas-phase formation routes (e.g. Vasyunin & Herbst 2013, Balucani, Ceccarelli & Taquet 2015, Skouteris et al. 2019 and Vazart et al. 2020).

Approaching the YSO, the H_2 volume density reaches a few $\times 10^4$ cm^{-3}, and the temperature rises to about 30 K (Figure 2.8b). More and more species are trapped on the icy mantles of the grains, and react with one another to form increasingly complex molecules. The most volatile species such as CH_4, CO and N_2 start to sublimate due to the slightly raised temperatures, inducing also changes in the gas-phase chemistry where now more species are free to interact.

It is closer to the YSO (Figure 2.8c) that the star formation process has its more profound effect on the chemistry of the surrounding medium (e.g. Kurtz et al. 2000; van der Tak 2005; Beuther et al. 2007; Jorgensen, Belloche & Garrod 2020). Here, an accretion disk can form around the YSO, and the gas and dust that surround newborn stars experience radiative heating and shock waves. Grains are heated up to ~100-300 K, resulting in the sublimation of the ices. Around the YSOs, hot molecular cores are developed, i.e. compact regions of ~0.1 pc (Cesaroni 2005), where the evaporation and the sputtering of icy grain mantles allow the return to the gas-phase the *inheritance* of grain-surface chemistry. The conditions, and therefore the time and the location, at which the various chemical species thermally desorb are not always the same. The desorption rate coefficient depends exponentially on the ratio between the binding energy, that defines how strongly an *i*th species is chemically bound to the grain's surface, and the temperature of the grains, i.e. $k_i^{des,th} \propto \exp(-T_b^i/T_d)$, where T_b^i is the binding energy of the *i*th species (see also Chapter 5). When the conditions that favour the evaporation of a given species are reached, this defines the evaporation front of that species (Choudhury et al. 2015). This results in an onion-shell internal structure for clumps, through which the environmental chemical richness changes and increases near the YSO. Hot cores are thus particularly rich in COMs and are characterised by enhanced deuterium fractions (Caselli & Ceccarelli 2012 and Ceccarelli et al. 2014).

2.3 Star formation chemical clocks

In the previous sections we have introduced the chemical processes taking place during the star formation process and described the variety of reaction paths leading to the for-

mation of new chemical species and COMs.

Chemical tracers that show empirical relations between their abundances and the different phases of the star formation process are commonly considered chemical clocks. Some molecules become abundant (or disappear) after a certain evolutionary phase, while others tracers may show a constant increase (or decrease) throughout the evolutionary sequence. This is the confirmation that the star formation process and the chemical state of the parent clumps are deeply connected, and that the latter is the *echo* of a process that we cannot directly observe. Things become more complicated when we try to connect the different tracers into a wider evolutionary context, where each molecule is linked to the others by a huge network of chemical reactions. Astrochemical models help us to explain how the distribution and the abundance of different molecular tracers can change throughout the evolutionary sequence, and can provide a constraint on the timescales of the star formation process. Several authors have searched for a connection between the abundances and the distribution of various tracers with the evolutionary stages of massive star-forming regions, reporting different, and sometimes contradictory results (see e.g. Doty et al., 2002; Miettinen et al., 2006; Beuther et al., 2009; Fontani et al., 2011; Hoq et al., 2013; Molinari et al., 2016).

Relevant for this Thesis work are the recent results reported by Giannetti et al. (2017b), who employed three chemical clocks: methyl acetylene (CH_3CCH), acetonitrile (CH_3CN) and methanol (CH_3OH), to derive and compare the physical properties of TOP100 sources along the evolutionary sequence. By simultaneously fitting the large number of lines that these molecules have in the observed band, Giannetti et al. (2017b) noted that CH_3CN and CH_3OH are particularly sensitive to the thermal state of the clumps, and the most evolved sources require two separate thermal components to reproduce all observed lines. The first component is cold and traces the gas at temperatures < 100 K, associated to the outermost part of the clump. The second component is hot at $T_g > 100$ K, and follows the thermal evolution of the clumps closer to the YSOs. The molecular abundances and the temperatures obtained in this way are found to increase with the evolution of the clumps.

To compare the temperatures traced by each chemical species, and by both the thermal components found for CH_3CN and CH_3OH, with those expected by the warm-up induced by a newborn star, Giannetti et al. (2017b) constructed a simple toy-model. They considered a spherical clump with power-law density and temperature profiles. The results of the comparison are shown in Fig. 2.10. At a given L/M, assumed in their model, the line-of-sight and beam-averaged temperature obtained by the model (the black lines in Fig. 2.10) are found to be consistent with those traced by all the species. Only the cold component of methanol exhibits a different behaviour, underestimating the temperatures provided by the model, always showing an increasing trend with L/M. This result suggests that CH_3OH is less sensitive to the warm-up process with respect to acetonitrile and methyl acetylene.

Formaldehyde (H_2CO), was also seen to correlate with the thermal properties of the

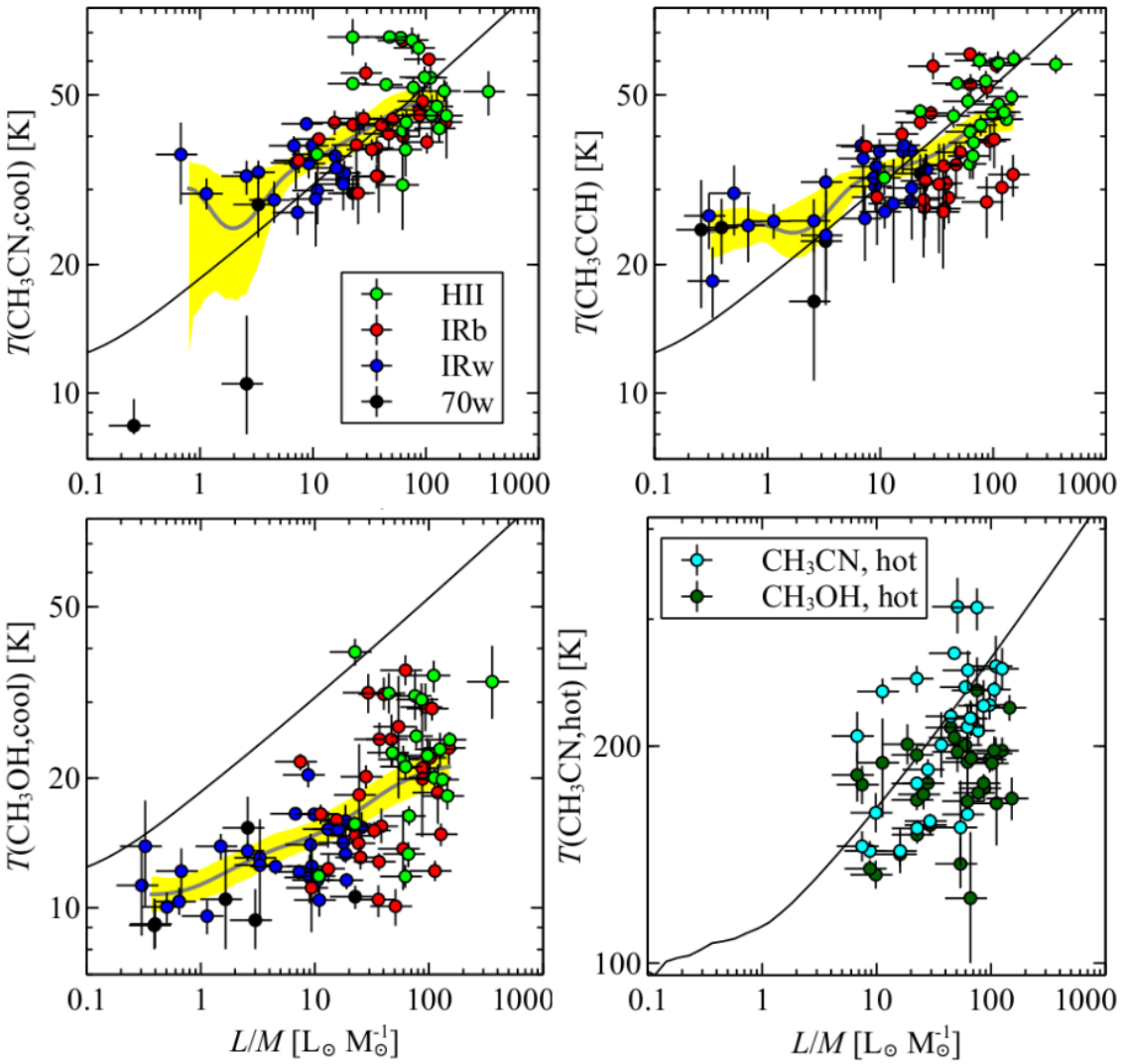

Figure 2.10: L/M ratio vs. the temperature traced by the different chemical species analysed in Giannetti et al. (2017b). The solid black line indicates the line-of-sight- and beam-averaged temperature for a spherical clump model passively heated by a central object accounting for the total luminosity. The thick grey line shows the non-parametric regression of the data, with uncertainties indicated by the yellow-shaded area. This plot is adapted from Fig. 10 of Giannetti et al. (2017b).

TOP100 massive clumps by Tang et al. (2018). Ten transitions of ortho- and para-H_2CO (J = 3-2 and 4-3) were observed between ~210 and ~290 GHz, with the aim to determine the kinetic temperatures and relative abundances of H_2CO in the TOP100 sources as well as to prove its reliability as chemical clock of the high-mass star formation process. The average abundance of para-H_2CO is seen increasing by a factor of 3 from the less to the most evolved sources, while the average gas kinetic temperatures traced by formaldehyde increase with time through the evolution of the clumps (see Fig. 2.11).

Other tracers have been suggested to be good chemical clocks of the high-mass star formation process. Taniguchi et al. (2018) reported that the fractional abundance of HC_3N

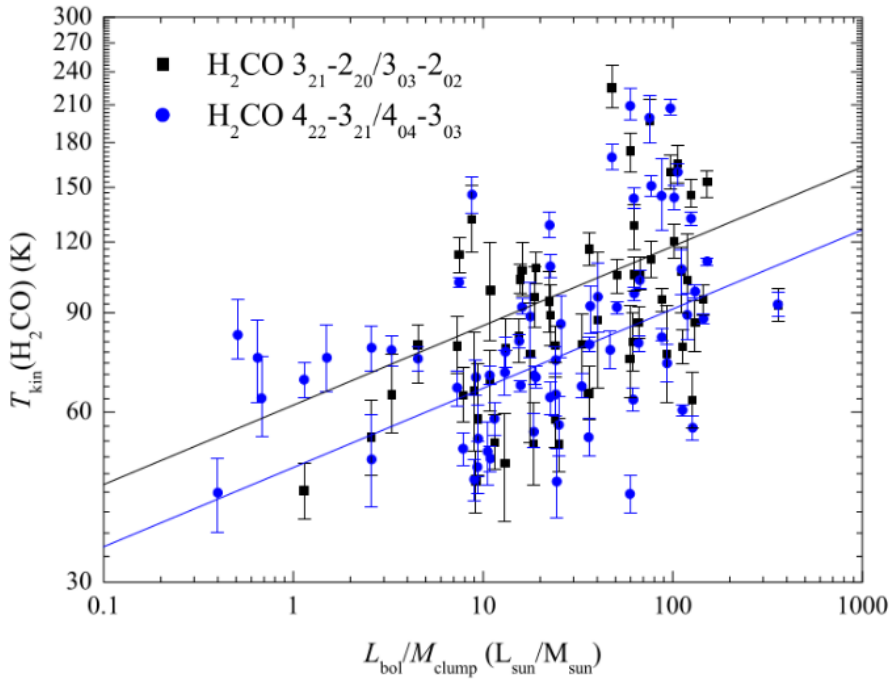

Figure 2.11: The same as Fig. 2.10, derived from p-H_2CO in Tang et al. (2018). Different colours and symbols correspond to different molecular transitions used to derive the temperatures, as indicated in the figure. The straight lines are the results from unweighed linear fits to the two ratios for clumps with $L/M > 10$ $L_\odot$ $M_\odot^{-1}$. This plot is taken from Fig. 12 of Tang et al. (2018).

decreases, on average, from $(6.6\pm0.8)\times10^{-11}$ computed in a sample of 17 prestellar cores, to $(3.6\pm0.5)\times10^{-11}$ in 35 massive protostellar objects. The same decreasing trend is observed between the HC_3N column densities and L/M, as indicator of the evolutionary state of the clumps. The authors explained this behaviour suggesting that HC_3N might be particularly sensitive to the stellar activity, in agreement with the recent results of Yu, Wang & Xu (2019). Subsequently, several emission lines of HC_3N, N_2H^+, CCS and *cyclic*-C_3H_2 were observed between 81 and 94 GHz in a reduced sample of 17 prestellar cores and 28 protostellar objects by Taniguchi et al. (2019). The authors found that the column density ratio $N(N_2H^+)/N(HC_3N)$ decreases from prestellar to protostellar objects, suggesting it as potential chemical indicator for the evolution of massive cores.

Similarly, Urquhart et al. (2019), focused on specific ratios in a sample of 570 high-mass star-forming clumps, selected from the ATLASGAL survey, and observed with Mopra between 85 and 94 GHz. The authors identified thirteen line intensity ratios as potential chemical clocks for the high-mass regime. Among them, six line ratios (i.e. $^{13}CS/N_2H^+$, $^{13}CS/H^{13}CO$, $H^{13}CN/N_2H^+$, $H^{13}CN/HN^{13}C$, C_2H/N_2H^+ and HC_3N/N_2H^+) were also found sensitive to the thermal-state of the clumps traced with the dust temperature T_d, confirming their high usability as evolutionary tracer.

As discussed in Sect. 2.2.2, the ratio between *o*-H_2D^+ and N_2D^+ was also proposed as a reliable chemical clock to follow the high-mass star formation process by Giannetti

Ratio	Pre-stellar	Proto-stellar	HII	Trends	References
N_2D^+/N_2H^+	0.26	0.04	0.04	↘	Fontani et al. (2011)
N_2H^+/HCO^+	0.07	0.08	0.11	↗	Hoq et al. (2013)
N_2H^+/HC_3N	20.40	3.70	–	↘	Taniguchi et al. (2019)
	20.00	5.26	2.94	↘	Urquhart et al. (2019)
o-H_2D^+ /N_2D^+	13.5	12.0	∼1	↘	Giannetti et al. (2019)
$^{13}CS/N_2H^+$	0.02	0.08	0.16	↗	Urquhart et al. (2019)
$^{13}CS/H^{13}CO$	0.35	0.40	0.60	↗	Urquhart et al. (2019)
$H^{13}CN/HN^{13}C$	1.70	2.16	3.20	↗	Urquhart et al. (2019)
C_2H/N_2H^+	0.30	0.71	1.15	↗	Urquhart et al. (2019)

Table 2.1: Summary of some of the observed ratios between column densities of different chemical tracers in massive star forming regions (column 1). The ratios are divided by evolutionary class (i.e. pre-/proto-stellar and HII; columns 2, 3 and 4, respectively). Column 5 indicates the observed trend with time (i.e. ↗: increasing; ↘: decreasing and →: constant). The symbols '–' means 'not observed'. The references are given in column 6.

et al. (2019). Both these tracers show a trend in their abundances with the evolution of the clumps. Even more clear appears the trend shown by the ratio o-H_2D^+ /N_2D^+ that varies by more than a factor of 10 from the less to the most evolved clump.

Finally, Coletta et al. (2020) focused their studies on whether COMs can be used as chemical clocks. They observed methyl formate (CH_3OCHO), dimethyl-ether (CH_3OCH_3), formamide (NH_2CHO) and ethyl-cyanide (C_2H_5CN) towards a sample of 39 high-mass star-forming regions in different evolutionary stages. They found that the abundances of each tracer seem to be strongly connected to the evolutionary stage of their parent clump, increasing by at least ∼2 orders of magnitude over ∼4 orders of magnitude in L/M. Those results lead the authors to propose a general scenario for all COMs, where they are expected to be formed in the cold, earliest phases of star formation and desorbed into the gas-phase with the progressive thermal/shock-induced heating of the evolving cores.

These are just a few recent examples of works attempting to connect the chemical evolution of an astrophysical environment, such as massive clumps, with the physical process that takes place within them, the formation of high-mass stars. Tables 2.1 and 2.2, respectively, summarise some of the observed trends of selected ratios and specific abundances of different chemical tracers. It is evident that the role of the star formation process in driving the chemical evolution of the parent clump is important. The number of chemical features detected in the spectra of massive clumps is increasing as a function of their evolutionary stage (e.g. Tan et al. 2014; Urquhart et al. 2019; Coletta et al. 2020). In particular, HMCs show the widest diversity of observed COMs, due to the large number of chemical species that are released into the gas-phase during warm-up. As the luminosity

of a newborn star harboured in a massive clump increases, gas-phase reactions involving activation barriers become also more efficient (see Eq. 2.1), leading to an increasing variety and number of chemical features. It is also clear that it is necessary to investigate this connection from a broader perspective, to combine the different pieces of the puzzle in a single framework. The large number of tracers that can be used for this purpose and the different samples studied up to now, complicates finding a solution to this problem. For this reason, a systematic study of these tracers in a single sample of objects, if considered representative of a larger population, seems to be a reasonable starting point to select the molecules that can be used as reliable chemical clocks.

Tracer	Transition (J)	Temperature [K]	Abundances			Trend	References
			Pre-stellar	Proto-stellar	HII		
SiO	3-2; 2-1	~5 (T_{ex})	–	⊗	$>1\times10^{-10}$	↗	Miettinen et al. (2006)
HCO^+	1-0	14-26 (T_d)	4.0×10^{-9}	6.4×10^{-9}	7.5×10^{-9}	↗	Hoq et al. (2013)
H_2D^+	$1_{1,0}$-$1_{1,1}$	13-20 (T_d)	3.2×10^{-11}	2.5×10^{-11}	2.0×10^{-12}	↘	Giannetti et al. (2019)
N_2H^+	1-0	14-26 (T_d)	3.0×10^{-10}	5.0×10^{-10}	7.7×10^{-10}	↗	Hoq et al. (2013)
	1-0	3-10 (T_{ex})	2.1×10^{-10}	1.2×10^{-10}	–	↘	Taniguchi et al. (2019)
N_2D^+	3-2	13-20 (T_d)	1.2×10^{-12}	1.4×10^{-12}	1.9×10^{-12}	↗	Giannetti et al. (2019)
H_2CO	several	52-110 (T_{kin})	1.0×10^{-9}	1.3×10^{-9}	1.1×10^{-9}	→	Tang et al. (2018)
HC_3N	37-36	150-340 (T_{ex})	–	⊗	$>7\times10^{-11}$	↗	Beuther et al. (2009)
	10-9; 9-8	6-51 (T_{ex})	1.0×10^{-11}	3.1×10^{-11}	–	↗	Taniguchi et al. (2019)
	10-9	19-32 (T_d)	–	–	$>10^{-10}$	↗	Yu, Wang & Xu (2019)
	5-4	13-25 (T_{kin})	6.6×10^{-11}	3.6×10^{-11}	–	↘	Taniguchi et al. (2018)
CH_3CN	19-18; 6-5; 5-4	20-51 (T_{ex})	5.2×10^{-10}	3.3×10^{-10}	2.1×10^{-10}	↘	Giannetti et al. (2017b)
CH_3OH	7-6	10-22 (T_{ex})	5.7×10^{-8}	3.6×10^{-8}	1.7×10^{-8}	↘	Giannetti et al. (2017b)
CH_3CCH	6-5; 5-4	29-39 (T_{ex})	–	⊗	$>3\times10^{-9}$	↗	Miettinen et al. (2006)
	20-19; 6-5; 5-4	24-46 (T_{ex})	7.2×10^{-8}	8.9×10^{-9}	5.9×10^{-9}	↘	Giannetti et al. (2017b)
CH_2CHCN	38-37; 37-36; 36-35	150-340 (T_{ex})	–	⊗	6.0×10^{-11}	↗	Beuther et al. (2009)
CH_3OCHO	several	21-184 (T_{ex})	⊗	7.0×10^{-10}	2.3×10^{-7}	↗	Coletta et al. (2020)
CH_3OCH_3	several	37-168 (T_{ex})	⊗	9.4×10^{-10}	1.5×10^{-8}	↗	Coletta et al. (2020)
C_2H_5CN	several	29-193 (T_{ex})	⊗	⊗	7.5×10^{-10}	↗	Coletta et al. (2020)

Table 2.2: Summary of some of the relative abundances (with respect to H_2) observed for different tracers (column 1) and samples. Each transition is in column 2. Column 3 provides the temperature assumed to derive the column densities. The abundances are in columns 4 to 6, for different evolutionary stages (as in Tab. 2.1). The symbols '–' and '⊗' mean 'not observed' and 'observed but not detected', respectively. Column 5 indicates the observed trend (as in Tab. 2.1). Column 6 collects the references. We note that the abundances have been derived from observations with different angular resolutions. Therefore caution should be applied when directly comparing the reported values.

2.4 Aims of this thesis

In spite of the efforts made in recent years to understand the formation of massive stars (see Chapter 1), there are still some fundamental questions that remain unanswered. One of these is related to the timescales of the different evolutionary phases of the massive star formation process, which represents the main aim of this Thesis. Obtaining an estimate of these timescales would provide crucial information in distinguishing among competing theories (Sect. 1.3.1), and in reaching a comprehensive view of the chemical evolution of gas that resides in the parental clumps of massive stars.

One way to determine the timescales of the various phases in the high-mass star formation process is to look at the effects that massive YSOs have on their surroundings. The increasing/decreasing trends in the relative abundances of specific chemical tracers (as well as in their ratios), with the different phases of the star formation process, suggest that the chemistry can be used as a powerful tool to infer how long each evolutionary phase lasts. Therefore, the detection of chemical clocks in large samples of massive clumps at different evolutionary stages might also help to quantify the timescales of each stage.

The main goal of this Thesis is to **derive the duration of the different evolutionary stages of the star formation process** by comparing the abundances obtained from the observed TOP100 sample with those obtained by time-dependent chemical models, which simultaneously describe the physical (e.g. temperature, density, extinction, etc.) and chemical evolution of massive clumps. By employing as tracers formaldehyde (H_2CO), methyl acetylene (CH_3CCH), acetonitrile (CH_3CN) and methanol (CH_3OH), **we aim to establish a timeline for the high-mass star formation process from a statistically relevant sample**.

The aforementioned tracers normally show a monotonic increase in abundance during the star formation process. To study in detail the earliest stages, which are also the most challenging to observe, we need to explore tracers that are abundant where the gas is cold and dense. Deuterated molecules are particularly suitable for this purpose, since the deuteration efficiently occurs in cold and dense environments, enhanced by the depletion of several C-, N-, and O-bearing species (in particular CO) onto the surface of dust grains (Sections 2.2.1 and 2.2.2). We focus on *o*-H_2D^+, one of the key molecules driving deuteration, and on the processes which most affect it. One of the most relevant is CO depletion. Determining the physical scale on which the CO-depletion process becomes efficient is one of the first steps in understanding deuterium fractionation, and its role as chemical clock of the high-mass star formation process, and represents another important aspect of the work described here.

This Thesis is structured to follow the star formation process from the early to the late evolutionary stages:

- In Chapter 3 we explore the hypothesis that the size of the so-called depletion radius might be an additional tool to study the evolution of massive clumps with ongoing star formation, as pointed out in Sect. 2.2.1. To do that, we map the depletion factor over the entire filamentary structure of the nearest and most massive filament of the ATLASGAL-survey, the IRDC G351.77-0.51, covering a broad range of densities and temperatures. We model the volume density distribution of the H_2 and we estimate the size of the depletion radius by assuming different analytical functions to describe the abundance profile of $C^{18}O$ with respect to H_2. Finally we explore how much the estimates of the depletion radius are affected by the models.

- Chapter 4 is focused on the early evolutionary stages of the massive star formation process, to probe the role of o-H_2D^+ as chemical clock of these regions, as discussed in Sect. 2.2.2. We have analysed a sample of 106 high-mass star-forming regions where the ground state transition of o-H_2D^+ was observed. We find 16 new detections and obtain o-H_2D^+ detection limits for the rest of the sample. The goal of the study is to test the trend reported by Giannetti et al. (2019) for o-H_2D^+, on a sample that is statistically representative of regions with ongoing massive star formation activity, in order to exclude possible bias induced by particular initial chemical conditions that may have affected previous (much smaller) samples. Several correlations are found between the o-H_2D^+ abundance and the other physical quantities (e.g. T_d, f_D, L/M, and many others) and we use the survey to provide new estimates for the cosmic-ray ionization rate (CRIR) for hydrogen molecules, ζ_2, based on the recent results of Bovino et al. (2020).

- In Chapter 5, we present our method used to derive accurate evolutionary timescales for the individual phases and for the high-mass star formation process as a whole. We compare the column densities of a carefully selected set of molecules derived from unbiased spectral line surveys, with those predicted by a comprehensive set of general chemical models of massive clumps. With this pilot study we demonstrate how the information provided by a small set of reliable chemical clocks observed in the circum(proto-)stellar material of a representative sample of high-mass star forming regions, allows us to confidentially estimate the timescales on which the star formation process takes place. The guidelines provided in this chapter can be strengthened by increasing the number of tracers and can be applied to different astrophysical contexts, such as low-mass stars or planet formation, to put additional constraints on their timescales.

Chapter 3

The early stages of the high-mass star formation process (I)

This chapter is based on the published work of Sabatini et al. (2019).

Abstract of the Chapter

An estimate of the degree of CO*-depletion (f_D) provides information on the physical conditions occurring in the innermost and densest regions of molecular clouds. A key parameter in these studies is the size of the depletion radius, i.e. the radius within which the C-bearing species, and in particular* CO*, are largely frozen onto dust grains. A strong depletion state (i.e. $f_D > 10$) is highly favoured in the innermost regions of dark clouds, where the temperature is* < 20 K *and the number density of molecular hydrogen exceeds a few* $\times 10^4$ cm^{-3}*. In this chapter, we estimate the size of the depleted region by studying the Infrared Dark Cloud* (IRDC) G351.77-0.51*. Continuum observations performed with the Herschel Space Observatory and the LArge* APEX *BOlometer CAmera, together with* APEX $C^{18}O$ *and* $C^{17}O$ $J = 2 - 1$ *line observations, allowed us to recover the large-scale beam- and line-of-sight-averaged depletion map of the cloud. We built a simple model to investigate the depletion in the inner regions of the clumps in the filament and the filament itself. The model suggests that the depletion radius ranges from* 0.02 *to* 0.15 pc*, comparable with the typical filament width* (*i.e.*$\sim$0.1 pc)*. At these radii, the number density of* H_2 *reaches values between* 0.2 *and* 5.5×10^5 cm^{-3}*. These results provide information on the approximate spatial scales on which different chemical processes, such as the deuterium fractionation, operate in high-mass star forming regions and also suggest caution when using* CO *for kinematical studies in* IRDC*s.*

3.1 The IRDC G351.77-0.51

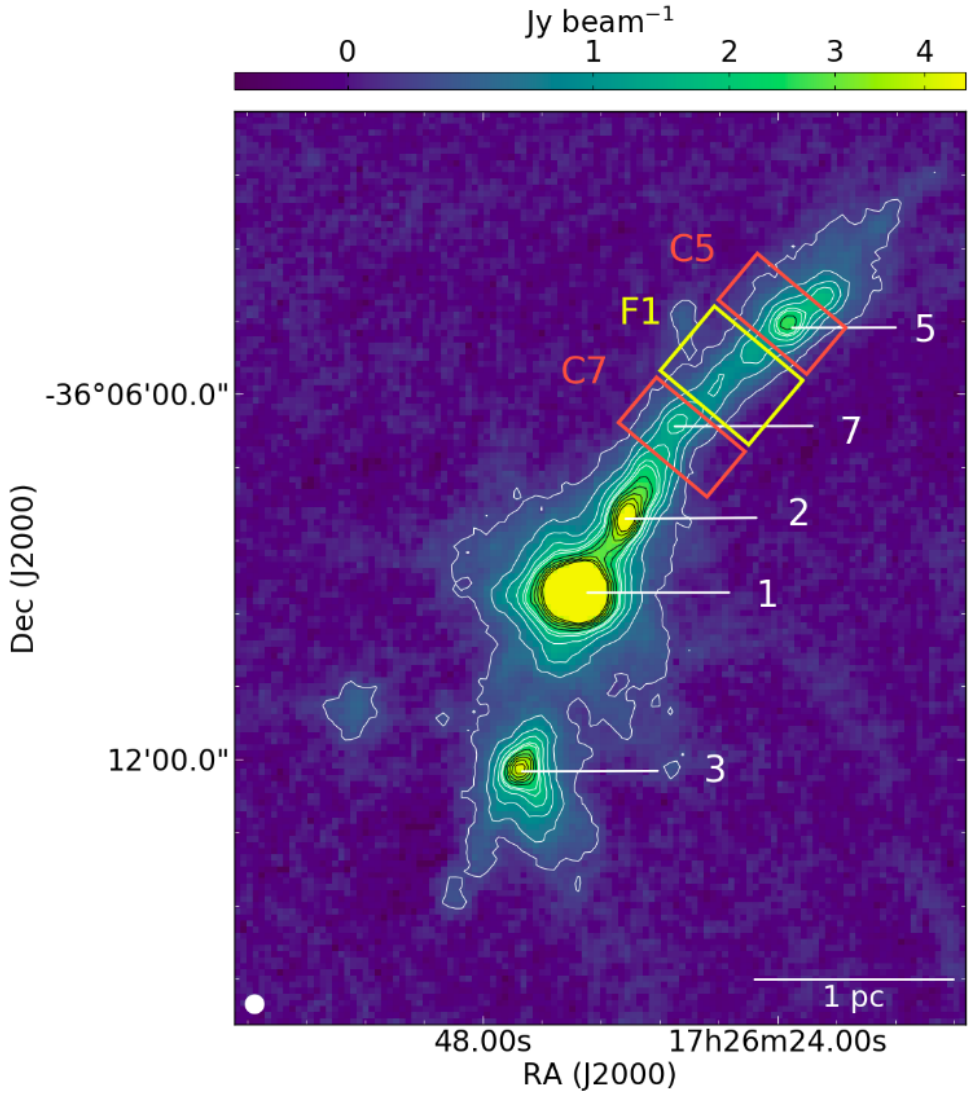

Figure 3.1: *LArge APEX BOlometer CAmera* (LABOCA) map of the 870μm dust continuum emission from the IRDC G351.77-0.51. The white contours are from 5σ (0.2 Jy beam^{-1}) to 2 Jy beam^{-1} while the black contours are from 2.4 Jy beam^{-1} to 4 Jy beam^{-1} in steps of 10σ. The white labels indicate some of the dust clumps identified by Leurini et al. (2011b). Orange and yellow boxes (i.e., regions C5, C7 and F1) are linked to the models discussed in Sect. 3.4.

In Fig. 3.1 we show a continuum image at 870 μm of G351.77-0.51 from the ATLASGAL survey (Schuller et al. 2009; see also Sect. 1.5), in which the clumps (labels 1, 2, 3, 5 and 7) appear well-pronounced along the filamentary structure of the star-forming region. Among the twelve original clumps defined in Leurini et al. (2011b), we selected only the five clumps for which data of both $C^{17}O$ and $C^{18}O$ are available (see Sect. 3.2.2).

G351.77-0.51 is the most massive filament in the ATLASGAL survey within 1 kpc

from us. Leurini et al. 2019 estimate M~ 2600 $M_\odot$ (twice that listed by Li et al. 2016 who used different values for dust temperature and opacity). Following the evolutionary sequence of massive clumps defined by König et al. (2017) - originally outlined by Giannetti et al. (2014); Csengeri et al. (2016); see also Sect. 1.5.1 - the main clump of G351.77-0.51 was classified as an infrared-bright (IRb; Sect. 1.5.1) source, and revealed hot-cores features (e.g. Giannetti et al. 2017b).

Leurini et al. (2011a, 2019) studied the velocity field of the molecular gas component. They discussed the velocity gradients by considering whether they might be due to rotation, or outflow(s) around clump-1, or indicative of multiple velocity components detected in several $C^{18}O$ spectra.

In this chapter we assume the same nomenclature as Leurini et al. (2019) to identify the structures that compose the complex network of filaments of G351.77-0.51: below, we will refer to the central filament as "main body" or "main ridge". It is well-visible in Fig. 3.1 as the elongated structure that harbours the five clumps, identified by white labels. The gas that constitutes the main body is cold and chemically young, in which Giannetti et al. (2019) find high abundances of o-H_2D^+ and N_2D^+ and an age $\lesssim 10^5$ yr for the clump-7 (see Sect 2.2.2). The northern part of the main body appears in absorption against the mid-IR background of our Galaxy up to 70 μm and in emission in the (sub)millimetre range (Faúndez et al. 2004). The analysis of the $J = 2 - 1$ $C^{18}O$ molecular line (see Sect. 3.2.2) confirmed the presence of less dense molecular sub-filaments, linked to the main body: we refer to them as "branches", following the naming convention of Schisano et al. (2014).

3.1.1 $C^{18}O$ map

The $C^{18}O$ $J = 2 - 1$ observations were carried out with APEX[1] between 2014 August and November. The observations were centred at 218.5 GHz, with a velocity resolution of 0.1 km s^{-1}. The whole map covers an approximate total area of 234 $arcmin^2$. We refer to Leurini et al. (2019), for a more detailed description of the dataset.

The root mean square (rms) noise was calculated iteratively in each pixel because of the non uniform distribution in the map, induced by the different exposure times. The first estimate of rms noise was obtained from the unmasked spectra, then, any emission higher than 3σ was masked and the rms recomputed. The cycle continues until the difference between the rms noise of two consecutive iterations (i.e. $\sigma_{i+1} - \sigma_i$) is equal to 10^{-4} K. We estimate a typical-final noise of 0.30 K (between 0.15 and 0.45 K) per velocity channel.

[1] Güsten et al. (2006)

3.2 Derived quantities

3.2.1 Temperature and H_2 column density maps

Calculating $C^{18}O$ column densities requires the gas excitation temperature, $T_{ex}^{C^{18}O}$. We derived $T_{ex}^{C^{18}O}$ from the dust temperature map, which can be obtained from the *Herschel* data through pixel-by-pixel SED fits. This allows one to determine also the H_2 column density map from the emission between 160 and 500 μm.

In this work we adopt the dust temperature map presented in Leurini et al. (2019). To estimate the dust temperature of the whole filamentary structure of G351.77-0.51, these authors used the *Herschel*[2] Infrared Galactic Plane Survey (Hi-GAL, Molinari et al. 2010) images at 500, 350 and 250 μm from SPIRE[3] and at 160 μm from PACS[4]. The authors adopted a model of two emission components (see Schisano et al. 2020), splitting the fluxes observed by *Herschel* in each pixel into the filament and background contributions, as discussed by Peretto et al. (2010). Then, they fitted the filament emission pixel-by-pixel with a greybody model deriving the dust temperature and the H_2 column density. They assumed a dust opacity law $\kappa_0(\nu/\nu_0)^\beta$ with $\beta = 2$, $\kappa_0 = 0.1$ cm^2 g^{-1} at $\nu_0 = 1250$ GHz (Hildebrand 1983). This prescription assumes a gas-to-dust ratio equal to 100. A detailed description of the procedure is given in Leurini et al. (2019).

Their final maps are shown in Fig. 3.2 at an angular resolution of 36″, i.e. the coarsest resolution in *Herschel* bands, that is comparable to the resolution of our $C^{18}O$ data (29″). In most of the map, the dust temperature (panel (a) of Fig. 3.2) ranges between 10 and 30 K, while along the main body and in the region to the south of clump-3 in Fig. 3.1 typically $\lesssim$ 15 K are found.

$T_{ex}^{C^{18}O}$ was derived from the dust temperature T_d, by applying the empirical relation defined by Giannetti et al. (2017b), $T_{ex}^{C^{18}O} = 1.46 \times T_d - 12.9$ with an intrinsic scatter of 6.7 K. However, this relation is only valid for dust temperatures between ~10 and ~45 K, while at low T_d it may underestimate $T_{ex}^{C^{18}O}$. Our data are concentrated in this low temperature regime, where $T_d \lesssim 9$ K would be translated into negative $T_{ex}^{C^{18}O}$ following the relation mentioned earlier. Therefore we decided to use the flattest curve allowed by the uncertainties at 2σ of the original fit to limit this issue: $T_{ex}^{C^{18}O} = T_d - 2.6$ and imposing a lower limit of $T_{ex} = 10.8$ K, equal to the separation between the levels of the $J = 2 - 1$ transition of the $C^{18}O$ line.

$T_{ex}^{C^{18}O}$ typically ranges between ~11 and 28 K with peaks of up to 40 K. Most of the pixels along the main ridge show a temperature of 12-13 K. At these temperatures CO is efficiently removed from the gas phase and frozen onto dust grains for densities exceeding few $\times 10^4$ cm^{-3} (see Giannetti et al. 2014). This suggests that the depletion factor in these regions will have the highest values of the entire IRDC (as confirmed by the analysis of

[2] Pilbratt et al. (2010)
[3] Griffin et al. (2010)
[4] Poglitsch et al. (2010)

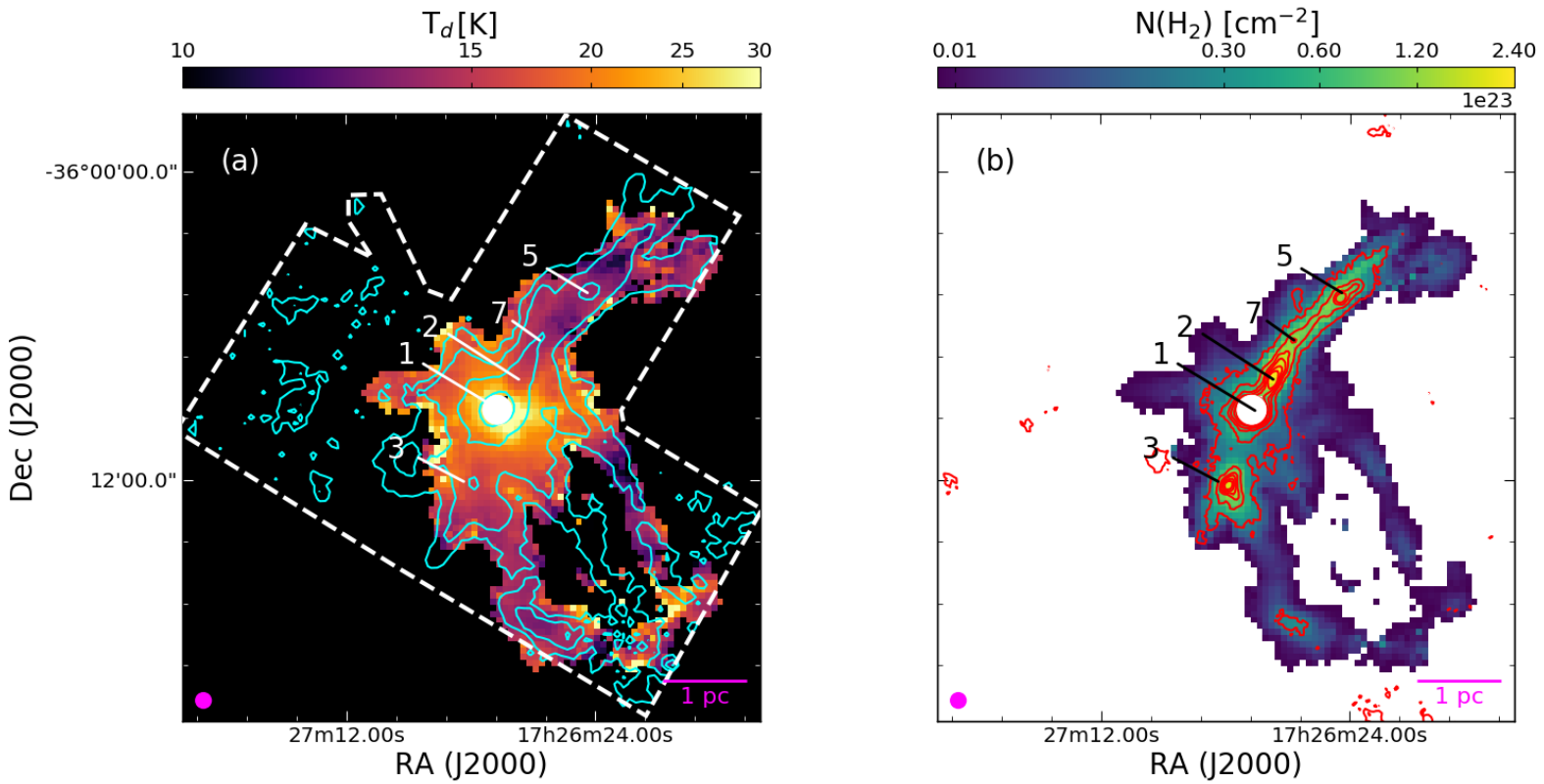

Figure 3.2: (a) Dust temperature map from Leurini et al. (2019), generated by a pixel-by-pixel SED-fitting of the 160-500 μm continuum fluxes of the Hi-Gal Survey. The white dashed lines mark the region observed in $C^{18}O$ with APEX (Sect. 3.1.1). Cyan contours are defined on the integrated intensity map of the APEX $C^{18}O$ $J = 2 - 1$ line in Fig. 3.3 (a) at 3, 9, 27 and 81σ contours, where $3\sigma = 0.9$ K km s^{-1}; (b) H_2 column density map from the pixel-by-pixel SED-fitting from Leurini et al. (2019), scaled to a gas mass-to-dust ratio equal to 120 (see text). Red contours are the same of the LABOCA map of the 870μm dust continuum emission shown in Fig. 3.1. In both panels, we masked the saturated hot-core region in *Herschel* data (white circle).

the depletion map - Sect. 3.3).

We modified the H_2 column density map of Leurini et al. (2019) by using a different value of gas mass-to-dust ratio, γ, equal to 120 as derived by Giannetti et al. (2017a), by assuming a galactocentric distance D_{GL} = 7.4 kpc (Leurini et al. 2019). The resulting H_2 column density map is shown in Fig. 3.2 (*b*). The saturated region in *Herschel* data, corresponding to the hot-core clump, was masked out and excluded from our analysis. In the coldest regions of Fig. 3.2 (a), the molecular hydrogen reaches a column density of 2.4×10^{23} cm^{-2}, decreasing to 4.2×10^{21} cm^{-2} in the regions in which the dust is warmer. Fig. 3.2 (b) shows that the high-column density material (i.e. $\gtrsim 1 \times 10^{23}$ cm^{-2}) is distributed in a single filamentary feature and in clump-like structures, the so called "main-body" in Leurini et al. (2019).

3.2.2 Column densities of $C^{18}O$

Under the assumption of local thermal equilibrium (LTE), we derived the $C^{18}O$ column density, $N(C^{18}O)$, from the integrated line intensity of the (2-1) transition. Following Kramer & Winnewisser (1991), the general formula for $N(C^{18}O)$ is:

$$N(C^{18}O) = \frac{C_\tau}{\eta_c}\frac{3h}{8\pi^3\mu^2}\frac{Q(T_g)}{2}e^{\frac{E_{low}}{k_B T_{ex}}}\left[1-e^{-\frac{h\nu_{J,J-1}}{k_B T_{ex}}}\right]^{-1}[J(T_{ex},\nu_{C^{18}O})-J(T_{bg},\nu_{C^{18}O})]^{-1}\int T_{mb}d\upsilon$$
$$= C_\tau\, f(T_{ex})\int T_{mb}\ d\upsilon, \tag{3.1}$$

where C_τ is the optical depth correction factor defined as $\tau/[1-\exp(-\tau)]$ and the τ is the optical depth of the $J = 2-1$ transition of $C^{18}O$ (see Sect. 3.2.2); h and k_B are the Planck and Boltzmann constants, respectively; η_c is the *filling factor*; μ is the dipole moment of the molecule; $Q(T_g)$ is the partition function; E_{low} is the energy of the lower level of the transition and $\nu_{J,J-1}$ is the frequency of the $J \rightarrow J-1$ transition of the considered molecule (in this case, the $C^{18}O$ $J = 2-1$, equal to ~219.5 GHz); $J(T_{ex},\nu) = (h\nu/k_B)[\exp(h\nu/k_BT)-1]^{-1}$; $T_{bg} \cong 2.7$ K, is the background temperature; T_{mb} is the main beam temperature, and its integral over the velocity range covered by the $C^{18}O$ line is shown in Fig. 3.3 (*a*). In the last line of Eq. (3.1), $f(T_{ex})$ incorporates all the constants and the terms depending on T_{ex}. We further considered possible saturation effects of the continuum *Herschel* maps, in the hot-core region (i.e. clump-1 in Fig. 3.1).

In the following paragraphs, we discuss the steps that allowed us to derive the C_τ map (Fig. 3.3*b*) and the final $C^{18}O$ column density map (Fig. 3.3*c*), necessary to produce the depletion map discuss in Sect. 3.3.

• Opacity correction: We estimate the optical depth of $C^{18}O$ $J = 2-1$ in the clumps by means of the detection equation (see Hofner et al. 2000 for more details). We use the

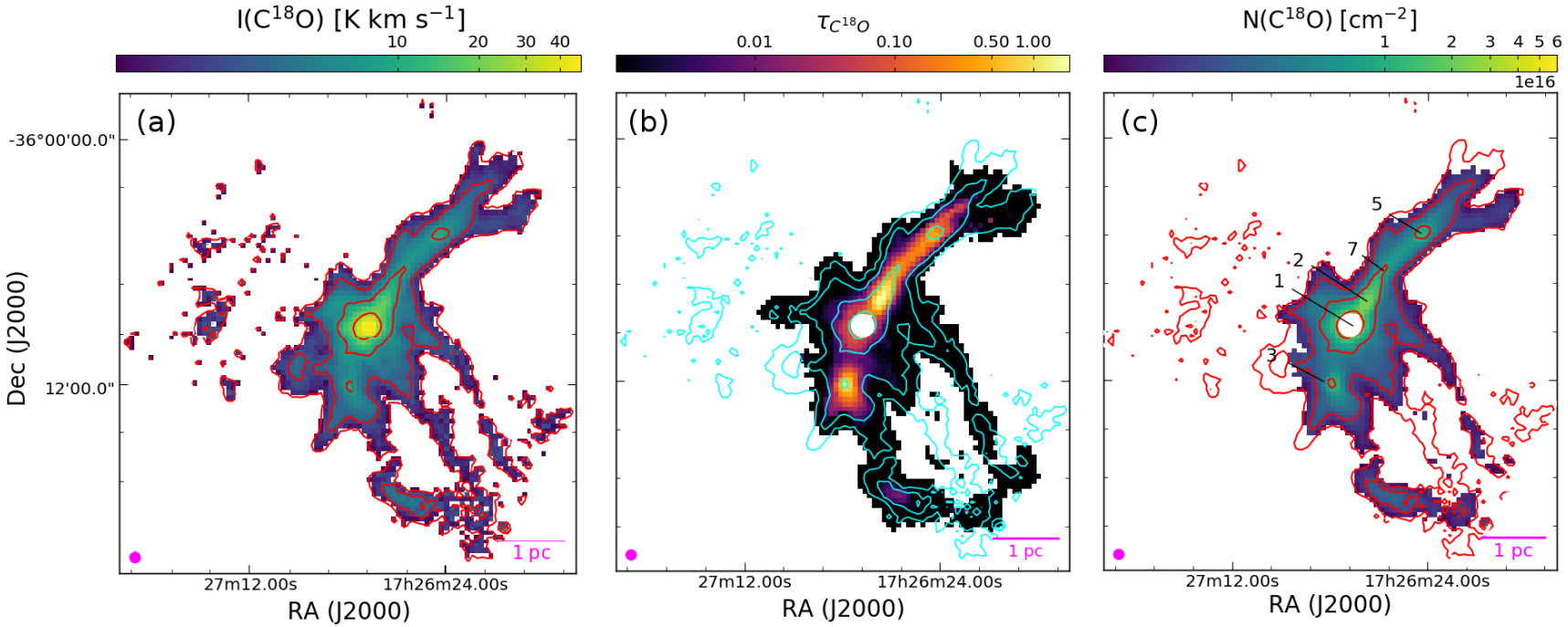

Figure 3.3: (a) Integrated intensity distribution of the $C^{18}O$ $J = 2 - 1$ molecular line transition detected with *APEX* in units of K km s^{-1}. (b) Map of the opacity from the $N(H_2)$ map, following the procedure described in Sect. 3.2.2. The final τ-$N(H_2)$ relation used is: $\log_{10}(\tau_{C^{18}O})$=2.5 $\log_{10}(N(H_2))$-57.4. (c) $C^{18}O$ column density map obtained from the integrated intensity distribution, shown in Fig. 3.3 (a), using Eq. (3.1). The red and cyan contours are the 3, 9, 27 and 81σ contours of the integrated flux density distribution in panel (a), where 3σ = 0.9 K km s^{-1}.

$C^{18}O$ and $C^{17}O$ $J = 2-1$ APEX observations presented in Leurini et al. (2011b), assuming the relative abundance, $\phi_{18/17}$, equal to 4.16 as found by Wouterloot et al. (2008)[5]. Both transitions were observed in seven single-pointing observations, centered at the coordinates of the clumps. We did not consider the data of clumps 9 and 10 (defined in Leurini et al. 2011b), because they are not part of the source, nor the data of the hot-core (i.e. clump-1), due to the saturation of the continuum observations (especially at 250 micron).

The ratio between the peaks of the two CO isotopologues line intensities in each clump is then equal to:

$$R_{C^{18}O,C^{17}O} = \frac{T_{mb,C^{18}O}}{T_{mb,C^{17}O}} = \frac{\eta_{C^{18}O}[J(T_{ex,C^{18}O}, \nu_{C^{18}O}) - J(T_{bg}, \nu_{C^{18}O})](1 - e^{-\tau_{C^{18}O}})}{\eta_{C^{17}O}[J(T_{ex,C^{17}O}, \nu_{C^{17}O}) - J(T_{bg}, \nu_{C^{17}O})](1 - e^{-\tau_{C^{17}O}})}, \quad (3.2)$$

where the considered transition of $C^{18}O$ and $C^{17}O$ is the $J = 2 - 1$, at frequencies $\nu_{C^{18}O} = 219.5$ GHz and $\nu_{C^{17}O} = 224.7$ GHz, respectively; τ is the optical depth of the two considered CO isotopologues, with $\tau_{C^{18}O} = \phi_{18/17}\, \tau_{C^{17}O}$.

In Eq. (3.2) we assumed that the $J = 2 - 1$ transition of the two CO isotopologues correct under the same excitation conditions (i.e. $T_{ex,C^{17}O} = T_{ex,C^{18}O}$; e.g. Martin & Barrett 1978; Ladd 2004), are tracing the same volume of gas. This assumption has the consequence that the *filling factor*, η, is the same for both the transitions and therefore it can be eliminated from Eq. (3.2).

We fitted a linear relation between $\log_{10}(\tau)$ and $\log_{10}[N(H_2)]$[6], and then computed C_τ to correct $N(C^{18}O)$ following the schematic diagram in Fig. 3.4. The best linear fit was computed with a least squares regression by extracting the $C^{17}O$ and $C^{18}O$ peak fluxes 10^4 times (step 2 in Fig. 3.4). The basic assumption - see step 1a - was that the rms (ε) is normally distributed[7]. For each clump, the τ value was calculated following Eq. (3.2) including the contribution of the rms and assuming a 10% calibration uncertainty (σ), summed in quadrature to ε (i.e. $T_{mb,C^{18}O} = T_{mb}^{C^{18}O} + \sqrt{\varepsilon^2 + \sigma^2}$ and the same for $T_{mb,C^{17}O}$). We note that when applying this method to select different values of T_{mb} in the $C^{18}O$ and $C^{17}O$ spectra, the errors do not change significantly if the distributions are built by 10^3 elements or more. Applying this procedure to each clump - steps 1(a-d) - we derived the linear fit $\log_{10}(\tau)$-$\log_{10}[N(H_2)]$. We repeated this procedure 10^4 times and we generated a cube of CO column density maps, one for each solution $\log_{10}(\tau)$-$\log_{10}[N(H_2)]$ found. At the end of this procedure, a distribution of 10^4 values of $N(C^{18}O)$ has been associated to each pixel, used to compute the average value of $N(C^{18}O)$ and its relative error bar in each pixel. We finally imposed an upper limit for the correction at $\tau = 2$. This value would only be achieved in the saturated hot-core region and thus no pixel included in our analy-

[5]Assuming a galactocentric distance of the source equal to D_{GL} =7.4 kpc as in Leurini et al. (2019).

[6]We used a τ-$N(H_2)$ relation and not a τ-$N(C^{18}O)$ one because $N(H_2)$ is not affected by opacity (i.e. κ_ν correction already applied in Sect. 3.2.1) and the first relation has less scatter than the second.

[7]This step was performed using the numpy.random.normal function of NumPy (Oliphant 2006) v1.14.

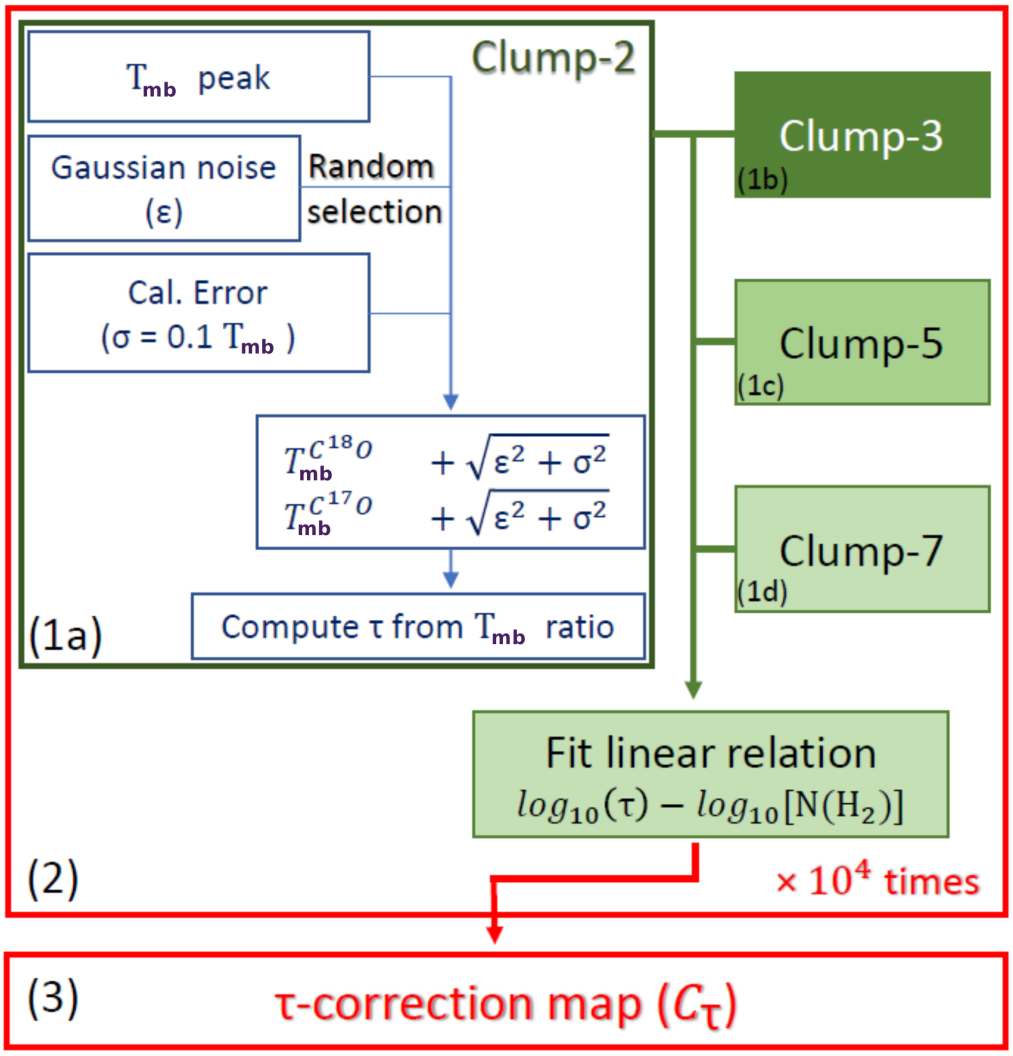

Figure 3.4: Schematic diagram showing how the correction C_τ was calculated starting from the $C^{18}O$ and $C^{17}O$ single pointing observation of clumps 2, 3, 5 and 7 in Leurini et al. (2011b).

sis is affected by this condition. We then generated the opacity map from the $N(H_2)$ map. The results of this procedure are shown in Fig. 3.3 (*b*), together with the $C^{18}O$ integrated intensity map - Fig. 3.3 (*a*). We note that the estimated τ values range between $\sim$ 0.1 and $\sim$ 1.9 along the main body of G351.77-0.51.

The final $C^{18}O$ column density map was derived by including the opacity correction using the bestfit final $\log_{10}(\tau)$-$\log_{10}[N(H_2)]$ relation - i.e. $\log_{10}(\tau_{C^{18}O}) = 2.5\log_{10}[N(H_2)]$-57.4. The opacity-corrected column density map of $C^{18}O$ is shown in Fig. 3.3 (c). Our correction has increased the $C^{18}O$ column densities by up to a factor of 2.3 along the main body, while $N(C^{18}O)$ remains almost constant in the branches. Over the whole structure, the column densities of $C^{18}O$ range between 1 and 6×10^{16} cm^{-2}.

To summarise, the $C^{18}O$ column density map was generated under the assumption of LTE following Eq. (3.1). Possible saturation in the continuum data and opacity effects have been considered. The column density map was then used to evaluate the depletion map, as discussed in the next section.

3.3 The large-scale depletion map in G351.77-0.51

The final CO-depletion factor map - Fig. 3.5 - was generated by taking the ratio between the expected and the observed CO emission, using an abundance of $C^{18}O$ with respect to H_2 (i.e. $\chi^E_{C^{18}O}$) equal to 2.1×10^{-7} (see Giannetti et al. 2017a).

As visible in Fig. 3.5, in almost half of the map the depletion factor is <1.5, highlighting the absence of strong CO depletion around the core and along the branches in the southern directions with respect to the main ridge. This effect can have two different causes for the region of the central clump and for the branches, respectively: in the first case (i.e. the surroundings of clump-1 in Fig. 3.1), the absence of depletion can be linked to the intense star formation activity, demonstrated in previous papers (e.g. Leurini et al. 2011b, König et al. 2017, Giannetti et al. 2017b and Leurini et al. 2019). The increase in temperature induced by the forming stars is able to completely desorb the ice mantles around dust grains in which the CO molecules are locked. This effect lowers the observed depletion factor, until it reaches unity. Following Eq. (2) in Schuller et al. (2009), we computed $N(H_2)$ in the hot-core (saturated) region from the ATLASGAL peak flux at 870 μm, using $\kappa_{870\mu m} = 0.76$ cm^2 g^{-1} consistent with the dust opacity law discussed in Sect. 3.2.1. To obtain a depletion of 1 here, consistent with the surroundings, the dust temperature should be 80 K.

To evaluate the effects of $N(H_2)$ and T_d on f_D, Fig. 3.6 was obtained by the pixel-by-pixel combination of Figures 3.2 (b), 3.2 (a) and 3.5. Each point represents a pixel shared between the three maps, where dots and squares are used to distinguish the pixels of the main body from those of the branches, respectively.

Instead, in the branches (square markers in Fig. 3.6) we notice that depletion reaches values ~3.5. This result suggests that even in these structures the depletion process can start to occur. On the other hand, where f_D is close to 1, the lower density disfavours a high degree of depletion (e.g. see Caselli et al. 1999). Furthermore, we should consider that the observed depletion factor is averaged along the line-of-sight and in the beam.

Along the main ridge and in the surroundings of clump-3 (Fig. 3.1) the depletion factor ranges between 1.5 and 6 and reaches its maximum in clump-3. Both regions appear in absorption at 8 μm in the Leurini et al. (2011b) maps, showing H_2 column densities of a few 10^{22} cm^{-2}. In particular, it may appear counter-intuitive that we observe the highest depletion factor of the whole structure in clump-3, as an HII region has been identified in Wide-field Infrared Survey Explorer (WISE[8]) at a distance of only $\approx 8''$ from the centre

[8] https://irsa.ipac.caltech.edu/applications/wise/

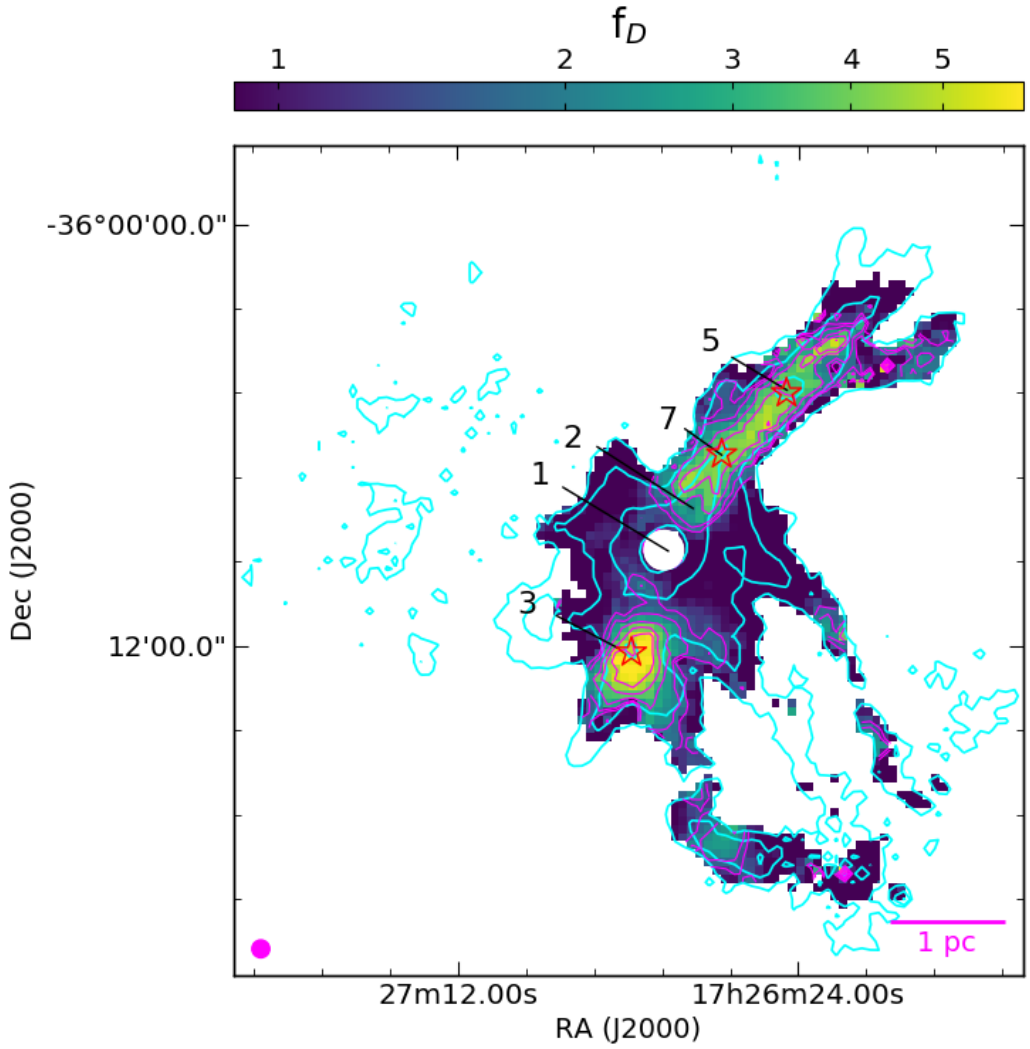

Figure 3.5: Depletion factor (f_D) map obtained by taking the ratio between the expected and observed $N(C^{18}O)$. We assumed a canonical abundance of 2.1×10^{-7} of $N(C^{18}O)$ relative to $N(H_2)$. The cyan contours are the same as in Fig. 3.3, while the ones in magenta are the defined at f_D = 1.5, 2, 3, 4 and 5. Red stars in clump-5 and -7 indicate locations of the clumps reported in Leurini et al. (2019), while the one in clump-3 is set on the coordinates of the candidate HII region reported in Anderson et al. (2014).

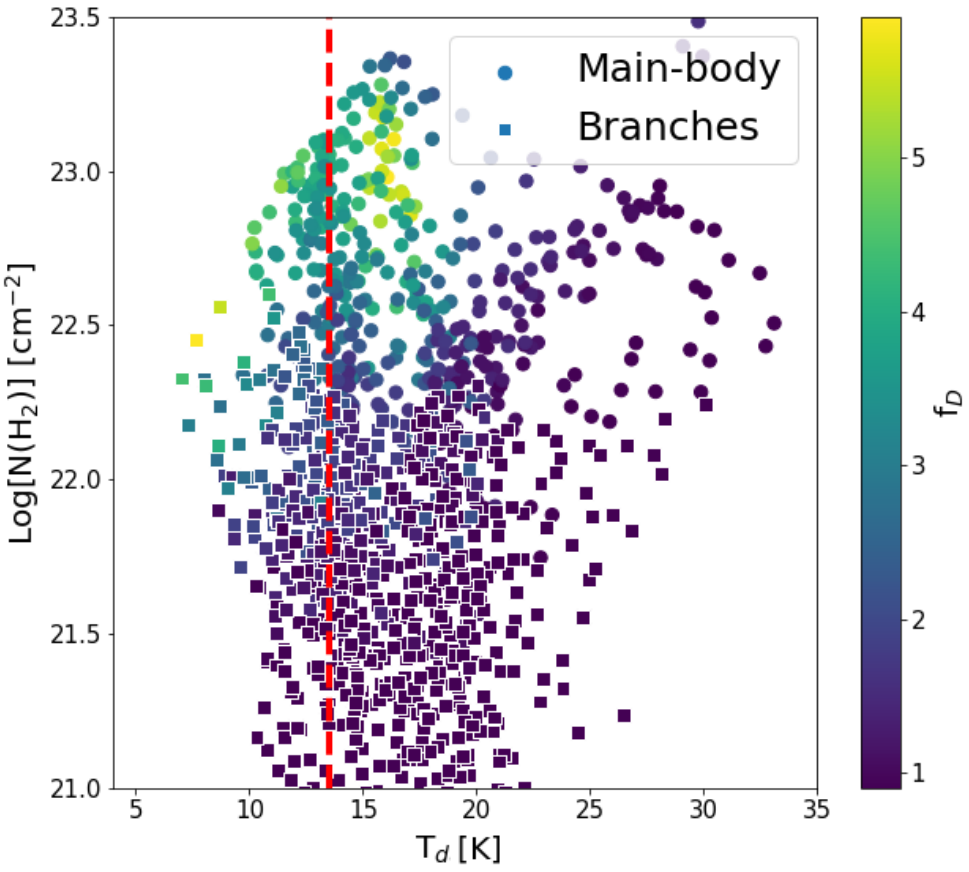

Figure 3.6: Pixel-by-pixel scatter plot of the whole structure detected in Fig. 3.5. Circles represent the pixels of the main body, while the squares indicate those of the branches. The red-dashed line represents the 10.8 K lower limit imposed on the $T_{\rm ex}^{\rm C^{18}O}$ (as discussed in Sect. 3.2.1) that corresponds to $T_{\rm d} \sim 13.4$ K.

of clump-3 (Anderson et al. 2014). For these reasons, clump-3 should show similar depletion conditions to those in clump-1. However, such a high depletion factor suggests dense and cold gas close to the HII region contained in the clump. Within the cloud, the degree of depletion could be maintained if self-shielding would attenuate the effects of the radiation field of the HII region. These ideas are supported by the analysis of the 8 and 24 μm maps shown in Leurini et al. (2019). At the location of the clump-3, a region slightly offset with respect to the bright spot associated with the HII region, is clearly visible in absorption at both wavelengths.

In clumps 5 and 7 (i.e. regions C5 and C7 in Fig. 3.1), along the main ridge of G351.77-0.51, the average depletion factors are $f_{D,\rm C7} = 3.4^{+0.4}_{-0.5}$ (in clump-7; peaking at 4.5) and $f_{D,\rm C5} = 3.1^{+0.5}_{-0.6}$ (in clump-5; peaking at 4.3), respectively (see Sect. 3.4.1 for

more details about error estimation). Compared to the average values of the samples mentioned in Sect. 2.2.1, our values are slightly lower. However, we should consider that the observed depletion factors are affected by many factors such as the opacity correction applied and the $C^{18}O/H_2$ abundance ratio, which can vary up to a factor of 2.5 (e.g. Ripple et al. 2013).

Along the main ridge the depletion factors reach values of ~6. This is comparable to what is found in Hernandez et al. (2011), where the authors studied depletion factor variations along the filamentary structure of IRDC G035.39-00.33 with IRAM 30-m telescope observations, and also with Feng et al. (2020). Along the filament, they estimate a depletion factor of up to 5. Of course, also in this case the considerations made earlier hold, but if we consider the different opacity corrections applied and $C^{18}O/H_2$ relative abundance assumed, we note the difference between our results and those of Hernandez et al. (2011) are not larger than $\Delta f_D \sim 1$.

To summarise, the final depletion map of G351.77-0.51 - Fig. 3.5 - shows widespread CO-depletion in the main body, as well as at various locations in the branches. Comparing our results with those of Hernandez et al. (2011) and Feng et al. (2020), we note that in both cases the phenomenon of depletion affects not only the densest regions of the clumps, but also the filamentary structures that surround them. This result suggests that CO-depletion in high-mass star forming regions affects both small and large scales.

3.4 Depletion modeling

High densities and low gas temperatures favour CO-depletion. Given this, it is reasonable to think that $C^{18}O/H_2$ is not constant within a cloud. This quantity varies as a function of location, following the volume density and gas temperature distributions/profiles.

The regions with a high depletion degree are those where the dust surface chemistry becomes more efficient, due to the high concentration of frozen chemical species on the dust surface. To understand how the efficiency of the various types of chemical reactions change, we need to understand how the depletion degree varies within the dark molecular clouds.

In order to reproduce the average depletion factor observed in G351.77-0.51 we built a simple 1D model describing the distributions of $C^{18}O$ and H_2. We focus our attention on the main ridge, identifying three distinct regions: clumps 5 and 7, and the filamentary region between them (i.e. regions C5, C7 and F1 in Fig. 3.1, respectively).

The model assumes that both profiles have radial symmetry with respect to the centre of the ridge, i.e. the "spine". $R_{\rm flat}$ is the distance relative to the spine within which the density profiles remain roughly flat; i.e. $n(H_2, R_{\rm flat}) = 0.5\, n(H_{2,\rm spine})$, if $p = 2$ (see Eq. 3.3). We normalise the volume density profiles of the model with respect to $R_{\rm flat}$, while $n(H_{2,\rm spine})$ is the central volume density of H_2.

Our H_2 volume density profile is described by:

$$n(H_2) = n(H_{2,spine}) \left[1 + \left(\frac{R}{R_{flat}} \right)^{\alpha} \right]^{-p/2}, \tag{3.3}$$

up to a maximum distance $R_{max} \sim 0.2$ pc, two times larger than the filament width estimated in Leurini et al. (2019), that encompasses the entire filamentary structure.

For the clumps (i.e. regions C5 and C7), we assume $p = 2$ and Eq. (3.3) takes the functional form described by Tafalla et al. (2002). In this case, the free parameters of the model are α, $n(H_{2,spine})$ and R_{flat}. Starting from R_{flat}, both the distributions of $C^{18}O$ and H_2 scale as a power-law with the same index: $\alpha_{C5} = 1.9$ and $\alpha_{C7} = 1.8$, for the clump-5 and clump-7, respectively. The solutions for α was found by exploring the parameter space defined by the results of Beuther et al. (2002a), i.e. $1.1 < \alpha < 2.1$. The best-fitting models returns the values of $n(H_{2,spine})$ and R_{flat} equal to $8{\times}10^6$ cm^{-3} and $5.5{\times}10^{-3}$ pc in the case of clump-5 and $5.8{\times}10^7$ cm^{-3} and $2{\times}10^{-3}$ pc for the clump-7.

In the case of the main body (i.e. region F1), α is equal to 2 and the volume density profile of H_2 is described by a Plummer-like profile (see Plummer 1911 for a detailed discussion). For this model, the free parameters are p and $n(H_{2,spine})$, while R_{flat} corresponds to the thermal Jeans length, λ_J, for an isothermal filament in hydrostatic equilibrium, calculated at $R = 0$:

$$\lambda_J = \frac{c_s^2}{G\,\mu\,m_H\,N(H_2)_{R=0}}, \tag{3.4}$$

where c_s^2 is the isothermal sound speed for the mean temperature in the region F1 (equal to 13.8 K), G is the gravitational constant in units of [cm^3 g^{-1} s^{-2}], μ is the mean molecular weight of interstellar gas ($\mu = 2.34$), m_H is the hydrogen mass in grams and $N(H_2)_{R=0}$ is the average H_2 column density calculated at $R = 0$. Applying Eq. (3.4) to region F1, we obtained $R_{flat} \sim 0.008$ pc, while the value of p index was found equal to 1.9 after exploring the range of values reported in Arzoumanian et al. (2011). We note that an R_{flat} variation of 30% does not significantly change the $N(H_2)$ profile.

To simulate depletion effects, we introduced the depletion radius, R_{dep}, that is the distance from the spine where the degree of depletion drastically changes, following equations:

$$\left.\begin{array}{lll} R \leqslant R_{dep} & f_D = (10, \infty); & \chi_{C^{18}O} = \dfrac{\chi^E_{C^{18}O}}{f_D} \\ R > R_{dep} & f_D = 1; & \chi_{C^{18}O} = \chi^E_{C^{18}O} \end{array}\right\}, \tag{3.5}$$

where the canonical abundance for $C^{18}O$ ($\chi^E_{C^{18}O} = 2.1{\times}10^{-7}$) was set by using the findings of Giannetti et al. (2017a) and references therein. The lower limit of $f_D = 10$ for $R < R_{dep}$, is motivated by the observational constraints for high-mass clumps, which in most cases

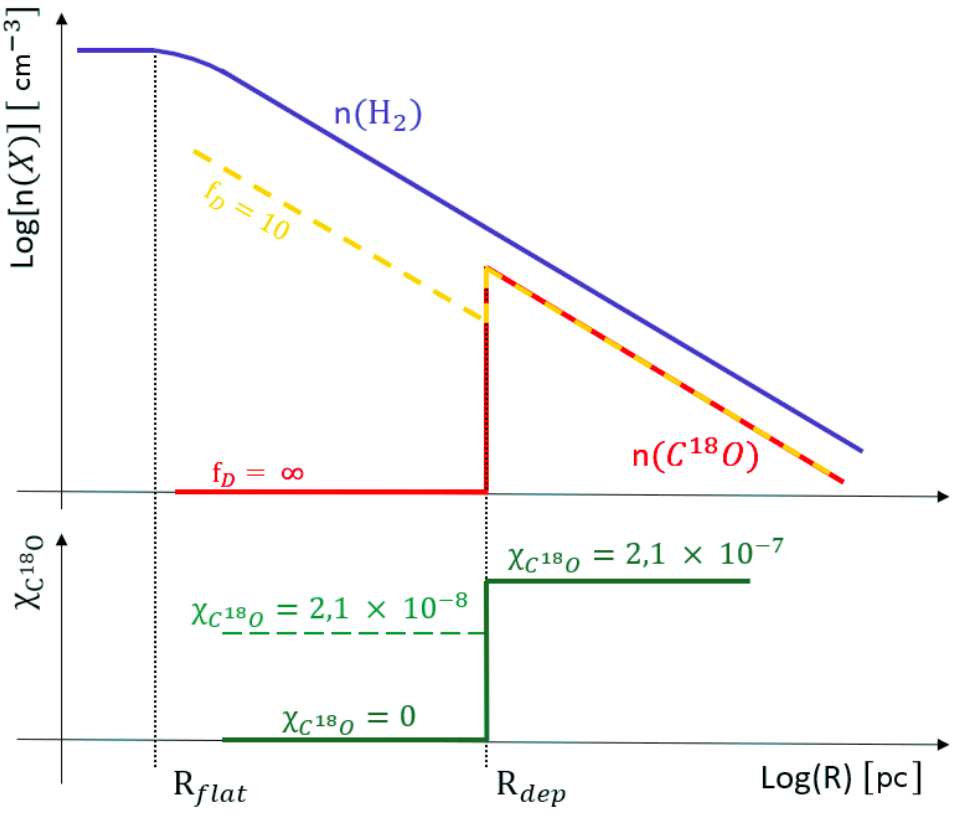

Figure 3.7: Schematic representation of the model applied to simulate the depletion effect inside/outside the depletion radius. In blue we show the number density profile of H_2; in green the step-function reproducing the variation of the degree of depletion inside/outside $R_{\rm dep}$. The $C^{18}O$ profiles are obtained by the product of the other two.

show depletion values between 1 and 10 (e.g. Thomas & Fuller 2008; Fontani et al. 2012; Giannetti et al. 2014, and as also discussed in Sect. 2.2.1). Instead, the choice of extending the model until reaching the theoretical condition of full depletion (i.e., $f_D = \infty$) is done to study the effect that such a drastic variation of f_D has on the size of $R_{\rm dep}$.

In Fig. 3.7 we present a sketch of the model: the profile for $n(H_2)$ is indicated in blue, that for $\chi_{C^{18}O}$ is shown in green (bottom panel), while profiles for $n(C^{18}O)$ appear in orange and red (two values of f_D are considered). Radial profiles are then convolved with the *Herschel* beam at $\lambda = 500$ μm (i.e. FWHM= 36″). For each profile we calculate the set of values that best represent the observed data in each considered region (i.e. C5, C7 and F1 of Fig. 3.1): we generate a set of profiles of $n(H_2)$ by exploring the parameter space defined by the free parameters of each modelled region, and we select the curve that maximises the total probability of the fit; i.e. $P_{\rm TOT} = \prod_i p(r_i)$, where $p(r_i)$ are the probability density values corresponding to the error bars in Fig. 3.8, and calculated at the various distances (r_i) from the spine. The same was done for the profile of $n(C^{18}O)$, selecting the best profile from a set of synthetic profiles generated by assuming different sizes of $R_{\rm dep}$.

In the next section, we present the results of the model applied to the clumps and the filament region, providing an estimate for both $R_{\rm dep}$ and $n(H_2)$ calculated at $R_{\rm dep}$.

3.4.1 Estimate of $R_{\rm dep}$

In Fig. 3.8 (*a*) and (*b*) we show the best-fitting mean-radial column density profiles of $N(H_2)$ and $N(C^{18}O)$ of clump-5, respectively. In blue, we show the column density profiles before the convolution while in orange the convolved profiles. Black-dashed lines in the central panels are the convolved undepleted-profile of $N(C^{18}O)$, assuming a constant $\chi_{C^{18}O} = 2.1 \times 10^{-7}$ for all R. The observed data are plotted as red points. In Fig. 3.8 (*c*) we report (in green) the depletion factor as a function of the distance from the projected centre in clump-5 assuming an abundance of $C^{18}O$ for $R < R_{\rm dep}$ equal to $2.1{\times}10^{-8}$, while in magenta the same profile for $\chi_{C^{18}O} = 0$. Figures 3.8 (*d*, *e*, *f*) are the same for the clump-7. In both cases, we estimate $R_{\rm dep}$ by varying the abundance $\chi_{C^{18}O}$ for $R < R_{\rm dep}$ considering the two limiting cases, i.e. with $f_D = 10$ and $f_D = \infty$.

All error bars have been defined taking the range between the $5th$ and $95th$ percentiles. In the case of $N(H_2)$, the uncertainties are generated by computing the pixel-to-pixel standard deviation of the H_2 column density along the direction of the spine, and assuming a Gaussian distribution of this quantity around the mean value of $N(H_2)$. The errors associated with $N(C^{18}O)$ are computed propagating the uncertainty on the line integrated intensity and on the optical depth correction (as discussed in Sect. 3.2.2), using a Monte Carlo approach. Finally, the uncertainties on f_D are estimated computing the abundance, randomly sampling the probability distributions of $N(H_2)$ and $N(C^{18}O)$.

In clump-5, the estimated R_{dep}, ranges between a minimum of ~ 0.07 pc and a maximum of ~ 0.10 pc. In clump-7 instead, which shows a larger average depletion factor compared to clump-5, $R_{\rm dep}$ ranges between ~ 0.07 and 0.12 pc (for $f_D = \infty$ and f_D =10, respectively). Comparing Fig. 3.8 (c) and (f), both synthetic profiles are able to well reproduce the observed ones up to a distance of ~0.15 pc. In the case of total depletion regime (i.e. $f_D = \infty$), at $R_{\rm dep}$ we estimate a volume density $n(H_2)= 5.7 \times 10^4$ cm^{-3} and 5.3×10^4 cm^{-3}, for clump-5 and 7, respectively. Our results are comparable with those found in low-mass prestellar cores (i.e. Caselli et al. 2002; Ford & Shirley 2011) and in high-mass clumps in early evolutionary stages (i.e. Giannetti et al. 2014). Furthermore, it is important to note that both values should be considered as upper limits due to the approximation in Eq. (3.5) for all the models discussed. For the case with $f_D = 10$, in clump-5 at a distance equal to $R_{\rm dep} = 0.10$ pc, the model provides a volume density of $n(H_2)= 3.1 \times 10^4$ cm^{-3}. Similarly, in the case of clump-7, $R_{\rm dep} = 0.12$ pc is reached at a density of $n(H_2)= 2.2 \times 10^4$ cm^{-3}.

In Fig. 3.8 (*g*) and (*h*) we show the mean-radial column density profiles of $N(H_2)$ and $N(C^{18}O)$, respectively, in the region between the two clumps (i.e. F1 region in Fig. 3.1), while the depletion profile is shown in the panel (*i*). The results are in agreement with

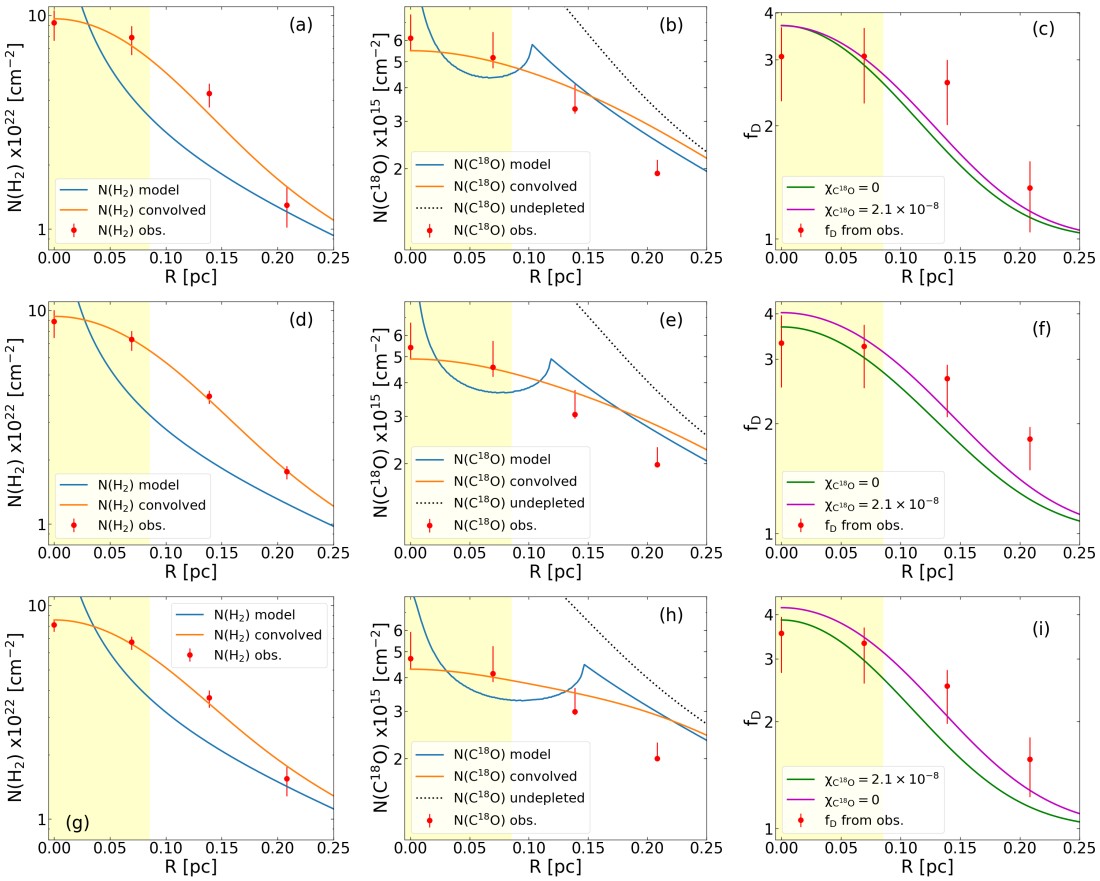

Figure 3.8: From top: Column density profiles of H_2 and $C^{18}O$ in panel (a) and (b), respectively, while panel (c) shows the los- and beam-averaged depletion factor profile from region C5 in Fig. 3.1. Panels (d), (e) and (f) are the same as (a), (b) and (c), respectively, for region C7, while panels (g), (h) and (i) refer to the region F1. Blue profiles in panels (a), (d) and (g) are the synthetic column densities profiles obtained by the integration of the number densities distribution; solid-blue lines in panels (b), (e) and (h) are the same for $C^{18}O$ assuming $\chi_{C^{18}O} = 2.1 \times 10^{-8}$ for $R < R_{dep}$, and the estimated position of R_{dep} corresponds to the cusp. Orange profiles in panel (a), (b), (d), (e), (g) and (h) are the convolution results of the blue profiles and the *Herschel* beam at 500μm, while black-dotted line in panel (b), (e) and (h) indicates the convolved undepleted-profile of $N(C^{18}O)$ assuming a constant $\chi_{C^{18}O} = 2.1 \times 10^{-7}$. Green and magenta profiles in panel (c), (f) and (i) show the depletion factor profiles obtained by assuming different abundances for $R < R_{dep}$ (see Eq. 3.5). Red points are the observed values for each quantity plotted. The yellow- shaded area shows the dimensions of the *Herschel* HWHM at $\lambda = 500$ μm (i.e. FWHM= 36″; 0.17 pc at the source distance). All profiles are plotted as a function of the radial distance from the spine.

those estimated for clumps 5 and 7, showing a 0.08 pc < R_{dep} < 0.15 pc (comparable with the typical filament width of ~0.1 pc; e.g. Arzoumanian et al. 2011), which corresponds to densities of 4.2×10^4 cm^{-3} > $n(H_2)$ > 2.3×10^4 cm^{-3} for $f_D = \infty$ and 10 respectively. The summary of the R_{dep} estimates is shown in Tab. 3.1 (Sect. 3.4.3).

Once the model presented in Sect. 3.4 has been calibrated with the data, it allows one to estimate how the size of R_{dep} can vary according to the obtained f_D. An example of these predictions is shown in Fig. 3.9. Starting from the best fitting synthetic profile obtained for clump-5 (i.e. the blue profile in Fig. 3.8a), we generated a family of $N(C^{18}O)$ profiles, convolved with different beams, i.e. FWHM between 10″ and 40″. The value of f_D reported in Fig. 3.9 is calculated at $R = 0$ according to Eq. (2.6), by varying the dimensions of R_{dep} from 0 up to the maximum size of the model radius for both f_D = (10, ∞) at $R < R_{dep}$, i.e. full and dash-dotted lines, respectively. We stress that this type of prediction is strongly dependent on the model assumptions (i.e. the assumed $C^{18}O/H_2$ abundance) and valid only for sources for which clump-5 is representative.

Nevertheless, a variation in R_{dep} is notable and it changes by varying the value of f_D for $R < R_{dep}$: at FWHM = 10″for example, a factor of 1.5 in f_D (from 4 to 6) corresponds

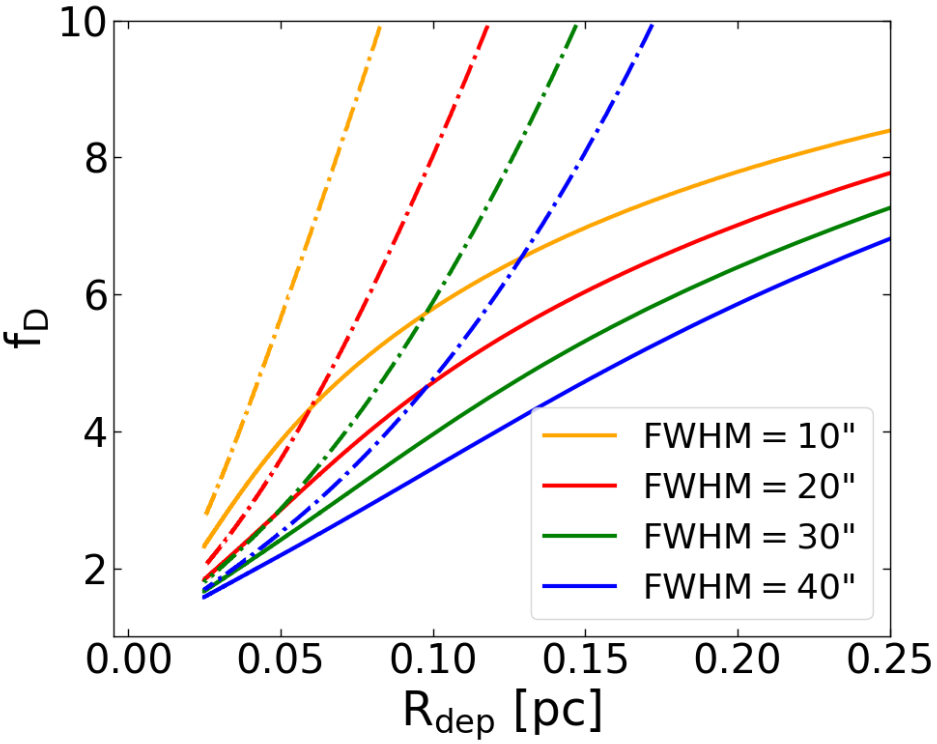

Figure 3.9: f_D variation as a function of R_{dep} for the best-fitting model applied to clump-5. Full lines represent the f_D-R_{dep} relation assuming $f_D = 10$ for $R < R_{dep}$, while the dash-dotted line are the ones for $f_D = \infty$. The blue profile in Fig. 3.8 (a) was convolved at different angular resolutions in order to predict how the same relation changes with different observational conditions. Colours are linked to the different FWHMs used for the convolution.

to a factor of 1.6 in the $R_{\rm dep}$ size if we assume $f_D = \infty$ for $R < R_{\rm dep}$, while for $f_D = 10$ it corresponds to a factor of 2.2. These variations are slightly lowered if FWHM increases: a factor of 1.4 and 1.6 for $f_D = (\infty, 10)$ at $R < R_{\rm dep}$, respectively for FWHM = 40″.

3.4.2 Comparison with 3D models

As a first step, we compared the estimated values of $R_{\rm dep}$ with the results obtained by Körtgen et al. (2017). These authors imposed a condition of total depletion on the whole collapsing region of their 3D magneto-hydrodynamical simulations, to model the deuteration process in prestellar cores. The initial core radius range is $0.08 < R_c < 0.17$ pc, depending on the simulation setup. This assumption is necessary to reduce the computational costs of the 3D-simulations but the outcome can strongly deviate from reality when we move to larger scales.

For G351.77-0.51, the largest estimated depletion radius is $\lesssim$ 0.15 pc, comparable with the initial core size assumed by Körtgen et al. (2017). We note also that, imposing the condition of total depletion on a scale of 0.17 pc, Körtgen et al. (2017) find a deuterium fraction of H_2D^+ of 10^{-3} (100 times the canonical values reported in Oliveira et al. 2003) on scales comparable with our size of $R_{\rm dep}$, after $\sim 4{\times}10^4$ yr (see their Fig. 3). This result is in agreement with the ages suggested by the depletion timescale estimated following Caselli et al. (1999) (see also Eq. 2.2), assuming the volume densities at $R_{\rm dep}$ provided in our models and a sticking coefficient $\mathcal{S}$ =1.

A similar picture is reported by Körtgen et al. (2018), who show that the observed high deuterium fraction in prestellar cores can be readily reproduced in simulations of turbulent magnetised filaments. After $\sim 10^4$ yr, the central region of the filament shows a deuterium fraction 100 times higher than the canonical value, and has a radius of ~0.1 pc (see their Fig. 1). The dimensions of the regions showing a high deuterium fraction can be compared with our estimates of $R_{\rm dep}$ since the two processes are connected, as shown in Sect. 2.2.

The comparison of our findings with both small- and large-scale simulations, suggests that the total-depletion assumption could provide reliable results within reasonable computational times.

3.4.3 The influence of a different abundance profile on $R_{\rm dep}$

It is important to stress that the model discussed in this chapter presents a number of limitations:

- A first caveat is represented by how we obtain the final values from observations (i.e. red dots in Fig. 3.8), with which our models were then optimised. The main assumption we made was the canonical $C^{18}O/H_2$ abundance, assumed equal to 2.1×10^{-7}, which actually can vary by up to a factor of 2.5 (Sect. 3.3). While knowing

these limits, we note that the results shown in the final f_D map of G351.77-0.51 (i.e. Fig. 3.5) are in good agreement with what was found by Hernandez et al. (2011, 2012) in the IRDC G035.39-00.33 and in Feng et al. (2020), the only other cases in which a filamentary-scale depletion factor map has been shown to date;

- A second limitation comes from the H_2 volume density profile we assumed in Sect. 3.4 to reproduce the data, which ignores all the substructures of clumps. Although this assumption is strong, it seems reasonable because the best fitted $N(H_2)$ profiles shown in Fig. 3.8 - panels (a), (d) and (g) - are in agreement with the data points;

- The third and strongest assumption we made is the shape of the $\chi_{C^{18}O}$ profile to describe the $C^{18}O$ variation with respect to H_2 within the main ridge of G351.77-0.51. In Fig. 3.10 the step-function profile assumed in Sect. 3.4 is shown as an orange-dashed line. All the results of Figures 3.8 and 3.9, and discussed in Sections 3.4.1 and 3.4.2, assume this abundance profile.

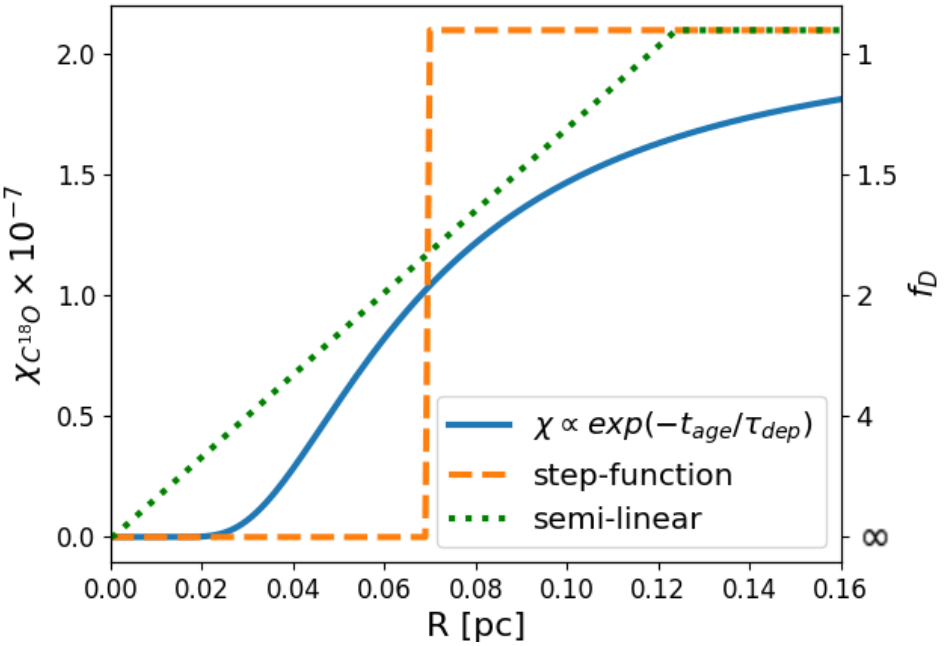

Figure 3.10: Sketch of the abundance profiles assumed to describe the $\chi_{C^{18}O}$ variation in our tests. The orange-dashed profile shows the step function described in the Sect. 3.4. In this plot, only the position of the step-function discontinuity has a physical meaning, equal to the estimated size of R_{dep} in the C5 model with $\chi_{C^{18}O} = 0$ for $R < R_{dep}$; the green-dotted profile was assumed to be constant up to a certain distance, called R_{drop}, while describes a more gradual variation of $\chi_{C^{18}O}$ for $R < R_{drop}$ (i.e. a linear profile between canonical $C^{18}O/H_2$ abundance and the value of $\chi_{C^{18}O}$ at the centre of the model); blue profile reproduces the trend $\chi_{C^{18}O} \propto \exp(-t_{age}/\tau_{dep})$, assuming only the CO absorption as dominant process (see Sect.3.4.3).

We therefore examined how much the estimates of the R_{dep} size, reported in the previous paragraphs, depend on the abundance profile assumed. For these tests, we studied the three $\chi_{C^{18}O}$ profiles shown in Fig. 3.10. In addition to the already discussed step function, we tested two profiles describing a smoother variation of $\chi_{C^{18}O}$ as a function of R. In the first test, the $C^{18}O/H_2$ abundance was assumed to be constant up to a certain distance, called R_{drop}, at which the abundance starts to linearly decrease (the green-dotted profile in Fig. 3.10). In the second test, we considered the physical case in which $\chi_{C^{18}O} \propto \exp(-t_{age}/\tau_{dep})$, where t_{age} is the age of the source, as expected integrating the evolution of $n(C^{18}O)$ over time, $dn(C^{18}O)/dt$, only considering depletion (i.e. $dn(C^{18}O)/dt = -k^{ads}n(C^{18}O)$, ignoring desorption and where k^{ads} is the adsorption rate coefficient; see Sect. 5.2 for further details). The latter case is shown as a blue profile in Fig. 3.10, which reaches zero in the innermost regions of the model (i.e., in this example, for $R \lesssim 0.02$ pc), due to the high H_2 volume densities(see Eq. 2.2).
Since we no longer have a strong discontinuity in the new abundance profiles, we have to redefine the concept of R_{dep}. In the new tests, R_{dep} will correspond to the distance at which $\chi_{C^{18}O} = 2.1 \times 10^{-8}$ (i.e. the distance at which the 90% of the CO molecules are depleted) and its new estimates are shown in Tab. 3.1.

We note that the exponential profile describes the physically more realistic case, whereas the other two are simpler approximations, easier to model. Therefore, we compare the results of the exponential case with those described by the other two.

As visible in Tab. 3.1, the new R_{dep} estimated from the model with the exponential profile are within a factor of about 2-3 of those found from the semi-linear and the step-function models, respectively. These results suggest that, although the assumption of a step-function to describe the $\chi_{C^{18}O}$ profile may seem too simplistic, it actually succeeds reasonably well in reproducing the results predicted by the other profiles. In addition, even if the relative 3σ level errors on R_{dep}, calculated by probability distributions, are always less than 15% of the values reported in Tab. 3.1, the uncertainty of a factor of 2-3 on the estimates of R_{dep} from the exponential profile seems more realistic.

Finally, it is worth discussing the predictions of the depletion timescales in the three models. In clump-5, for example, the R_{dep} of the model with the semi-linear profile, provides $n(H_2) = 5.5 \times 10^5$ cm^{-3}. With this result, we estimated $\tau_{dep} = 1.8 \times 10^3$ yr, smaller by a factor of about 10-20 than those obtained using the model with the step-function profile, for which $\tau_{dep} = 1.8 \times 10^4$ yr ($f_D = \infty$ for $R < R_{dep}$) or 3.2×10^4 yr ($f_D = 10$), respectively. This large discrepancy can be explained once again by noting that these tests represent the two extreme cases with respect to the exponential profile and therefore define the lower and upper limits for the estimate of τ_{dep}.
For the exponential-model the best fit t_{age} is 2.1×10^4 yr, while at R_{dep}, $n(H_2) =$

Regions	**C5**	**C7**	**F1**
Profiles	(pc)	(pc)	(pc)
Step-function ($f_D = 10$; $R < R_{\rm dep}$)	0.10	0.12	0.15
Step-function ($f_D = \infty$; $R < R_{\rm dep}$)	0.07	0.07	0.08
Exponential	0.04	0.04	0.05
Semi-linear	0.02	0.02	0.03

Table 3.1: Summary of the estimated $R_{\rm dep}$ assuming the profiles in Fig. 3.10 and discussed in Sect. 3.4.3. For all the estimates of $R_{\rm dep}$ shown in the table, we estimate the relative 3σ level error to be less than 15%.

1.7×10^5 cm^{-3} which corresponds to a $\tau_{\rm dep} = 5.8 \times 10^3$ yr. As mentioned above, for this model only the adsorption process of CO is included, neglecting desorption at T $\gtrsim$ 20 K. Nevertheless, the estimated $t_{\rm age}$ and the corresponding lower limit of $\tau_{\rm dep}$ suggest that the contribution of the adsorption process is negligible at these scales. Furthermore, both ages seem in good agreement with the values of $\tau_{\rm dep}$ provided by the step-function profiles, suggesting again that this is a reasonable approximation to describe the profile of $\chi_{\rm C^{18}O}$.

3.5 Chapter conclusions

The main results of this chapter are the following:

1. We used Herschel and LABOCA continuum data, together with APEX $J(2-1)$ line observations of C^{18}O, to derive the f_D map (Fig. 3.5). Along the main body and close to clump-3, the observed depletion factor reaches values as large as ~6, whereas in lower density higher temperature structures f_D is close to 1. In many regions of the spine and the branches of G351.77-0.51, the depletion factor reaches values larger than ~2.5. This highlights that even in the less prominent structures, as the branches, the depletion of CO can start to occur, altering the chemistry of the interstellar medium and making it more difficult to study the gas dynamics and to estimate the mass of cold molecular gas (using CO);

2. We find that CO-depletion in high-mass star forming regions affects not only the densest regions of the clumps, but also the filamentary structures that surround them;

3. The model assumed to estimate the size of the depletion radius suggests that it ranges between 0.15 and 0.08 pc, by changing the depletion degree from 10 to the

full depletion state. These estimates agree with the full depletion conditions used in the 3D-simulations of Körtgen et al. (2017) and Körtgen et al. (2018). This results highlights that such assumptions are not so far from the limits constrained by the observations. We also show that under certain assumptions, it is possible to estimate the size of the depleted region from f_D (see Fig. 3.9);

4. We verified that by changing the shape of the profile assumed to describe the $C^{18}O/H_2$ variations inside the filament and clumps, the estimates of the size of R_{dep} do not change more than a factor of 3. This difference was interpreted as the final uncertainty associated with the new estimates of R_{dep} from the physically more realistic case of the exponential profile, where $R_{dep} \sim 0.04 - 0.05$ pc;

5. At $R = R_{dep}$ the model shows a number densities of H_2 between 0.2 and $5.5{\times}10^5$ cm^{-3}. Following Caselli et al. (1999), we estimated the characteristic depletion timescale for the clump-5, the clump-7 and the filament region included between them (i.e. $\tau_{dep} \sim 5$ and $0.2{\times}10^4$ yr, respectively). It is interesting to note that at similar ages, in Körtgen et al. (2017, 2018) the simulated deuterium degree also suggest a regime of high depletion on scales that are consistent with the R_{dep} estimated by our model.

Chapter 4

The early stages of the high-mass star formation process (II)

This chapter is based on the published work of Sabatini et al. (2020).

Abstract of the chapter

In this chapter, we present a large sample of o-H_2D^+ observations in high-mass star-forming regions and discuss possible empirical correlations with relevant physical quantities to assess its role as a chemical clock. APEX *observations of the ground-state transition of o-H_2D^+ were analysed in a sample of high-mass clumps selected from the* ATLASGAL *survey at different evolutionary stages. We detected 16 sources in o-H_2D^+ and find clear correlations between its beam-averaged abundances with respect to H_2, X(o-H_2D^+), and the clump bolometric luminosity and the dust temperature, while only a mild correlation is found with the* CO*-depletion factor. In addition, we see a clear correlation with L/M; this indicates that the deuterated forms of H_3^+ are more abundant in the very early stages of the star formation process and that deuteration is influenced by the time evolution of the clumps. We have employed these findings together with observations of $H^{13}CO^+$, DCO^+, and $C^{17}O$ to provide an estimate of the cosmic-ray ionisation rate in a sub-sample of eight clumps based on recent analytical work. Our study presents the largest sample of o-H_2D^+ in star-forming regions to date. The results suggests that o-H_2D^+ can be considered a reliable chemical clock during the star formation processes, as proved by its strong temporal dependence.*

4.1 Observed o-H_2D^+ sample in ATLASGAL

We present new observations of the o-H_2D^+ ground-state transition $J_{K_a,K_c} = 1_{10} - 1_{11}$, based on two spectral line surveys of the ATLASGAL sources. The first set of observations[1] is composed of a sample of young massive clumps, selected from previous observations that showed a high degree of deuterated ammonia (Wienen et al. 2021). The second set of observations were part of a survey of [CI] $^3P_1 - ^3P_0$ fine structure line at 492 GHz (Lee et al., in prep.) carried out on the TOP100 sample (see Sect. 1.5.1). These were observed simultaneously with the dual-band FLASH345/460 GHz receiver on APEX, tuned to 372 GHz to observe the N_2H^+ $J = 4-3$ molecular line, which is very close to the o-H_2D^+ transition. Since the main goal of that survey was the observation of the [CI] line, the noise levels were often not adequate for an o-H_2D^+ detection. For this reason many of the sources observed in this project were used here to calculate o-H_2D^+ detection limits (we postpone further discussions to Sect. 4.3).

In Table 4.1, we summarise the physical parameters, for each clump with detection, from the dust continuum taken from Giannetti et al. (2017b), König et al. (2017), and Urquhart et al. (2018). The heliocentric distances, $d_\odot$, of our sample are taken from König et al. (2017) and Urquhart et al. (2018), and range between 1 and 5.5 kpc; the galactocentric radii, D_{GC}, are between 4.5 and $\sim$ 7.5 kpc with a mean associated error of 0.3 kpc, and are taken from Urquhart et al. (2018) assuming a distance to the Galactic Centre of 8.35 kpc (Reid et al. 2014); the effective radii, $R_{\rm eff}$, range from 0.1 to 1.3 pc; the radial velocities, $v_{\rm lsr}$, are taken from Giannetti et al. (2014) and Wienen et al. (2012), based on the $C^{17}O$ $J = 3-2$ and the NH_3 (1, 1) lines, respectively[2]. We note that a few TOP100 sources show an offset of $\sim$1 km s^{-1} between the $v_{\rm lsr}$ derived from $C^{17}O$ and the central velocity of the o-H_2D^+ lines. However, the $v_{\rm lsr}$ derived from o-H_2D^+ are consistent with the values from N_2H^+ $J = 4-3$ at $\sim$ 372.672 GHz. Such offsets are potential indications of strong internal motions driven by contraction, consistent with our finding that these clumps are on the verge of collapse (see Sect. 4.3.2).

The dust temperatures, $T_{\rm d}$, cover the typical range of values between $\sim$10 and $\sim$25 K, while $\Delta T_{\rm d}$ is the error derived from the covariance matrix of the Spectral Energy Distribution (SED) fit performed using a Levenberg-Marquardt least-squares minimisation (see Appendix D in Urquhart et al. 2018). We note that a wider exploration of the parameter space used to describe the dust properties, would lead to an error of $\sim$ 10% in $T_{\rm d}$ (e.g. Schisano et al., 2020).

The SED fit assumes a dust absorption coefficient at 870 μm, $\kappa_{870} = 1.85$ cm^2 g^{-1}, and a dust emissivity index $\beta = 1.75$ (König et al., 2017). The clump masses, $M_{\rm clump}$, and bolometric luminosities, $L_{\rm bol}$, are within $\sim$100-5400 $M_\odot$ and $\sim 10^{-1} - 10^4$ $L_\odot$, respectively. The molecular hydrogen column density, $N(H_2)$, spans less than an order of magnitude

[1] Sixteen objects in total, a part of which is also contained in the TOP100.

[2] See also the ATLASGAL Database Server at https://atlasgal.mpifr-bonn.mpg.de/cgi-bin/ATLASGAL_DATABASE.cgi

ATLASGAL-ID	$d_\odot$ (kpc)	D_{GC} (kpc)	$R_{\rm eff}^{(b)}$ (pc)	$T_{\rm d}$ (K)	$\Delta T_{\rm d}$ (K)	$v_{\rm lsr}$ (km s^{-1})	$M_{\rm clump}$ ($10^2 M_\odot$)	$L_{\rm bol}$ ($10^2 L_\odot$)	$N(H_2)$ $\log_{10}$(cm^{-2})	$n(H_2)$ $\log_{10}$(cm^{-3})	Class$^{(b,c)}$
G08.71–0.41$^{(a)}$	4.8	4.0	1.0	11.8	0.3	39.4	16.6	5.0	22.8	4.0	IRw
G13.18+0.06$^{(a)}$	2.4	5.9	0.5	24.2	0.8	49.9	3.7	83.2	22.9	4.3	70w
G14.11–0.57$^{(a)}$	2.6	6.9	0.5	22.4	0.8	20.8	3.5	31.8	22.9	4.4	IRw
G14.49–0.14$^{(a)}$	3.9	5.4	0.8	12.4	0.4	39.5	19.2	7.5	23.1	4.4	70w
G14.63–0.58$^{(a)}$	1.8	6.9	0.4	22.5	0.4	18.5	2.5	27.8	23.0	4.6	IRw
G18.61–0.07$^{(a)}$	4.3	5.3	0.7	13.8	0.3	46.6	8.8	5.9	22.8	4.2	IRw
G19.88–0.54$^{(a)}$	3.7	5.3	0.6	24.2	1.4	44.9	8.0	124.0	23.1	4.5	IRb
G28.56–0.24$^{(a)}$	5.5	4.8	1.3	11.7	0.1	87.3	54.1	17.7	23.1	4.2	IRw
G333.66+0.06$^{(a)}$	5.3	4.5	1.1	17.8	0.3	-85.1	14.2	42.7	22.7	3.9	70w
G351.57+0.76$^{(a)}$	1.3	7.0	0.3	17.0	0.1	-3.2	1.6	4.3	22.7	4.4	70w
G354.95–0.54$^{(a)}$	1.9	7.4	0.4	19.1	1.3	-5.4	1.5	4.8	22.6	4.2	70w
G12.50-0.22$^{(b)}$	2.6	5.9	0.4	13.0	0.2	35.5	1.2	1.6	22.8	4.4	70w
G14.23–0.51$^{(b)}$	1.5	6.9	0.5	17.2	2.7	19.5	7.2	14.8	23.3	4.8	HII
G15.72–0.59$^{(b)}$	1.8	6.6	0.2	12.1	0.5	17.8	1.7	0.3	22.8	4.7	IRw
G316.76-0.01$^{(b)}$	2.5	6.7	0.2	18.9	0.4	-39.9	4.7	24.2	23.1	5.0	IRb
G351.77-CL7	1.0	7.4	0.1$^{(d)}$	13.0	0.3$^{(e)}$	-3.2$^{(f)}$	1.2	0.2	22.9	5.2	70w

Table 4.1: Pysical properties (observed and derived) of our sample. *Top row*: TOP100 sources; *Bottom row*: ATLASGAL sources not in the TOP100; T_{dust} and $L_{\rm bol}$ are obtained from a SED fit described in Appendix D in Urquhart et al. 2018, while their uncertainties are derived as discussed in Sect. 4.1; $R_{\rm eff}$ values are evaluated from the dust continuum at 870 μm, while $N(H_2)$ and $M_{\rm clump}$ come from the 870 μm flux peaks and the integrated flux density over $R_{\rm eff}$, respectively; $n(H_2)$ values are lower limits derived here as $n(H_2) = N(H_2)/2R_{\rm eff}$; [(a)] $d_\odot$, M_{clump}, $L_{\rm bol}$, and $N(H_2)$ from Giannetti et al. (2017b); $R_{\rm eff}$ and $T_{\rm d}$ from König et al. (2017); D_{GC} from Urquhart et al. (2018); $v_{\rm lsr}$ from Giannetti et al. (2014) and derived from the $C^{17}O$ $J = 3 - 2$; [(b)] distances, $R_{\rm eff}$, $v_{\rm lsr}$, $M_{\rm clump}$, $L_{\rm bol}$, $N(H_2)$, and classification from Urquhart et al. (2018): the error associated with $N(H_2)$ is 20% for each source; $v_{\rm lsr}$ values are derived from the NH_3 (1, 1) data published in Wienen et al. (2012); [(c)] Classification scheme from König et al. (2017); [(d)] $R_{\rm eff}$ and $v_{\rm lsr}$ are from Leurini et al. (2011b); $v_{\rm lsr}$ is derived from the $C^{18}O$ $J = 2 - 1$; [(e)] T_{dust} standard deviation computed on the clump region defined in Leurini et al. (2019).

among the different sources, with values of $\log_{10}(N(H_2)\ [cm^{-2}])$ between 22.7 and 23.3. The sample includes clumps associated with all evolutionary classes. We refer to Giannetti et al. (2017b), König et al. (2017), and Urquhart et al. (2018) for the details on how each parameter and the relative error were derived.

Since more than 2/3 of the sample in Table 4.1 is part of the TOP100, we employ the evolutionary classes defined in Giannetti et al. (2014), Csengeri et al. (2016), and König et al. (2017) assigned in the TOP100, where four stages were defined (see also Sect. 1.5.1). We extended the same classification to the sources not included in the TOP100 by following the classes defined in Urquhart et al. (2018) and double-checking the 70 μm and 24 μm maps[3,4] according to the TOP100 classification: (*i*) Quiescent → 70w; (*ii*) Protostellar → IRw; (*iii*) YSOs → IRb; (*iv*) MSF → HII.

4.2 Observations and data reduction

4.2.1 *o*-H_2D^+ detections

The *o*-H_2D^+ $J_{K_a,K_c} = 1_{10} - 1_{11}$ spectra (rest frequency 372.4214 GHz; Amano & Hirao 2005) were observed using the on-off (ONOFF) observing mode with the FLASH+ dual-frequency MPIfR principal investigator (PI) receiver (Klein et al. 2014), mounted at the APEX telescope (Güsten et al. 2006).

The complete set of observations consists of 106 sources belonging to the ATLASGAL survey. Among them, 99 targets are also contained in the TOP100 sample, observed in three projects (IDs: 0101.F-9517, M-097.F-0039-2016, and M-098.F-0013-2016; PI: F. Wyrowski), from July 2017 to December 2018, with a total of 11568 spectra (with about 20-30 s of integration time each). The receiver bandpass is covered with overlapping FFTS backends of 2.5 GHz widths (Klein et al. 2012), one of them covering from 372.250 to 374.750 GHz with a native spectral resolution $\Delta\nu = 0.038$ km s^{-1} and with final mean system temperatures, T_{sys}, ranging between ~ 700 and 1300 K, depending on the weather conditions. At ~372 GHz, the APEX telescope has an effective beam size of 16.8″, with a corresponding main-beam efficiency of ~0.73[5]. The observations were reduced using a GILDAS class[6] Python interface pipeline.

To discard the low-quality spectra throughout the survey, the first step of the procedure was to evaluate the T_{sys} in each spectrum. We generated T_{sys} distributions source-by-source to avoid possible anomalies between different APEX projects. Only a few spectra (138 in total) are associated with clear T_{sys} outliers and were therefore discarded.

As a second step the main-beam efficiency was applied to the spectra after setting the reference frequency to 372.421 GHz. From each spectrum we subtracted a first-order

[3] https://atlasgal.mpifr-bonn.mpg.de/cgi-bin/ATLASGAL_DATABASE.cgi
[4] http://www.alienearths.org/glimpse/glimpse.php
[5] http://www.apex-telescope.org/telescope/efficiency/
[6] https://www.iram.fr/IRAMFR/GILDAS/

Tracer	Quantum numbers	ν (GHz)	Telescopes	Beam ($arcsec$)	$\Delta\nu$ (km s^{-1})	rms[(a)] (K)	Ref.[(b)]
o-H_2D^+	1_{10} - 1_{11}	372.4	APEX	17	0.5-0.6	0.02-0.05	[1]
$H^{13}CO^+$	1-0	86.8	IRAM-30m[a]	24	0.75	0.02	[1]
DCO^+	1-0	72.0	IRAM-30m	28	0.75	0.02	[1]
$C^{17}O$[(c)]	1-0	112.4	IRAM-30m	21	0.5	0.05	[2]
$C^{17}O$	1-0	112.4	IRAM-30m	21	0.6	0.06	[3]

Table 4.2: Summary of the observations. [(a)] The temperatures are reported on the main-beam temperature scale; [(b)] [1] Sabatini et al. (2020); [2] Giannetti et al. (2014); [3] Csengeri et al. (2016); [(c)] TOP100 sample;

[a] IRAM is supported by INSU/CNRS (France), MPG (Germany), and IGN (Spain).

polynomial baseline around the line, properly masked. In some rare cases, where a first order was not enough, a second-order (or third-order) baseline was used. In the last step, all spectra related to each source were averaged (stitched) with a spectral resolution between 0.5 and 0.6 km s^{-1} to reduce the noise, generating one final spectrum per source.

Despite the relatively high Einstein-A coefficient (i.e. $\log_{10}(A_{ul}\,[s^{-1}])= -3.96$; CDMS[7]; Müller et al. 2001), the $J_{K_a,K_c} = 1_{10} - 1_{11}$ transition of o-H_2D^+ is usually faint due to the low relative abundance of this species with respect to H_2. Typical observed abundances range between $\sim 10^{-10} - 10^{-12}$ in many low-mass star-forming regions (e.g. Vastel et al. 2006, Harju et al. 2008, Caselli et al. 2008, Friesen et al. 2010, 2014 and Miettinen 2020) and few high-mass regime counterparts (e.g. Harju et al. 2006, Swift 2009, Pillai et al. 2012 and Giannetti et al. 2019). For this reason, in this chapter we evaluated the significance of each detection considering the integrated signal-to-noise ratio of the line (iS/N) and taking each line with iS/N > 3 as a detection. For reference, the integrated main-beam temperature, $\int_{ch} T_{mb}\, d\upsilon$, and the integrated root mean square noise, $i$rms[8], of each line is reported in the spectra shown in Figures 4.1 and 4.2.

Sixteen ATLASGAL sources with o-H_2D^+ detection[9] are reported, 11 of which belong to the TOP100 sample. All the evolutionary classes are contained in our sample allowing us to study (within statistical limits) how the o-H_2D^+ emission changes through the massive star formation process, and how these changes correlate with the other observed or derived quantities.

[7] *Cologne Database for Molecular Spectroscopy* (CDMS): https://cdms.astro.uni-koeln.de/cdms/portal/

[8] $i$rms = $\sqrt{N_{ch}}\, d\upsilon\, \sigma$, where N_{ch} is the number of channels included in the integration, $d\upsilon$ is the velocity resolution, and σ is the rms per channel.

[9] One of them showing a double velocity component.

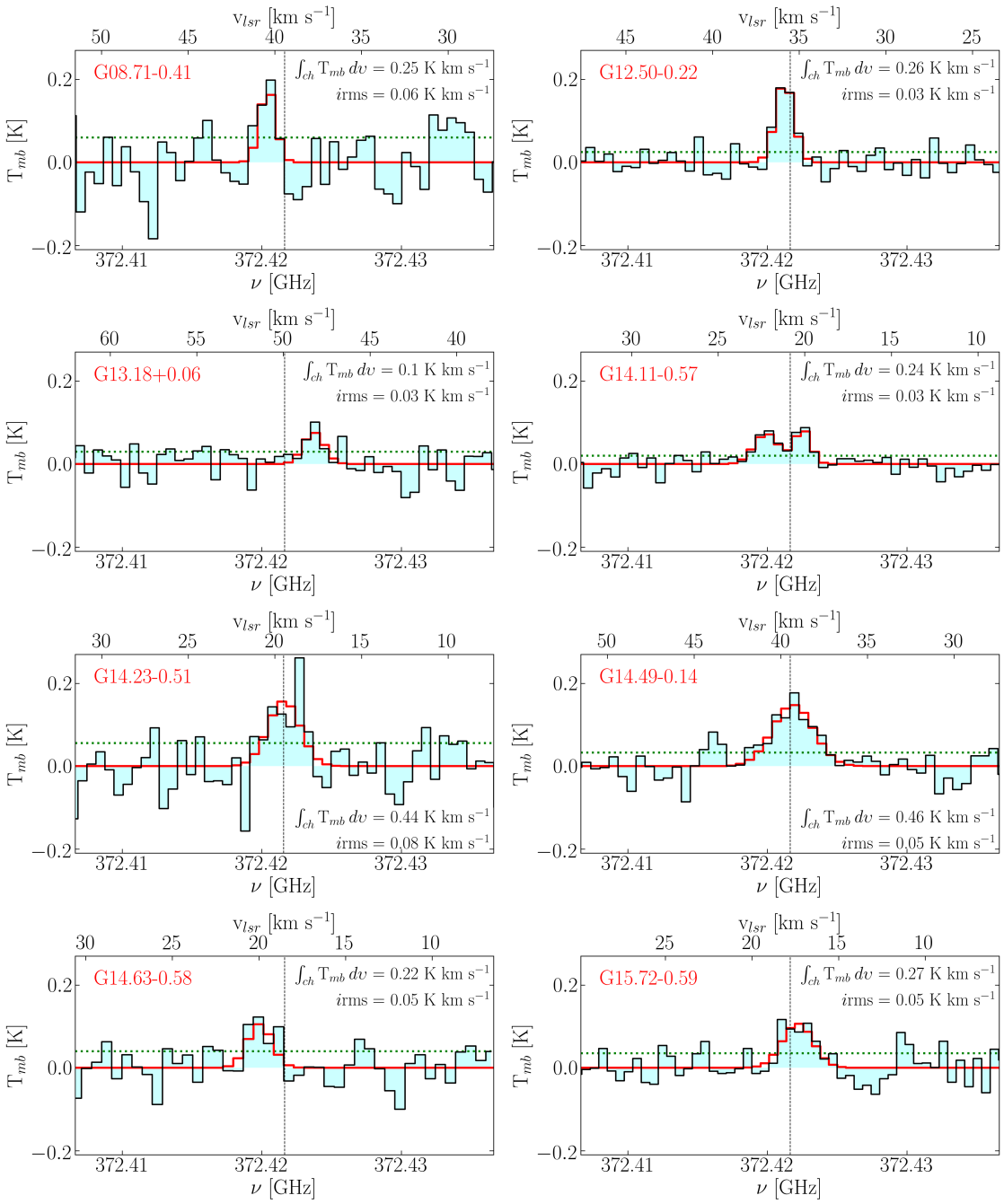

Figure 4.1: o-H_2D^+ spectra for half the sources of our sample (cyan). In each panel the source name is shown on the left in red, while the integrated main-beam temperature, $\int_{ch} T_{\rm mb}\,dv$, and rms (*i*rms) are on the right in black. The green dotted line represents the 1σ noise levels in $T_{\rm mb}$. The frequency and velocity axes are reported and cut around the line, while the intensity axis is fixed to the highest $T_{\rm mb}$ detected (observed in G14.23-0.51). The black dashed lines indicate the $v_{\rm lsr}$ of each source derived from the $C^{17}O$ $J = 3-2$ and the NH_3 (1, 1) lines (see Table 4.1). The model obtained with MCWeed is superimposed in red.

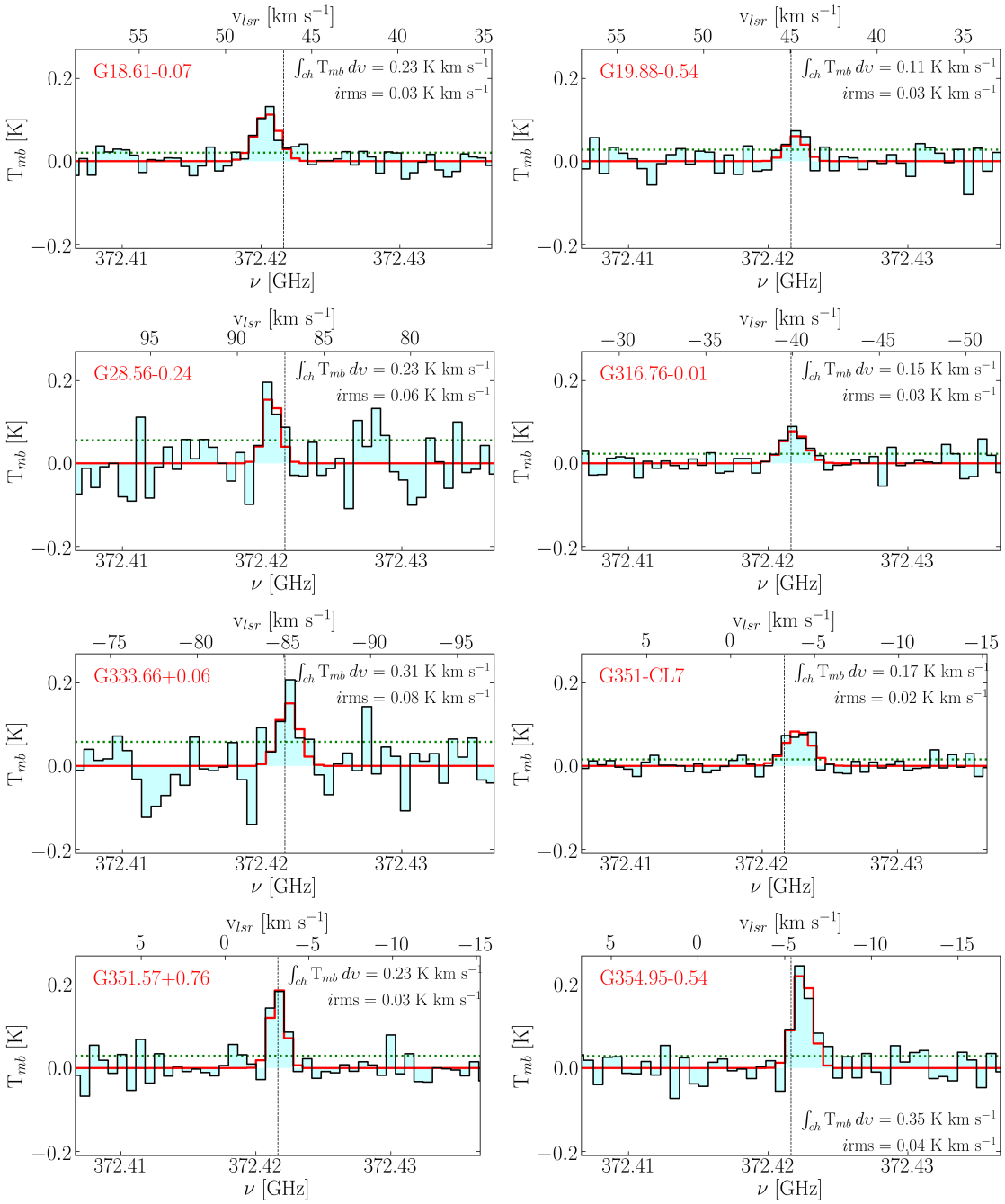

Figure 4.2: Same as Fig. 4.1, but for the second half of the sample.

4.2.2 Additional tracers: $H^{13}CO^+$, DCO^+ and $C^{17}O$

For one-half of the sources with *o*-H_2D^+ detection, we collected further observations of $H^{13}CO^+$, DCO^+, and $C^{17}O$ that we employ to estimate the cosmic-ray ionisation rate (explained in Sect. 4.5). We simultaneously observed the rotational transition of both $H^{13}CO^+$ and DCO^+ in 30 ATLASGAL sources, using the *Eight MIxer Receiver* E90 (EMIR; Carter et al. 2012) at the IRAM-30m single-dish telescope. (IRAM project-id 107-15). From the study of Wienen et al. (2021), the 30 brightest clumps in deuterated ammonia were selected. The receiver was tuned at 75 GHz, allowing the simultaneous detection of $H^{13}CO^+$ and DCO^+ at 87 and 72 GHz, respectively, in 2×8 GHz sidebands with a velocity resolution of 0.75 km s^{-1}. The sources were observed with an average integration time of $\sim$ 1 h in February 2016. The average $T_{\rm sys}$ are $\sim$ 130 K for the $H^{13}CO^+$ lines and 210 K for the DCO^+ lines. The data were originally calibrated to antenna temperature, T_A^*, and then converted to $T_{\rm mb}$ using the tabulated values for the IRAM-30m forward efficiency, $\eta_{\rm eff} = 0.95$, and main-beam efficiency, $\eta_{\rm mb} = 0.74$. The final averaged rms noise was 0.02 K for both lines.

The $C^{17}O$ $J = 1 - 0$ observations at 112 GHz, were taken from Giannetti et al. (2014) for the TOP100 sources and from Csengeri et al. (2016) for the two sources not included in the TOP100, to which we refer for a full description of the data sets.

The full set of observations used in this chapter is summarised in Table 4.2, where the details on the observed molecules (columns 1-3), the telescope setup used (columns 4-7) and the references from which the data were selected (column 7) are reported.

4.3 Deriving the physical properties of the sample

The spectra for all the sources with a reliable *o*-H_2D^+ detection are presented in Figures 4.1 and 4.2, where the 1σ noise level is indicated by the green dotted line, and the integrated main-beam temperature and rms of each line are shown for each spectrum.

The *o*-H_2D^+ line emission is clearly visible with a single velocity component in all the sources, with the exception of G14.11-0.57 (see Fig. 4.1), which shows two line components separated by $\sim$ 2.5 km s^{-1}. Both the components have iS/N $\gg$ 3, and for this reason in the following analysis they are treated separately (i.e. C1 and C2). The average noise of the spectra is almost constant at $\sim 0.02 - 0.03$ K ($T_{\rm mb}$-scale) throughout the sample. In a few cases the noise is slightly higher (i.e. $\sim$ 0.05 K). The *full width at half maximum* (FWHM) line widths are between 1 and 3 km s^{-1}, in agreement with those reported by Giannetti et al. (2014) and Wienen et al. (2012) for the $C^{17}O$ $J = 3 - 2$ and the NH_3 (1, 1), respectively.

ATLASGAL-ID	$L/M^{(a)}$ $L_\odot M_\odot^{-1}$	$[f_D]^{(b)}$ $(\chi^{obs}_{C^{17}O}/\chi^{exp}_{C^{17}O})$	$N(o\text{-}H_2D^+)^{(c)}$ $\log_{10}(cm^{-2})$	$X(o\text{-}H_2D^+)$ $\log_{10}(N[o\text{-}H_2D^+]/N[H_2])$	$\Delta\upsilon^{(c)}_{obs}$ (km s^{-1})	$\alpha^{(d)}$	$\mathcal{M}^{(e)}$
G08.71–0.41	0.3	8.2	$12.7^{+0.2}_{-0.2}$	-10.1 ± 0.2	$1.0^{+0.7}_{-0.7}$	0.1	1.9
G13.18+0.06	22.5	6.6	$12.2^{+0.3}_{-0.3}$	-10.6 ± 0.3	$1.4^{+1.2}_{-1.0}$	0.5	1.9
G14.11–0.57 (C1)	9.1	2.1	$12.1^{+0.3}_{-0.3}$	-10.8 ± 0.3	$1.0^{+0.7}_{-0.7}$	0.3	1.3
G14.11–0.57 (C2)	9.1	2.1	$12.3^{+0.3}_{-0.3}$	-10.6 ± 0.3	$1.6^{+1.1}_{-0.9}$	0.9	2.3
G14.49–0.14	0.4	9.0	$13.0^{+0.1}_{-0.1}$	-10.1 ± 0.1	$2.7^{+0.9}_{-0.8}$	0.6	5.4
G14.63–0.58	11.1	2.0	$12.4^{+0.2}_{-0.2}$	-10.6 ± 0.2	$1.7^{+1.0}_{-0.7}$	1.0	2.4
G18.61–0.07	0.7	4.2	$12.6^{+0.1}_{-0.1}$	-10.2 ± 0.1	$1.6^{+0.7}_{-0.6}$	0.5	3.0
G19.88–0.54	15.5	2.2	$12.1^{+0.3}_{-0.3}$	-11.0 ± 0.3	$1.3^{+1.0}_{-0.9}$	0.3	1.7
G28.56–0.24	0.3	6.3	$12.7^{+0.3}_{-0.4}$	-10.4 ± 0.4	$1.0^{+0.6}_{-0.7}$	0.1	1.9
G333.66+0.06	3.0	4.0	$12.6^{+0.3}_{-0.4}$	-10.1 ± 0.4	$1.3^{+1.0}_{-1.0}$	0.2	2.1
G351.57+0.76	2.7	2.4	$12.6^{+0.2}_{-0.2}$	-10.1 ± 0.2	$1.0^{+0.5}_{-0.4}$	0.4	1.5
G354.95–0.54	3.2	3.8	$12.7^{+0.1}_{-0.1}$	-9.9 ± 0.1	$1.2^{+0.4}_{-0.4}$	0.8	1.8
G12.50-0.22	1.3	–	$12.8^{+0.1}_{-0.1}$	-10.0 ± 0.1	$1.2^{+0.4}_{-0.3}$	1.0	2.2
G14.23–0.51	2.1	–	$12.8^{+0.2}_{-0.2}$	-10.5 ± 0.2	$2.1^{+0.8}_{-0.8}$	0.6	3.5
G15.72–0.59	0.2	–	$12.8^{+0.2}_{-0.2}$	-10.0 ± 0.2	$1.9^{+1.0}_{-1.0}$	1.0	3.8
G316.76-0.01	5.2	–	$12.3^{+0.2}_{-0.2}$	-10.8 ± 0.2	$1.6^{+0.6}_{-0.5}$	0.2	2.5
G351.77-CL7	0.7[f]	3.4[g]	$12.6^{+0.1}_{-0.1}$	-10.3 ± 0.1	$1.8^{+0.6}_{-0.6}$	0.5	3.5

Table 4.3: Summary of derived properties of the ATLASGAL-sources in our sample. Top panel: TOP100 sources; Bottom panel: ATLASGAL sources not in the TOP100. The symbol "–" means or that this value is not available from the literature (for f_D) or that it could not be calculated (for G14.11-0.57). [(a)]data from columns (5) and (6) in Table 4.1; [(b)]we associate with the CO depletion factor, f_D, a conservative error of 15%, considering recent results discussed in Sabatini et al. (2019) (see discussion in Sect. 3.3), where we presented a detailed CO-depletion study in a local filament of massive star-forming regions; [(c)]computed using MCweeds (Giannetti et al. 2017b) assuming $T_{ex} = T_{dust}$; [(d)]virial parameters derived following MacLaren, Richardson & Wolfendale (1988), see Sect. 4.3.2; [(e)]Mach numbers derived following Eq. 4.3 in Sect. 4.3.2, where $m(H_2D^+) = 6.692 \times 10^{-24}$ g is the mass of the H_2D^+ molecule; [(f)]this value was calculated from Giannetti et al. (2019) as the integrated values on the APEX-beam at 230 GHz (i.e. 28 arcsec); [(g)]data from Sabatini et al. (2019) (see also Sect. 3.3); see their Sect. 4.1 for more details on how this value is computed.

4.3.1 Column densities

Column densities are obtained by fitting the observed spectra by employing MCWeeds (Giannetti et al. 2017b), an external interface between Weeds (Maret et al. 2011), simple and fast in building synthetic spectra assuming LTE, and the Bayesian statistical models and adaptation algorithms of PyMC (Patil, Huard & Fonnesbeck 2010). Model fits are shown in red in Figures 4.1 and 4.2. For the 16 sources with an o-H_2D^+ detection, the results of the fit are summarised in Table 4.3, computed assuming that the excitation temperature of the transition $T_{ex} = T_d$ (e.g. Giannetti et al. 2019). This assumption implies that the dust and o-H_2D^+ are well mixed, which might not be the case for every source as the average volume densities (see Table 4.3) are slightly lower than the o-H_2D^+ critical density, $n_{cr} \sim 10^5$ cm^{-3} (e.g. Caselli et al. 2008 and Vastel et al. 2012). By following Caselli et al. (2008), who reported excitation temperatures up to 40% lower than the dust values, we re-performed the column density calculations taking $T_{ex} = 0.6\ T_d$. We find that in most of the cases, the o-H_2D^+ column densities are well within the uncertainties reported in Table 4.3, so that our assumption of $T_{ex} = T_d$ does not affect the final results. In this range of temperatures the line optical depths, τ, computed as in Eq. (5) of Caselli et al. (2008), are $\lesssim 0.1$, implying optically thin regimes.

We also assumed extended emission relative to the APEX beam size to compute the column densities, supported by the results of Pillai et al. (2012), where the emission scale of the o-H_2D^+ was observed to be close to the typical clump size. We note that for all the clumps with a clear o-H_2D^+ detection, the angular size corresponding to the effective radii (R_{eff}; König et al. 2017 and Urquhart et al. 2018) is always larger than the APEX beam sizes, except for G316.76-0.01 (Pillai et al. 2012).

Fits were performed assuming the o-H_2D^+ molecular line parameters provided by the CDMS database: an Einstein-A coefficient $\log_{10}(A_{ul}\ [s^{-1}]) = -3.96$, a statistical weight of $g_{ul} = 9$, an energy gap between the two quantum levels $E_{ul} = 17.9$ K, and the partition function, $Q(T_g)$ for o-H_2D^+ in the relevant temperature range [9.375 K: 10.3375, 18.750 K: 12.5068, 37.500 K: 15.5054] (see also Giannetti et al. 2019).

Throughout the sample the o-H_2D^+ column density varies by less than an order of magnitude, from 1.3×10^{12} to 10^{13} cm^{-2}, while its abundance $X(o\text{-}H_2D^+) = N[o\text{-}H_2D^+]/N[H_2]$ is in the range $\sim$ (1–12.6) $\times\ 10^{-11}$. The abundances were derived with respect to the H_2 column densities reported by Giannetti et al. (2017b) and Urquhart et al. (2018). The final $X(o\text{-}H_2D^+)$ values are slightly lower than those found in many low-mass pre- and protostellar regions (e.g. Vastel et al. 2006; Harju et al. 2008; Caselli et al. 2008; Friesen et al. 2010, 2014; and Miettinen 2020), but in agreement with those derived for the few high-mass star-forming regions observed by Harju et al. (2006), Pillai et al. (2012), and Giannetti et al. (2019); the only exception is the significantly lower values found by Swift (2009), who reports an abundance of $\sim (3-5)\times10^{-13}$ in the infrared dark clouds (IRDCs) G030.88+00.13 and G028.53-00.25.

The o-H_2D^+ detection limits were calculated for sources with no detection. For each evolutionary class, we selected the observations with rms within a factor of 2 with respect to the average noise of the spectra with detection. This selection criteria leads to a loss of 44 sources, while the remaining 46 sources are summarised in Table 4.4.

Each class has a similar number of objects (70w:8, IRw:16, IRb:9, and HII:13 sources), which makes detection limits comparable through the progressive evolution of the clumps, with a o-H_2D^+ detection rate of 47% in 70w, 27% in IRw, 18% in IRb, and 7% in HII. The detection limits were calculated for each selected source in order to obtain the column density value that would correspond to a 3σ detection. The assumptions made for the column density calculations are the same as for the sources with detection. We used the typical FWHM line width of 1.5 km s^{-1}, which corresponds to the average value of the detection sources. We note that this approach provides only qualitative information on the upper limits of column densities of o-H_2D^+ throughout the sample as it is mainly influenced by the quality of observations rather than by the physics of the sources. Nevertheless, we limit this problem by selecting spectra with noise levels comparable to those of detected sources. The (upper limit) relative abundances retrieved in this way are in the range of the typically observed values.

4.3.2 Estimates of dynamical quantities

In addition to the $N(o\text{-}H_2D^+)$ value, MCWeeds also provides an estimate of the FWHMs ($\Delta\upsilon_{\rm obs}$ in Table 4.3) of each line, allowing us to derive two important dynamical quantities: the virial parameter, α, and the Mach number, $\mathcal{M}$. The first gives us an assessment of the dynamical state of each source, hence how far a source is from virial equilibrium (Chandrasekhar & Fermi 1953), while the second tells us about the turbulence of the gas, hence whether the clump is supported mainly by dynamical or by thermal motions.

The estimates of α were made following the classic definition of Bertoldi & McKee (1992):

$$\alpha = \frac{M_{\rm vir}}{M_{\rm clump}} \quad \text{with} \quad M_{\rm vir} = k_2 \left(\frac{\rm R_{clump}}{\rm pc}\right)\left(\frac{\Delta\upsilon_{\rm dyn}}{\rm km\ s^{-1}}\right)^2. \tag{4.1}$$

Here $M_{\rm vir}$ and $M_{\rm clump}$ are the source virial mass and total mass, respectively; $\rm R_{clump}$ is the clump size in parsec, here assumed equal to the $R_{\rm eff}$; and $k_2 = 210$ for the homogeneous core assumption we made (MacLaren, Richardson & Wolfendale 1988). The line width $\Delta\upsilon_{dyn}$ in Eq. 4.1 is the combination of the thermal motions of the particle of mean mass, $\Delta\upsilon_{\rm th,\langle m\rangle}$; the thermal gas motions, $\Delta\upsilon_{\rm th}$; and the observed FWHM. Using the values of $\Delta\upsilon_{\rm obs}$ in Table 4.3, we derived $\Delta\upsilon_{\rm dyn}$ by following Kauffmann, Pillai & Goldsmith (2013) (see their Sect.4):

$$\Delta\upsilon_{\rm dyn} = \sqrt{\Delta\upsilon^2_{\rm obs} + \Delta\upsilon^2_{\rm th,\langle m\rangle} - \Delta\upsilon^2_{\rm th}}. \tag{4.2}$$

ATLASGAL-ID	T_d (K)	$N(H_2)$ $\log_{10}(cm^{-2})$	rms noise$^{(c)}$ (K)	$N(o\text{-}H_2D^+)^{(d)}$ $\log_{10}(cm^{-2})$	$X(o\text{-}H_2D^+)^{(d)}$ $\log_{10}(N[o\text{-}H_2D^+]/N[H_2])$	Class
G30.85-0.08$^{(a)}$	16.7	22.72	0.07	< 12.77	$< -9.95 \pm 0.14$	70w
G320.88-0.40$^{(a)}$	16.8	22.69	0.09	< 12.86	$< -9.83 \pm 0.13$	70w
G337.28+0.01$^{(a)}$	10.7	22.92	0.09	< 13.07	$< -9.86 \pm 0.12$	70w
G338.07+0.01$^{(a)}$	18.5	22.51	0.07	< 12.75	$< -9.76 \pm 0.15$	70w
G338.78+0.48$^{(a)}$	12.2	22.85	0.07	< 12.90	$< -9.95 \pm 0.13$	70w
G351.13+0.77$^{(a)}$	18.6	22.49	0.06	< 12.69	$< -9.80 \pm 0.14$	70w
G351.57+0.76$^{(a)}$	17.0	22.67	0.03	< 12.40	$< -10.27 \pm 0.14$	70w
G353.42-0.08$^{(a)}$	17.1	22.38	0.04	< 12.49	$< -9.89 \pm 0.14$	70w
G08.68-0.37$^{(a)}$	24.2	22.94	0.07	< 12.67	$< -10.27 \pm 0.16$	IRw
G10.45-0.02$^{(a)}$	20.7	22.70	0.05	< 12.58	$< -10.12 \pm 0.15$	IRw
G18.73-0.23$^{(a)}$	21.9	22.90	0.06	< 12.66	$< -10.24 \pm 0.15$	IRw
G18.89-0.47$^{(a)}$	14.4	23.17	0.02	< 12.25	$< -10.92 \pm 0.13$	IRw
G23.21-0.38$^{(a)}$	22.1	23.20	0.03	< 12.36	$< -10.84 \pm 0.15$	IRw
G24.63+0.17$^{(a)}$	18.1	22.57	0.06	< 12.71	$< -9.86 \pm 0.14$	IRw
G317.87-0.15$^{(a)}$	19.3	23.03	0.06	< 12.68	$< -10.35 \pm 0.14$	IRw
G318.78-0.14$^{(a)}$	24.9	22.60	0.07	< 12.68	$< -9.92 \pm 0.15$	IRw
G331.71+0.60$^{(a)}$	21.0	22.88	0.05	< 12.56	$< -10.32 \pm 0.15$	IRw
G335.79+0.17$^{(a)}$	24.7	23.18	0.06	< 12.61	$< -10.57 \pm 0.16$	IRw
G337.26-0.10$^{(a)}$	21.7	22.72	0.06	< 12.62	$< -10.10 \pm 0.16$	IRw
G338.92+0.56$^{(a)}$	24.2	23.48	0.07	< 12.68	$< -10.80 \pm 0.16$	IRw
G340.78-0.10$^{(a)}$	26.2	22.82	0.08	< 12.70	$< -10.12 \pm 0.16$	IRw
G342.48+0.18$^{(a)}$	23.6	22.82	0.05	< 12.50	$< -10.32 \pm 0.16$	IRw
G343.75-0.16$^{(a)}$	24.3	23.30	0.06	< 12.57	$< -10.73 \pm 0.16$	IRw
G351.45+0.66$^{(a)}$	21.4	23.88	0.03	< 12.34	$< -11.54 \pm 0.16$	IRw
G19.88-0.54$^{(a)}$	24.2	23.12	0.03	< 12.30	$< -10.82 \pm 0.16$	IRb
G34.41+0.23$^{(a)}$	26.1	23.25	0.05	< 12.50	$< -10.75 \pm 0.16$	IRb
G37.55+0.20$^{(b)}$	28.4	22.73	0.05	< 12.50	$< -10.23 \pm 0.16$	IRb
G53.14+0.07$^{(a)}$	25.4	22.90	0.04	< 12.42	$< -10.48 \pm 0.16$	IRb
G332.09-0.42$^{(a)}$	30.8	22.94	0.05	< 12.46	$< -10.48 \pm 0.16$	IRb
G333.31+0.11$^{(a)}$	25.9	22.76	0.04	< 12.44	$< -10.32 \pm 0.16$	IRb
G341.22-0.21$^{(b)}$	27.0	22.86	0.06	< 12.55	$< -10.31 \pm 0.16$	IRb
G351.16+0.70$^{(a)}$	21.9	23.67	0.06	< 12.62	$< -11.05 \pm 0.15$	IRb
G351.25+0.67$^{(a)}$	32.5	23.34	0.06	< 12.56	$< -10.78 \pm 0.16$	IRb
G10.62-0.38$^{(a)}$	34.5	23.57	0.08	< 12.67	$< -10.90 \pm 0.17$	HII
G12.81-0.20$^{(a)}$	35.1	23.56	0.07	< 12.60	$< -10.96 \pm 0.17$	HII
G14.33-0.64$^{(a)}$	21.6	23.51	0.06	< 12.65	$< -10.86 \pm 0.15$	HII
G31.41+0.31$^{(a)}$	26.3	23.57	0.06	< 12.62	$< -10.95 \pm 0.16$	HII
G34.26+0.15$^{(a)}$	31.0	23.85	0.07	< 12.66	$< -11.19 \pm 0.16$	HII
G34.40+0.23$^{(a)}$	22.8	23.11	0.06	< 12.63	$< -10.48 \pm 0.15$	HII
G330.95-0.18$^{(a)}$	33.0	23.76	0.04	< 12.37	$< -11.39 \pm 0.16$	HII
G333.13-0.43$^{(a)}$	35.2	23.40	0.05	< 12.50	$< -10.90 \pm 0.16$	HII
G333.28-0.39$^{(a)}$	30.4	23.24	0.05	< 12.51	$< -10.73 \pm 0.16$	HII
G333.60-0.21$^{(a)}$	41.1	23.49	0.07	< 12.60	$< -10.89 \pm 0.17$	HII
G337.40-0.40$^{(a)}$	31.8	23.36	0.07	< 12.70	$< -10.65 \pm 0.16$	HII
G337.70-0.02$^{(a)}$	25.6	23.35	0.08	< 12.66	$< -10.70 \pm 0.16$	HII
G343.13-0.06$^{(a)}$	30.9	23.39	0.05	< 12.48	$< -10.91 \pm 0.16$	HII

Table 4.4: Summary of the ATLASGAL sources used to compute the o-H_2D^+ detection limits. Derived physical properties of the sample of massive clumps used to calculate the o-H_2D^+ detection limits described in Sect. 4.3.1. The sources are separated, from top to bottom, into their evolutionary classes. The error associated with $N(H_2)$ is 20% for each source; $^{(a)}$ T_d, $N(H_2)$, and Class from Giannetti et al. (2017b); $^{(b)}$ T_d, $N(H_2)$, and Class from Urquhart et al. (2018); $^{(c)}$ The temperatures are reported on the main-beam temperature scale;$^{(d)}$derived in Sabatini et al. (2020).

The values of α reported in Table 4.3 suggest that the sample in this study is heterogeneous, but mainly composed of marginal to highly unstable sources ($\alpha < 1$). We also found agreement with the results of Kauffmann, Pillai & Goldsmith (2013), who discussed the implications of the low virial parameters estimated in a large sample of high-mass star-forming regions, finding that for $M_{\rm clump} \gtrsim 10\,M_\odot$, $\alpha \ll 2$ (see their Fig. 1). Such low α values might be the symptomatic consequence of a rapid collapse.

The Mach number has been computed from the o-H_2D^+ line width provided by MCWeeds as

$$\mathcal{M} = \frac{\sigma_{\rm turb}}{\sigma_{\rm H_2}}, \tag{4.3}$$

with

$$\sigma_{\rm turb} = \sqrt{\sigma^2_{\rm obs} - \sigma^2_{\rm th}} \quad \text{and} \quad \sigma_{\rm H_2} = \frac{\Delta\upsilon_{\rm th,\langle m\rangle}}{2\sqrt{2ln2}}. \tag{4.4}$$

In Eq. 4.4, $\sigma_{\rm obs} = \Delta\upsilon_{\rm obs}/(2\sqrt{2ln2})$ is the observed velocity dispersion of o-H_2D^+ lines, while $\sigma_{\rm th} = [k_B\,T_{\rm d}/m(H_2D^+)]^{1/2}$ is the gas thermal component at $T_{\rm d}$, where k_B is the *Boltzmann constant* and $m(H_2D^+)$ is the mass of the H_2D^+ molecule. The Mach number range is $1.3 - 5.4$, implying supersonic turbulent motions for the gas involved in all the sources.

4.4 New correlations with $X(o\text{-}H_2D^+)$

We have explored possible correlations between the o-H_2D^+ relative abundance obtained from our survey and the other physical quantities in Tables 4.1 and 4.3. In the following, we divide the discussion into three main parts: correlations with (*i*) the source parameters, to examine the possible connection between the amount of o-H_2D^+ and the quantities that characterise the clump structures and their distribution in the Galactic plane; (*ii*) the evolutionary parameters, to highlight trends during the star formation process; and (*iii*) the dynamical parameters, for possible dependencies on factors such as the concentration and the turbulence of the gas of the source.

Each data set comprises N measurements $\mathrm{D} \equiv \{x_k, \delta x_k, y_k, \delta y_k\}$, with $k = 1, ..., N$. Each point k of a data set corresponds to a different massive star-forming region of the ATLASGAL survey; x_k and y_k are the quantities over which we look for a linear correlation, while δx_k and δy_k are the corresponding errors. We test our data sets against linear correlations of the form $y = mx + q$, where m and q are, respectively, the slope and the normalisation. In our specific case we always consider the y_k as measures of $\log_{10}X(o\text{-}H_2D^+)$, while the x_k can be the target's galactocentric radius, $D_{\rm GC}$; effective radius, $R_{\rm eff}$; clump total mass, $\log_{10} M_{\rm clump}$; bolometric luminosity, $\log_{10} L_{\rm bol}$; dust temperature, $T_{\rm d}$; CO-depletion factor, $f_{\rm D}$; H_2 column density, $\log_{10} H_2$; luminosity-to-mass ratio, $\log_{10} L/M$; virial factor, α; or the Mach number, $\mathcal{M}$ (see Tables 4.1 and 4.3).

We assume that the errors δx_k and δy_k follow a Gaussian distribution, and the model's

likelihood $\ln \mathscr{L}(\boldsymbol{\theta}|\mathrm{D})$, defined by the parameters $\boldsymbol{\theta} \equiv \{m, q\}$ given the data D, is

$$\ln \mathscr{L} = \frac{1}{2} \sum_{k=1}^{N} \left[-\left(\frac{y_k - m x_k - q}{\sigma_k} \right)^2 - \ln(2\pi\sigma_k^2) \right], \tag{4.5}$$

where $\sigma_k = \sqrt{m^2 \delta x_k^2 + \delta y_k^2}$.

We rely on a Bayesian approach and we run a parameter space search using a Markov chain Monte Carlo (MCMC) algorithm, using uninformative flat priors over the models' free parameters. For each correlation, we run ten chains, each evolved for 10000 steps, using a classical Metropolis-Hasting sampler to sample from the posterior (Metropolis et al. 1953 and Hastings 1970). All the chains converge quickly towards the highest probability region of the parameter space and we eliminate, as a conservative choice, the first 3000 steps from each chain, considered to be a reliable burn-in. The remaining steps are used to compute the posterior distribution over the model's free parameters. We test the correlations we obtain from the MCMC also considering smaller burn-in and we note that our results do not depend significantly on this choice as long as we eliminate at least the first 1000 steps.

To test the performances of our models in fitting data, for each correlation we evaluate the Bayesian evidence $\mathcal{E}$, defined as the average of the likelihood (Eq. 4.5) under the considered priors. Specifically,

$$\mathcal{E} = \int \mathscr{L}(\boldsymbol{\theta}|\mathrm{D}) P(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}, \tag{4.6}$$

where P denotes the priors that we used. Afterwards, we compute the Bayesian evidence after fitting each data set with a model where the slope has been fixed to zero. In $\mathscr{M}_1$ and $\mathscr{M}_2$, which are respectively the models with $m = 0$ and the general model with two free parameters, the Bayes factor,

$$\mathcal{B}_{2,1} = \frac{\mathcal{E}_2}{\mathcal{E}_1}, \tag{4.7}$$

describes whether $\mathscr{M}_1$ provides a better description of the data than $\mathscr{M}_2$ ($\mathcal{B}_{2,1} < 1$), or the other way around ($\mathcal{B}_{2,1} > 1$). The Bayesian evidence is computed using the software package PyMultiNest (Buchner et al., 2014), as implemented in its publicly available version.

Figure 4.3 shows an example of the posterior distribution obtained for each correlation. The black curves in the two-dimensional distributions indicate regions enclosing 68%, 95%, and 99.7% of the total probability (i.e. 1σ, 2σ, and 3σ, respectively). The black vertical lines in the one-dimensional marginalised distributions instead indicate the 16th, 50th, and 84th percentiles, used to estimate the 1σ error bars. To derive the 2σ error bars we use instead the interval between the 5th and 95th percentiles of the corresponding one-dimensional marginalised posterior distributions.

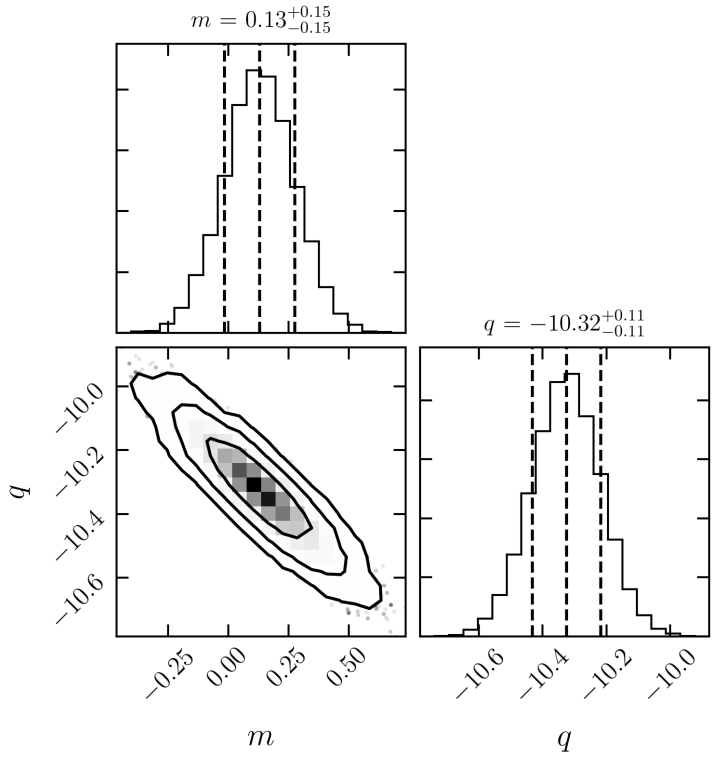

Figure 4.3: Example of one- and two-dimensional marginalised posterior distributions of the free parameters in our model. The black curves in the two-dimensional distribution correspond to regions enclosing 68%, 95%, and 99.7% of the total probability, while the black vertical lines in the one-dimensional marginalised distributions correspond to the 16th, 50th, and 84th percentiles, used to estimate the uncertainties on the models' parameters.

Figure 4.4 shows the linear correlations we find between the $\log_{10}X(o\text{-}H_2D^+)$ and the set of physical quantities of each clump listed in Tables 4.1 and 4.3. In each panel of Fig. 4.4 we report the best fit and the models within 3σ. According to the Jeffreys scale (Jeffreys 1961) if $\mathcal{B}_{2,1} < 1$, there is no statistical evidence to justify the use of a model with two free parameters (yellow shaded regions in Fig. 4.4 and Tab. 4.5). Alternatively, if $\mathcal{B}_{2,1} > 1$, a model with two free parameters is justified and we can consider the correlation reliable (green shaded regions in Fig. 4.4 and Tab. 4.5). We tested the same fit approach on the TOP100 sub-sample, always finding an agreement with the correlations shown in Fig. 4.4. This result excludes possible bias depending on the sample selection or on the procedures followed to derive the other values reported in Tables 4.1 and 4.3.

In the following paragraphs we discuss each group of correlations, paying particular attention to whether the dependence of each pair can be physically explained or if it is simply fortuitous.

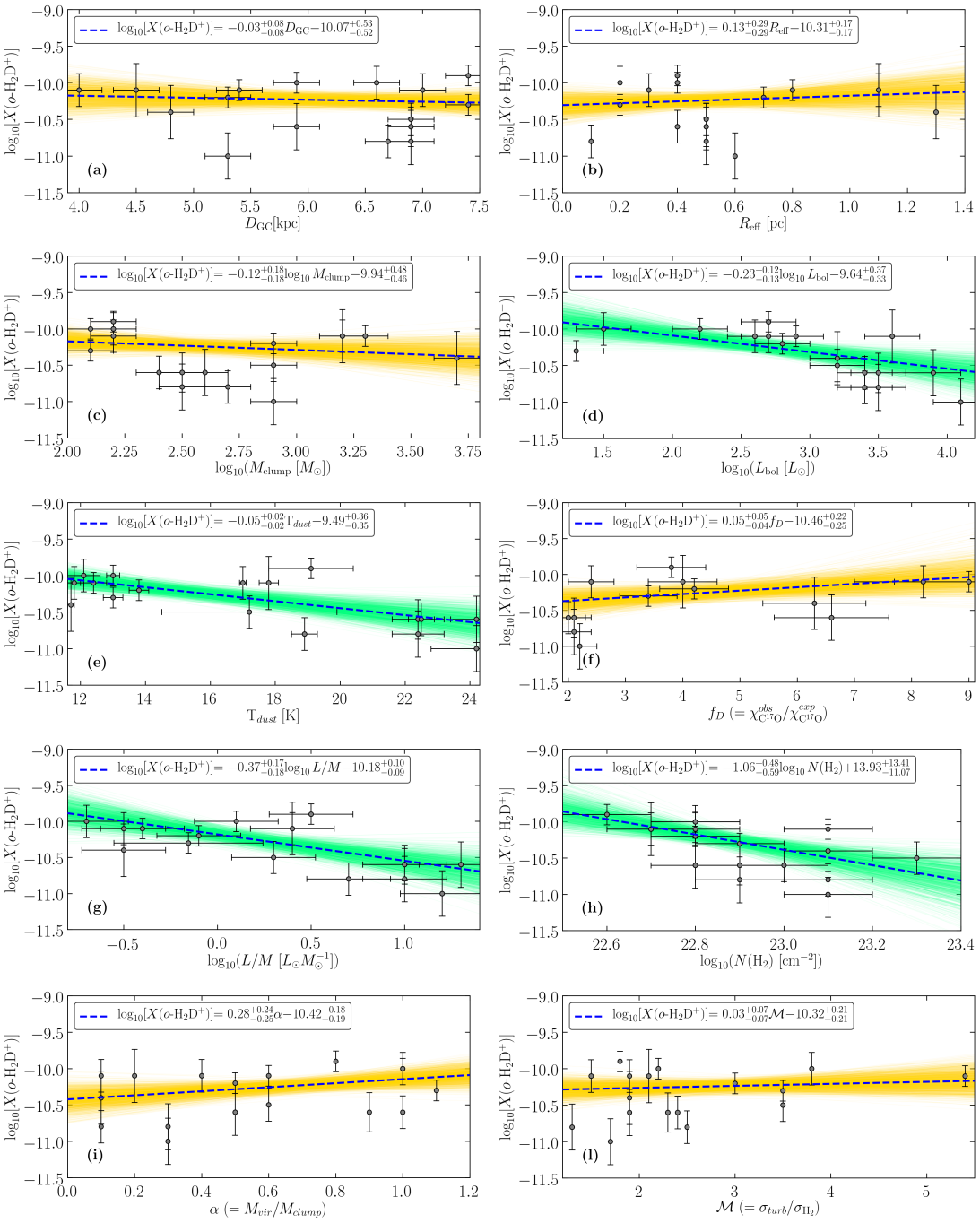

Figure 4.4: Collection of the different correlations between $X(o\text{-}H_2D^+)$ and the quantities summarised in Tables 4.1 and 4.3. Grey dots are associated with each source, while uncertainties are shown as black bars. Orange and green shaded regions are the 3σ results of the MCMC linear fit respectively for the unreliable ($\mathcal{B}_{2,1} < 1$) and reliable ($\mathcal{B}_{2,1} > 1$) correlations we found for the whole sample. Blue dashed lines represent the *fiducial model* (i.e. with the highest likelihood among the ones explored by MCMC). The fit parameters are shown in the caption of each correlation. All plots are in log-linear scale, except for panels (c), (d), (g), and (h) where the axes are set in log-log scale.

Panel	Correlation with	Bayes factors
(a)	D_{GC}	5.5×10^{-5}
(b)	$R_{\rm eff}$	1.0×10^{-3}
(c)	$\log_{10}(M_{\rm clump})$	1.0×10^{-3}
(d)	$\log_{10}(L_{\rm bol})$	9.7
(e)	$\log_{10}(T_{\rm d})$	4.5
(f)	$f_{\rm D}$	4.1×10^{-4}
(g)	$\log_{10}(L/M)$	51.1
(h)	$\log_{10}[N(H_2)]$	1.9×10^{4}
(i)	α	2.4×10^{-1}
(l)	$\mathcal{M}$	4.5×10^{-5}

Table 4.5: Bayes factors, $\mathcal{B}_{2,1}$, found for each correlation in Fig. 4.4. The correlations are computed with respect to the abundance of o-H_2D^+ derived from the column densities in Tables 4.1 and 4.3 (i.e. N[o-H_2D^+]/N[H_2]).

4.4.1 Source parameters

The dependence of $X(o\text{-}H_2D^+)$ on the physical properties of clumps does not seem particularly pronounced. In the first three plots in Fig. 4.4 (panels a, b, and c), $X(o\text{-}H_2D^+)$ appears roughly uniform as the other quantities vary. In these cases the correlations we found are in agreement with a purely flat regime, with small fluctuations of the fiducial models slopes, probably caused by regions with sparse density of observed points.

Variations in the deuterium abundance with respect to hydrogen, [D/H], through the Galactic disk are expected considering that all deuterium atoms were formed at the birth of the Universe and then progressively consumed by thermonuclear reactions within the stellar cores (e.g. Lubowich et al. 2000 and Ceccarelli et al. 2014; see also Galli & Palla 2013 for a review). Following this scenario, [D/H] should vary with location and the same would be expected for deuterated species. On the other hand, it is well known that the major boost in deuterium fractionation is driven by the star formation process itself: the temperatures, the global chemistry, and energetics of the stars can play a central role, (dis-)favouring the deuteration process (e.g. Caselli et al. 2002; Bacmann et al. 2003; Lis et al. 2006; Pillai et al. 2012; Ceccarelli et al. 2014 and Giannetti et al. 2019). The last dependence seems particularly evident in panel (d) of Fig. 4.4, which shows that $X(o\text{-}H_2D^+)$ clearly decreases as $L_{\rm bol}$ rises (i.e. $\mathcal{B}_{2,1} \gg 1$). The o-H_2D^+ abundance is found to change by ~ 0.5 orders of magnitude for a factor of ~ 4 change in $\log_{10}(L_{\rm bol})$, with a fiducial model's slope of ~ -0.23.

4.4.2 Evolutionary parameters

Among the quantities we defined as evolutionary tracers we have included the following: the CO-depletion factor, f_D, as the indicator of the depletion degree of each source, and defined as the ratio of the expected CO abundance with respect to H_2 to the observed value (see Caselli et al. 1999; Fontani et al. 2012); the dust temperature, which is determined by the activity of the protostars that are forming in each source; and L/M as the main indicator for the evolutionary stage of star formation processes in the high-mass regime (e.g. Saraceno et al. 1996; Molinari et al. 2008; Körtgen et al. 2017; Urquhart et al. 2018; Giannetti et al. 2019). The values for each clump are summarised in Tables 4.1 and 4.3. Depletion factors are derived from Giannetti et al. (2014) and Sabatini et al. (2019) (see also and Chapter 3) using the $C^{17}O$ column density distributions generated by taking into account optical depth effects, while T_d and L/M values are taken from Giannetti et al. (2017b), König et al. (2017), Urquhart et al. (2018), and Giannetti et al. (2019) to which we refer for a more detailed discussion of how these quantities were estimated. Depletion factors are also the only incomplete values due to the lack of observational data for the sources not in the TOP100. Therefore, we note that panel (f) of Fig. 4.4 is the only one obtained with a reduced number of data points (i.e. we could not include G12.50-0.22, G14.23–0.51, G15.72–0.59, G316.76-0.01).

In panel (e) of Fig. 4.4, $X(o\text{-}H_2D^+)$ correlates with T_d with small slopes, changing by ~ 0.5 orders of magnitude in the range of T_d between 12 and 24 K. The fit with the model with two free parameters is statistically supported by a $\mathcal{B}_{2,1} > 1$. However, we note that the scale chosen for the plot and the different ranges covered by the quantities involved may confuse the interpretation of these correlations, leading one to conclude that small slopes could mean soft correlation.

The correlation between $X(o\text{-}H_2D^+)$ and f_D in Fig. 4.4 (f) appears similar to that found for T_d (i.e. a clear correlation with a small slope). However, we find that $\mathcal{B}_{2,1} < 1$, which excludes any statistical evidence to justify the use of a model with two free parameters to establish a correlation between $X(o\text{-}H_2D^+)$ and f_D. We note in this case that the reduced number of sources may complicate the fit with two free parameters, while a correlation between $X(o\text{-}H_2D^+)$ and f_D is expected since the highly CO-depleted environments are those in which deuteration is favoured (e.g. Dalgarno & Lepp 1984 and Caselli & Ceccarelli 2012). The log-lin scale in Fig. 4.4 (e) and (f) was chosen to allow direct comparison with Caselli et al. (2008), where a survey of the $o\text{-}H_2D^+$ towards a sample of 16 low-mass sources - between starless and protostellar cores - was presented. This study gives us the opportunity to test the connection between low- and high-mass star-forming regions, with two samples statistically comparable. Caselli et al. (2008) found a progressively decreasing trend of $X(o\text{-}H_2D^+)$ at increasing temperatures. Here, we find the same behaviour even if less pronounced, but supported by an inverse correlation with f_D (even if not statistically confirmed), also in agreement with Vastel et al. (2006).

We also find a strong anti-correlation (i.e. $\mathcal{B}_{2,1} \gg 1$) of $X(o\text{-}H_2D^+)$ with L/M, reported

in Fig. 4.4 (g); $X(o\text{-}H_2D^+)$ is found to decrease by one order of magnitude for one order of magnitude change in L/M, with a slope of -0.37. This confirms the hypothesis outlined by Giannetti et al. (2019) that $o\text{-}H_2D^+$ can be considered a good chemical clock during the star formation process (see Sect. 4.6 for a detailed discussion).

4.4.3 Dynamical parameters

We analyse here the possible correlations between $X(o\text{-}H_2D^+)$ and the total H_2 column density, the virial parameter, and the Mach number, which we consider quantities related to the dynamical stage of the clumps. The correlation we find between $X(o\text{-}H_2D^+)$ and $N(H_2)$, shown in Fig. 4.4 (h), suggests that an increased density could also play a relevant role in the deuteration process. The $o\text{-}H_2D^+$ abundance is found to change by one order of magnitude for less than one order of magnitude change in $\log_{10}[N(H_2)]$, with a fiducial model slope of ~ -1.06. Frequent collisions are expected in denser environments (i.e. fast chemical kinetics) that would accelerate the formation of $o\text{-}H_2D^+$. On the other hand, when moving to a denser environment, the conversion of H_2D^+ in heavier isotopologues (D_2H^+ and D_3^+) or the deuterium transfer from H_2D^+ to other chemical species (e.g. N_2 and CO) can also be boosted, as suggested by Giannetti et al. (2019) .

The last two panels of Fig. 4.4, (i) and (l), present the correlations with α and $\mathcal{M}$ calculated from the $o\text{-}H_2D^+$ detected lines, as described in Sect. 4.3.2. Although the virial parameter and the Mach number depend on quantities for which a correlation with $X(o\text{-}H_2D^+)$ was established, such as the line FWHM, on which both M_{vir} and σ_v depend, or the dust temperature, we do not find any particularly relevant trend with $X(o\text{-}H_2D^+)$. We find that $X(o\text{-}H_2D^+)$ does not show a significant variation with α and $\mathcal{M}$, which means that the slopes are in agreement (within the uncertainties) with zero in these log-linear plots.

To summarise, we find strong correlations between $X(o\text{-}H_2D^+)$ and different physical quantities of the clumps, some expected and others less trivial. In Fig. 4.4 the correlations with $\mathcal{B}_{2,1} > 1$ are highlighted in green. To avoid any possible bias we repeated the MCMC fit approach looking for correlations with the $o\text{-}H_2D^+$ column densities, and we found that only the correlation between $N(o\text{-}H_2D^+)$ and $N(H_2)$ is not confirmed. All the other parameters in which a correlation is confirmed for both $X(o\text{-}H_2D^+)$ and $N(o\text{-}H_2D^+)$ are more or less directly connected to the evolution of the clumps. We conclude that the deuterium fractionation seems to be driven more by a temporal evolution than by a dynamical evolution of the star-forming regions, once the minimum physical conditions required to boost deuteration are fulfilled (i.e. cold, dense, and highly CO-depleted regions).

4.5 Estimates of the CRIR

Recently, a new way to estimate the cosmic-ray ionisation rate (CRIR) of hydrogen molecules, ζ_2, was presented by Bovino et al. (2020), based on H_2D^+ and other H_3^+ isotopologues. Our survey provides the opportunity to test, for the first time, this new method on a large sample of massive star-forming regions with simultaneous detections of o-H_2D^+ , DCO^+, $H^{13}CO^+$, and $C^{17}O$. According to Bovino et al. (2020) the CRIR can be obtained from

$$\zeta_2 = \bar{\alpha}\, k_{\rm CO}^{\rm H_3^+}\, \frac{\chi({\rm CO}) N({\rm H_3^+})}{\ell}, \tag{4.8}$$

where $\bar{\alpha} = 0.77$ is the correction factor that takes into account the errors due to approximations in the method (see Bovino et al., 2020, for details); $k_{\rm CO}^{\rm H_3^+}$ is the rate at which CO destroys H_3^+; ℓ is the path length over which the column densities are estimated; and χ(CO) is the relative abundance of CO with respect to H_2, which we derive by the simple formula:

$$\chi({\rm CO}) = \phi_{18/17} \frac{N({\rm C^{17}O})}{N({\rm H_2})} \frac{[^{16}{\rm O}]}{[^{18}{\rm O}]}\,. \tag{4.9}$$

Here $\phi_{18/17}$ is the isotopic ratio of $[^{18}O]/[^{17}O] = 4.16$ adopted from Wouterloot et al. (2008). For the TOP100 sources, the $C^{17}O$ column densities were taken from Giannetti et al. (2014), while for ATLASGAL sources not contained in the TOP100 we used the data described by Csengeri et al. (2016) (bold values in Table 4.6).

In Eq. 4.8, $N(H_3^+)$ is the column density of H_3^+ derived from the o-H_2D^+ and the HCO^+ deuterium fractionation, $R_{\rm D}$, that we calculate from the DCO^+ and $H^{13}CO^+$ column densities reported in Table 4.6. Column densities were derived consistently with the procedure employed for o-H_2D^+ (i.e. using MCWeeds and by considering how the $^{12}C/^{13}C$ ratio varies as a function of D_{GC}; see discussion in Giannetti et al. 2014): $[^{12}C]/[^{13}C] = 6.1^{+1.1}_{-1.8}\, D_{GC} + 14.3^{+7.7}_{-7.2}$. Since the real size of the region associated with the o-H_2D^+ emission is unknown, we assume three multiples of $R_{\rm eff}$ as representative cases to calculate ζ_2: (a) $\ell = 0.5R_{\rm eff}$, which approximately covers a range of ℓ comparable with the mean core size associated with the o-H_2D^+ emission reported by Pillai et al. (2012) (i.e. ~0.2 pc); (b) $\ell = R_{\rm eff}$ and (c) $\ell = 2R_{\rm eff}$, which implicitly means that the o-H_2D^+ is emitting on the full clump size defined at 870 μm.

Our estimates of ζ_2 are summarised in Table 4.6. In Fig. 4.5 we present these results arranged, from left to right, according to the evolutionary stage of each source. The blue line represents the mean value of $\zeta_2 = 2.5 \times 10^{-17} s^{-1}$ for $\ell = R_{\rm eff}$, while the blue shaded area is its variation with ℓ. It is worth noting that the range of values spans more or less an order of magnitude, which is in line with typical errors reported from models (see e.g. Padovani et al. 2018 and reference therein). We compare this result with recent estimates

ATLASGAL-ID	$N(C^{17}O)^{(a)}$	$\chi(CO)^{(a)}$	$N(H^{13}CO^+)$	$N(DCO^+)$	R_D	ζ_2^A $\ell = 0.5R_{\rm eff}$	$\zeta_2^{\rm ref}$ $\ell = R_{\rm eff}$	ζ_2^B $\ell = 2R_{\rm eff}$
	$\log_{10}(\rm cm^{-2})$	(10^{-5})	$\log_{10}(\rm cm^{-2})$	$\log_{10}(\rm cm^{-2})$		$(10^{-17}\ [\rm s^{-1}])$	$(10^{-17}\ [\rm s^{-1}])$	$(10^{-17}\ [\rm s^{-1}])$
G13.18+0.06	15.5	8.0	13.6±0.1	12.4±0.2	0.002±0.001	6.92	3.46	1.73
G14.11–0.57	15.7	11.5	13.8±0.1	13.6±0.1	0.011±0.004	3.18	1.59	0.80
G14.23–0.51	**15.7**	**4.6**	14.0±0.1	13.8±0.1	0.011±0.004	3.67	1.84	0.92
G14.49–0.14	15.6	4.7	13.3±0.1	13.0±0.1	0.011±0.003	6.68	3.34	1.67
G14.63–0.58	15.5	5.8	13.6±0.1	12.8±0.1	0.003±0.001	11.62	5.81	2.91
G15.72–0.59	**15.6**	**11.1**	13.2±0.1	13.2±0.1	0.018±0.006	3.06	1.53	0.77
G18.61–0.07	15.3	4.6	13.5±0.1	12.9±0.1	0.005±0.002	1.33	0.67	0.34
G19.88–0.54	15.7	5.8	13.8±0.1	13.2±0.1	0.005±0.002	3.00	1.50	0.75

Table 4.6: Summary of the quantities to calculate the CRIR. ζ_2 values are calculated following the recent results of Bovino et al. (2020) and assuming three values of ℓ in Eq. 4.8; $^{(a)}$ data and calculations from Giannetti et al. (2014). We assumed a constant $[^{18}O]/[^{17}O] = 4.16$, as in Wouterloot et al. (2008). Values in boldface were taken from Csengeri et al. (2016)

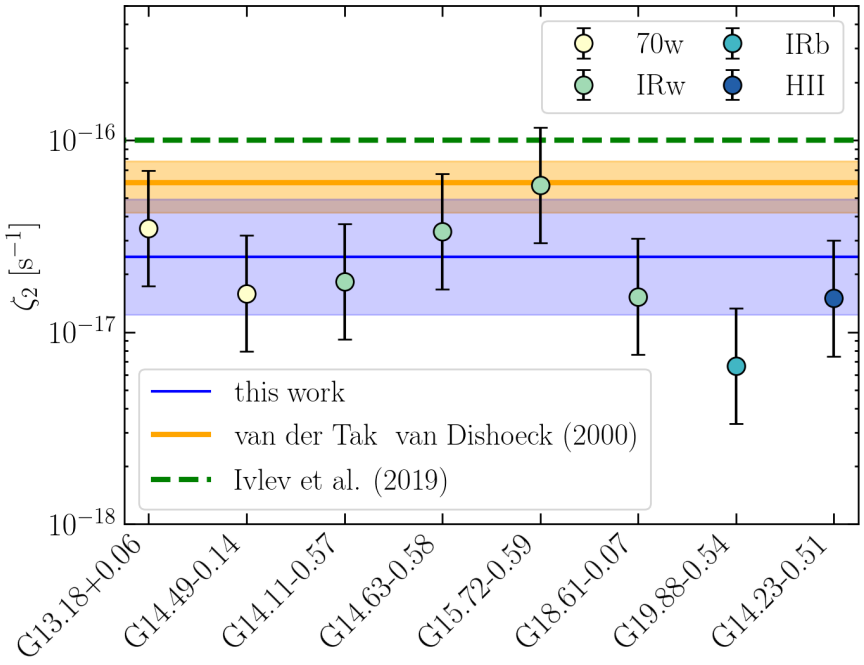

Figure 4.5: Estimates of the CRIR of hydrogen molecules, ζ_2, derived for a sub-sample of sources for which H_2D^+, DCO^+, $H^{13}CO^+$, and $C^{17}O$ observations are available, and assuming $\ell = R_{\rm eff}$ in Eq. 4.8, representing our reference case (dots). The sources are arranged from left to right following their evolutionary phase (bottom left legend). The bars associated with each point represent the variability of our results with ℓ (i.e. $\ell = 0.5R_{\rm eff}$ and $\ell = 2R_{\rm eff}$ for the upper and lower limit, respectively). The blue line is our mean value for ζ_2 for the reference case, while the blue shaded area is its variability with ℓ. The orange line and shaded area represents the estimates of van der Tak & van Dishoeck (2000), obtained by converting the CRIR for hydrogen atom, $\zeta_{\rm H}$, via $\zeta_2 = 2.3 \times \zeta_{\rm H}$. The green line is the recent upper limit for L1544 reported by Ivlev et al. (2019).

from Ivlev et al. (2019), who derive an upper limit of $\zeta_2 \sim 10^{-16} s^{-1}$ from a self-consistent model for the equilibrium gas temperature and size-dependent dust temperature in the prestellar core L1544. A second comparison is made with the estimate obtained by van der Tak & van Dishoeck (2000) in a sample of seven young massive stars using models based on $H^{13}CO^+$ observations. They found an average value of $\zeta_2 \sim (6.0 \pm 1.8) \times 10^{-17} s^{-1}$. Looking at our reference case, most of the sources show values lower than the average obtained by van der Tak & van Dishoeck (2000).

The analysis we report has the advantage of being model-independent, but it is purely qualitative, first because the method proposed in Bovino et al. (2020) includes strong approximations, and second because its validity has been shown for very small regions (i.e. $R = \ell/2 \leq 0.05$ pc, where R is the distance from the centre of the clump). However, as the error here is driven by the uncertainties in the column density calculations, we can consider the CRIR estimates valid and robust within the variability shown by the different ℓ. In addition, this represents the only viable and model-independent way to estimate the CRIR in dense regions.

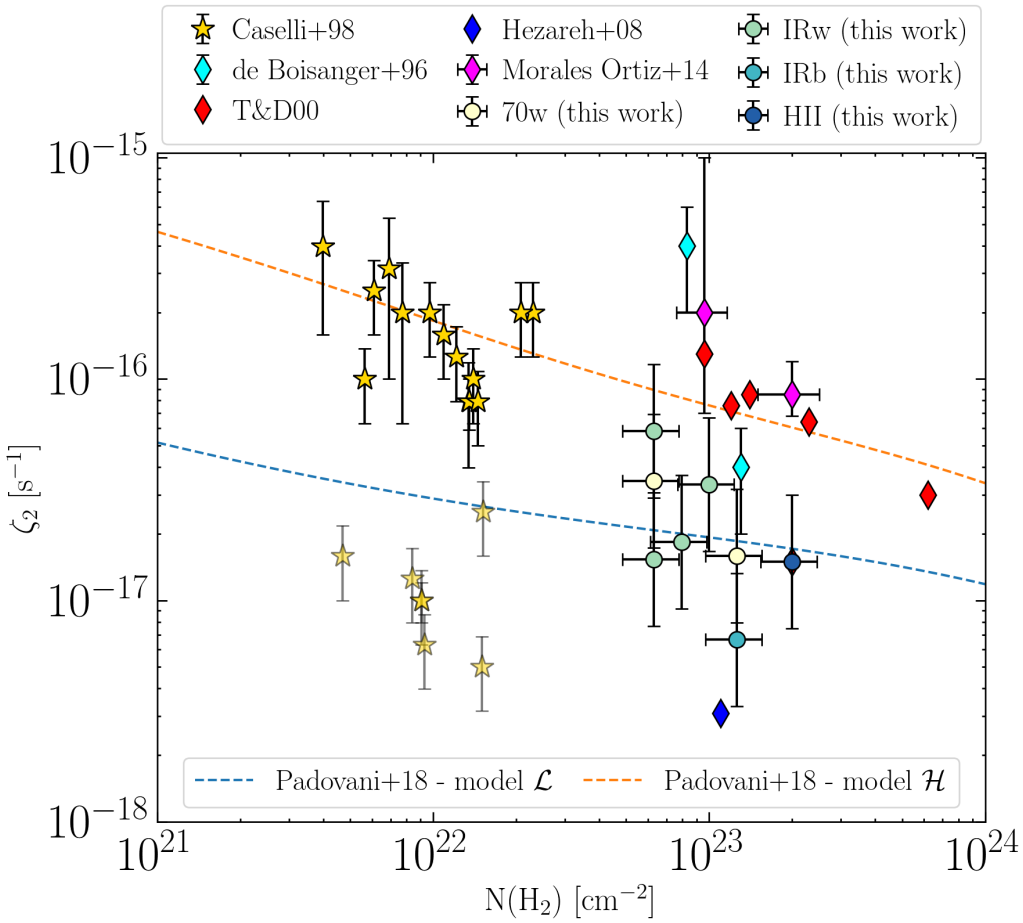

Figure 4.6: ζ_2 variation as a function of $N(H_2)$. Circles refer to our estimates of ζ_2 in high-mass star-forming regions in different evolutionary stages (same colour-coding as in Fig. 4.5). *Yellow stars* show the estimates of Caselli et al. (1998) for a large sample of low-mass cores, while *diamonds* represent high-mass cores (*cyan* from de Boisanger, Helmich & van Dishoeck 1996; *red* from van der Tak & van Dishoeck 2000 (T&D00) for ζ_2 estimates and from van der Tak et al. 2000 and Doty et al. 2002 for $N(H_2)$ values; *blue* from Hezareh et al. 2008; *magenta* from Morales Ortiz et al. 2014). Dashed lines show the models discussed in Padovani et al. (2018) assuming different slopes for the CR proton spectrum. Different levels of transparency have been used to separate what we considered outliers (pale color) from the other estimates (full colour). This plot was readapted from Padovani, Galli & Glassgold (2009).

Figure 4.6 shows how the cosmic-ray ionisation rate varies as a function of $N(H_2)$ for a sample of low- and high-mass star-forming regions. Considering the associated uncertainties, the data cover a range of ζ_2 between $\sim 3 \times 10^{-18}$ and $\sim 10^{-15}$ s^{-1}, while the molecular hydrogen column densities are in the range $\sim (0.4 - 80) \times 10^{22}$ cm^{-2}. The yellow stars are the ζ_2 values computed in low-mass cores from Caselli et al. (1998), while the values for $N(H_2)$ are taken from Butner, Lada & Loren (1995). Diamonds indicate massive protostellar envelopes (see the caption). The circles represent the estimates obtained in this chapter for the reference ℓ (same colour-coding as in Fig. 4.5). The dashed

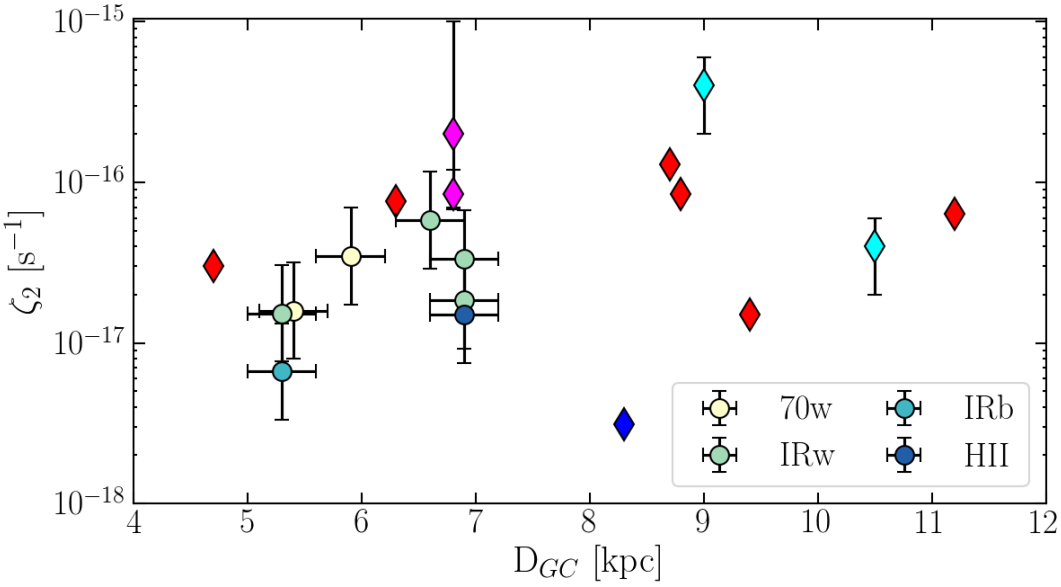

Figure 4.7: ζ_2 variation as a function of D_{GC} for the high-mass cores shown in Fig. 4.6. Circles refer to our estimates of ζ_2 in high-mass star-forming regions in different evolutionary stages (see legend). The same colour-coding as in Fig. 4.6 is assumed for diamonds. The galactocentric radii are taken from Table 4.1 (circles), van der Tak & van Dishoeck (2000) (*red*), Morales Ortiz et al. 2014 (*magenta*), Winkel et al. 2017 (*blue*), Spina et al. 2017, and Yan et al. 2019 (*cyan*). D_{CG} are scaled assuming a distance to the Galactic Centre of 8.35 kpc.

lines are the models discussed in Padovani et al. (2018), considering a single cosmic-ray (CR) electron spectrum and two different CR proton spectra, namely model $\mathcal{L}$ (in blue) and model $\mathcal{H}$ (in orange), and depending on the slope of the CR proton spectrum (see also Ivlev et al. 2015 for more details about the CR spectrum model).

The distribution of ζ_2 in Fig. 4.6, and in particular that of the full-coloured data points, seems to suggest the general trend where the CRIR decreases while the H_2 column density increases. We note that the estimates of ζ_2 in high-mass star-forming regions are smaller by a factor of $\sim 3-4$ with respect to their low-mass counterparts, which is in agreement with the fact that IRDCs are usually much denser compared to the low-mass regime. Our estimates fall between the two models of Padovani et al. (2018), while a small sub-sample of low-mass cores ($\sim$ 1/3 of the total sample) and a few high-mass star-forming regions find agreement with the model with lower CR proton energy. The apparent bimodality (yellow and pale yellow stars in Fig. 5) in the low-mass regime can be explained by the different f_D values assumed by Caselli et al. (1998) to estimate ζ_2. Here, by following Padovani, Galli & Glassgold (2009), we report the best ζ_2 estimates from Caselli et al. (1998), obtained from HC_3N/CO data assuming $f_D = 3-5$. In particular, values of ζ_2 around $\sim 10^{-16}$ s^{-1} are produced by both of the f_D values, while those around $\sim 10^{-17}$ s^{-1} correspond to $f_D = 3$. On the other hand, the morphology of the magnetic field lines can also play a major role in determining the CRIR. Padovani & Galli (2011) and Padovani,

Hennebelle & Galli (2013) have shown that even in high-density regions, if the magnetic field lines are very concentrated around the accreting protostar, ζ_2 can be reduced by up to a factor of 10 or more.

Finally, in Fig. 4.7 we report the cosmic ray ionisation rate as a function of the galactocentric distances for different high-mass star-forming clumps. We found that there is no significant variation of ζ_2 with D_{CG} for distances between 4.7 and 11.2 kpc. This result is in agreement with those of Indriolo et al. (2015), suggesting that the CRIR is uniform for sources located at distances above 5 kpc from the Galactic centre. Therefore, the observed spread in Fig. 4.6 could be the signature of local effects, for instance different magnetic field topology of each source.

4.6 o-H_2D^+ as an evolutionary tracer for massive clumps

Figure 4.8 shows the abundance of o-H_2D^+ as a function of the evolutionary class of the TOP100 clumps (red markers and error-bars). To mitigate the possible bias caused by extreme sources with peculiar initial conditions, and to make our result as general as possible, we report the median $X(o\text{-}H_2D^+)$ values for each evolutionary class. $X(o\text{-}H_2D^+)$ varies by more than one order of magnitude between the least evolved and the most advanced stage of evolution, suggesting that the deuterium fractionation of H_3^+ is favoured during the initial phases of star formation, while its chemical products, in this case o-H_2D^+, slowly disappear as massive clumps evolve. We repeated this procedure for the detection limits discussed in Sect. 4.3 to exclude possible sensitivity effects. The results are reported in Fig. 4.8 as grey markers. It is clear that the trend mentioned above holds: o-H_2D^+ decreases as the evolution progresses.

Between the two solutions, we found a discrepancy (always in agreement within the error bars) of at most a factor of 6, which is expected if we consider the different noise levels of the spectra. The trend found in Fig. 4.8 is a different manifestation of the result already shown in panel (g) of Fig. 4.4, where the evolution of the clumps is indicated by a progressive increase of the luminosity-to-mass ratio (e.g. Saraceno et al. 1996; Molinari et al. 2008) and $X(o\text{-}H_2D^+)$ shows the same decreasing behaviour passing from the lowest to the highest L/M values (i.e. from less to more evolved sources). The same result also emerges from Fig. 4.4 (d) since the bolometric luminosity traces the evolution of clumps if we assume that during the star formation process the clump mass stays constant. This latter assumption seems reliable as $X(o\text{-}H_2D^+)$ correlates with all the other evolutionary indicators, while no correlation was found between $X(o\text{-}H_2D^+)$ and $M_{\rm clump}$ (Fig. 4.4c), in agreement with the results of König et al. (2017) and Urquhart et al. (2018), who find no trend in the clump mass between the evolutionary classes.

We compared the results found in Fig. 4.8 with those of Giannetti et al. (2019), where the same clear downward trend was observed in the $X(o\text{-}H_2D^+)$ of three massive clumps associated with different evolutionary stages and harboured in the IRDC G351.77–0.51.

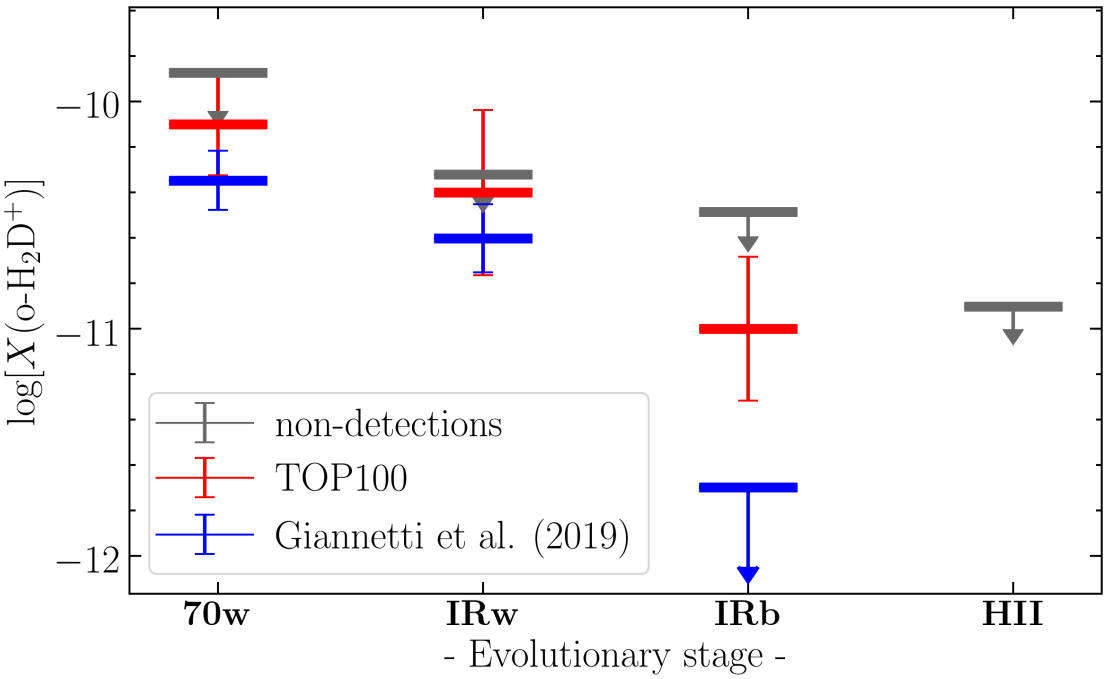

Figure 4.8: Average $X(o\text{-}H_2D^+)$ abundance as a function of the evolutionary classes in our sample. Blue markers indicate the recent estimates of Giannetti et al. (2019), while the red and grey ones indicate the mean of detection and the median $X(o\text{-}H_2D^+)$ derived from detection limits, respectively. Uncertainties are derived as mean and median as well, considering the errors on the individual source.

Although less pronounced, similar results were recently reported by Miettinen (2020) for three prestellar and three protostellar low-mass cores in Orion-B9. In the less evolved sources with detection (i.e. 70w and IRw), the agreement with Giannetti et al. (2019) is within the associated error limits, while comparing our IRb detection with the upper limit provided by Giannetti et al. (2019), we found $X(o\text{-}H_2D^+)$ higher by a factor of ∼5, which might depend on different initial conditions (both physical and chemical). These results confirm that through the evolutionary sequence of massive star-forming regions a general downward trend for $o\text{-}H_2D^+$ is clearly observable with upper limits associated with the HII class. The same trend is also visible if the average abundances reported by Miettinen (2020) are considered.

Figure 4.8 suggests that the observed $X(o\text{-}H_2D^+)$ trend is not influenced by the nature of the samples as a similar behaviour has been reported both within the same star formation complex (Giannetti et al., 2019) and within a heterogeneous sample (Sabatini et al. 2020). Our new confirmation of the same trend is completely independent from the initial chemical conditions of the ISM in which the star-forming complexes were formed and from any particular or stochastic episodes during the star formation process (i.e. possible outflows or different mass accretion rates).

4.6.1 o-H_2D^+ in a broader scenario of star formation

According to Urquhart et al. (2018), each evolutionary class of the ATLASGAL survey shows a clear separation in temperature that increases from ∼10 to ∼40 K, in line with the expected evolutionary sequence. The same behaviour was also noted in the TOP100, a further confirmation of the statistical relevance of this flux-limited sub-sample as representative of the whole ATLASGAL (e.g. König et al. 2017). Urquhart et al. (2018) also pointed to positive strong correlations that link the gas temperature of the clumps to their embedded massive protostar luminosity or to the luminosity-to-mass ratio of each clump. Our results resemble a similar evolutionary behaviour in the $X(o\text{-}H_2D^+)$ -T_d correlation reported in panel (e) of Fig. 4.4, where it is visible that higher values of $X(o\text{-}H_2D^+)$ are reached in colder environments, while as T_d increases, $X(o\text{-}H_2D^+)$ become progressively lower. The temperature is also directly responsible for the evolution of CO in these regions. Low concentrations of CO molecules in the gas phase have been observed in cold environments (e.g. Roberts & Millar 2000; Bacmann et al. 2003; Ceccarelli et al. 2014), and correlations of the CO abundance with the evolutionary stages have been reported in the low-mass regime (e.g. Caselli et al. 1998; Bacmann et al. 2002). The same behaviour was also confirmed in high-mass star-forming regions, where f_D was found to decrease with the increase of L/M, as the indirect confirmation that the clumps become warmer with time (see Giannetti et al. 2014, 2017b).
In highly CO-depleted environments the formation of H_2D^+ is made more efficient by two events. On the one hand, the H_3^+ reservoir is kept available for reaction (2.5) as the following proton-transfer reaction is slowed down (see also Sect. 2.2):

$$H_3^+ + CO \rightarrow HCO^+ + H_2. \tag{4.10}$$

On the other hand, due to the same effect, H_2D^+ is prevented from reacting with CO and, to a lesser extent, with N_2 (which normally depletes slowly) to form DCO^+ and N_2D^+, remaining abundant and observable in gas phase. The f_D-$X(o\text{-}H_2D^+)$ correlation shown in Fig. 4.4 (f), qualitatively supports this scenario. We note that in this case the highest $X(o\text{-}H_2D^+)$ correspond to the highest f_D values, with an opposite trend respect to what is found for T_d. This anti-correlation is expected as cold (and dense) environments are those in which CO-depletion is boosted (e.g. Kramer et al. 1999; Caselli et al. 1999; Crapsi et al. 2005; Fontani et al. 2012; Wiles et al. 2016; Sabatini et al. 2019).

It is also clear, from the obtained results, that the answer to how (and if) the observed $X(o\text{-}H_2D^+)$ are linked to the clumps' dynamical picture across their evolution is far from being exhaustive. In agreement with the results of König et al. (2017), we do not find any evident trend of α and $\mathcal{M}$ with the variation of the evolutionary stage in the TOP100 sources. We only note that for the entire sample α is found to be $\lesssim 1$, which indicates dynamically unstable sources and very fast collapses, if magnetic fields are negligible (see Kauffmann, Pillai & Goldsmith 2013 for a more detailed description). Interestingly,

similar results have been obtained by numerical simulations (Körtgen et al., 2017, 2018) in which the influence of dynamical parameters like the Mach number ($\mathcal{M}$), the magnetic field magnitude and distribution, and the cloud gas surface density have been shown to affect the results much less than the time evolution of the clumps. This prevents us from determining whether the correlations found in panels (i) and (l) of Fig. 4.4 support the scenario in which $X(o\text{-}H_2D^+)$ and the clumps' dynamical quantities are not empirically correlated, and the possibility that our sample may be affected by some type of selection bias. However, the latter is not supported by the other correlations we report, while it is more likely that the link between the emitting region associated with $o\text{-}H_2D^+$ and the clump dynamics is not so trivial. In addition, we should also consider that the H_2D^+ velocity dispersion cannot give us full information on the clump dynamics because close to the centre of the clumps it is quickly replaced by N_2D^+ or D_2H^+ or destroyed by the presence of freshly desorbed CO. High spectral resolution observations are necessary in order to estimate with sufficient accuracy the line FWHMs, and consequently α and $\mathcal{M}$, while high spatial resolution observations (< 0.2 pc; e.g. Pillai et al. 2012) are needed to be able to distinguish the dynamics of the sub-clumps associated with the $o\text{-}H_2D^+$ emission.

Overall, our results suggest that the $o\text{-}H_2D^+$ evolution follows the physical conditions associated with the different evolutionary stages: high abundances in the cold starless or prestellar stages (70w and IRw), and a clear decrease while the evolution of the protostellar object proceeds from its young phases to more evolved situations, including HII regions. This is in particular confirmed by the correlations reported in Fig. 4.4, which support the scenario in which deuteration proceeds faster in the first stages favoured by a high degree of CO freeze-out and drops at later stages mainly due to the increase in temperature induced by the presence of a luminous YSO and the subsequent release of CO back into gas-phase.

4.7 Chapter conclusions

The results of this chapter can be summarised as follows:

1. We confirm, also in the high-mass regime of star formation, the empirical correlation between $X(o\text{-}H_2D^+)$ and T_d, and we find new correlations with L_{bol} and L/M. $X(o\text{-}H_2D^+)$ has been found to correlate also with $N(H_2)$, but we interpret this result as less relevant for the H_3^+ deuterium fractionation process since it is not supported by an evident correlation with $N(o\text{-}H_2D^+)$, which instead has been found with f_D and with all the other quantities connected to the evolutionary stage of the clump(s). It is, however, likely that a denser clump might have already enabled the conversion of $o\text{-}H_2D^+$ into other deuterated species (e.g. D_2H^+ and N_2D^+) due to fast kinetics, with the effect of reducing the abundance of $o\text{-}H_2D^+$.

2. From the additional detections of DCO^+, $H^{13}CO^+$, and $C^{17}O$ available in our sam-

ple, we have added six new estimates of ζ_2 in massive star-forming regions by following the analytical formulae recently reported by Bovino et al. (2020). We find a variation in the estimated ζ_2 with the H_2 column density. Nevertheless, while being connected to the morphology of the magnetic field lines of each source, as discussed by Padovani et al. (2018), we did not find any signature that the ζ_2 can be interpreted as an evolutionary indicator of the star formation activity. We estimate a mean $\zeta_2 = 2.5 \times 10^{-17}$ s^{-1}, assuming that the region associated with the o-H_2D^+ emission is comparable to the clump's effective radius in Table 4.1.

3. We confirm that $X(o\text{-}H_2D^+)$ shows a general downward trend with massive clumps evolution, as pointed out by Giannetti et al. (2019). We extend this trend to the most advanced phases of HII regions through upper limit estimations. This new result, together with that of Giannetti et al. (2019), establish the role of this tracer as a chemical clock, and provides a useful reference for future observations of o-H_2D^+ in massive star-forming regions.

Chapter 5

The late stages of the high-mass star formation process

This chapter is based on Sabatini et al. (2021, subm.) to A&A

Abstract of the Chapter

In this chapter, we present a method to derive accurate timescales of the different evolutionary phases of the high-mass star formation process. We model a representative number of massive clumps of the ATLASGAL-TOP100 *sample which cover all the evolutionary stages. The models describe an isothermal collapse and the subsequent warm-up phase, for which we follow their chemical evolution. The timescale of each phase is derived by comparing the results of the models with the properties of the sources of the* TOP100 *sample, taking into account the mass and luminosity of the clumps, and the column densities of methyl acetylene* (CH_3CCH)*, acetonitrile* (CH_3CN)*, formaldehyde* (H_2CO) *and methanol* (CH_3OH)*. We find that the chosen molecular tracers are affected by the thermal evolution of the clumps, showing steep ice evaporation gradients from distances to the YSO of* 10^3 *to* 10^5 AU *during the warm-up phase. We succeed in reproducing the observed column densities of* CH_3CCH *and* CH_3CN*, while* H_2CO *and* CH_3OH *show less agreement with the observed values. The total (massive) star formation time is found to be* $\sim 5.2\times10^5$ yr*, which is defined by the timescales of the individual evolutionary phases:* $\sim 5 \times 10^4$ yr *for* 70w, $\sim 1.2 \times 10^5$ yr *for* IRw, $\sim 2.4 \times 10^5$ yr *for* IRb *and* $\sim 1.1 \times 10^5$ yr *for* HII.

5.1 Physical model

To estimate the timescales of the evolutionary stages defined in the TOP100 sample, we have developed a two-phase physical model (see sketch in Fig. 5.1). The first, namely the collapse (left panel of Fig. 5.1), describes the density evolution of an isothermal clump with a single-zone approximation, exploring different conditions as in e.g. Viti & Williams (1999) and Garrod & Herbst (2006). In the second phase, the warm-up (central panel of Fig. 5.1), 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; Ercolano, Barlow & Storey 2005), while 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.

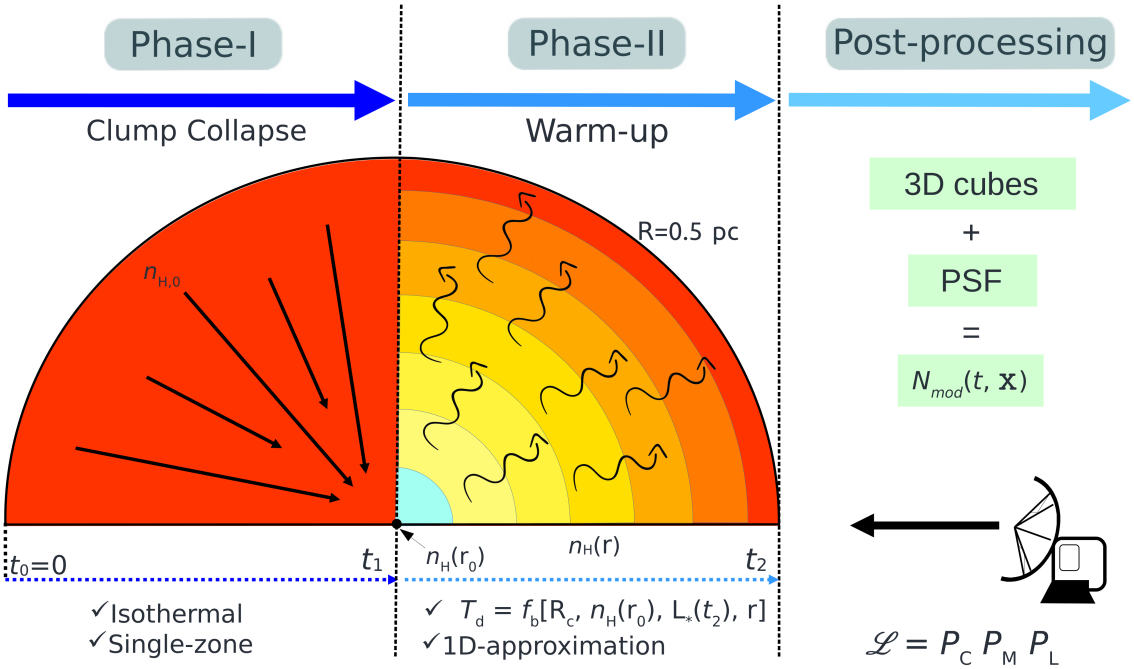

Figure 5.1: Sketch of the physical model employed in this work. *Left panel:* collapse phase (see Sect. 5.1.1) solved in a single-zone approximation. The chemical output of Phase-I is used as initial condition of the Phase-II. *Central panel:* warm-up phase solved in a 1D-approximation from 1 to 10^5 AU (see Sect. 5.1.2). *Right panel:* post-processing applied to compare the final chemical outputs of Phase-II with the column densities observed in the TOP100 (see Sections 5.3.1 and 5.3.3). The physical parameters assumed in Phase-I and II are summarised in Tab. 5.1. The time evolution of the model is indicated by the blue dotted arrows.

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

Table 5.1: Parameters space explored in our models; i.e. Phase-I (Sect. 5.1.1) and Phase-II (Sect. 5.1.2).

5.1.1 Phase-I: isothermal collapse

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

$$\frac{{\rm d}n_{\rm H}(t)}{{\rm d}t} = b \cdot \left(\frac{n_{\rm H}(t)^4}{n_{\rm H,0}}\right)^{1/3} \left\{24\pi G m_{\rm H}\, n_{\rm H,0} \left[\left(\frac{n_{\rm H}(t)}{n_{\rm H,0}}\right)^{1/3} - 1\right]\right\}^{1/2}, \tag{5.1}$$

where $n_{\rm H,0}$ is the initial central gas number density, b is a factor which, in our case, aims to mimic a slower collapse compared to the ideal free-fall time (i.e. $b = 1$), G is the gravitational constant and $m_{\rm H}$ the hydrogen mass. Eq. 5.1 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, $n_{\rm H,0} = n({\rm H}) + 2n({\rm H_2}) = 10^4$ cm^{-3} (Bergin & Tafalla, 2007; Gerner et al., 2014). The collapse assumes a constant temperature of 15 K (both for gas and dust temperatures; e.g. König et al. 2017) and a visual extinction $A_{\rm v} = 10$ mag (e.g. Semenov et al. 2010; Reboussin et al. 2014). The cosmic ray ionisation rate of hydrogen molecules is set to $\zeta_2 = 5 \times 10^{-17}$ s^{-1}, as observationally constrained by van der Tak & van Dishoeck (2000) in high-mass star-forming regions, and also in agreement with the results of Section 4.5, reported by Sabatini et al. (2020). The specific density of the dust grains is $\rho_0 = 3$ g cm^{-3}, typical of silicates (e.g. Draine & Lee 1984), the dust-to-gas ratio $\mathcal{D} = 10^{-2}$, and we use a constant grain size $\langle a \rangle = 0.1$ μm. The gas mean molecular weight is $\mu = 2.4$.

To simulate various physical conditions for the environment where the seeds of massive protostars are formed, the collapse is stopped at different final densities $n_{\rm H}(r_0)$ (see Fig. 5.1), observationally constrained (Tab. 5.1; e.g. Mueller et al. 2002; Sabatini et al. 2019 and Section 3.4) and then used as central densities during the warm-up phase.

5.1.2 Phase-II: the warm-up phase

The second phase of the model, sketched in the central panel of Fig. 5.1, simulates the warm-up induced by a protostar at the center of a spherical clump by using a 1D-approximation of 100 logarithmic radial steps from 1 to 10^5 AU (i.e. up to ~0.5 pc; e.g. Motte, Bontemps & Louvet 2018).

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

$$n_{\rm H}(r) = n_{\rm H}(r_0)\frac{R_{\rm c}^{5/2}}{R_{\rm c}^{5/2} + r^{5/2}}\,, \tag{5.2}$$

where $n_{\rm H}(r_0)$ is the central number density at r_0, at which the collapse phase is stopped (see second row in Tab. 5.1). This ensures the continuity between the two physical phases of the model. In Eq. 5.2, $R_{\rm c}$ is the core radius, i.e. the radius at which $n_{\rm H}(R_{\rm c}) = 0.5\, n_{\rm H}(r_0)$, and the slope 5/2 is in agreement with the observations in massive star-forming regions (e.g. Mueller et al. 2002; Schneider et al. 2015).

The physical parameters of Phase-II (i.e. ζ_2, $\mathcal{D}$, μ, and $\langle a \rangle$) are assumed the same as those of the collapse phase, while the visual extinction at each radius r is defined as (e.g. Tielens, 2010; Zhu et al., 2017)

$$A_{\rm v}(r) = \frac{N({\rm H}, r)}{(2 \times 10^{21} {\rm cm}^{-2})}\,, \tag{5.3}$$

where $N({\rm H}, r)$ is the column density of the hydrogen nuclei, obtained integrating Eq. 5.2 from the edge of the clump. We take the final abundances of Phase-I as initial chemical conditions for each radial grid point of the Phase-II rescaled by the 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 generate 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 $T_{\rm d}(r, t_2) = f_b[R_c, n_{\rm H}(r_0), L_*(t_2), r]$, where $L_*(t_2)$ 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 $L_*(t_2)$ evolves in time is derived by Hosokawa & Omukai (2009) and depends on the assumed protostar's mass accretion rate, $\dot{M}$. Based on the very few observational estimates (e.g. Beuther et al. 2002b; Herpin et al. 2012; Wang et al. 2013) we explore two typical cases, i.e. $\dot{M} = 10^{-5}$ M$_\odot$ yr^{-1} and $\dot{M} = 10^{-3}$ M$_\odot$ yr^{-1} (see Fig. 5.2a, rearranged from Hosokawa & Omukai 2009). These values are also in agreement with the results of Peters et al. (2011), that simulated the collapse of a magnetised rotating molecular cloud of 1000 M$_\odot$.

During Phase-II, we assume that the gas radial density profile of the clump and its dust content remain constant over time so that the mass accretion rate of the protostar is com-

pletely 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).

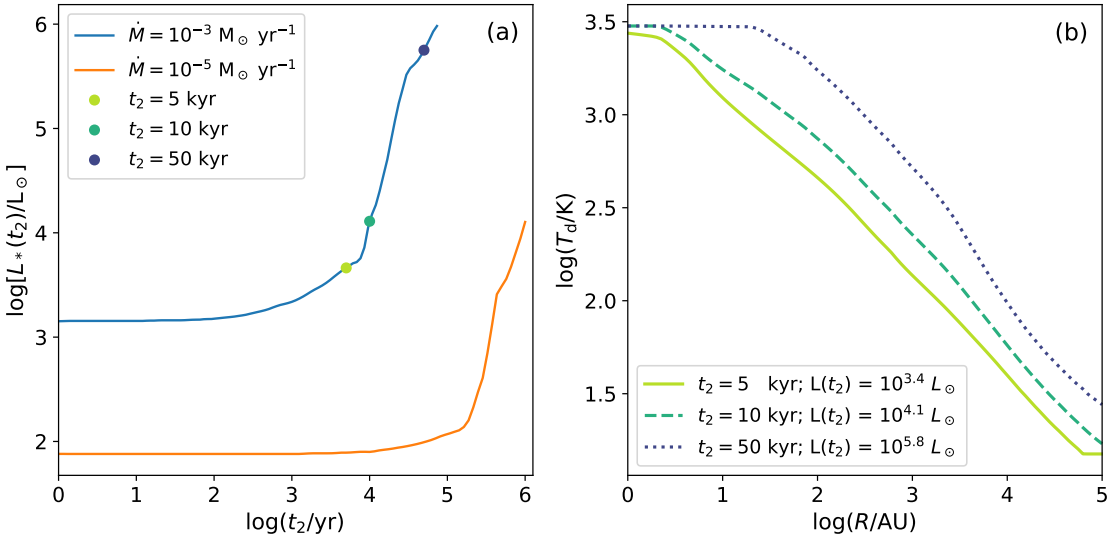

Figure 5.2: Panel (a): $L_*(t_2)$ function employed in our models and based on Hosokawa & Omukai (2009), assuming two typical cases of mass accretion rate, i.e. $\dot{M} = 10^{-5}$ M$_\odot$ yr^{-1} and $\dot{M} = 10^{-3}$ M$_\odot$ yr^{-1}. Figure rearranged from Figures 4 and 12 of Hosokawa & Omukai (2009); Panel (b): Dust temperature radial profile at different times, for an arbitrary model with $\dot{M} = 10^{-3}$ M$_\odot$ yr^{-1} and gas number density profile described by Eq. 5.2 with $R_c = 10^{4.5}$ AU and $n_{\rm H}(r_0) = 10^7$ cm^{-3}. With time, the $T_{\rm d}$ increases driven by the protostar's 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.

Analogous to Phase-I, we assume that the gas and dust temperatures are 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, both these regions are not relevant in our analysis and do not affect our findings. An example of how the $T_{\rm d}$ radial profile evolves is shown in Fig. 5.2b, for three (typical) times, $\dot{M} = 10^{-3}$ M$_\odot$ yr^{-1} and for the following parameters $R_c = 10^{4.5}$ AU and $n_{\rm H}(r_0) = 10^7$ cm^{-3}.

To compare our results with the observed properties of the TOP100 clumps (Giannetti et al., 2014; König et al., 2017), we evolve one model for each mass in Tab. 5.2 (derived from $n_{\rm H}(r_0)$ and $R_{\rm c}$ in Tab. 5.1), varying the values of b and $\dot{M}$ in Tab. 5.1. We consider only clumps with a total mass larger than 15 M$_\odot$, that is the smallest mass associated with a TOP100 source, for a total of 54 models.

$n_{\rm H}(r_0)$	$R_{\rm c}$/AU		
[cm^{-3}]	10^4	$10^{4.5}$	10^5
10^5	**0.4** M$_\odot$	**3.8** M$_\odot$	18.5 M$_\odot$
10^6	**3.5** M$_\odot$	38.4 M$_\odot$	184.6 M$_\odot$
10^7	34.7 M$_\odot$	383.7 M$_\odot$	1846.8 M$_\odot$
10^8	346.6 M$_\odot$	3836.7 M$_\odot$	18468.2 M$_\odot$

Table 5.2: Summary of the total masses of each model. The models with masses smaller than the lowest mass associated with the TOP100 sources (König et al., 2017), and marked in boldface, are excluded.

5.2 Chemical model

The chemical evolution of the clumps in the two physical phases is described by employing the publicly available time-dependent code KROME[1] (Grassi et al., 2014).

Adsorption and desorption processes are included as in Hasegawa, Herbst & Leung (1992) and Hasegawa & Herbst (1993). The depletion (or adsorption) rate coefficient, in units of s^{-1}, that involves an ith chemical species is described by

$$k_i^{\rm ads} = \mathcal{S}\sigma_{\rm d}\upsilon_i n_{\rm dust}\,, \tag{5.4}$$

where $\mathcal{S}$ is the sticking coefficient of the gas species on a dust particle, here assumed equal to unity, $\sigma_{\rm d} = \pi\langle a\rangle^2$ the grain's cross-section, $\upsilon_i = (8k_B T_{\rm g}/\pi m_i)^{1/2}$ the thermal velocity of the species with mass m_i and temperature $T_{\rm g}$, where k_B is the Boltzmann constant. The dust number density is defined as $n_{\rm d} = \mathcal{D} n_{\rm H} m_{\rm H}\mu/M_{\rm d}$, where $M_{\rm d} = 4/3\pi\rho_0\,\langle a\rangle^3$ is the dust mass.

The adsorbed species can return into the gas-phase through the *thermal* and *cosmic-ray* induced desorption. The chemical evaporation (namely, the desorption of the products of an exothermic surface reaction) has not been included because of the large uncertainties on the probability associated to this process and the long timescale needed to take place (e.g. Garrod, Wakelam & Herbst 2007). The total desorption rate coefficient, in units of s^{-1}, is

$$k_i^{des} = k_i^{des,th} + k_i^{des,cr}\,, \tag{5.5}$$

where the rate coefficients for thermal and CR-induced desorption, $k_i^{des,th}$ and $k_i^{des,cr}$, are

$$k_i^{des,th} = \nu_0^i \exp\left(-\frac{T_{\rm b}^i}{T_{\rm d}}\right), \tag{5.6}$$

and

[1] https://bitbucket.org/tgrassi/krome/wiki/Home

$$k_i^{des,cr} = f(70\,\mathrm{K}) k_i^{des,th}(70\,\mathrm{K})\,, \tag{5.7}$$

respectively.

In Eq. 5.6, T_d is the dust temperature, $\nu_0^i = (2 n_S E_b^i / \pi^2 m_i)^{1/2}$ the characteristic Debye vibration frequency for the ith adsorbed species, $n_S = 1.5 \times 10^{15} \mathrm{cm}^{-2}$ the surface density of binding sites, and $E_b^i = k_B T_\mathrm{b}^i$ the binding energy of the ith species on the binding site. In Eq. 5.7 $f(70\mathrm{K})$ is the fraction of time spent by grains at 70 K, equal to $\sim 3.16 \times 10^{-19}$ for a CRs ionisation rate of hydrogen molecules $\zeta_2 = 1.3 \times 10^{-17}$ s^{-1}, and $k_i^{des,th}(70\,\mathrm{K})$ the rate coefficient of thermal desorption at $T_\mathrm{d} = 70$ K.

Surface reactions between two species i and j are treated via Langmuir-Hinshelwood mechanism (see Sect. 2.2.3) and follows Semenov et al. (2010), where the rate coefficient, in units of cm^3 s^{-1}, is

$$k_{ij}^s = \frac{P_{ij}\,(k_\mathrm{diff}^i + k_\mathrm{diff}^j)}{N_\mathrm{sites} n_\mathrm{d}}\,, \tag{5.8}$$

with $k_\mathrm{diff}^i = \nu_0^i \exp(-T_\mathrm{diff}^i / T_\mathrm{d})$ the diffusion via thermal hopping ($T_\mathrm{diff}^i = 0.77 T_\mathrm{b}^i$; Semenov et al. 2010), while the total number of binding sites of a grain $N_{sites} = 4\pi \langle a \rangle^2 / a_{pp}^2$ assumes an average distance between two contiguous sites of $a_{pp} = 3$ Å (Hocuk & Cazaux 2015).

The probability for a reaction to occur is

$$P_{ij} = \alpha_{ij} \exp\left(\frac{-E_\mathrm{a}}{k_B T_\mathrm{d}}\right)\,, \tag{5.9}$$

where E_a is the activation energy of the reaction and α_{ij} a parameter that depends on the number and on the type of species in the reaction. In the case of exothermic reactions (i.e. $E_\mathrm{a} = 0$): (I) if $i \neq j$, $P_{ij} = 1$; (II) if $i = j$, $P_{ij} = \alpha_{ij} = 1/2$. For endothermic reactions (i.e. $E_\mathrm{a} \neq 0$), α_{ij} is the inverse of the number of paths in the branching ratios.

The final chemical network is derived from Semenov et al. (2010)[2], and contains 654 chemical species and 5869 gas-phase, gas-grain and grain surface reactions. The photochemical rate coefficients follow the A_v formalism as in Draine (1978) (see KIDA[3]). In order to ensure that each depleted species is 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 Tab. A.1 and Tab. A.2 for the reactions, and Appendix A.2 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 a few $\times 10^3$ cm^{-3}. Following these findings we assume H, C, and O to be in molecular form. In particular the abundances of H_2, H_3^+, He, N, O, CO, and N_2 are taken from

[2]Ohio StateUniversity (OSU) chemical network, version March 2008 (http://www.mpia.de/homes/semenov/Chemistry_benchmark/model.html)

[3]http://kida.astrophy.u-bordeaux.fr/

species	$(n_i/n_{\rm H})_{t=0}$	species	$(n_i/n_{\rm H})_{t=0}$
H_2	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_2	5.25(-6)	Cl	1.40(-8)
Mg	2.55(-6)	H_3^+	3.18(-9)

Table 5.3: Summary of fiducial initial chemical abundances, n_i, with respect to the abundance of H-nuclei, $n_{\rm H}$. The "A(-B)" notation assumed in the table, means "A$\times 10^{-\rm B}$".

Bovino et al. (2019), while for the other elements we refer to Garrod & Herbst (2006), as reported in Table 5.3. The remaining species are initialised to 10^{-40} $n_{\rm H}$, assuming total charge neutrality. We set the relative and absolute tolerances of the solver to `RTOL` $= 10^{-6}$ and `ATOL` $= 10^{-40}$, respectively.

5.3 Results

Fig. 5.3 shows the evolution of the fractional abundances of the chemical clocks selected from the TOP100 survey as a function of the central density of the clump during the isothermal collapse.

We first test the effect of the collapse velocity (parameter b in Eq. 5.1) on the fractional abundances evolution, delaying the collapse by 50% and the 90%, i.e. $b = 0.5$ and $b = 0.1$, respectively (colours in Fig. 5.3). Despite the effects induced by varying b, the relative importance of the chemical reactions involving some of the main tracers remains unaltered. In particular, H_2CO is mainly formed via $CH_3 + O \rightarrow H_2CO + H$, and it is destroyed by ion-molecule reactions which involve H^+, C^+ and S^+. For densities higher than $\sim 10^5$ cm^{-3}, ion-molecule reactions become predominant and the abundance of H_2CO decreases. At higher densities, some relative abundance profiles return to increase since the dominant reaction becomes $H_3CO^+ + e^- \rightarrow H_2CO + H$.

Methyl acetylene is formed by the dissociative recombination:
(*i*) $C_3H_5^+ + e^- \rightarrow CH_3CCH + H$, the main gas-phase source of CH_3CCH (see Hickson, Wakelam & Loison 2016). For densities above $\sim 10^6$ cm^{-3} the relative abundance is almost constant, since CH_3CCH reacts with C with a rate coefficient similar to (*i*).

Conversely, the abundance of acetonitrile is more sensitive to b. The main source of gas-phase CH_3CN in our model is (*ii*) $HCN + CH_3^+ \rightarrow CH_3CNH^+$, followed by the dissociative recombination (*iii*) $CH_3CNH^+ + e^- \rightarrow CH_3CN + H$. We found that at beginning of the collapse CH_3CN is rapidly formed, in agreement with what estimated by Loison, Wakelam & Hickson (2014) and then slow-down by (*iv*) $CH_3CN + H_3^+ \rightarrow CH_3CNH^+ + H_2$. The combined effect of reactions (*iii*) and (*iv*) is also sensitive to b since for fast collapse

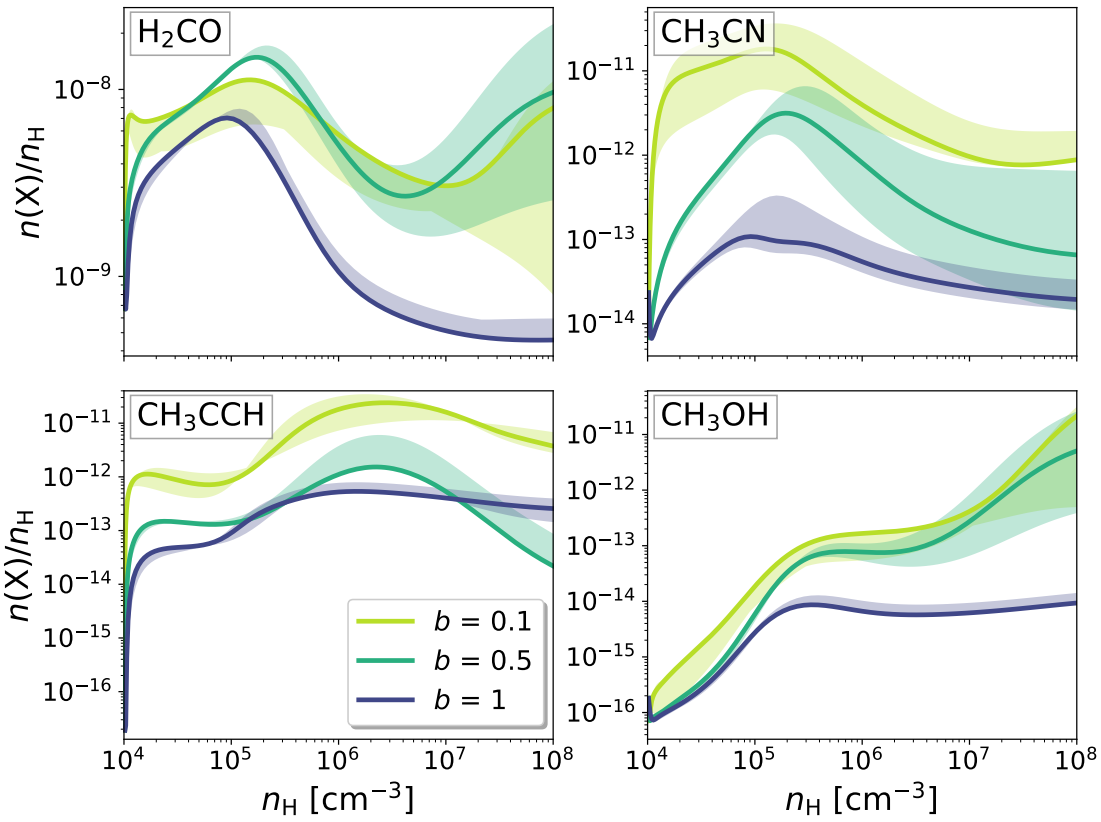

Figure 5.3: Evolution of the fractional abundance of H_2CO, CH_3CN, CH_3CCH and CH_3OH as a function of the central density of the clump during the collapse phase. Colours indicate the value of b assumed for the collapse described by Eq. 5.1. Solid-lines show the results employing the canonical binding energies from KIDA, while shaded-areas represent the solutions by varying T_b by ±10%.

(b = 1) reactions (*ii*) and (*iii*) have a shorter amount of time to form CH_3CN before reaction (*iv*) becomes effective. For densities between 10^5 and 10^6 cm^{-3}, CH_3CN starts to react efficiently with $H_3^+ \rightarrow CH_3CNH^+ + CO$. The effect of this reaction in combination with (*iv*) produces a slow decrease in the CH_3CN profiles at high densities.

The abundance of methanol increases with density, driven by reaction $CH_3^+ + H_2O \rightarrow CH_3OH_2^+$ followed by $CH_3OH_2^+ + e^- \rightarrow CH_3OH + H$ (see also Viti & Williams 1999).

To determine other sources of uncertainty, we analysed the variation of the chemical evolution by changing the binding energies reported in Tab. A.2 by ±10%, repeating the collapse with the same physical conditions. The effects on the fractional abundances are represented by the colour-shaded areas in Fig. 5.3. In some cases, such as methanol, the discrepancy is more than one order of magnitude at the highest densities. We note that even larger discrepancies are reported by Penteado, Walsh & Cuppen (2017), when the binding energies are randomly selected from a normal distribution.

These tests show that the uncertainties due to the variation of the binding energy are not negligible (see Fig. 5.3), but they are not larger than variations caused by changing b, suggesting 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 have tested our collapse-phase with a set-up similar to Garrod & Herbst (2006), but employing the chemical network and binding energies of Semenov et al. (2010). Even considering that the network and some key chemical parameters (e.g. binding energies) differ, the abundances we found are within an order of magnitude compared to Garrod & Herbst (2006).

In summary, during Phase-I, the fractional abundances are considerably influenced by the dynamics of the collapse, showing a difference of 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. Since the outcome of Phase-I is the initial chemical conditions for Phase-II, the final effect that b has on the models is to provide different inputs for the warm-up phase, which translates in exploiting the effect of the chemical initial conditions on our final results.

Fig. 5.4 shows the abundances of H_2CO, CH_3CCH, CH_3CN and CH_3OH, as a function of time and distance both 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 indicate the location of the *evaporation fronts*, namely the distance from the protostar at which the desorption timescale of a given species (defined as the inverse of the desorption rate of Eq. 5.5; Semenov et al. 2010) is smaller than the dynamical timescale of the model. As a consequence, at distances larger than the evaporation front, the abundances of the tracers on dust (see Fig. 5.4) rapidly increase.

Each evaporation front grows as a function of time following the temperature evolution and ranges between ~10^3 and ~10^4.2 AU. Our findings are in agreement with Choudhury et al. (2015), where the efficiency of the evaporation as a function of the thermal

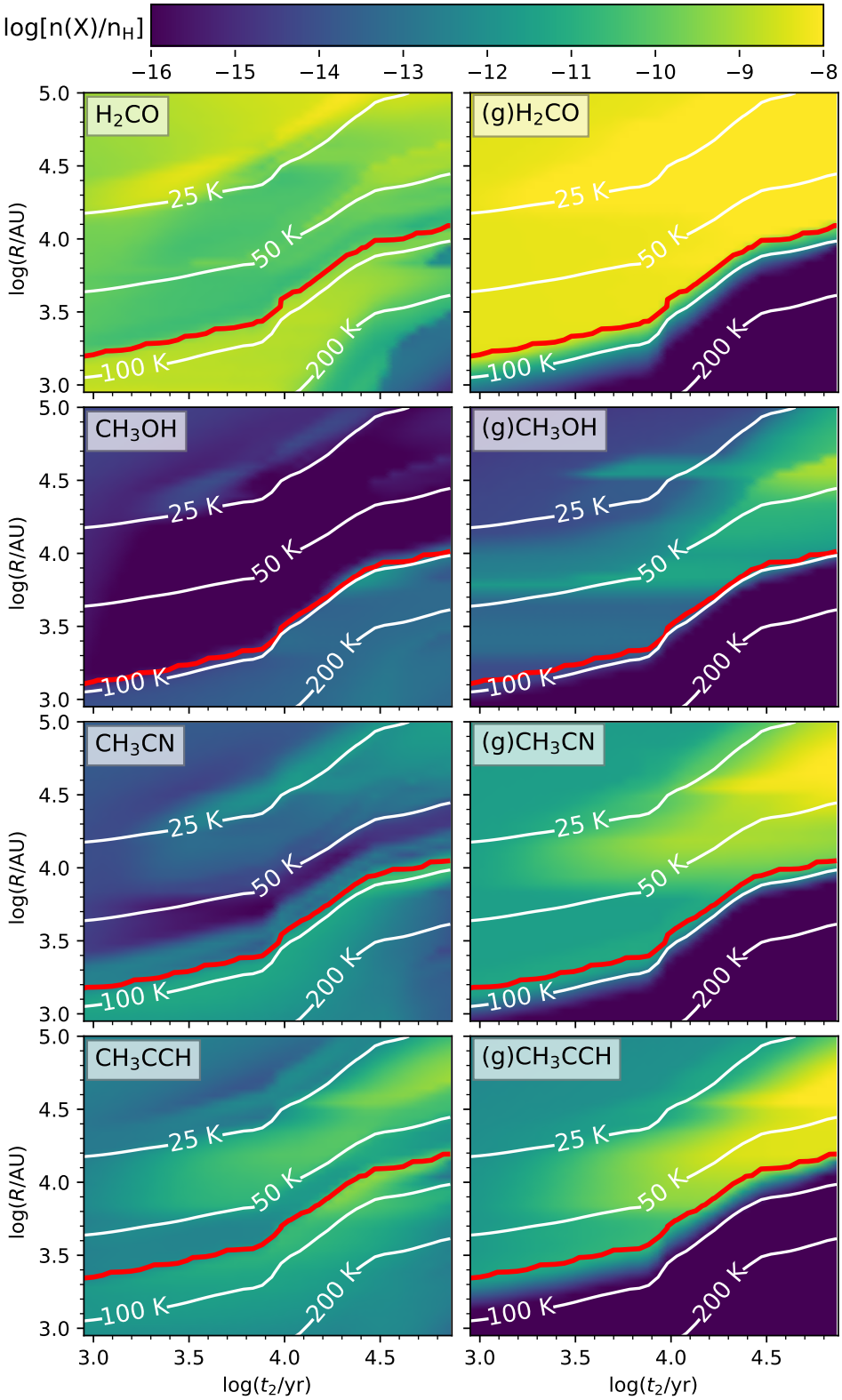

Figure 5.4: Temporal and radial evolution of the abundances of the tracers observed in the TOP100 (Giannetti et al., 2017b; Tang et al., 2018) during the warm-up phase, both in gas-phase (*left panels*) and on dust (*right panels*), for the same reference model as in Fig. 5.2. White contours indicate the temperature computed with Mocassin. Red curves correspond to the evaporation front of the given tracer.

evolution of the clump is responsible for evaporation spatial scales similar to ours (see their Fig. B1).

5.3.1 Post-processing of the models' outputs

The chemical structure of each clump has been reconstructed assuming that each radial profile from Phase-II (i.e. each column in the relative abundance maps in Fig. 5.4) represents a radius of a spherically symmetric clump. For each profile we generated a data cube of 150×150×150 pixels at a resolution of $\sim 10^3$ AU/pixel, which corresponds to one-half of the size of the smallest evaporation front radius in Fig. 5.4, 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 are then obtained by integrating the abundances along the line-of-sight (LOS), applying for each tracer a convolution[4] to the APEX resolution of the observed transitions (see Section 5.3.2), and taking into account the different distances of the TOP100 objects. The final column density profile of a given tracer X, $N_{\rm mod}$(X), is the value at the center of each map, that corresponds to a LOS through the center of the clump.

With respect to the aforementioned procedure, CH_3OH and CH_3CN require an additional constraint to reproduce the observed higher-K lines. In fact, in their spectral fit, Giannetti et al. (2017b) 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_3CN and CH_3OH, we applied a 100 K temperature-mask, taking into account only the regions where $T > 100$ K, to generate the averaged column densities profiles of the hot components only.

As an example of the overall procedure, we report the final gas-phase CH_3CCH profiles in Fig. 5.5*a* for G353.41-0.36. With a mass of 3000 $M_\odot$ and a distance of 4 kpc, G353.41-0.36 can be considered a prototypical source from our sample (see Tab. 5.4). We report the results for $\dot{M} = 10^{-3}$ $M_\odot$ yr^{-1} and for the standard free-fall collapse model (i.e. $b = 1$), and we note that the $n_{\rm H}(r_0)$ is the parameter that mostly influences the evolution of $N_{\rm mod}(CH_3CCH)$.

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 larger values of $n_{\rm H}(r_0)$ have larger initial column densities (see Fig. 5.3). Second, the different thermal structure of the clumps has a strong impact on the chemical evolution. At a given time, the average temperature along the LOS is lower in clumps with larger $n_{\rm H}(r_0)$, as the radiation of the protostar is more attenuated. Since the luminosity of the protostar increases with time and affects the temperature of the surrounding gas,

[4]This step was done using the `fft2` function of `numpy.fft`: https://numpy.org/doc/stable/reference/generated/numpy.fft.fft2.html

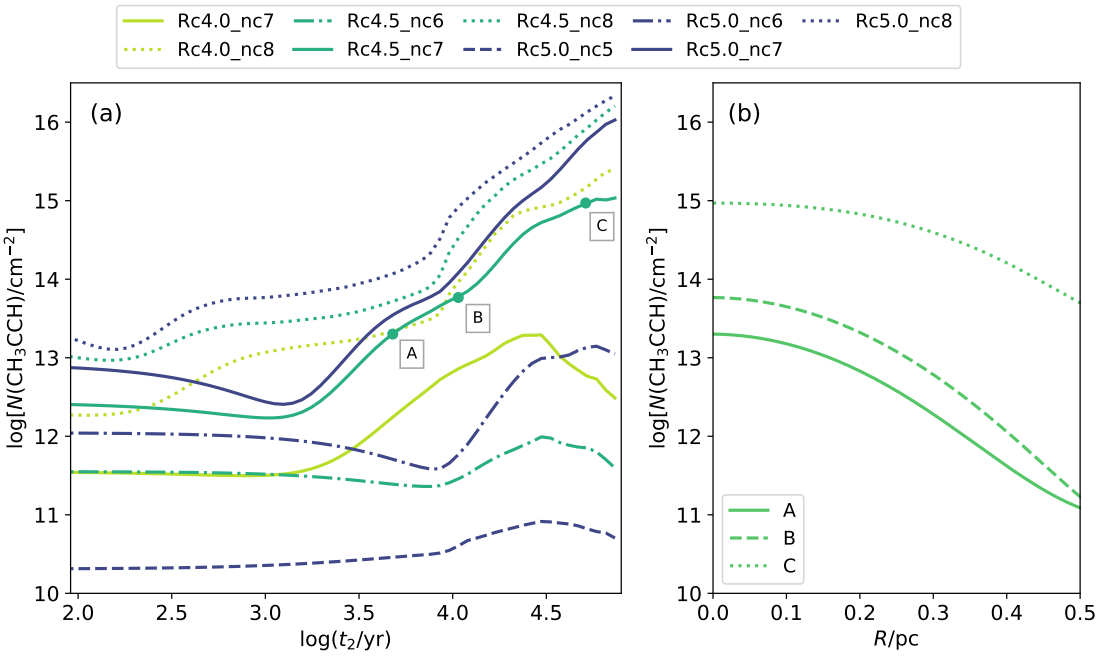

Figure 5.5: *Panel* (*a*): Example of the line-of-sight gas-phase column density profiles obtained at the end of the post-processing described in Sect. 5.3.1 for models with $b = 1$ and $\dot{M} = 10^{-3} M_{\odot}$ yr^{-1} and for an arbitrarily chosen source (G353.41-0.36; see Tab. 5.4). For the `U_W` items in the legend `U` and `W` are respectively R_c (colours) and $n_{\rm H}(r_0)$ (styles). *Panel* (*b*): Radial distribution of gas-phase column densities extracted at the positions A, B and C in Panel (*a*), i.e. the same times of Fig. 5.2.

the evaporation front grows and consequently the amount of chemical species released into the gas. In Fig. 5.5*b* we show the radial profile of gas-phase $N(CH_3CCH)$ at three times (A,B, and C in Fig. 5.5*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 of the CH_3CCH 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. 5.4). This also provides an explanation for the observed increasing trend in the CH_3CCH abundance with the evolutionary stage of massive clumps (e.g. Molinari et al. 2016; Giannetti et al. 2017b).

5.3.2 Notes on the column densities derived for the TOP100 sources

In this Section we report some notes on how the column densities of the molecular tracers observed in the TOP100 sample (see Sect. 2.3) are treated in order to make a comparison with the outputs of the models (as discussed in Sect. 5.3.3).

- Formaldehyde (H_2CO)
  Formaldehyde data are taken from Tang et al. (2018). For simplicity, we assume a

(1) Source-ID	(2) Class	(3) $d_\odot$	(4) L_{bol}	(5) M_{clump}	(6) T_d	(7)	(8)	(9)	(10)	(11)	(12)
						$\log_{10}[N_{obs}(X(o\text{-}H_2D^+)/cm^{-2})]$					
		kpc	$10^2\ L_\odot$	$10^2\ M_\odot$	K	H_2CO	CH_3CCH	$(CH_3CN)_c$	$(CH_3CN)_h$	$(CH_3OH)_c$	$(CH_3OH)_h$
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	13064.99	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	12416.71	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

Table 5.4: Summary of the observed properties of the selected TOP100 sources (column 1): columns from (2) to (6) are related to the physical properties of each source (König et al., 2017). Columns from (7) to (12) report the final beam- and LOS-average column densities, obtained applying the procedure described in Sec. 5.3.2 to the values reported in Giannetti et al. (2017b) and Tang et al. (2018). The subscripts 'c' and 'h' separate the two thermal components (*cold* and *hot*, respectively) of CH_3CN and CH_3OH (see text in Sect. 5.3.2).

single *half power beam width* (HPBW) beam size of 29″to convolve the modelled column densities (i.e. the APEX-telescope resolution corresponding to the $J = 3 - 2$ transition of both the H_2CO *ortho* and the *para* forms at ~211 and ~218 GHz, respectively.) For the sources where both isomers are observed, the total column density, $N_{obs}(H_2CO)$, is derived as $N(o\text{-}H_2CO)+N(p\text{-}H_2CO)$. For sources not detected in *o*- and/or *p*-H_2CO, we estimate the corresponding detection limit column density using WEEDS (Maret et al. 2011), with a 3σ detection with respect to the average noise level (Tang et al. 2018; 0.06 K main-beam temperature scale) and assuming the H_2CO molecular line parameters provided by the CDMS[5] database (Müller et al. 2001). For each evolutionary class we assume the average values of T_{kin} (i.e. 52, 73, 81 and 110 K for 70w, IRw, IRb and HII, respectively) and a FWHM of 5 km s^{-1} for the line reported by Tang et al. (2018), and that the source size is equal to the APEX beam at the corresponding frequency.

- Methyl acetylene (CH_3CCH), methanol (CH_3OH) and acetonitrile (CH_3CN)
  Column densities of methyl acetylene, methanol and acetonitrile are taken from Giannetti et al. (2017b), who observed with APEX the $J = 20_K - 19_K$, $J = 7_K - 6_K$, and $J = 19_K - 18_K$ bands for CH_3CCH, CH_3OH, and CH_3CN, respectively. Since these lines are very close to each other in frequency, we assume the same angular resolution for APEX, i.e. 18″. The final column densities in Giannetti et al. (2017b) are corrected for the APEX-beam dilution, $\eta = \theta_S^2/(\theta_S^2 + \theta_{beam}^2)$, where θ_S and θ_{beam} are the source- and beam (HPBW) sizes. We have (re-)applied this correction to each source where the detections were not resolved in the APEX-beam angular size (i.e. when $\theta_S < \theta_{beam}$) to obtain the beam-averaged column densities, whereas for sources with $\theta_S > \theta_{beam}$, $\eta = 1$. Detection limits have been estimated assuming the mean noise of the source's spectra, the average line width and the median excitation temperatures reported in Tables 6 and 7 of Giannetti et al. (2017b) for each evolutionary class. We treated the hot components observed for these tracers (see Sect. 5.3.1) in the same way as the colder ones, considering the appropriate temperatures reported by Giannetti et al. (2017b), in the detection limits calculation.

5.3.3 Comparison of the modelled column densities with sources from TOP100 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 (Section 5.3.2 and Tab. 5.4) with those from our post-processed models. For a given clump the dataset comprises eight measurements $D = \{L, M, N(\mathtt{X})\}$, with $\mathtt{X}$ ranging over the 6 observed chemical tracers. At each time-step of the warm-up phase, with the free

[5] https://cdms.astro.uni-koeln.de/cdms/portal/

parameters $\boldsymbol{\theta} = \{b, \dot{M}, R_c, n_H(r_0)\}$ given the data D, the likelihood is

$$\mathscr{L}(\boldsymbol{\theta}|\mathrm{D}) = P_C(\boldsymbol{\theta}|\mathrm{D})\, P_M(\boldsymbol{\theta}|\mathrm{D})\, P_L(\boldsymbol{\theta}|\mathrm{D})\,, \tag{5.10}$$

where $P_C(\boldsymbol{\theta}|\mathrm{D})$, $P_M(\boldsymbol{\theta}|\mathrm{D})$ and $P_L(\boldsymbol{\theta}|\mathrm{D})$ are probability density distributions of the chemical abundances, the mass and the bolometric luminosity of each clump, respectively. The first factor in Eq. 5.10 is

$$P_C = \prod_{\mathtt{X}} P_{C,\mathtt{X}}\,, \tag{5.11}$$

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

$$P_{C,\mathtt{X}} = \mathrm{erfc}\left\{\frac{|\log_{10}[N_{obs}(\mathtt{X})] - \log_{10}[N_{mod}(\mathtt{X})]|}{\sqrt{2}\sigma}\right\}, \tag{5.12}$$

with 'erfc' the complementary error function. In Eq. 5.12, $N_{obs}(\mathtt{X})$ and $N_{mod}(\mathtt{X})$ are the observed and modelled column densities, respectively. As in Garrod, Wakelam & Herbst (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 Phase-I, and on the shape of the density profile assumed in Phase-II (see Sect. 5.1.2). For sources without a detection we assume a detection limit (see Section 5.3.2) and set $P_{C,\mathtt{X}} = 1$ when $N_{mod}(\mathtt{X}) < N_{obs}(\mathtt{X})$, while we use Eq. 5.12 when $N_{mod}(\mathtt{X}) > N_{obs}(\mathtt{X})$.

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 are computed as follows: we add in quadrature a 50% uncertainty associated with the models, due to the approximation of a single clump size of 0.5 pc and a constant accretion rate over time (Sect. 5.1.2), on top of 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 about 50% of M for P_M. We note that an additional variation of 10% on these uncertainties does not produce significant variations in our final results.

Fig. 5.6a shows an example of how $\mathscr{L}$ varies as a function of time when considering models with the same mass accretion rate ($\dot{M} = 10^{-3}\,M_\odot\,yr^{-1}$). The models are sorted in ascending order by the difference in mass between the values reported in König et al. (2017) and those of the models (i.e. Tab. 5.2), while t_2 is the time spent in the warm-up phase defined in Sect. 5.1.2. In both panels, P_L defines the range of time spent by the proto-star in the warm-up phase (i.e. the range of time in between the vertical cyan lines), that only depends on the uncertainties associated to 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 hence, it is possible

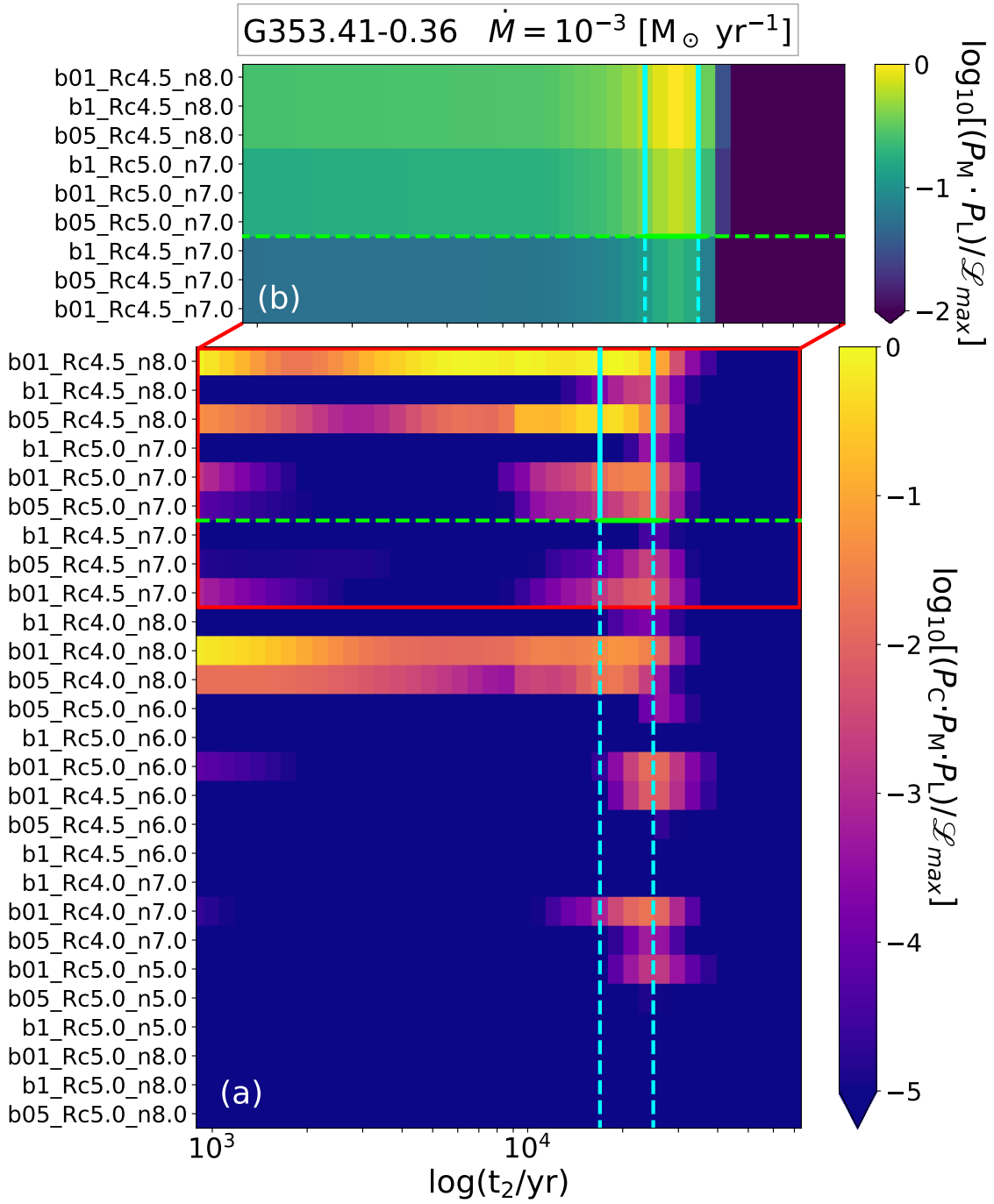

Figure 5.6: *Panel* (*a*): Example of the $\mathscr{L}$ matrix calculated for G353.41-0.36, in all the models with $\dot{M} = 10^{-3} M_\odot$ yr^{-1}. Models on the y-axis are sorted from the top to the bottom, by the difference between the mass of the models (see Tab. 5.2) and the value associated in the TOP100 (see Tab. 5.4). For the sake of clarity, the colorbar is limited to five orders of magnitude, and normalised to maximum likelihood ($\mathscr{L}_{max}$). *Panel* (*b*): Zoom around the age limits identified by P_L and P_M showing how the likelihood would appear without taking into account P_C in Eq. 5.10. Cyan vertical lines indicate the range of time that corresponds to the uncertainties on the observed luminosity in the TOP100, while the green horizontal line shows the same limit associated to the clump mass. Times and $\mathscr{L}$ are shown in log-scale.

to define a mass lower limit indicated by the green horizontal line. The combined information given by $P_{\rm L}$ and $P_{\rm M}$ limits the age range of each source. In this area, the models with the same density profile defined by Eq. 5.2, but different values of b, are degenerate showing the same $\mathscr{L}$ as a function of t_2 (see Fig. 5.6b). $P_{\rm C}$ is fundamental to break this degeneracy (see Fig. 5.6a), providing a further information on the dynamics of the collapse that led to the observed chemical properties of each clump, and put additional constraints for 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 (0.4-5) $\times$ 10^6 yr depending on b and $n_{\rm H}(r_0)$, while the warm-up time depends on the mass accretion rate of Phase-II (see Sect. 5.1.2).

The absolute age for the ℓth source is the weighted mean

$$t_{{\rm A},\ell} = \left[\sum_{i,j} \frac{\log_{10}[t_{1,i,\ell} + t_{2,j,\ell}]}{\left(\log_{10}\mathscr{L}_{i,j,\ell}\right)^2}\right]\left[\sum_{i,j} \frac{1}{\left(\log_{10}\mathscr{L}_{i,j,\ell}\right)^2}\right]^{-1}, \tag{5.13}$$

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

We performed the comparison described above only for the sources in the TOP100 that are in agreement, within the uncertainties, with the masses of the models in Tab. 5.2. This reduces the sample to 48 objects, i.e. 6 sources in the 70w stage, 16 IRw, 15 IRb and 11 HII regions. We summarise their observed properties in Tab. 5.4. Nevertheless, this limitation does not affect our approach, since 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. 2017b; König et al. 2017).

5.4 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 $\mathcal{N}$ objects, $\langle t_{\rm A}\rangle = \mathcal{N}^{-1}\sum_{n=1}^{\mathcal{N}} t_{{\rm A},n}$, we define the "age-factor" of the ℓth source as

$$f_{{\rm A},\ell} = \frac{t_{{\rm A},\ell}}{\langle t_{\rm A}\rangle}\,. \tag{5.14}$$

In this context, the models provide $\langle t_{\rm A}\rangle \sim 10^6$ yr starting from $t_1 = 0$. The average $f_{\rm A}$ are estimated as the median of the age-factor distributions in each evolutionary class, computed via Eq. 5.14 and reported as purple markers in Fig. 5.7. From the same distributions we quantify the duration of the phases $\Delta t_{\rm phase}$, taking the values in between the 5th and 95th percentiles (purple shaded-areas in Fig. 5.7). The sum of the four $\Delta t_{\rm phase}$ is the total time of the high-mass star formation process $t_{\rm MSF}$. The purple percent values in Fig. 5.7 (M_1) indicate the contribution of each phase to the total time $t_{\rm MSF}$.

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 first phase (70w), while IRw stage is associated to 19% of $t_{\rm MSF}$. This suggests a fast evolution during the early stages of the massive star formation process, which in total correspond to ~30% of $t_{\rm MSF}$. Advanced stages (i.e. IRb sources) have the longest duration (39%), while the remaining 30% of $t_{\rm MSF}$ is spent in the final stage (HII).

Assuming that the number of objects in an evolutionary stage is also representative of its duration, we compare our findings with the relative number of objects per evolutionary class in ATLASGAL, i.e. black percentages in Fig. 5.7 ("O"; from Urquhart et al. 2018). The duration found for the early stages are in agreement with the observed classification, while 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 different sensitivities (see Urquhart et al. 2014b for more details), so that the final number of objects in the advanced stages might be underestimated. Since 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 give a better agreement.

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), while when the protostars are close to the main-sequence, their radiation pressure might slow down or quench the mass accretion rate (e.g. Nakano, Hasegawa & Norman, 1995; Stahler & Palla, 2004; Klassen, Peters & Pudritz, 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 $\dot{M}$ produce different durations and $t_{\rm MSF}$. 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 orange symbols and percentages in Fig. 5.7; M_2).

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 $f_{\rm A}$ of each class increases with the evolution of the sources (circles and stars in Fig. 5.7), so that the typical object of each class appears statistically older than the average from the previous evolutionary phase. We therefore assume M_2 to be the most representative model, with a total massive star formation time $t_{\rm MSF,M_2} \sim 5.2 \times 10^5$ yr (and $t_{\rm MSF,M_1} \sim 1.1\ t_{\rm MSF,M_2}$).

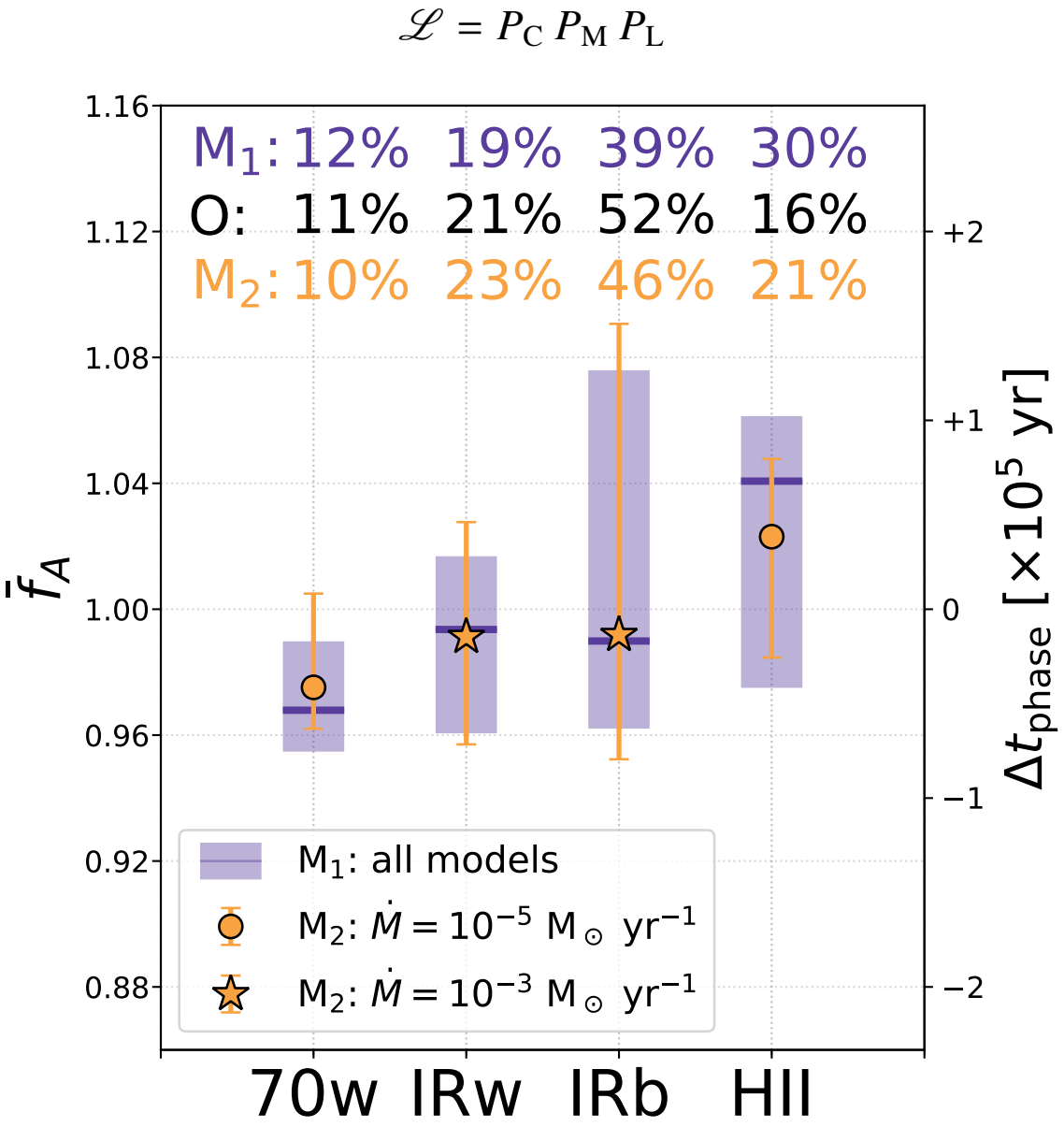

Figure 5.7: Summary of the final durations estimated in each evolutionary class defined for the TOP100. The relative number of massive clumps in each phase, and observed in ATLASGAL, are shown in black ("O"; Urquhart et al., 2018). Purple shaded-areas indicate the 5th and 95th percentiles of the age factor, f_A, in our sample (model "M_1"), while purple markers ('−') represent their median values. The numbers associated to the model "M_2" (orange), indicate the same quantities, obtained 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 accretion rates that best match the observed relative number of object in each phase (see legend).

In this context the chemistry plays a key role in breaking the model degeneracy when the likelihood is defined only by P_M and P_L, and in matching the results from ATLASGAL (see Fig. 5.7). In fact, if P_C is not included in $\mathscr{L}$ (Eq. 5.10), the duration of the advanced stages show a difference of up to a factor of ~8 when compared to those predicted by M_2. This corresponds to a total time, t_{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 (see Fig. 5.8).

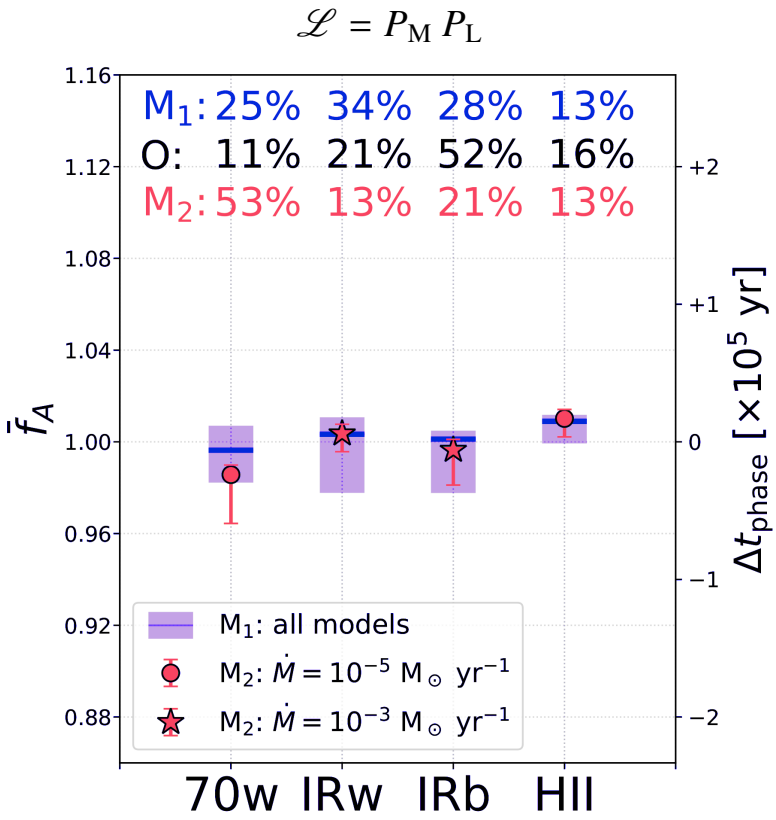

Figure 5.8: Same plot as Fig. 5.7, obtained by removing P_C from the calculation of $\mathscr{L}$ in Eq. 5.10.

5.5 Selected tracers and chemical clocks

Four molecules were selected in this work: formaldehyde, methyl acetylene, acetonitrile and methanol (see Sections 1.5.1 and 2.3), all of them manifesting an observed upward trend in their abundances with increasing evolutionary stage, indicated by the luminosity-to-mass ratio of the clumps (Giannetti et al., 2017b; 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. (2017b).

To evaluate to what extend the models are able to reproduce the observed column densities, the same procedure used to calculate Δt_{phase} has been employed (Equations 5.13 and 5.14). Figure 5.9 summarises the comparison between the modelled and the observed

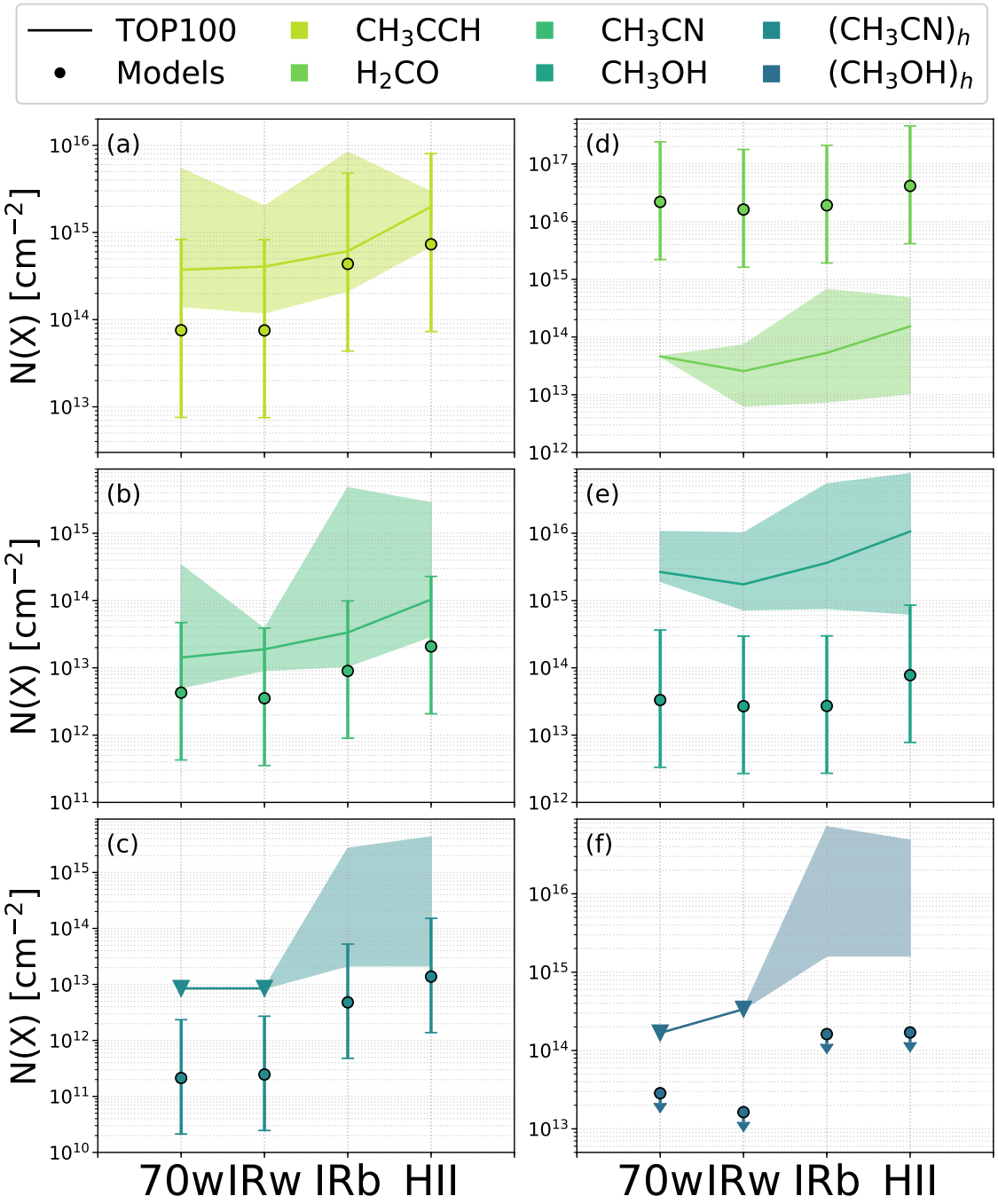

Figure 5.9: Comparison of the median column densities observed in the TOP100 (solid-lines) with those predicted by our models (circles), obtained applying the same procedure used to quantify $\Delta t_{\rm phase}$ (see Sect. 5.3.3). Panels are separated by tracers (colours). The shape of the shaded-areas indicates the minimum and maximum column density observed, while the errorbars associated to each circle incorporate the uncertainty of one order of magnitude assumed in Eq. 5.12 for the comparison. The modelled column densities of the hot component of methanol (circles in panel f) are multiplied by 10^3. Triangles indicate the upperlimits.

column densities, grouped by tracer type (colours) and evolutionary class. Shaded areas represent the range of the observed beam-averaged column densities (Sections 1.5.1 and 2.3 and Section 5.3.2), with an order of magnitude uncertainty (Eq. 5.12), 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.

5.5.1 Methyl acetylene and acetonitrile

The reliability of CH_3CCH and CH_3CN as thermometers has been widely discussed in the past (e.g. Zhang, Ho & Ohashi 1998; Molinari et al. 2016; Giannetti et al. 2017b). In particular, Giannetti et al. (2017b) detected an increasing trend of the temperature and column density of CH_3CCH from the less to the more evolved sources in the TOP100 (i.e. shaded area in panel a of Fig. 5.9). The same behaviour is found for CH_3CN if the contribution of a hot and a cold component are taken into account. In Fig. 5.9, 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, reproducing the observed values. The discrepancy between models and observations are within the uncertainties.

The robustness of CH_3CCH and CH_3CN as chemical clocks is finally tested by removing these two tracers from P_C; this has the same effect as removing P_C from $\mathscr{L}$ altogether (see Sect. 5.4 and Fig. 5.8), and we conclude that methyl acetylene and acetonitrile are effective chemical clocks to characterise the evolution of massive star-forming clumps, and contribute to constrain our findings.

5.5.2 Formaldehyde and methanol

Unlike CH_3CCH and CH_3CN, the model column densities of methanol and formaldehyde (panels d, e and f in Fig. 5.9) present a less accurate agreement with those observed. The slightly increasing trend in the observed column densities with evolution is reproduced by the models, but the modelled column densities under-/over-estimate by at least one order of magnitude those observed for the TOP100, 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 in H_2CO and CH_3OH on grains, since hydrogenation is made very efficient by the large CO-depletion occurring 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 also in the cold phases of massive star-forming regions (e.g. Cosentino et al. 2018).

Alternative formation paths and mechanisms are proposed to solve this issue (e.g. Viti & Williams 1999; Garrod, Wakelam & Herbst 2007; Vasyunin & Herbst 2013), but this was found to be relatively inefficient in reproducing the observed abundances of methanol (Geppert et al., 2006), and to no longer accurately predict the abundances of other chemical species (Garrod, Wakelam & Herbst, 2007). The same issue also concerns H_2CO, a molecular precursor of methanol. In the advanced evolutionary stages, the amount of methanol is underestimated (although closer to the observed values). This might happen because 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 (Charnley, Tielens & Rodgers 1997),

$$\mathrm{CO} \xrightarrow{+\mathrm{H}} \mathrm{HCO} \xrightarrow{+\mathrm{H}} \mathrm{H_2CO} \xrightarrow[E_a]{+\mathrm{H}} \mathrm{CH_3O} \xrightarrow{+\mathrm{H}} \mathrm{CH_3OH}\,. \tag{5.15}$$

Recently, a study of CO hydrogenation on water ice by Rimola et al. (2014), reviewed the activation energy of the hydrogenation reaction between H_2CO and H producing CH_3O, the precursor of methanol (i.e. third hydrogenation in reaction 5.15). The authors proposed a barrier of 1300 K, which would favour the formation of CH_3OH compared to the value commonly employed (E_a ~2500 K, as in this work and derived for the gas-phase; Woon 2002). An even lower activation energy of ~500 K has been predicted by Fuchs et al. (2009). However, we have tested both scenarios proposed above, without finding significant changes with respect to the modelled column densities of the tracers in Fig. 5.9. The most substantial change concerns the CH_3OH (see Fig. 5.10; panels *e* and *f*) when assuming the extreme activation energy proposed by Fuchs et al. (2009). In this case, methanol column densities come closer to the values observed in the two most advanced evolutionary stages, while those of formaldehyde remains always largely overestimated, without showing substantial variations. This suggests that the activation energy has secondary effects in reproducing the observed abundances of methanol and formaldehyde, and other processes or observational strategies, e.g. different or multiple transitions, should be taken into account.

We finally note that the discrepancies in the abundances of H_2CO and CH_3OH are not relevant for our results. If these two tracers are removed from P_C and $\mathscr{L}$, the duration of each phase and the final t_{MSF} remain unaffected, showing a variation of ~10% with respect to the numbers reported in Fig. 5.7.

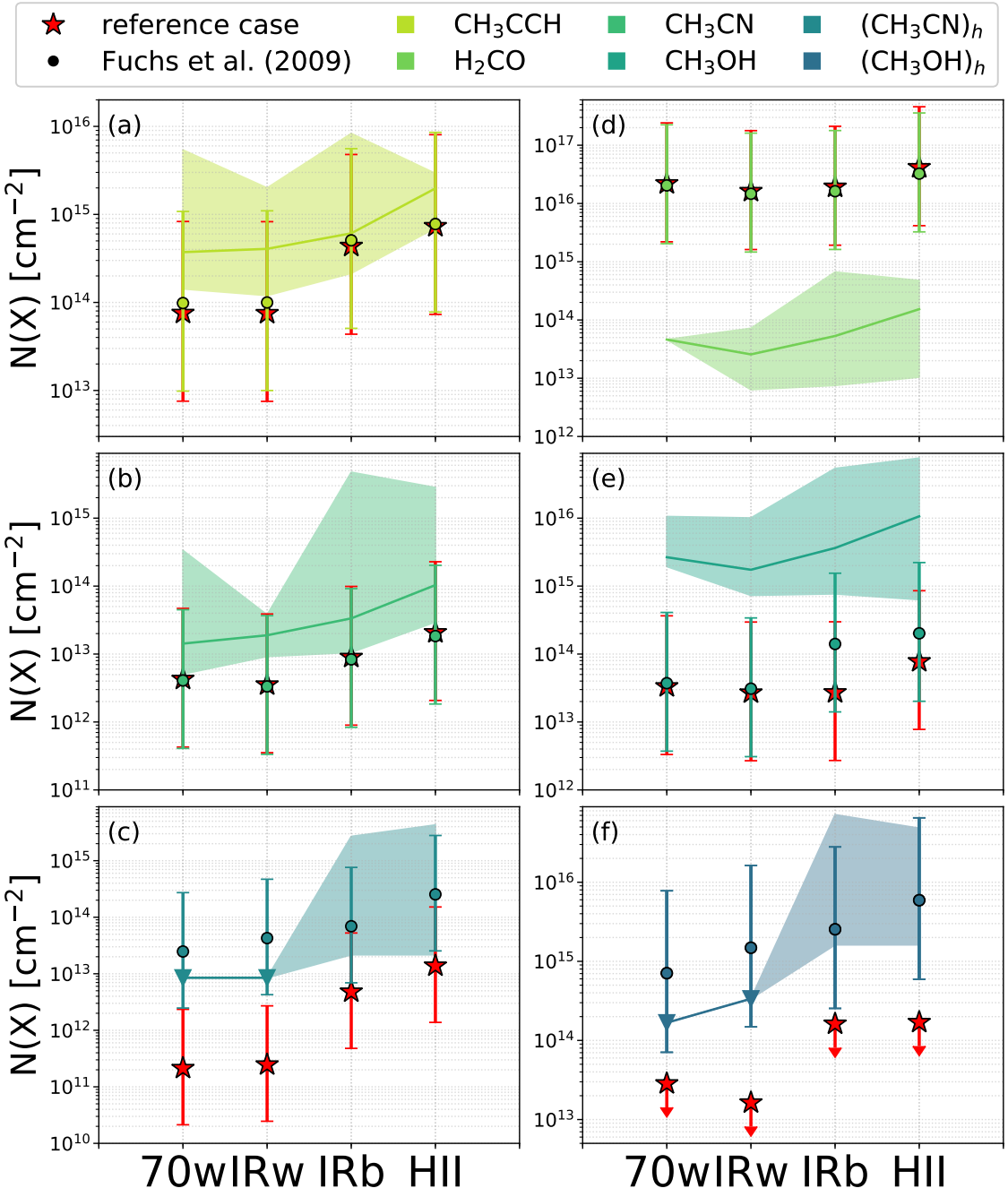

Figure 5.10: Same of Fig. 5.9, obtained by assuming a barrier of 500 K for the reaction $H_2CO + H \rightarrow CH_3O$ to occur (Fuchs et al. 2009). Red stars indicate the solutions reported in Fig. 5.9 as coloured-circles.

5.6 Chapter conclusions

The main results of this chapter are the following:

- We have built a set of 54 models to represent the entire population of massive clumps of the ATLASGAL-TOP100 sample (Sect. 1.5.1). During the two physical phases of the model (i.e. an isothermal collapse and a warm-up phase), we have modelled the chemical evolution of the clumps up to a maximum distance of 0.5 pc from the central massive protostar, i.e. the average radius of the TOP100 sources, and describing the thermal evolution of the clumps by employing the radiative transfer code Mocassin (Ercolano et al. 2003; Ercolano, Barlow & Storey 2005). The chemistry was solved during both the collapse and the warm-up phases, with the time-dependent code Krome (Grassi et al., 2014). We then compared the observed column densities of formaldehyde (H_2CO), methyl acetylene (CH_3CCH), acetonitrile (CH_3CN) and methanol (CH_3OH) observed in the ATLASGAL-TOP100 sources (Giannetti et al., 2017b; Tang et al., 2018) by post-processing the outputs of the models at the same angular resolution as the observed data (see Sect. 5.3.2).

- Based on this comparison, we developed a generalised method to derive 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, Bontemps & Louvet 2018; Urquhart et al. 2018), where $t_{\rm MSF}$ is assumed *a priori*, in our method we obtain $t_{\rm MSF}$ as the sum of the individual $\Delta t_{\rm phase}$ (Sect. 5.4). We find individual $\Delta t_{\rm phase}$ of $\sim 5 \times 10^4$ 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 (model M_2 in Fig. 5.7 which has a variable $\dot{M}$). The total star formation time, derived by the sum of the $\Delta t_{\rm phase}$ is $t_{\rm MSF} \sim 5.2 \times 10^5$ yr. This is in agreement with that assumed by the statistical methods, and provides a validation of the latter. The relative duration of each phase (i.e. $\Delta t_{\rm phase}/t_{\rm MSF}$) are found in agreement with the relative number of objects observed in the ATLASGAL survey (see Fig. 5.7).

- The chemical constraint, $P_{\rm C}$, included in the likelihood to determine the duration of the different phases, is necessary to achieve results in agreement with the observed relative number of objects in the ATLASGAL survey. In fact, without this additional constraint the $\Delta t_{\rm phase}$ become 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 (see Fig. 5.8). Among the selected molecular tracers, CH_3CCH and CH_3CN are those that are best reproduced by our models, while a discrepancy is found for H_2CO and CH_3OH with respect to the observed values (although they do follow the observed trends).

- We have also found that the radii of the evaporation fronts of the discussed molecular tracers vary between 10^3 and 10^5 AU during the warm-up phase (Sect. 5.3 and

Fig. 5.4), marking in our models the regions where each tracer becomes abundant in the gas-phase.

These results suggest to compare our findings with the observations provided by the new astronomical facilities. In particular, ALMA (Wootten & Thompson 2009) offers the perfect opportunity to achieve 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 the physical sizes of the modelled evaporation fronts. Moreover, ALMA can also reach sensitivities to detect a large number of components in a J_{K_a,K_c} band. This would allow to remove possible opacity effects, and to accurately define the thermal state of the clump that can be compared with the results of this work.

Chapter 6

Conclusions and future perspectives

In this Thesis, we have investigated the high-mass star formation process with the aim to provide accurate timescales of the different evolutionary stages characterising this complex phenomenon. Through the comparison of a statistically relevant number of observations, and accurate chemical models, we have explored the chemical evolution of the gas that resides in the parental clumps of massive stars. A number of chemical clocks have been carefully selected both for the early as well as for the late evolutionary stages from the ATLASGAL-TOP100 sample (Sections 1.5.1 and 2.3).

We have developed a comprehensive set of models, built to represent the entire population of massive clumps in the TOP100 sample (Chapter 5). The models consist of two physical phases: an initial isothermal collapse, followed by a warm-up phase induced by a massive YSO at the center of a spherical clump. We focused our analysis on four chemical tracers: formaldehyde, H_2CO; methyl acetylene, CH_3CCH; acetonitrile, CH_3CN; and methanol, CH_3OH, and compared their column densities with the output of the models properly post-processed to match the same angular resolution of the observed data (see Sect. 5.3.1).

The timescales of the evolutionary stages associated with the massive star formation process are derived by considering both the physical properties of the clumps, i.e. their mass and luminosity, and the observed abundances of each selected molecular tracer. Considering a time-varying mass accretion rate among the evolutionary phases, we found a total star formation time $t_{\rm MSF} \sim 5.2 \times 10^5$ yr, in good agreement with that assumed from statistical methods, providing a new and significant validation of the latter. The duration of the individual phases $\Delta t_{\rm phase}$, that define $t_{\rm MSF}$ 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. Knowing $t_{\rm MSF}$ and $\Delta t_{\rm phase}$, we derived the relative duration of each phase, finding agreement with the relative number of objects observed in those phases in the whole ATLASGAL survey. This provides a well-constrained confirmation of the established timeline of the high-mass star formation process proposed by Giannetti et al. (2014); König et al. (2017),

based on the luminosity-to-mass ratio (L/M) of the clumps, and also confirms the unbiased nature of the ATLASGAL survey. Our results have also clearly shown that the chemistry plays a fundamental role in the evolution of these regions and in determining their timescales. The chemical constraint included in the likelihood to determine the duration of the different phases (see Eq. 5.10) is in fact crucial to achieve agreement with the ATLASGAL survey (Sect. 5.4). Among the selected molecular tracers, CH_3CCH and CH_3CN are those that are best reproduced by our models, while H_2CO and CH_3OH show a discrepancy with respect to the observed values, although they do follow the observed trends (Sect. 5.5).

While the above tracers are important for the late stages of the star formation process, for the early stages we need a tracer which shows an opposite behaviour, i.e. it is abundant at early times and decreases when the environment is warmed up. In this Thesis we focused on o-H_2D^+, with the goal to establish its role as a valid chronometer of the first stages of the star formation process. In Chapter 4 we presented the first survey of o-H_2D^+ in high-mass star-forming regions, observing 106 sources of the ATLASGAL-survey in different evolutionary stages and almost all belonging to the TOP100 sample. We found 16 sources with reliable detections, from which we retrieved column densities and abundances, while for the other sources we only provided upper limits. Among the empirical correlations discussed in Chapther 4 between $X(o\text{-}H_2D^+)$ and the source-, evolutionary- and dynamical-parameters of the clumps (Sections 4.4.1, 4.4.2 and 4.4.3), that between $X(o\text{-}H_2D^+)$ and L/M appears the most reliable and robust (Tab. 4.5), highlighting a connection between the amount of formed o-H_2D^+ and the evolutionary stage of a massive clump. Finally, the decreasing trend in the average $X(o\text{-}H_2D^+)$ with the evolutionary stages (Fig. 4.8) suggests that the inclusion of deuterium fractionation chemistry (e.g. Sipilä, Caselli & Harju 2015; Bovino et al. 2019) in the chemical network employed in Chapter 5, might help to improve the accuracy of the determination of the timescales of the different phases. In this context, it could be also interesting to verify the anti-correlation between o-H_2D^+ and N_2D^+ pointed out by Giannetti et al. (2019), and extending the number and complexity of the considered molecular tracers, including for instance HC_3N (Taniguchi et al., 2018); CH_3OCHO and CH_3OCH_3 (Coletta et al., 2020); see also Urquhart et al. 2019; Belloche et al. 2020. In addition, the extended molecular survey carried out on the ATLASGAL-TOP100 sample provides an excellent possibility to test alternative ways for the identifications of new chemical clocks. A machine learning framework to detect the possible existence of spectral-classes comparable with the stages identified by L/M (Sect. 1.5.1) is a viable example. In such an approach, each spectral channel can be interpreted as a feature directly tied to a molecular tracer. The framework would take advantage of a stochastic unsupervised dimensionality reduction based on covariance matrix eigendecomposition (such as principal component analysis (PCA); Jolliffe & Cadima 2016) and classification of the spectra, to investigate possible spectral clusters. This approach allows to extract the most relevant evolutionary chemical features

of the classes, corresponding to new and reliable chemical clocks (e.g. Platt, 1999; Li & Chen, 2016). This challenging future perspective would provide an additional tool to follow the star formation process, and increase the number of constraints on the $\Delta t_{\rm phase}$ reported in this work.

Finally, to further explore the role of chemical clocks like o-H_2D^+, it would be necessary to investigate the behaviour of this tracer at smaller scales than the clumps reported in this work. High sensitivities observations at the angular resolutions accessible with ALMA represent an ambitious perspective and a natural continuation of the Thesis work. This study cannot be pursued without a clear comprehension of the processes affecting deuteration and its evolution, such as for instance the depletion of CO on the surface of grains. The results presented in Chapter 3 provide a good basis for this investigation, discussing the depletion factor fluctuations over the whole observed structure of the IRDC G351.77-0.51 (Fig. 3.5). We aim in the future to extend this study to more sources, and better assess the depletion radius through high-resolution observations. An accurate estimate of this quantity is fundamental to understand at which scales a full depletion assumption is a reasonable approximation. This is crucial, for instance, to save computational time when performing expensive three-dimensional magneto-hydrodynamical simulations (see Chapter 3).

To conclude, the results presented in this Thesis confirm the evolutionary sequence based on the L/M of the object in ATLASGAL survey, and employed to define the different phases of the high-mass star formation process. With the help of a reliable set of chemical clocks, sensitive to the changes in density and temperature affecting the different evolutionary stages, we established a timeline of the whole process and provided a duration for each of the proposed phases. Our approach shows that chemistry is a powerful tool to understand the timescales of a complex process like massive star formation.

6.1 Final remarks

We stress that the approach and the results discussed in this Thesis are based on one-dimensional models, which cannot capture in full the impact of dynamical processes like magnetic fields and turbulence on the density evolution of the collapsing clumps and the subsequent (proto-)stellar accretion process. On the other hand, due to the chemical complexity, which requires a huge number of chemical processes, and the computational costs of highly complex dynamical environments, we have to rely on low-dimensional models. These allow to distinguish among the different chemical processes, and to build a large number of models to be compared with observations in a statistically relevant approach. Improvements on the treatment of the chemistry in three-dimensional simulations have been recently reported, but in the presence of a (proto-)stellar object, modelling the chemistry and the thermal evolution of the gas in three-dimensions still represents a challenge.

Chapter A

Notes on the chemical network

A.1 Reaction types and binding energies

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

Reaction	Example	#included
Recomb. on grain	$A^+ + grain^- \rightarrow A + grain$	13
CR ionisation	$AB + cr \rightarrow AB^+ + e^-$	} 220
CR photodissociation	$AB + cr \rightarrow A + B$	
Gas-phase reactions	(see the note below)[1]	4016
Photo-ionisation	$AB + \gamma_{UV} \rightarrow AB + e^-$	} 153
Photo-dissociation	$AB + \gamma_{UV} \rightarrow A + B$	
Grain-surf. reactions	$A_{dust} + B_{dust} \rightarrow AB_{dust}$	266
Thermal desorpion	$AB_{dust} + heat \rightarrow AB$	195
Desorp. induced by CR	$AB_{dust} + cr \rightarrow AB$	195
$(\text{CR photodiss.})_{dust}$	$AB_{dust} + cr \rightarrow A_{dust} + B_{dust}$	185
$(\text{CR photoion.})_{dust}$	$AB_{dust} + cr \rightarrow AB^+_{dust} + e^-$	56
$(\text{FUV photodiss.})_{dust}$	$AB_{dust} + \gamma_{UV} \rightarrow A_{dust} + B_{dust}$	204
$(\text{FUV photoion.})_{dust}$	$AB_{dust} + \gamma_{UV} \rightarrow AB^+_{dust} + e^-$	171
$(\text{Freeze-out})_{dust}$	$AB \rightarrow AB_{dust} + \gamma$	195

Table A.1: Summary of the reactions in the chemical network employed in this work. In column (2): A and B are two generic reactants; column (3) shows the total number of reactions of each type, contained in the chemical network.

[1] The gas-phase processes included are ion-molecule, neutral-neutral, charge exchange, radiative association, radiative recombination and dissociative recombination.

Chemical species - T_b/K							
C	10000	C_6H_6	7590	CO	1300	HNC_3	4580
C_{10}	8000	C_7	5600	CO_2	2600	HNCO	4400
C_2	10000	C_7H	6140	CS	3200	HNO	3000
C_2H	3000	C_7H_2	6590	Fe	4200	HS	2700
C_2H_2	2590	C_7H_3	7040	FeH	4650	HSS	2650
C_2H_3	2800	CH_3C_6H	7490	H	650	Mg	5300
C_2H_4	2500	C_7N	6400	H_2	440	MgH	5750
C_2H_5	3100	C_8	6400	**H_2C_3N**	**5030**	MgH_2	6200
CH_3CH_2OH	5400	C_8H	6940	**H_2C_3O**	**3650**	N	720
C_2H_6	1600	C_8H_2	7390	**H_2C_5N**	**6630**	N_2	1100
C_2N	2400	C_8H_3	7840	**H_2C_7N**	**8230**	**N_2H_2**	**4760**
CCO	1950	C_8H_4	8290	**H_2C_9N**	**9830**	Na	11800
C_2S	2700	C_9	7200	H_2CN	2400	NaH	12300
C_3	2500	C_9H	7740	H_2CO	4500	NaOH	14700
C_3H	4000	C_9H_2	8190	H_2CS	4400	NH	2600
C_3H_2	3390	C_9H_3	8640	H_2O	5600	NH_2	3200
C_3H_3	3300	C_9H_4	9090	H_2O_2	6000	NH_2CHO	6300
C_3H_3N	5480	C_9N	8000	H_2S	2700	**NH_2OH**	**6810**
CH_3CCH	3800	CH	925	HSSH	3100	NH_3	5500
C_3N	3200	CH_2	1400	**H_3C_5N**	**7080**	NO	1600
C_3O	2750	H_2CCN	4230	**H_3C_7N**	**8680**	NS	1900
C_3S	3500	H_2CCO	2800	**H_3C_9N**	**10300**	O	1600
C_4	3200	CH_3N	5530	**H_4C_3N**	**5930**	O_2	1200
C_4H	3740	CH_2NH_2	5530	**H_5C_3N**	**6380**	O_2H	5000
C_4H_2	4190	CH_3O	4400	HCCNC	4580	O_3	2100
C_4H_3	4640	CH_3	1600	HC_2O	2400	OCN	2400
C_4H_4	5090	CH_3C_3N	6480	HC_3N	4580	OCS	2400
C_4N	4000	CH_3C_4H	5890	**HC_3O**	**3200**	OH	4600
C_4S	4300	CH_3C_5N	7880	HC_5N	6180	S	2600
C_5	4000	CH_3C_6H	7490	HC_7N	7780	S_2	2200
C_5H	4540	CH_3C_7N	9480	HC_9N	9380	Si	11600
C_5H_2	4990	CH_3CHO	5400	**HCCN**	**3780**	SiC	3500
C_5H_3	5440	CH_3CN	4680	HCN	3700	SiH	13000
CH_3C_4H	5890	CH_2NH_2	5530	HNCCC	4580	SiH_2	3600
C_5N	4800	CH_3OCH_3	3150	HCO	2400	SiH_3	4050
C_6	4800	CH_3OH	5000	$HCOOCH_3$	6300	SiH_4	4500
C_6H	5340	CH_4	9600	CH_2O_2	5570	SiO	3500
C_6H_2	5790	CH_3NH_2	6580	HCS	2900	SiS	3800
C_6H_3	6240	H_2CN	2400	He	100	SO	2800
C_6H_4	6690	CN	2800	HNC	3800	SO_2	3400
Added in this work							
CH_3COCH_3	3500	H_2SiO	4050	NO_2	2400	SiC_3	5100
C_3P	5900	HCCP	4750	P	1100	SiC_3H	5550
C_4P	7500	HCL	5170	PH	1550	SiC_4	5900
CCl	1900	HCP	2350	PH_2	2000	$SiCH_2$	3750
CCP	4300	HCSi	3630	PN	1900	$SiCH_3$	4200
CH_2PH	2600	HNSi	5080	PO	1900	SiN	3500
Cl	3000	HPO	2350	SiC_2	4300	SiNC	4300
ClO	1900	N_2O	2400	SiC_2H	4700	SiO_2	4300
CP	1900	NH_2CN	5560	SiC_2H_2	5200		

Table A.2: Complete list of binding energies used in this work. The values were taken from the KIDA-database, and updated to the latest available estimates (mainly from Wakelam et al. 2017). When this was not possible, the values from Semenov et al. (2010) were used. The latter case has been highlighted in boldface.

In Tab. 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. 5.2).

A.2 Benchmark of the chemical network

We benchmarked our network against Semenov et al. (2010). The initial conditions are summarised in Tab. A.3. All the elements are initially atomic, with the exception of hydrogen which is assumed completely in molecular form. Except for He, N and O, all the elements are also ionised, while grains are initially neutral.

- **TMC1-model:** We benchmark the physical case named, "TMC1-model". The gas has a constant temperature of 10 K, a constant hydrogen nuclei density $n_{\rm H} = 2\times10^4$ cm^{-3}, and visual extinction $A_{\rm v} = 10$ mag. The sticking coefficient is $\mathcal{S} = 1$, and the cosmic ray ionisation rate of hydrogen molecules is set to $\zeta_2 = 1.3 \times 10^{-17}$ s^{-1} (Spitzer & Tomasko, 1968; Glassgold & Langer, 1974). We have assumed an average grain size of $\langle a \rangle = 0.1$ μm and bulk density $\rho_0 = 3$ g cm^{-3} corresponding to silicate grains, as in Sect. 5.1. Figures A.1 shows the perfect agreement with the results reported by Semenov et al. (2010).

species	$(n_i/n_{\rm H})_{t=0}$	species	$(n_i/n_{\rm H})_{t=0}$
He	9.00(-2)	Si^+	8.00(-9)
H_2	5.00(-1)	Na^+	2.00(-9)
C^+	1.20(-4)	Mg^+	7.00(-9)
N	7.60(-5)	Fe^+	3.00(-9)
O	2.56(-4)	P^+	2.00(-10)
S^+	8.00(-8)	Cl^+	1.00(-9)

Table A.3: Summary of initial abundances used for the "TMC1-model". All the species, n_i, are weighted on the initial density of hydrogen nuclei ($n_{\rm H} = n[{\rm H} + 2{\rm H_2}]$). The notation A(-B) means A$\times10^{-{\rm B}}$.

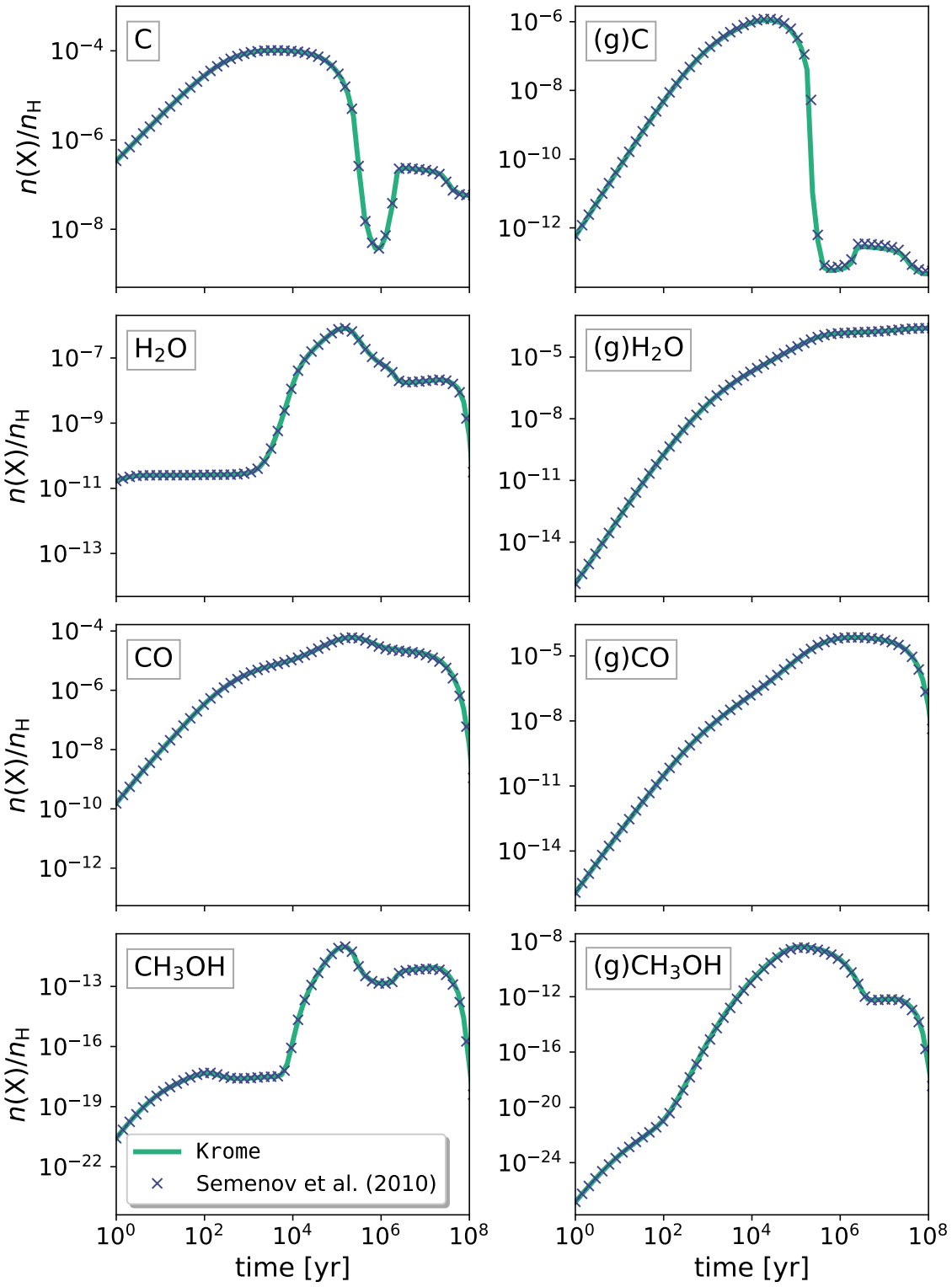

Figure A.1: Comparison of the time-dependent variation in the abundances of four arbitrary species (i.e. C, H_2O, CO and CH_3OH), both in gas-phase (*left panels*) and on dust (*right panels*), in the "TMC1-model". The green line shows the results obtained with KROME, while the blue crosses are the results of Semenov et al. (2010) with ALCHEMIC.

List of Figures

1.1 View of the Galactic plane of the Milky Way, here seen looking toward its centre . . . 2
1.2 Hierarchical cloud structures in the Perseus molecular cloud . . . 4
1.3 Evolutionary stages defined in the low-mass star formation process . . . 6
1.4 Evolutionary diagram proposed by Tigé et al. (2017) to describe the formation of high-mass stars . . . 10
1.5 Mass-Luminosity for low- (grey) and high-mass (black) regimes. . . . 13
1.6 Artist's conception of our Galaxy on which the region covered by the ATLASGAL survey is reported . . . 15
1.7 Illustrative three-colour images for each evolutionary class defined in the TOP100 sample: (a) 70w; (b) IRw; (c) IRb and (d) HII region. . . . 16
1.8 Sketch of the different evolutionary phases identified for the TOP100 (see also Fig. 1.7), and an example of a typical spectra in each evolutionary stages . . . 19

2.1 Cumulative number of confirmed detected interstellar molecules over time 22
2.2 Full final GUAPOS spectrum from 84 to 116 GHz . . . 23
2.3 Tomographic view of a molecular filament revealed by different chemical tracers . . . 26
2.4 Multiwavelength image of the molecular core Barnard 68 . . . 27
2.5 Sketch of the chemical structure of a molecular cloud from the UV exposed cloud edge (left) to the protected cloud core (right), where cloud depth is parameterised by extinction (measured in visual magnitudes A_V) 28
2.6 Panels (a) show CO depletion factors from $^{12}C^{17}O$ (on the left) and $^{12}C^{18}O$ (on the right) as a function of T_{ex} for the TOP100 sample. Panels (b) show the CO depletion factor (from $^{12}C^{17}O$) as a function of $N(H_2)$, for the TOP100 sample . . . 31
2.7 Observed anti-correlation between the abundances of o-H_2D^+ (in blue) and N_2H^+ (in orange) as a function of the evolutionary stage of three massive clumps located in the IRDC G351.77-0.51 . . . 34

2.8 Inner structure of a clump during the late stages of the star formation process 35
2.9 Sketch of the three major mechanisms for surface reactions on a regular grain surface (blue-rectangular surface): Langmuir-Hinshelwood, Eley-Rideal and hot-atom mechanisms 36
2.10 L/M ratio vs. the temperature traced by the different chemical species analysed in Giannetti et al. (2017b) 39
2.11 L/M ratio vs. the temperature traced by para-H_2CO analysed in Tang et al. (2018) 40

3.1 *LArge APEX BOlometer CAmera* (LABOCA) map of the 870μm dust continuum emission from the IRDC G351.77-0.51 48
3.2 Dust temperature and H_2 column density maps generated by a pixel-by-pixel SED-fitting of the 160-500 μm continuum fluxes of the Hi-Gal Survey 51
3.3 Panel (a): Integrated intensity distribution of the $C^{18}O$ $J = 2 - 1$ molecular line transition detected with *APEX*: (b) Map of the opacity; Panel (c): $C^{18}O$ column density map obtained from the integrated intensity distribution, shown in panl (a) 53
3.4 Schematic diagram showing how the correction C_τ was calculated starting from the $C^{18}O$ and $C^{17}O$ single pointing observation of clumps 2, 3, 5 and 7 in G351.77-0.51 55
3.5 Large scale depletion factor (f_D) map of G351.77-0.51, obtained by taking the ratio between the expected and observed $N(C^{18}O)$ 57
3.6 Pixel-by-pixel scatter plot of the whole structure detected in Fig. 3.5 . . . 58
3.7 Model applied to simulate the depletion effect inside/outside the depletion radius in G351.77-0.51 61
3.8 Column density profiles of H_2 and $C^{18}O$ in panel (a) and (b), respectively, while panel (c) shows the los- and beam-averaged depletion factor profile from region C5 in Fig. 3.1 63
3.9 f_D variation as a function of $R_{\rm dep}$ for the best-fitting model applied to clump-5 in G351.77-0.51 64
3.10 Sketch of the abundance profiles assumed to describe the $\chi_{C^{18}O}$ variation testing their influence on $R_{\rm dep}$ 66

4.1 *o*-H_2D^+ spectra for half the sources in the sample high-mass star-forming regions analysed in Sabatini et al. (2020). In each panel the source name is shown on the left in red, while the integrated main-beam temperature, $\int_{ch} T_{\rm mb}\, d\upsilon$, and rms (*i*rms) are on the right in black 76
4.2 Same as Fig. 4.1 for the second half of the sample 77

4.3 Example of one- and two-dimensional marginalised posterior distributions of the free parameters in our model 85
4.4 Collection of the different correlations between $X(o\text{-}H_2D^+)$ and the quantities summarised in Tables 4.1 and 4.3 86
4.5 Estimates of the CRIR of hydrogen molecules, ζ_2, derived for a subsample of sources for which H_2D^+, DCO^+, $H^{13}CO^+$, and $C^{17}O$ observations are available 92
4.6 ζ_2 variation as a function of $N(H_2)$ 93
4.7 ζ_2 variation as a function of D_{GC} for the high-mass cores shown in Fig. 4.6 94
4.8 Average $X(o\text{-}H_2D^+)$ abundance as a function of the evolutionary classes in our sample 96

5.1 Sketch of the physical model employed to describe the evolution of a massive clumps 102
5.2 Panel (a): $L_*(t_2)$ function employed in our models and based on Hosokawa & Omukai (2009); Panel (b): Example of the dust temperature radial profile at different times, employed in the physical model 105
5.3 Evolution of the fractional abundance of H_2CO, CH_3CN, CH_3CCH and CH_3OH 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. 5.1. Solid-lines show the results employing the canonical binding energies from KIDA, while shaded-areas represent the solutions by varying T_b by ±10%. 109
5.4 Temporal and radial evolution of the abundances of the tracers observed in the TOP100 (Giannetti et al., 2017b; Tang et al., 2018) during the warm-up phase, both in gas-phase and on dust 111
5.5 Example of the outputs of the models after the post-processing 113
5.6 Example of the $\mathscr{L}$ matrix calculated for G353.41-0.36, in all the models with $\dot{M} = 10^{-3} M_\odot$ yr^{-1} 117
5.7 Summary of the final durations estimated in each evolutionary class defined for the TOP100 120
5.8 Same plot as Fig. 5.7, obtained by removing P_C from the calculation of $\mathscr{L}$ in Eq. 5.10. 121
5.9 Comparison of the median column densities observed in the TOP100 with those predicted by our models 122
5.10 Same of Fig. 5.9, obtained by assuming a barrier of 500 K for the reaction $H_2CO + H \rightarrow CH_3O$ to occur (Fuchs et al. 2009) 125

A.1 Results of the benchmark between Krome and Alchemic 136

List of Tables

1.1 Summary of physical properties of giant molecular clouds and their substructures . . . 3

2.1 Some of the observed ratios between column densities of different chemical tracers in massive star forming regions. The ratios are divided by evolutionary class (i.e. pre-/proto-stellar and HII) . . . 41
2.2 Some of the relative abundances (with respect to H_2) observed for different tracers and samples divided by evolutionary class . . . 43

3.1 Estimated R_{dep} assuming the profiles in Fig. 3.10 and discussed in Sect. 3.4.3 68

4.1 Physical properties (observed and derived) of our sample. *Top row*: TOP100 sources; *Bottom row*: ATLASGAL sources not in the TOP100 . . . 73
4.2 Summary of the observations on which Chapter 4 . . . 75
4.3 Derived properties of the ATLASGAL-sources in our sample . . . 79
4.4 ATLASGAL sources used to compute the o-H_2D^+ detection limits . . . 82
4.5 Bayes factors, $\mathcal{B}_{2,1}$, found for each correlation in Fig. 4.4 . . . 87
4.6 Summary of the quantities to calculate the CRIR . . . 91

5.1 Parameters space explored in our models; i.e. Phase-I (Sect. 5.1.1) and Phase-II (Sect. 5.1.2). . . . 103
5.2 Summary of the total masses of each model . . . 106
5.3 Summary of fiducial initial chemical abundances . . . 108
5.4 Summary of the observed properties of the selected TOP100 sources . . . 114

A.1 Summary of the reactions in the chemical network employed in this Thesis 133
A.2 List of binding energies used in this Thesis . . . 134
A.3 Summary of initial abundances used for the benchmark of the chemical network . . . 135

Bibliography

Adams S. M., Kochanek C. S., Beacom J. F., Vagins M. R., Stanek K. Z., 2013, ApJ, 778, 164

Aguirre J. E. et al., 2011, ApJS, 192, 4

Amano T., Hirao T., 2005, Journal of Molecular Spectroscopy, 233, 7

Anderson L. D., Bania T. M., Balser D. S., Cunningham V., Wenger T. V., Johnstone B. M., Armentrout W. P., 2014, The Astrophysical Journal Supplement Series, 212, 1

Andre P., Ward-Thompson D., Barsony M., 1993, ApJ, 406, 122

Arzoumanian D. et al., 2011, A&A, 529, L6

Bacmann A., Lefloch B., Ceccarelli C., Castets A., Steinacker J., Loinard L., 2002, A&A, 389, L6

Bacmann A., Lefloch B., Ceccarelli C., Steinacker J., Castets A., Loinard L., 2003, ApJ, 585, L55

Bacmann A., Taquet V., Faure A., Kahane C., Ceccarelli C., 2012, A&A, 541, L12

Balucani N., Ceccarelli C., Taquet V., 2015, MNRAS, 449, L16

Belloche A. et al., 2020, A&A, 635, A198

Beltrán M. T., Cesaroni R., Neri R., Codella C., Furuya R. S., Testi L., Olmi L., 2005, A&A, 435, 901

Beltrán M. T., Codella C., Viti S., Neri R., Cesaroni R., 2009, ApJ, 690, L93

Bergin E. A., Alves J., Huard T., Lada C. J., 2002, ApJ, 570, L101

Bergin E. A., Tafalla M., 2007, ARA&A, 45, 339

Bertoldi F., McKee C. F., 1992, ApJ, 395, 140

Beuther H., Churchwell E. B., McKee C. F., Tan J. C., 2007, in Protostars and Planets V, Reipurth B., Jewitt D., Keil K., eds., p. 165

Beuther H., Schilke P., Menten K. M., Motte F., Sridharan T. K., Wyrowski F., 2002a, ApJ, 566, 945

Beuther H., Schilke P., Sridharan T. K., Menten K. M., Walmsley C. M., Wyrowski F., 2002b, A&A, 383, 892

Beuther H., Zhang Q., Bergin E. A., Sridharan T. K., 2009, AJ, 137, 406

Bianchi S., Gonçalves J., Albrecht M., Caselli P., Chini R., Galli D., Walmsley M., 2003, A&A, 399, L43

Black J. H., Hartquist T. W., Dalgarno A., 1978, ApJ, 224, 448

Bok B. J., 1948, Dimensions and Masses of Dark Nebulae, Vol. 7, p. 53

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 2001, MNRAS, 323, 785

Bonnor W. B., 1956, MNRAS, 116, 351

Bovino S., Ferrada-Chamorro S., Lupi A., Sabatini G., Giannetti A., Schleicher D. R. G., 2019, ApJ, 887, 224

Bovino S., Ferrada-Chamorro S., Lupi A., Schleicher D. R. G., Caselli P., 2020, MNRAS: Letters, 495, L7

Brown P. D., Charnley S. B., Millar T. J., 1988, MNRAS, 231, 409

Brünken S. et al., 2014, Nature, 516, 219

Buchner J. et al., 2014, A&A, 564, A125

Burke J. R., Hollenbach D. J., 1983, ApJ, 265, 223

Busch L. A., Belloche A., Cabrit S., Hennebelle P., Commerçon B., 2020, A&A, 633, A126

Butner H. M., Lada E. A., Loren R. B., 1995, ApJ, 448, 207

Cami J., Bernard-Salas J., Peeters E., Malek S. E., 2010, Science, 329, 1180

Carey S. J. et al., 2009, PASP, 121, 76

Carter M. et al., 2012, A&A, 538, A89

Caselli P., Ceccarelli C., 2012, A&ARv, 20, 56

Caselli P., Vastel C., Ceccarelli C., van der Tak F. F. S., Crapsi A., Bacmann A., 2008, A&A, 492, 703

Caselli P., Walmsley C. M., Tafalla M., Dore L., Myers P. C., 1999, ApJ, 523, L165

Caselli P., Walmsley C. M., Terzieva R., Herbst E., 1998, ApJ, 499, 234

Caselli P., Walmsley C. M., Zucconi A., Tafalla M., Dore L., Myers P. C., 2002, ApJ, 565, 344

Ceccarelli C., Caselli P., Bockelée-Morvan D., Mousis O., Pizzarello S., Robert F., Semenov D., 2014, in Protostars and Planets VI, Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds., p. 859

Ceccarelli C. et al., 2017, ApJ, 850, 176

Ceccarelli C., Caselli P., Herbst E., Tielens A. G. G. M., Caux E., 2007, Protostars and Planets V, 47

Ceccarelli C., Dominik C., Lefloch B., Caselli P., Caux E., 2004, ApJ, 607, L51

Cernicharo J., Marcelino N., Roueff E., Gerin M., Jiménez-Escobar A., Muñoz Caro

G. M., 2012, ApJ, 759, L43

Cesaroni R., 2005, in IAU Symposium, Vol. 227, Massive Star Birth: A Crossroads of Astrophysics, Cesaroni R., Felli M., Churchwell E., Walmsley M., eds., pp. 59–69

Cesaroni R., 2019, A&A, 631, A65

Chandrasekhar S., Fermi E., 1953, ApJ, 118, 116

Charnley S. B., Tielens A. G. G. M., Rodgers S. D., 1997, ApJ, 482, L203

Choudhury R., Schilke P., Stéphan G., Bergin E., Möller T., Schmiedeke A., Zernickel A., 2015, A&A, 575, A68

Christie H. et al., 2012, Monthly Notices of the Royal Astronomical Society, 422, 968

Cimatti A., Fraternali F., Nipoti C., 2019, Introduction to Galaxy Formation and Evolution: From Primordial Gas to Present-Day Galaxies. Cambridge University Press

Clark P. C., Glover S. C. O., Ragan S. E., Duarte-Cabral A., 2019, MNRAS, 486, 4622

Coletta A., Fontani F., Rivilla V. M., Mininni C., Colzi L., Sánchez-Monge Á., Beltrán M. T., 2020, A&A, 641, A54

Contreras Y. et al., 2013, A&A, 549, A45

Cosentino G. et al., 2018, MNRAS, 474, 3760

Crapsi A., Caselli P., Walmsley C. M., Myers P. C., Tafalla M., Lee C. W., Bourke T. L., 2005, ApJ, 619, 379

Crowther P., 2012, Astronomy & Geophysics, 53, 4.30

Csengeri T. et al., 2018, A&A, 617, A89

Csengeri T. et al., 2016, A&A, 586, A149

Csengeri T. et al., 2014, A&A, 565, A75

Cuppen H. M., Walsh C., Lamberts T., Semenov D., Garrod R. T., Penteado E. M., Ioppolo S., 2017, Space Sci. Rev., 212, 1

Dalgarno A., Lepp S., 1984, ApJ, 287, L47

de Boisanger C., Helmich F. P., van Dishoeck E. F., 1996, A&A, 310, 315

Di Francesco J., Johnstone D., Kirk H., MacKenzie T., Ledwosinska E., 2008, ApJS, 175, 277

Doty S. D., van Dishoeck E. F., van der Tak F. F. S., Boonman A. M. S., 2002, A&A, 389, 446

Douglas A. E., Herzberg G., 1941, ApJ, 94, 381

Draine B. T., 1978, ApJS, 36, 595

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton Series in Astrophysics

Draine B. T., Lee H. M., 1984, ApJ, 285, 89

Ebert R., 1955, Z. Astrophys., 37, 217

Egan M. P., Shipman R. F., Price S. D., Carey S. J., Clark F. O., Cohen M., 1998, ApJ, 494, L199

Ehrenfreund P., Charnley S. B., 2000, ARA&A, 38, 427

Elmegreen B. G., 1998, in Astronomical Society of the Pacific Conference Series, Vol. 148, Origins, Woodward C. E., Shull J. M., Thronson Harley A. J., eds., p. 150

Ercolano B., Barlow M. J., Storey P. J., 2005, MNRAS, 362, 1038

Ercolano B., Barlow M. J., Storey P. J., Liu X. W., 2003, MNRAS, 340, 1136

Faúndez S., Bronfman L., Garay G., Chini R., Nyman L.-Å., May J., 2004, A&A, 426, 97

Feng S., Beuther H., Zhang Q., Henning T., Linz H., Ragan S., Smith R., 2016, A&A, 592, A21

Feng S. et al., 2020, ApJ, 901, 145

Foing B. H., Ehrenfreund P., 1994, Nature, 369, 296

Fontani F., Giannetti A., Beltrán M. T., Dodson R., Rioja M., Brand J., Caselli P., Cesaroni R., 2012, MNRAS, 423, 2342

Fontani F. et al., 2011, A&A, 529, L7

Fontani F., Pascucci I., Caselli P., Wyrowski F., Cesaroni R., Walmsley C. M., 2007, A&A, 470, 639

Ford A. B., Shirley Y. L., 2011, ApJ, 728, 144

Fraser H. J., McCoustra M. R. S., Williams D. A., 2002, Astronomy and Geophysics, 43, 2.10

Friesen R. K., Di Francesco J., Bourke T. L., Caselli P., Jørgensen J. K., Pineda J. E., Wong M., 2014, ApJ, 797, 27

Friesen R. K., Di Francesco J., Myers P. C., Belloche A., Shirley Y. L., Bourke T. L., André P., 2010, ApJ, 718, 666

Fuchs G. W., Cuppen H. M., Ioppolo S., Romanzin C., Bisschop S. E., Andersson S., van Dishoeck E. F., Linnartz H., 2009, A&A, 505, 629

Galli D., Palla F., 2013, ARA&A, 51, 163

Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485

Garrod R. T., Herbst E., 2006, A&A, 457, 927

Garrod R. T., Pauly T., 2011, ApJ, 735, 15

Garrod R. T., Wakelam V., Herbst E., 2007, A&A, 467, 1103

Geppert W. D. et al., 2006, Faraday Discuss., 133, 177

Gerlich D., Herbst E., Roueff E., 2002, Planetary and Space Science, 50, 1275 , special issue on Deuterium in the Universe

Gerner T., Beuther H., Semenov D., Linz H., Vasyunina T., Bihr S., Shirley Y. L., Henning

T., 2014, A&A, 563, A97

Giannetti A. et al., 2019, A&A, 621, L7

Giannetti A. et al., 2017a, A&A, 606, L12

Giannetti A., Leurini S., Wyrowski F., Urquhart J., Csengeri T., Menten K. M., König C., Güsten R., 2017b, A&A, 603, A33

Giannetti A. et al., 2014, A&A, 570, A65

Gibb A. G., Davis C. J., 1998, MNRAS, 298, 644

Glassgold A. E., Langer W. D., 1974, ApJ, 193, 73

Grassi T., Bovino S., Schleicher D. R. G., Prieto J., Seifried D., Simoncini E., Gianturco F. A., 2014, MNRAS, 439, 2386

Griffin M. J. et al., 2010, A&A, 518, L3

Guenthner K., Kauffmann J., 2009, in American Astronomical Society Meeting Abstracts, Vol. 213, American Astronomical Society Meeting Abstracts #213, p. 413.14

Guilloteau S. et al., 1992, A&A, 262, 624

Güsten R., Nyman L. Å., Schilke P., Menten K., Cesarsky C., Booth R., 2006, A&A, 454, L13

Harju J. et al., 2006, A&A, 454, L55

Harju J., Juvela M., Schlemmer S., Haikala L. K., Lehtinen K., Mattila K., 2008, A&A, 482, 535

Hasegawa T. I., Herbst E., 1993, MNRAS, 261, 83

Hasegawa T. I., Herbst E., Leung C. M., 1992, ApJS, 82, 167

Hastings W. K., 1970, j-BIOMETRIKA, 57, 97

Herbst E., van Dishoeck E. F., 2009, ARA&A, 47, 427

Hernandez A. K., Tan J. C., Caselli P., Butler M. J., Jiménez-Serra I., Fontani F., Barnes P., 2011, ApJ, 738, 11

Hernandez A. K., Tan J. C., Kainulainen J., Caselli P., Butler M. J., Jiménez-Serra I., Fontani F., 2012, The Astrophysical Journal Letters, 756, L13

Herpin F. et al., 2012, A&A, 542, A76

Heyer M., Dame T. M., 2015, ARA&A, 53, 583

Heyer M., Gutermuth R., Urquhart J. S., Csengeri T., Wienen M., Leurini S., Menten K., Wyrowski F., 2016, A&A, 588, A29

Hezareh T., Houde M., McCoey C., Vastel C., Peng R., 2008, ApJ, 684, 1221

Hickson K. M., Wakelam V., Loison J.-C., 2016, Molecular Astrophysics, 3, 1

Hildebrand R. H., 1983, Quarterly Journal of the Royal Astronomical Society, 24, 267

Ho P. T. P., Moran J. M., Lo K. Y., 2004, ApJ, 616, L1

Hoare M. G. et al., 2012, PASP, 124, 939

Hocuk S., Cazaux S., 2015, A&A, 576, A49

Hocuk S., Cazaux S., Spaans M., Caselli P., 2016, MNRAS, 456, 2586

Hofner P., Wyrowski F., Walmsley C. M., Churchwell E., 2000, ApJ, 536, 393

Holland W. S. et al., 1999, MNRAS, 303, 659

Hollenbach D., McKee C. F., 1979, ApJS, 41, 555

Hoq S. et al., 2013, ApJ, 777, 157

Hosokawa T., Omukai K., 2009, ApJ, 691, 823

Hugo E., Asvany O., Harju J., Schlemmer S., 2007, in Molecules in Space and Laboratory, p. 119

Immer K., Li J., Quiroga-Nuñez L. H., Reid M. J., Zhang B., Moscadelli L., Rygl K. L. J., 2019, A&A, 632, A123

Indriolo N. et al., 2015, ApJ, 800, 40

Ioppolo S., Cuppen H. M., Romanzin C., van Dishoeck E. F., Linnartz H., 2008, ApJ, 686, 1474

Isokoski K., Bottinelli S., van Dishoeck E. F., 2013, A&A, 554, A100

Ivlev A. V., Padovani M., Galli D., Caselli P., 2015, ApJ, 812, 135

Ivlev A. V., Silsbee K., Sipilä O., Caselli P., 2019, ApJ, 884, 176

Jeffreys H., 1961, Theory of probability, 3rd edn. Oxford University Press, Oxford, England

Jiménez-Serra I. et al., 2016, ApJ, 830, L6

Johnston K. G. et al., 2020, A&A, 634, L11

Jolliffe I. T., Cadima J., 2016, Philosophical Transactions of the Royal Society of London Series A, 374, 20150202

Jorgensen J. K., Belloche A., Garrod R. T., 2020, arXiv e-prints, arXiv:2006.07071

Jørgensen J. K., Schöier F. L., van Dishoeck E. F., 2004, A&A, 416, 603

Kahn F. D., 1974, A&A, 37, 149

Kauffmann J., Pillai T., 2010, ApJ, 723, L7

Kauffmann J., Pillai T., Goldsmith P. F., 2013, ApJ, 779, 185

Kennicutt R. C., 2005, in Massive Star Birth: A Crossroads of Astrophysics, Cesaroni R., Felli M., Churchwell E., Walmsley M., eds., Vol. 227, pp. 3–11

Kippenhahn R., Weigert A., Weiss A., 2012, Stellar Structure and Evolution. Astronomy and Astrophysics Library; Springer

Klassen M., Peters T., Pudritz R. E., 2012, ApJ, 758, 137

Klein B., Hochgürtel S., Krämer I., Bell A., Meyer K., Güsten R., 2012, A&A, 542, L3

Klein T. et al., 2014, IEEE Transactions on Terahertz Science and Technology, 4, 588

Kobayashi C., Karakas A. I., Lugaro M., 2020, ApJ, 900, 179

Kolasinski K., 2012, Surface Science: Foundations of Catalysis and Nanoscience, Third Edition, Vol. 4

Kong S. et al., 2016, ApJ, 821, 94

König C. et al., 2017, A&A, 599, A139

Könyves V. et al., 2020, A&A, 635, A34

Körtgen B., Bovino S., Schleicher D. R. G., Giannetti A., Banerjee R., 2017, MNRAS, 469, 2602

Körtgen B., Bovino S., Schleicher D. R. G., Stutz A., Banerjee R., Giannetti A., Leurini S., 2018, MNRAS, 478, 95

Kramer C., Alves J., Lada C. J., Lada E. A., Sievers A., Ungerechts H., Walmsley C. M., 1999, A&A, 342, 257

Kramer C., Winnewisser G., 1991, Astronomy and Astrophysics Supplement Series, 89, 421

Kulterer B. M., Drozdovskaya M. N., Coutens A., Manigand S., Stéphan G., 2020, arXiv e-prints, arXiv:2008.04798

Kumar M. S. N., Palmeirim P., Arzoumanian D., Inutsuka S. I., 2020, A&A, 642, A87

Kurtz S., Cesaroni R., Churchwell E., Hofner P., Walmsley C. M., 2000, in Protostars and Planets IV, Mannings V., Boss A. P., Russell S. S., eds., pp. 299–326

Kwon W., Looney L. W., Crutcher R. M., Kirk J. M., 2006, ApJ, 653, 1358

Lacy J. H., Knacke R., Geballe T. R., Tokunaga A. T., 1994, ApJ, 428, L69

Lada C. J., 1987, in Star Forming Regions, Peimbert M., Jugaku J., eds., Vol. 115, p. 1

Ladd E. F., 2004, ApJ, 610, 320

Lee, M.-Y. ., et al., in prep.

Leurini S., Codella C., Zapata L., Beltrán M. T., Schilke P., Cesaroni R., 2011a, A&A, 530, A12

Leurini S., Pillai T., Stanke T., Wyrowski F., Testi L., Schuller F., Menten K. M., Thorwirth S., 2011b, A&A, 533, A85

Leurini S. et al., 2019, A&A, 621, A130

Li G.-X., Urquhart J. S., Leurini S., Csengeri T., Wyrowski F., Menten K. M., Schuller F., 2016, A&A, 591, A5

Li P., Chen S., 2016, CAAI Transactions on Intelligence Technology, 1, 366

Lis D. C., Gerin M., Roueff E., Vastel C., Phillips T. G., 2006, ApJ, 636, 916

Liu X.-L., Xu J.-L., Ning C.-C., Zhang C.-P., Liu X.-T., 2018, Research in Astronomy and Astrophysics, 18, 004

Loison J.-C., Wakelam V., Hickson K. M., 2014, Monthly Notices of the Royal Astronomical Society, 443, 398

Lubowich D. A., Pasachoff J. M., Balonek T. J., Millar T. J., Tremonti C., Roberts H., Galloway R. P., 2000, Nature, 405, 1025

MacLaren I., Richardson K. M., Wolfendale A. W., 1988, ApJ, 333, 821

Maret S., Hily-Blant P., Pety J., Bardeau S., Reynier E., 2011, A&A, 526, A47

Martin R. N., Barrett A. H., 1978, ApJS, 36, 1

Maud L. T. et al., 2019, A&A, 627, L6

McCall B. J., Geballe T. R., Hinkle K. H., Oka T., 1999, ApJ, 522, 338

McElroy D., Walsh C., Markwick A. J., Cordiner M. A., Smith K., Millar T. J., 2013, A&A, 550, A36

McGuire B. A., 2018, The Astrophysical Journal Supplement Series, 239, 17

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

McKee C. F., Tan J. C., 2002, Nature, 416, 59

McKee C. F., Tan J. C., 2003, ApJ, 585, 850

McKellar A., 1940, PASP, 52, 312

Menten K. M., Pillai T., Wyrowski F., 2005, in Massive Star Birth: A Crossroads of Astrophysics, Cesaroni R., Felli M., Churchwell E., Walmsley M., eds., Vol. 227, pp. 23–34

Mercimek S., Myers P. C., Lee K. I., Sadavoy S. I., 2017, AJ, 153, 214

Metropolis A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, Journal of Chemical Physics, 21, 1087

Miettinen O., 2020, A&A, 634, A115

Miettinen O., Harju J., Haikala L. K., Pomrén C., 2006, A&A, 460, 721

Mininni C. et al., 2020, arXiv e-prints, arXiv:2009.13297

Molinari S., Merello M., Elia D., Cesaroni R., Testi L., Robitaille T., 2016, ApJ, 826, L8

Molinari S., Pezzuto S., Cesaroni R., Brand J., Faustini F., Testi L., 2008, A&A, 481, 345

Molinari S. et al., 2010, PASP, 122, 314

Morales Ortiz J. L., Ceccarelli C., Lis D. C., Olmi L., Plume R., Schilke P., 2014, A&A, 563, A127

Motte F., Bontemps S., Louvet F., 2018, Annual Review of Astronomy and Astrophysics, 56, 41

Mottram J. C. et al., 2011, ApJ, 730, L33

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

Navarete F. et al., 2019, A&A, 622, A135

Neugebauer G. et al., 1984, ApJ, 278, L1

Oberg K. I., Bergin E. A., 2020, arXiv e-prints, arXiv:2010.03529

Öberg K. I., Bottinelli S., Jørgensen J. K., van Dishoeck E. F., 2010, ApJ, 716, 825

Oliphant T., 2006, NumPy: A guide to NumPy. [Online; accessed <today>]

Oliveira C. M., Hébrard G., Howk J. C., Kruk J. W., Chayer P., Moos H. W., 2003, ApJ, 587, 235

Padoan P., Pan L., Juvela M., Haugbølle T., Nordlund Å., 2020, ApJ, 900, 82

Padovani M., Galli D., 2011, A&A, 530, A109

Padovani M., Galli D., Glassgold A. E., 2009, A&A, 501, 619

Padovani M., Hennebelle P., Galli D., 2013, A&A, 560, A114

Padovani M., Ivlev A. V., Galli D., Caselli P., 2018, A&A, 614, A111

Pagani L. et al., 2009, A&A, 494, 623

Patil A., Huard D., Fonnesbeck C., 2010, Journal of Statistical Software, Articles, 35, 1

Penteado E. M., Walsh C., Cuppen H. M., 2017, ApJ, 844, 71

Perault M. et al., 1996, A&A, 315, L165

Peretto N. et al., 2013, A&A, 555, A112

Peretto N. et al., 2010, A&A, 518, L98

Peretto N., Lenfestey C., Fuller G. A., Traficante A., Molinari S., Thompson M. A., Ward-Thompson D., 2016, A&A, 590, A72

Peters T., Banerjee R., Klessen R. S., Mac Low M.-M., 2011, ApJ, 729, 72

Pilbratt G. L. et al., 2010, A&A, 518, L1

Pillai T., Caselli P., Kauffmann J., Zhang Q., Thompson M. A., Lis D. C., 2012, ApJ, 751, 135

Pillai T., Kauffmann J., Zhang Q., Sanhueza P., Leurini S., Wang K., Sridharan T. K., König C., 2019, A&A, 622, A54

Platt J. C., 1999, in ADVANCES IN LARGE MARGIN CLASSIFIERS, MIT Press, pp. 61–74

Plummer H. C., 1911, MNRAS, 71, 460

Poglitsch A. et al., 2010, A&A, 518, L2

Pokhrel R. et al., 2018, ApJ, 853, 5

Price S. D., Egan M. P., Carey S. J., Mizuno D. R., Kuchar T. A., 2001, AJ, 121, 2819

Purcell C. R. et al., 2013, ApJS, 205, 1

Rathborne J. M., Jackson J. M., Simon R., 2006, ApJ, 641, 389

Reboussin L., Wakelam V., Guilloteau S., Hersant F., 2014, MNRAS, 440, 3557

Reid M. J. et al., 2014, ApJ, 783, 130

Retes-Romero R., Mayya Y. D., Luna A., Carrasco L., 2020, ApJ, 897, 53

Rigby A. J. et al., 2019, A&A, 632, A58

Rimola A., Taquet V., Ugliengo P., Balucani N., Ceccarelli C., 2014, A&A, 572, A70

Ripple F., Heyer M. H., Gutermuth R., Snell R. L., Brunt C. M., 2013, Monthly Notices of the Royal Astronomical Society, 431, 1296

Rivilla V. M., Beltrán M. T., Cesaroni R., Fontani F., Codella C., Zhang Q., 2017, A&A, 598, A59

Roberts H., Millar T. J., 2000, A&A, 361, 388

Ruaud M., Loison J. C., Hickson K. M., Gratier P., Hersant F., Wakelam V., 2015, MNRAS, 447, 4004

Russeil D., Zavagno A., Motte F., Schneider N., Bontemps S., Walsh A. J., 2010, A&A, 515, A55

Sabatini G. et al., 2021, subm., A&A

Sabatini G. et al., 2020, A&A, 644, A34

Sabatini G., Giannetti A., Bovino S., Brand J., Leurini S., Schisano E., Pillai T., Menten K. M., 2019, MNRAS, 490, 4489

Sadavoy S. I. et al., 2014, ApJ, 787, L18

Salaris M., Cassisi S., 2005, Evolution of Stars and Stellar Populations

Salpeter E. E., 1955, ApJ, 121, 161

Sanna A. et al., 2019, A&A, 623, A77

Saraceno P., Andre P., Ceccarelli C., Griffin M., Molinari S., 1996, A&A, 309, 827

Schisano E. et al., 2020, MNRAS, 492, 5420

Schisano E. et al., 2014, The Astrophysical Journal, 791, 27

Schneider N., Csengeri T., Bontemps S., Motte F., Simon R., Hennebelle P., Federrath C., Klessen R., 2010, A&A, 520, A49

Schneider N. et al., 2015, A&A, 575, A79

Schuller F. et al., 2009, A&A, 504, 415

Sellgren K., Uchida K. I., Werner M. W., 2007, The Astrophysical Journal, 659, 1338

Semenov D. et al., 2010, A&A, 522, A42

Sipilä O., Caselli P., Harju J., 2015, A&A, 578, A55

Skouteris D., Balucani N., Ceccarelli C., Faginas Lago N., Codella C., Falcinelli S., Rosi M., 2019, MNRAS, 482, 3567

Smartt S. J., 2009, ARA&A, 47, 63

Smith N., 2014, ARA&A, 52, 487

Spina L. et al., 2017, A&A, 601, A70

Spitzer, Lyman J., Tomasko M. G., 1968, ApJ, 152, 971

Spitzer L., 1978, Physical processes in the interstellar medium

Stahler S. W., Palla F., 2004, The Formation of Stars. Wiley-VCH

Strömgren B., 1939, ApJ, 89, 526

Swift J. J., 2009, ApJ, 705, 1456

Swings P., Rosenfeld L., 1937, ApJ, 86, 483

Tafalla M., Myers P. C., Caselli P., Walmsley C. M., 2004, Ap&SS, 292, 347

Tafalla M., Myers P. C., Caselli P., Walmsley C. M., Comito C., 2002, ApJ, 569, 815

Tan J. C., Beltrán M. T., Caselli P., Fontani F., Fuente A., Krumholz M. R., McKee C. F., Stolte A., 2014, in Protostars and Planets VI, Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds., p. 149

Tan J. C., Kong S., Zhang Y., Fontani F., Caselli P., Butler M. J., 2016, ApJ, 821, L3

Tang X. D. et al., 2018, A&A, 611, A6

Taniguchi K., Saito M., Sridharan T. K., Minamidani T., 2018, ApJ, 854, 133

Taniguchi K., Saito M., Sridharan T. K., Minamidani T., 2019, ApJ, 872, 154

Thomas H. S., Fuller G. A., 2008, A&A, 479, 751

Tielens A. G. G. M., 2010, The Physics and Chemistry of the Interstellar Medium

Tigé J. et al., 2017, A&A, 602, A77

Tobin J. J. et al., 2016, ApJ, 818, 73

Urquhart J. S., Busfield A. L., Hoare M. G., Lumsden S. L., Clarke A. J., Moore T. J. T., Mottram J. C., Oudmaijer R. D., 2007, A&A, 461, 11

Urquhart J. S. et al., 2014a, A&A, 568, A41

Urquhart J. S. et al., 2019, MNRAS, 484, 4444

Urquhart J. S., Figura C. C., Moore T. J. T., Hoare M. G., Lumsden S. L., Mottram J. C., Thompson M. A., Oudmaijer R. D., 2013, MNRAS, 437, 1791

Urquhart J. S. et al., 2018, MNRAS, 473, 1059

Urquhart J. S. et al., 2014b, MNRAS, 443, 1555

Urquhart J. S. et al., 2013, MNRAS, 435, 400

van der Tak F. F. S., 2005, in IAU Symposium, Vol. 227, Massive Star Birth: A Crossroads of Astrophysics, Cesaroni R., Felli M., Churchwell E., Walmsley M., eds., pp. 70–79

van der Tak F. F. S., van Dishoeck E. F., 2000, A&A, 358, L79

van der Tak F. F. S., van Dishoeck E. F., Evans, Neal J. I., Blake G. A., 2000, ApJ, 537, 283

Vastel C. et al., 2012, A&A, 547, A33

Vastel C., Caselli P., Ceccarelli C., Phillips T., Wiedner M. C., Peng R., Houde M., Dominik C., 2006, ApJ, 645, 1198

Vastel C., Ceccarelli C., Lefloch B., Bachiller R., 2014, ApJ, 795, L2

Vasyunin A. I., Caselli P., Dulieu F., Jiménez-Serra I., 2017, ApJ, 842, 33

Vasyunin A. I., Herbst E., 2013, The Astrophysical Journal, 769, 34

Vaupré S., Hily-Blant P., Ceccarelli C., Dubus G., Gabici S., Montmerle T., 2014, A&A, 568, A50

Vazart F., Ceccarelli C., Balucani N., Bianchi E., Skouteris D., 2020, arXiv e-prints, arXiv:2010.02718

Viti S., Williams D. A., 1999, MNRAS, 310, 517

Wakelam V., Loison J. C., Mereau R., Ruaud M., 2017, Molecular Astrophysics, 6, 22

Walsh A. J., Burton M. G., Hyland A. R., Robinson G., 1998, MNRAS, 301, 640

Wang K. S., Bourke T. L., Hogerheijde M. R., van der Tak F. F. S., Benz A. O., Megeath S. T., Wilson T. L., 2013, A&A, 558, A69

Wang P., Li Z.-Y., Abel T., Nakamura F., 2010, ApJ, 709, 27

Wang Y. et al., 2020, A&A, 641, A53

Werner M. W., Uchida K. I., Sellgren K., Marengo M., Gordon K. D., Morris P. W., Houck J. R., Stansberry J. A., 2004, The Astrophysical Journal Supplement Series, 154, 309

Wienen M. et al., 2015, A&A, 579, A91

Wienen M., Wyrowski F., Schuller F., Menten K. M., Walmsley C. M., Bronfman L., Motte F., 2012, A&A, 544, A146

Wienen M., Wyrowski F., Walmsley C. M., Csengeri T., Pillai T., Giannetti A., Menten K. M., 2021, arXiv e-prints, arXiv:2102.04478

Wiles B., Lo N., Redman M. P., Cunningham M. R., Jones P. A., Burton M. G., Bronfman L., 2016, MNRAS, 458, 3429

Winkel B. et al., 2017, A&A, 600, A2

Woon D. E., 2002, ApJ, 569, 541

Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181

Wootten A., Thompson A. R., 2009, IEEE Proceedings, 97, 1463

Wouterloot J. G. A., Henkel C., Brand J., Davis G. R., 2008, A&A, 487, 237

Wyrowski F. et al., 2016, A&A, 585, A149

Yan Y. T. et al., 2019, ApJ, 877, 154

Young E. T. et al., 2012, ApJ, 749, L17

Yu N., Wang J.-J., Xu J.-L., 2019, MNRAS, 489, 4497

Zhang Q., Ho P. T. P., Ohashi N., 1998, ApJ, 494, 636

Zhang Q., Wang Y., Pillai T., Rathborne J., 2009, ApJ, 696, 268

Zhu H., Tian W., Li A., Zhang M., 2017, MNRAS, 471, 3494

Zinnecker H., 2011, in Astronomical Society of the Pacific Conference Series, Vol. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function?, Treyer M., Wyder T., Neill J., Seibert M., Lee J., eds., p. 3

Zinnecker H., Yorke H. W., 2007, ARA&A, 45, 481

Zucker C., Battersby C., Goodman A., 2018, ApJ, 864, 153