A song of ice and gas: the formation and evolution of complex organic molecules

Chen, Y.

1 Introduction

The chemical world on Earth is astonishingly diverse. From water, air, and minerals to organic compounds, from biotic molecules such as proteins and nucleic acids to human-made polymers and medicines, the variety of substances is virtually limitless. This richness of chemistry forms the basis of the natural environment on Earth, nurtures and sustains life, and has been further amplified by human ingenuity. In contrast, when we turn our eyes to the vast universe, the picture seems remarkably simple at first glance. The majority of baryonic matter is composed of only two elements, hydrogen and helium, synthesized in the first few minutes after the Big Bang (Peebles 1993). All heavier elements, grouped together by astronomers under the umbrella term "metals", are produced much later in the interiors of stars and dispersed through stellar winds and supernovae (Burbidge et al. 1957), which only constitute a small fraction of cosmic material. From this perspective, one might think the universe is a chemically barren space.

However, this impression exists only because we have not lifted the veil on the true complexity of interstellar chemistry. Over the past half century, advances in molecular spectroscopy and observational techniques have revealed that interstellar chemistry is far more diverse than initially expected. As of Nov 2025, around 340 molecules have been detected in the interstellar medium (ISM) and circumstellar environments[1]. These species range from simple diatomics to complex organics composed of ten or more atoms, including but not limited to alcohols, aldehydes, acids, ethers, esters, ketones, and even aromatic compounds.

An interesting aspect of these discoveries is their strong connection with star formation. The majority of known interstellar molecules have been identified in the dense, shielded environments of molecular clouds and in the warm, chemically rich zones surrounding protostars — the so-called **hot cores** for high-mass protostars and **hot corinos** for low-mass ones (see reviews by Herbst & van Dishoeck 2009; Jørgensen et al. 2020; Ceccarelli et al. 2023); here we collectively refer to both as "hot cores". These regions act as natural laboratories where low temperatures and high densities combine to drive unique chemical processes that lead to the formation of a variety of interstellar ices on dust grains (Sects. 1.2.1–1.2.2) followed by their subsequent release into the gas phase (Sect. 1.2.3). The recognition that the interstellar medium is not merely a dilute mixture of atoms and simple molecules but rather a chemically complex environment has transformed our understanding of cosmic evolution, linking the physics of star formation with the origins of molecular complexity and, ultimately, the ingredients necessary for life.

To deepen our understanding of how chemical complexity arises and evolves in the Universe, this thesis makes use of the most powerful astronomical facilities to date, the

[1] https://cdms.astro.uni-koeln.de/classic/molecules

Atacama Large Millimeter/submillimeter Array (ALMA; Sect. 1.3.3.2) and the *James Webb* Space Telescope (JWST; Sect. 1.3.4.4), to study **complex organic molecules (COMs)**[2] in protostellar systems. Particular focus is placed on a subset of COMs that are sufficiently abundant to be routinely detected in space (see Table 1.1). These observations provide critical insights into the processes that shape the chemistry of the interstellar medium and its evolution through the protostellar and protoplanetary stages, and potentially to the siblings of our own Solar System.

1.1 Star formation

Star formation is a process in which cold molecular gas assembles and collapses under self-gravity to form protostars, disks, and eventually planetary systems. In addition to self-gravity, other factors including turbulence, magnetic fields, thermodynamics (heating and cooling), angular momentum transport, and stellar feedback (e.g., winds, outflows, and radiation) also regulate the dynamical processes of star formation. Relevant studies are often conducted under two regimes: low-mass and high-mass, with a dividing line[3] of $\sim$8 $M_\odot$. However, the recent review by Beuther et al. (2025) has shown multiple similarities in the properties of low- and high-mass star-forming regions, which weaken the bimodality of star formation to some extent.

1.1.1 Prestellar phase

Molecular clouds are the cradles of star formation. Cloud complexes span a few to hundreds of parsecs (pc) in size and typically have total masses of 10^3–10^5 $M_\odot$ (e.g., Kennicutt & Evans 2012). Zooming from cloud scales to sub-parsec scales, the seeds of star formation are embedded in the coldest (10–20 K) and densest (10^4–10^6 cm^{-3}) regions of molecular clouds called prestellar or starless cores (e.g., Bergin & Tafalla 2007). In the low-mass regime, these cores are smaller (0.01–0.1 pc) and less massive ($\sim$1 $M_\odot$) than their high-mass analogs (0.1–1 pc, 10^2–10^3 $M_\odot$), which are often embedded in infrared dark clouds (IRDCs) and difficult to be identified due to their short lifetime, large distances, and crowded environments (e.g., Motte et al. 2018).

Prestellar cores have a centrally concentrated density profile and will meet the fate of gravitational collapse once the Jeans criterion (Jeans 1902) is met, that is, the density and mass are high enough for self-gravity to overcome the resistance of internal gas pressure and other support from turbulence and magnetic fields. During the collapse of a prestellar core, its inner regions contract faster than the outer layers, because the free-fall time $t_{\rm ff} \propto \rho^{-1/2}$ is shorter where the density is higher. As a result, the center quickly becomes opaque to its own cooling radiation, raising the internal pressure and forming a small hydrostatic object called the first hydrostatic core (FHSC; Larson 1969; Masunaga et al. 1998). As material continues to rain in, the

[2] COMs, or interstellar COMs (iCOMs), are initially defined by Herbst & van Dishoeck (2009) as molecules containing six or more atoms (i.e., complex), one of which is carbon (i.e., organic). In practice, however, some five-atom molecules such as formic acid (HCOOH) and ketene (CH_2CO) are also considered part of the broader class of COMs because of their strong chemical connections to larger organic molecules (e.g., Sect. 1.3.1).

[3] There is also a more stringent dividing line of 2 $M_\odot$ for low-mass stars, and those with masses between 2 and 8 $M_\odot$ are classified as intermediate-mass stars, also known as Herbig Ae/Be stars. Here we adopt a loosely defined dividing line of 8 $M_\odot$ between low- and high-mass stars.

central density and temperature rise to a point ($\rho_c \sim 10^{-7}\,\mathrm{g\,cm^{-3}}$ and $T_c \sim 2000\,\mathrm{K}$) where hydrogen molecules begin to dissociate and cool down the core, triggering a second, rapid collapse that leads to the formation of a protostar (Masunaga & Inutsuka 2000). At the same time, angular momentum naturally redirects some infalling gas into a rotating disk while bipolar outflows remove excess angular momentum and clear envelope cavities, marking the transformation from a cold prestellar core to an actively accreting protostellar system.

Given the differences in evolutionary timescales and physical properties (e.g., radiation intensity), the following protostellar and pre-main-sequence phases of star formation will be introduced separately for the low- and high-mass regimes.

1.1.2 Low-mass star formation

The processes of low-mass star formation after the prestellar phase are well described by the observational **Class** system and the physical **Stage** framework. Back to the 1970s, analytic models such as "inside-out collapse" of a singular isothermal sphere (Shu 1977) provided a canonical picture of how a molecular cloud core evolves into a protostar, which was further synthesized into a sequence of phases including the collapse of dense cores, the growth of protostars via accretion, the formation of circumstellar disks, and the eventual dispersal of envelopes (Shu et al. 1987). These theories were not tied to concrete observational diagnostics until the launch of the Infrared Astronomical Satellite (IRAS) in 1983, which provided unprecedented amounts of infrared (IR) photometry data of young stellar objects (YSOs) that are hardly visible at optical wavelengths. An observational classification system was then built by Lada (1987) upon these IR surveys using the slope of the spectral energy distribution (SED) between 2 and 25 μm:

$$\alpha_{\mathrm{IR}} = \frac{\mathrm{d}\log(\lambda F_\lambda)}{\mathrm{d}\log\lambda}. \tag{1.1}$$

In general, the IR sources were divided into three classes: Class I objects have rising SEDs longward of 2 μm that are apparently broader than blackbodies; Class II objects have flatter or declining SEDs with a still prominent IR excess; Class III objects have nearly photospheric SEDs with little IR excess. The IR excess in Class I, II, and III objects was suggested to arise from the infalling envelopes, the accreting circumstellar disks, and the leftover grains of the dispersed disks, respectively.

Soon afterward, submillimeter (sub-mm) continuum observations revealed sources with strong long-wavelength emission but essentially no near-IR counterpart, and a new "Class 0" category was introduced to describe these deeply embedded protostars (André et al. 1993). Class 0 objects are characterized by their large contribution of sub-mm luminosity ($L_{\mathrm{submm}}/L_{\mathrm{bol}} > 0.5\%$), implying that the bulk of the system mass still resides in a cold envelope rather than in the central protostar.

In the mid-1990s, the Class system was further refined by Greene et al. (1994), who carefully recalibrated the boundaries in spectral slope and introduced the "flat-spectrum" category to bridge the gap between Class I and Class II sources. They also codified the ranges of the IR spectral index α_{IR} for different classes: Class I ($\alpha_{\mathrm{IR}} > 0.3$), flat-spectrum ($-0.3 \leq \alpha_{\mathrm{IR}} \leq 0.3$), Class II ($-1.6 \leq \alpha_{\mathrm{IR}} \leq -0.3$), and Class III ($\alpha_{\mathrm{IR}} < -1.6$). Meanwhile, Chen et al. (1995) proposed to use bolometric

temperature ($T_{\rm bol}$) as another diagnostic of evolutionary stages, where $T_{\rm bol}$ was defined as the effective temperature of a blackbody with the same mean frequency as the observed SED. Their thresholds, <70 K for Class 0, 70–650 K for Class I, and 650–2800 K for Class II, became widely adopted in statistical studies such as the large surveys with the Infrared Space Observatory (ISO) and *Spitzer* Space Telescope (e.g., Evans et al. 2009; Enoch et al. 2009).

Although this empirical scheme offered a practical way to organize the newly accessible population of YSOs and mapped neatly onto the theoretical picture of evolving protostellar systems, a critical limitation was revealed that the observational classification did not always reflect the true physical state of a source. The apparent SED of a protostellar system is sensitive to its inclination, extinction, and outflow cavity geometry (e.g., Whitney et al. 2003b; Crapsi et al. 2008). For example, an edge-on disk around a T Tauri star (intrinsically Class II) could mimic the rising SED of a Class I protostar, while a pole-on Class I source could resemble a flat-spectrum or Class II object. This degeneracy motivated theoretical studies with radiative transfer models that could separate appearance from physical conditions. Based on Whitney et al. (2003a), Robitaille et al. (2006) created a large grid of modeled SEDs and introduced the **Stage** classification system, in which Stage 0/I objects are characterized by significant envelope accretion rates ($\dot{M}_{\rm env}/M_\star > 10^{-6}\,{\rm yr}^{-1}$), Stage II objects have little envelope but retain substantial disks ($M_{\rm disk}/M_\star > 10^{-6}$ at least, usually observed to be 10^{-3}–10^{-2}), and Stage III objects have negligible envelopes and disks ($M_\star \gg M_{\rm disk}, M_{\rm env}$). André et al. (2014) further separated Stage 0 from Stage I by the relative mass of envelope versus star and disk ($M_{\rm env} > M_\star + M_{\rm disk}$), giving a physical foundation to the empirical Class 0 definition. The timescales of different stages were estimated by large IR surveys with support and verification by theoretical models. Statistically speaking, the embedded protostellar Stages (0 and I) are relatively short (∼0.5 Myr in total; André et al. 2000; Enoch et al. 2009; Evans et al. 2009) compared to the disk-bearing Stages (II and III), which typically last for 1–3 Myr and 5–10 Myr, respectively (e.g., Haisch et al. 2001; Hernández et al. 2007; Fedele et al. 2010; Williams & Cieza 2011; Ribas et al. 2014).

In summary, the classification of YSOs into observational *Classes* and physical *Stages* converted a once qualitative picture of low-mass star formation into a coherent evolutionary sequence from cold dense cores to pre-main-sequence stars. Despite the distinctions between the two frameworks, they both contribute to advancing our understanding of low-mass star formation, with the Class system offering a practical observational taxonomy and the Stage system providing the physical context needed to interpret those observations.

1.1.3 High-mass star formation

Compared to the low-mass regime, massive stars form in denser, more distant, and highly clustered environments over shorter timescales ($\sim 10^5$ yr), making the observations of embedded prestellar cores very challenging (see reviews by Beuther et al. 2007; Zinnecker & Yorke 2007; Tan et al. 2014; Motte et al. 2018). Although the broad pictures of gravitational collapse, disk-mediated accretion, and bipolar outflows are shared with low-mass star formation, massive protostars typically reach the main sequence while still accreting, and their radiative and mechanical feedback rapidly reshapes their surroundings. As a result, observers often speak of an approximate

sequence from infrared-dark clumps and massive dense cores, through chemically enriched hot molecular cores, to hyper/ultracompact H II regions, and finally compact H II regions around emerging OB clusters.

How massive stars form has been under active debate and development. One of the dominant theories is **core accretion** (or **monolithic collapse**), where massive stars condense from high-surface-density cores and their envelopes collapse through massive, self-gravitating disks at high accretion rates (McKee & Tan 2003). This scenario is sometimes considered as a scale-up version of low-mass star formation, but only to the first order. Another popular theory is **competitive accretion**, where massive stars begin as lower-mass seeds within a common gravitational potential, and gain the bulk of their mass from the surrounding clump via Bondi–Hoyle accretion (Bonnell et al. 2001; Bonnell & Bate 2006). The competitive accretion theory was further extended into the **global hierarchical collapse** (GHC) theory to explain the observed "ridge and hub" structures in high-mass star-forming (HMSF) regions (Motte et al. 2018). In GHC, the classical Bondi–Hoyle accretion from a quasi-uniform background is upgraded to a non-homologous, multi-scale gravitational contraction of the entire cloud, in which collapse proceeds faster on smaller scales, and gas inflows from clouds through filaments (ridges) to clumps and cores (hubs; Vázquez-Semadeni et al. 2017, 2019; Motte et al. 2018).

The extreme physical conditions associated with high-mass star formation (e.g., high densities, intense radiation fields, strong turbulence, and rapid evolutionary timescales) can give rise to chemical environments that are very different from those of low-mass protostars. In particular, the high temperatures and ultraviolet (UV) radiation rates around massive protostars are expected to facilitate a wide range of chemical processes. In this context, high-mass star-forming regions can serve as a testbed for investigating how chemistry responds to different physical conditions during star formation.

1.2 Chemistry in early stages of star formation

In this section, chemical processes during the early stages of star formation are introduced. Considering the scope of this thesis and the fact that a large fraction of interstellar molecules were detected toward prestellar cores and hot cores around protostars, the chemistry occurring in other components (e.g., disks and outflows) and later stages (e.g., later than Class I) are left out. The chemical evolution will be divided into three phases, which are applicable in both low- and high-mass regimes, with the main difference lying in timescales.

1.2.1 Cold prestellar and collapsing phase

As the diffuse ISM evolves into dense clouds ($n \gtrsim 10^4$ cm^{-3}), the interstellar UV radiation is efficiently shielded. With the attenuation of photoionization and photodissociation, the dominant carbon species in the gas phase shifts from C^+ to C, and finally to CO (Leung et al. 1984; van Dishoeck & Black 1986; Snow & McCall 2006). Without photoelectric heating, the shielded interiors of dark clouds can be cooled down to ∼10 K by both dust and molecular gas (Goldsmith 2001).

At such low temperatures, gas-phase chemistry is mostly driven by cosmic rays (CRs), given that ion-neutral reactions are usually exothermic and barrierless (Herbst & Klemperer 1973). Neutral-neutral reactions are usually endothermic or have considerable activation barriers, which are very inefficient under cold conditions. A typical starting point of CR-induced gas-phase chemistry is H_3^+ produced by ionization of H_2 via $H_2 + CR \rightarrow H_2^+ + e^- + CR$ and $H_2^+ + H_2 \rightarrow H_3^+ + H$. More complex molecular ions are then formed through $H_3^+ + X \rightarrow XH^+ + H_2$. These proton-transfer reactions are exothermic and barrierless, and therefore can proceed very efficiently at ∼10 K. In particular, the proton transfer with HD leads to isotopic exchange (Watson 1974):

$$H_3^+ + HD \rightleftharpoons H_2D^+ + H_2, \tag{1.2}$$

which is exothermic in the forward direction by 232 K. The H_2D^+ ions can further transfer D to many other molecules via ion-neutral reactions, and hence promote the overall deuteration in the chemical network (see review by Ceccarelli et al. 2014). On the other hand, many hydrogenation reactions of molecular ions in the form of $X^+ + H_2 \rightarrow XH^+ + H$ are endothermic or possesses a potential barrier, thus are suppressed at low temperatures. As a result, many of the molecules present in cold dense clouds are unsaturated, which has been observed toward several prestellar cores (e.g., Spezzano et al. 2017; Xue et al. 2025).

As gas-phase chemistry evolves, a fraction of the gas-phase material begins to freeze onto dust grains at low temperatures; only the lightest species (e.g., H, H_2, and He) can sublimate efficiently at ∼10 K and thus remain in the gas phase. The adsorbed atoms and molecules, either via physisorption or chemisorption, then contribute to ice formation through three main reaction mechanisms: Langmuir-Hinshelwood (L-H), Eley-Rideal (E-R), and hot-atom. In the L-H mechanism, also known as the diffusive mechanism, reactants that are originally adsorbed at different potential minima (i.e., binding sites) have sufficient mobility to meet each other by thermal hopping or quantum tunneling over the energy barrier between binding sites. In the E-R mechanism, a gas-phase species is directly adsorbed on top of an existing adsorbate and reacts with it. In the hot-atom mechanism, a gas-phase species lands on a vacant binding site but is able to encounter a nearby adsorbate after brief diffusion. Among the three mechanisms, L-H (diffusive) has been the most studied and implemented one, but recent modeling studies call attention to the E-R (non-diffusive) mechanism in the ice formation of complex molecules, given the limited diffusion efficiency of heavy radicals at low temperatures (Jin & Garrod 2020; Garrod & Herbst 2006; also see Sect. 1.3.2).

Figure 1.1 summarizes the grain-surface chemistry in the prestellar and protostellar stages. Water (H_2O) is the first ice species that forms and accumulates on dust grains[4]. As the visual extinction (A_V) reaches the threshold of ∼1.6 mag (Whittet et al. 2001), H_2O ice can form efficiently via hydrogenation of O atoms (Tielens & Hagen 1982; Dulieu et al. 2010) and builds up a H_2O-rich layer on bare grains. A H_2O-rich environment is also referred to as a **polar** environment, indicating that the ice matrix is made of molecules with large permanent dipole moments. In the same epoch, NH_3 and CH_4 form by successive H-addition to the accreted N and C atoms (e.g., Hidaka et al. 2011; Qasim et al. 2020). Once CO forms abundantly in the gas

[4]Molecular hydrogen is the first molecule that forms on dust grains, but it promptly desorbs after forming exothermically and never builds up an ice mantle.

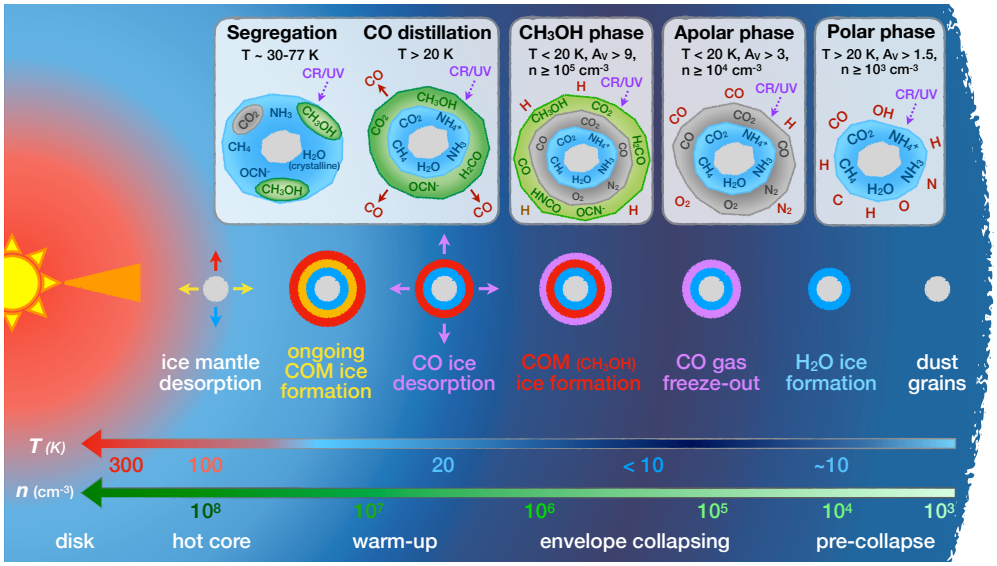

Figure 1.1: Schematic of the grain-surface chemistry in the prestellar and protostellar stages. From right to left, the diagram illustrates the formation and evolution of interstellar ices on dust grains from early to late stages. Different chemical phases are shown in greater detail at the top. Physical stages during star formation along with their typical temperatures and densities are indicated at the bottom. Adapted from Fig. 14 in Herbst & van Dishoeck (2009) and Fig. 10 in Boogert et al. (2015).

phase and gradually accretes onto dust grains, CO_2 also begins to form in the H_2O-rich layer via CO + OH $\rightarrow$ CO_2 + H[5] and CO + O (Watanabe & Kouchi 2002; Ioppolo et al. 2011; Minissale et al. 2013; Garrod & Pauly 2011). The freeze-out of gas-phase CO speeds up as density increases toward the cloud center, and when density exceeds $\sim 10^5$ cm^{-3}, CO freeze-out becomes so efficient that a CO-rich (**apolar**) ice layer is built upon the H_2O-rich one, with the gas-phase CO depleted significantly (Caselli et al. 1999). This process, known as "catastrophic CO freeze-out", leads to substantial accumulation of CO ices, which further boosts the formation of CO_2 (Pontoppidan 2006) and a series of hydrogenated products such as H_2CO and CH_3OH (e.g., Watanabe & Kouchi 2002; Fuchs et al. 2009; Cuppen et al. 2009). During this cold phase, the grain-surface chemistry is mostly driven by the diffusion of H atoms and CR-induced chemistry (e.g., bombardment of energetic particles and irradiation of secondary UV fields), which produce most of the commonly observed interstellar ices, except that ions such as NH_4^+, OCN^-, and $HCOO^-$ are formed via acid-base reactions between NH_3 and HNCO or HCOOH (Raunier et al. 2004; van Broekhuizen et al. 2004).

[5]Although this reaction has long been considered the main formation pathway of CO_2 ice, recent theoretical and laboratory studies show that it mainly produces HOCO and requires subsequent radical reactions to yield CO_2 (Molpeceres et al. 2023; Ishibashi et al. 2024).

1.2.2 Warm-up phase

Once a protostar is born at the center of the collapsed prestellar core (see Sect. 1.1.1), the envelope is gradually heated from the inside out. As temperature increases inward from ∼10 K to over 100 K, thermal desorption (or sublimation) of mixed ices proceeds in successive steps rather than in a one-off manner, which has been observed in temperature-programmed desorption (TPD) experiments[6] (e.g., Collings et al. 2004; Minissale et al. 2022; Kruczkiewicz et al. 2024). Species with higher binding energies[7] (E_b) tend to sublimate later at higher temperatures, although in some cases more volatile species (i.e., those with smaller E_b) can be trapped by H_2O—one of the least volatile common ices. As temperature reaches ∼20 K, apolar ices such as CO, CH_4, N_2, and O_2 start to leave their original ice matrices and sublimate from dust grains. In particular, the distillation of CO from CO_2 ice is observable via the CO_2 band at 15.2 µm, which develops a double-peak profile due to the emergence of the narrow band of pure CO_2. At higher temperatures (30 K to ∼77 K; Boogert et al. 2000; Öberg et al. 2009a), CO_2 further segregates from H_2O-rich ices but sublimates later at ∼45–90 K due to the trapping of H_2O (Fayolle et al. 2011).

While volatile species successively desorb from dust grains, the increased temperature also stimulates the formation of more complex molecules via the diffusion of heavier radicals. For example, CH_3OCHO can be formed through the recombination between HCO and CH_3O radicals (Garrod & Herbst 2006); similar routes are suggested for CH_3OCH_3 (via CH_3 + CH_3O) and HCOOH (via HCO + OH). However, the picture of COM formation is more complicated than diffusive grain-surface chemistry in the warm-up phase; a more detailed introduction is provided in Sect. 1.3.2. Regardless of how complex molecules are formed, they are likely trapped in mixtures with H_2O and CH_3OH (see Chapter 4). As a result, these trapped ices are released into the gas phase via co-desorption with H_2O and CH_3OH at ∼100 K, at which temperature all volatiles sublimate from dust grains.

1.2.3 Hot core phase

In compact regions around protostars where the temperature exceeds 100 K (∼10^2–10^3 au; i.e., hot cores), no volatile ices are left on dust grains, and the chemistry shifts decisively to the gas phase. This is the most popular phase for astrochemical observations in which a large fraction of gas-phase molecules, especially COMs, are discovered and studied to infer their formation mechanisms and evolution histories (see Sect. 1.3.3.3).

The high temperature in hot cores (100–300 K) permits a richer network of gas-phase reactions to occur than in cold prestellar cores, including those that are endothermic and have energy barriers. However, many gas-phase reactions proposed for COM formation were found to be inefficient to explain the observed abundances (e.g., Horn et al. 2004). There has also been a long-standing debate on the relative importance of gas-phase and grain-surface chemistries to COM formation, which is introduced in more detail in Sect. 1.3.2. Nevertheless, it is possible that the ice inven-

[6] In a TPD experiment, molecules are deposited first on a cold surface and then heated steadily, during which their desorption is monitored by a mass spectrometer or other detectors.

[7] Binding energy, also known as desorption energy (E_des), is the energy barrier that a molecule must overcome to leave the surface.

tories released from dust grains will be reprocessed by the gas-phase chemistry, which should be taken into account when interpreting the observational results of gas-phase molecules.

Hot cores are known for their richness of COMs. A few protostars, however, are found to harbor rich carbon-chain molecules that are dominant in prestellar clouds but deficient in hot cores (e.g., Sakai et al. 2008, 2009). This distinct chemical inventory is explained by the warm carbon chain chemistry (WCCC; Sakai & Yamamoto 2013), which is suggested to originate from a short prestellar phase. If the gravitational collapse proceeds fast enough, a considerable fraction of carbon atoms in the gas phase will be directly accreted onto dust grains before being converted into CO. These C atoms are then hydrogenated into CH_4 molecules on grain surfaces, which sublimate at 20–30 K early in the warm-up phase and trigger a series of gas-phase reactions to form carbon-chain molecules (Sakai & Yamamoto 2013). The lack of CO in the ingredients of grain-surface chemistry also lowers the production of common O-bearing COMs via hydrogenation of CO.

Toward the end of the hot core phase, materials in the protostellar envelope are either dissipated or accreted onto a circumstellar disk. With the development of observational techniques, many interstellar molecules including COMs, though with lower abundances and less complexity, have been observed toward protoplanetary disks (see Sect. 1.3.3.3). It is suggested that both inheritance from the protostellar phase and reset/reprocessing in later evolutionary stages are likely to contribute (Öberg et al. 2023).

1.3 Complex organic molecules

1.3.1 Categories

The earliest appearance of the term "complex organic molecules" in the astronomical context can be traced back to the 1970s (e.g., Sagan 1972), but the first explicit definition (carbon-bearing molecules with at least six atoms) was given by Herbst & van Dishoeck (2009) and was widely adopted by the following studies. So far, more than 150 COMs among ∼340 interstellar molecules have been detected in the gas phase, and the increasing rate of new detections has remained at a fairly high level since the 21st century, especially with the advent of ALMA (e.g., McGuire 2018, 2022). The detection of solid-phase COMs (or COM ices) was only considered robust for methanol (CH_3OH), the simplest and most abundant COM, before the latest JWST observations showcased the feasibility of detecting COM ices larger than CH_3OH (see Sect. 1.3.4.4 and Chapter 4). Table 1.1 lists a number of detected gas-phase COMs with up to three carbons categorized by their functional groups. The most studied COMs are the oxygen or nitrogen-bearing ones with one to two carbons, whereas the detection limit has gone far beyond this.

1.3.2 Formation theories

How COMs are formed is still an open question under active investigation. Early chemical models for simple molecules mainly consider gas-phase exothermic ion-molecule reactions initiated by cosmic-ray ionization (e.g., Herbst & Klemperer 1973). After

Table 1.1: Selected interstellar COMs with up to three carbon atoms.

Category	Functional group	1-carbon	2-carbon	3-carbon
Alcohol	C–OH	CH_3OH	C_2H_5OH C_2H_3OH $(CH_2OH)_2$ CH_3OCH_2OH	n/i-C_3H_7OH
Aldehyde	–C=O–H	H_2CO	CH_3CHO CH_2OHCHO	C_2H_5CHO
Ether	–C–O–C–		CH_3OCH_3	$C_2H_5OCH_3$
Ester	–O–C=O–		CH_3OCHO	C_2H_5OCHO
Acid	–C=O–OH	HCOOH	CH_3COOH	
Ketone	C–(C=O)–C		CH_2CO	CH_3COCH_3
Epoxide	cyclic –(C–O–C)–		c-C_2H_4O	c-C_3H_6O
Cyanide	–C≡N	HCN	CH_3CN	C_2H_5CN
Amine	$–NH_2$, –NH–	CH_3NH_2 NH_2CHO	$C_2H_3NH_2$ CH_3NHCHO	
Cyanate	–O–C≡N	HNCO	CH_3NCO	C_2H_5NCO
Alkane	saturated hydrocarbon	CH_4	C_2H_6	C_3H_8
Alkene	–C=C–, hydrocarbon		C_2H_4	C_3H_6
Alkyne	–C≡C–, hydrocarbon		C_2H_2	C_3H_4

Notes: (1) species in red are those that have been studied in this thesis (Chapters 2 & 3); (2) species in gray are simple molecules by definition (<6 atoms); HCOOH and CH_2CO are not strictly COMs, but they are usually studied together with O-COMs; (3) uncommon O/N-bearing COMs and all S-bearing species, carbon chains, and aromatic compounds are not included; (4) CH_2OHCHO is also considered as the simplest sugar; (5) CH_2CO is not strictly an ketone since one side of its carbonyl group is ended with a hydrogen.

the discovery of COMs other than CH_3OH in the 1970s (e.g., Snyder et al. 1974; Brown et al. 1975) and following wideband line surveys toward Sgr B2 and a few hot cores in Orion in the 1980s (e.g., Sutton et al. 1985; Cummins et al. 1986; Blake et al. 1987), the gas-phase network was extended to COMs, of which the formation is based on simple molecules released from grain mantles (e.g., H_2CO, NH_3, and at most CH_3OH; Charnley et al. 1992). The grain surface chemistry at that time only included cosmic-ray induced formation of radicals at low temperatures, followed by recombination reactions and thermal desorption in the warm-up phase.

However, both theoretical and experimental studies in the 2000s challenged the efficiency of gas-phase formation of COMs, and instead suggested significant contribution from grain-surface (solid-phase) chemistry. A study combining quantum chemical calculations and experiments showed that the previously adopted gas-phase formation route of CH_3OCHO (via $CH_3OH_2^+$+H_2CO) is inefficient due to the high activation energy barrier, and its product in laboratory experiments is prone to an isomer of CH_3OCHO; even with alternative gas-phase pathways, the observed abundances could not be achieved (Horn et al. 2004). Meanwhile, the gas-grain chemical model proposed by Garrod & Herbst (2006) provided results in closer agreement with

the observations, which emphasized the importance of including grain-surface chemistry and its coupling with gas-phase chemistry. More support for solid-phase COM formation came from cryogenic experiments. Watanabe & Kouchi (2002) showed the feasibility of efficiently forming CH_3OH by successive hydrogenation of CO at 10 K, which was confirmed by independent experiments and cross-checked with Monte Carlo simulations by another group (Fuchs et al. 2009). Energetic experiments also showed that a variety of larger O-COMs can be further produced by UV photolysis and warm-up starting from CH_3OH-rich ices (Öberg et al. 2009b).

Since the 2010s, grain-surface chemistry has been forced to undergo a series of revisions in response to the unexpected detection of gas-phase COMs in cold ($\lesssim 10$ K) prestellar cores (e.g., Bacmann et al. 2012; Vastel et al. 2014; Jiménez-Serra et al. 2016a; Soma et al. 2018). The presence of gas-phase COMs under such low temperatures, albeit in much lower abundances, casts doubt on the exclusive role of grain-surface chemistry and challenges the warm diffusion theory to form large COMs. Meanwhile, to explain the new observational results, Balucani et al. (2015) proposed a new gas-phase formation route for CH_3OCH_3 and CH_3OCHO, which is made up of low-barrier neutral–neutral and radiative-association reactions starting from non-thermally desorbed CH_3OH. Experimental studies also found that photodesorption of CH_3OH ices is far less efficient than previously assumed in chemical models ($\sim 10^{-5}$ molecules per photon; Bertin et al. 2016; Cruz-Diaz et al. 2016), potentially supporting the proposition of Balucani et al. (2015), whereas chemical/reactive desorption and cosmic-ray induced desorption seem to be more plausible explanations.

The questioning of the necessity of warm grains to form large COMs was mitigated by non-energetic experiments at low temperatures (10–15 K). Fedoseev et al. (2015) showcased the possibility of forming backbones (i.e., carbon–carbon bonds) of two-carbon O-COMs through recombination between two HCO radicals, with CH_2OHCHO and $(CH_2OH)_2$ observed in the product of their CO hydrogenation experiments at 13 K. More cryogenic experiments without energetic irradiation were able to produce CH_2CO, CH_3CHO, C_2H_3OH, and C_2H_5OH via hydrogenation of C, CO, H_2CO, and C_2H_2 (input ingredients varied among different experiments), with the potential contribution of H-abstraction and radical-radical recombination (Chuang et al. 2016, 2020; Fedoseev et al. 2022). However, CH_3OCH_3 has long been absent from the product of non-energetic experiments, and the yields of some species represented by CH_3OCHO in UV-photolysis experiments are significantly lower than the observed levels (Chuang et al. 2017).

Although our understanding of COM formation is still incomplete (also seen in *Chapter 4*), recent development of **non-diffusive chemistry** further addresses the questions posed by the detection of cold COM gas and reinforces the contribution of grain-surface chemistry (Jin & Garrod 2020; Garrod et al. 2022). Non-diffusive chemistry lifts the temperature constraint on the diffusive reactions involving radicals larger than H atoms, which is the case for most COM formation. Instead of waiting for heavy species to thermally diffuse and meet each other, reactions in non-diffusive chemistry can occur between reactants produced adjacent or very close to each other, hence enhancing the reaction rates. There are four main mechanisms: Eley–Rideal (E-R), three-body (3-B), three-body excited-formation (3-BEF), and photodissociation-induced (PDI) reactions. Three-body reactions are those that occur immediately after an initiating reaction forms a product adjacent to a potential partner. Three-body

excited-formation considers a special case of basic 3-B, where the product of an initiating reaction is excited and hence able to overcome the barrier of following reactions with other reactants. This mechanism plays an important role in producing enough CH_3OCHO in ice mantles to match the observed abundances in the gas phase. Photodissociation-induced reactions are similar to 3-B, except that the reactants are produced by UV photodissociation. The involvement of non-diffusive chemistry allows large COMs to form earlier at low temperatures without requiring mobility of heavy radicals (e.g., via 3-B reactions) and in more efficient ways (e.g., 3-BEF for CH_3OCHO). The reframed grain chemistry is now more promising to reconcile with the cold-COM detections and bridge the gap between simulations/experiments and observations.

There are other hotly debated topics in COM studies besides the formation routes in different phases; for example, the differentiation between O- and N-bearing COMs as suggested by differences in their emission distributions (e.g., Csengeri et al. 2019; Mininni et al. 2023; Nazari et al. 2023b; Busch et al. 2024). We do not dive deeper into them here considering that this thesis only involves O-COMs and the main focus is on gas-to-ice comparisons. Nevertheless, it is clear that our understanding of COM formation will continue to advance through the collaboration of observers, theorists, and experimentalists.

1.3.3 Observations of gas-phase COMs

1.3.3.1 Rotational spectroscopy

COMs in the gas phase are mostly observed at submillimeter (submm) or millimeter (mm) wavelengths for their rotational transitions. There are a few reasons to look at (sub)mm wavelengths: (1) pure rotational lines of molecules with permanent dipole moments have high Einstein A coefficients (10^{-6}–10^{-3} s^{-1}), making their emission bright in dense hot cores under quasi-local thermodynamic equilibrium (LTE); (2) the rotational transitions of large molecules such as COMs mainly fall at (sub)mm wavelengths, or equivalently tens to hundreds of GHz in frequency; (3) the Earth's atmosphere has good transmission windows in the 80–900 GHz range, making it possible for modern radio telescopes (see Sect. 1.3.3.2) to deliver spectra with high spectral and angular resolution as well as wide bandwidths.

An observed spectrum records three principal properties of each spectral line: (i) the central frequency, (ii) the integrated intensity, and (iii) the profile (width and shape). The rest frequency of a transition is determined by the energy difference between its upper and lower levels. However, the rest frequency usually deviates from the observed central frequency of this line because of the radial velocity of the emitting gas with respect to the observer (usually denoted as $v_{\rm lsr}$, where "lsr" refers to local standard of rest). The intensity is dependent on the column density of this molecule and its level populations, which are determined by the kinetic temperature ($T_{\rm kin}$) under strict LTE conditions. In practice, however, molecular emission in hot cores may still remain sub-thermally excited (i.e., $T_{\rm ex} < T_{\rm kin}$) at densities of $< 10^9$ cm^{-3} (e.g., Fig. 6 of Johnstone et al. 2003), and the LTE assumption usually indicates that the observed level populations can be characterized by a single temperature, which is not necessarily the same as $T_{\rm kin}$. The line profile reflects the velocity field (thermal and non-thermal) and kinematic components of the probed gas. The line width is

usually characterized by the full width half maximum (FWHM), which is typically smaller in prestellar cores (< 1 km s^{-1}) and low-mass protostars (< 3 km s^{-1}) than in HMSF regions with stronger turbulence ($\gtrsim 5$ km s^{-1} and up to a few tens of).

Due to the larger amount of atoms, COMs exhibit additional spectral complexity than simpler molecules. The two most relevant conceptions are torsion and conformer. **Torsion** is the internal rotation of a group that can rotate almost freely relative to the molecular frame through a periodic potential. The most common case in astrochemistry is the methyl group (–CH_3), which has a threefold rotational symmetry and vertical mirror planes (also denoted as C_{3v} internal rotors). When the internal rotation (i.e., torsion) couples to the overall rotation, energy levels are better described by a torsion–rotation Hamiltonian rather than a simple rigid rotor. In databases for rotational spectra, CH_3-bearing molecules are usually labeled with an additional torsion quantum number (v_t=0, 1, ...) and grouped into two symmetry species A and E, which represent the two lowest-energy torsional states. In observations, the A and E families of some COMs (e.g., CH_3OH) can be fitted separately, and their ratios can be used to infer the formation temperature given E/A ~ 1 at $T \gtrsim 20$ K and E/A ~ 0.6–0.7 at $T \sim 10$ K. Besides different torsional states, many COMs also have more than one **conformer** (also known as conformational isomer) that is distinguished by different rotations of a single bond in a molecule. Common cases in astrochemistry are the gauche- and trans-conformers of alcohols (e.g., a-$(CH_2OH)_2$ and g-$(CH_2OH)_2$ studied in Chapter 2) and ethers. In principle, the most abundant conformer corresponds to the lowest potential energy, but in interstellar environments, temperatures may not be high enough to overcome the interconversion barriers among the low-energy conformers, and the abundance of more than one conformer can remain at a detectable level. In astronomical databases, conformers are often listed separately and treated as different species.

For molecules that include nitrogen (^{14}N) or deuterium (D), their rotational transitions (e.g., $J+1 \rightarrow J$) are further split by interactions with nuclear spins (denoted by I) into **hyperfine** transitions, since the total angular momentum $\vec{F} = \vec{J} + \vec{I}$ is split into $F = |J-I|$, $|J-I|+1$, ..., $J+I$. The most relevant coupling in astrochemistry is with the electric quadrupole, which is the case for ^{14}N and D (both with $I = 1$ and each rotational transition usually splits into a triplet $J-1$, J, and $J+1$). Although the frequency separation of hyperfine lines is usually small (typically kHz–MHz), modern millimeter facilities have enough spectral resolution to resolve them.

Another case of transition splitting occurs in symmetric top molecules that have one special principal axis, also known as the symmetry axis. For these molecules, another quantum number K is introduced to describe the projection of the total rotational angular momentum $\vec{J}$ onto the symmetry axis. If the dipole is parallel to the symmetry axis (e.g., CH_3CN and CH_3CCH), the selection rule is $\Delta K = 0$; hence for a single $J+1 \rightarrow J$ transition, there is a series of lines for each $K = 0, 1, ..., J$, also known as the **K-ladder**. In astrochemical studies, K-ladders of CH_3CN and CH_3CCH are frequently used as thermometers because their K-ladder transitions sample a wide range of $E_{\rm up}$ while keeping almost all the other parameters (e.g., Einstein A coefficient) the same. Therefore, their relative intensities can provide robust constraints on the temperature assuming LTE (Boltzmann distributions). Most COMs are not perfect symmetric tops and two (pseudo–)quantum numbers K_a and K_c are introduced to track the prolate and oblate limits, respectively, but they do not strictly represent the

projections of $\vec{J}$.

Rotational transitions of interstellar molecules are measured in the laboratory under isolated and well-controlled conditions that allow individual transitions to be resolved with uncertainties far below those achievable in astronomical spectra. The measured spectroscopic parameters are compiled into publicly accessible molecular line databases. Two of the most widely used resources are the Cologne Database for Molecular Spectroscopy (CDMS[8]; Müller et al. 2001, 2005) and the Jet Propulsion Laboratory (JPL[9]) molecular spectroscopy catalog (Pickett et al. 1998). These databases provide standardized line lists including transition frequencies (with uncertainties), upper-state energies, line intensities (usually in the form of Einstein A coefficients), and partition functions over a range of temperatures. Such information is essential for line identification and for deriving physical parameters through spectral modeling. As astronomical observations continue to develop toward higher sensitivity, broader bandwidths, and more complex molecules, the accuracy and completeness of laboratory spectroscopic data remain a key limiting factor in observational studies.

1.3.3.2 Millimeter facilities

Observations of rotational lines emitted by gas-phase interstellar molecules at millimeter wavelengths were initiated by the first detection of CO J=1–0 line at 115 GHz back in 1970 (Wilson et al. 1970). Millimeter astronomy was further boosted by the development of single dishes and interferometer arrays in the following decades. Currently, many millimeter facilities are still in active service (see Table 1.2), but ALMA is undoubtedly the flagship among them for studying protostellar sources given its unmatched combination of high sensitivity and resolution (both spatial and spectral).

Sitting on the high, dry Chajnantor plateau in northern Chile, ALMA consists of the main 12-m array with fifty antennas, the Atacama Compact Array (ACA) with twelve 7-m antennas, and four total-power (TP) antennas. Since Cycle 10 (2023), nine out of ten receiver bands covering 35–950 GHz (0.3–8.5 mm) have been available for use, and the remaining Band 2 (67–90 GHz) will become available soon. ALMA has the longest baseline of 1.6 km among the ground-based interferometer arrays, providing sub-arcsecond to milli-arcsecond angular resolution in its longest configuration (C-10). On the other hand, the short spacings of ACA and TP can recover extended emission on arcminute scales that are filtered out by ALMA, which is important for low-$E_{\rm up}$ lines or abundant molecules. In terms of spectroscopic performance, ALMA offers multiple choices for channel spacing (down to 15 kHz) and bandwidth (up to 1.875 GHz per window, $4\times1.875 = 7.5$ GHz instantaneously), which can be customized flexibly for different targets and science cases. These features combine to make ALMA the most powerful facility for COM observations. Its high angular resolution alleviates beam dilution effects and separates different chemical and kinematical components, which is crucial for COMs emitting from compact hot core regions. Moreover, its high sensitivity enables detection of rare isotopologs and higher-excitation lines, yielding more accurate column density measurements for optically thick molecules (e.g., the main isotopologs).

Other interferometer arrays such as NOEMA and SMA have coarser spatial reso-

[8] `https://cdms.astro.uni-koeln.de/`
[9] `https://spec.jpl.nasa.gov/`

Table 1.2: Selected millimeter telescopes in service

Facility	Type & Aperture	Location	Frequency coverage	Angular resolution (beam size)	Spectral resolution	Instantaneous bandwidth
ALMA[a]	Interferometer array (50×12 m + 12×7 m + 4×TP)	Chajnantor plateau, Chile (∼5000 m)	35–950 GHz (0.3–8.5 mm)	down to 0.006–0.14″ (Config. 10)	15.3 kHz to 15.6 MHz	up to 7.5 GHz (4 × 1.875)
NOEMA[b]	Interferometer array (12×15 m)	Plateau de Bure, Hautes-Alps, French (∼2550 m)	80–370 GHz (0.8–3.7 mm)	down to ∼0.1–1.0″ (Config. A)	native 2 MHz, can zoom to 62.5 kHz	up to 31 GHz (PolyFiX)
SMA[c]	Interferometer array (8×6 m)	Maunakea, Hawaii, USA (∼4080 m)	180–420 GHz (0.7–1.7 mm)	∼0.25–5″ at 345 GHz (Very Extended Config.)	140 kHz	up to 44 GHz
GBT[e]	Single dish (100 m)	Green Bank, WV, USA (∼800 m)	0.29–50.5 & 67–116 GHz (2.6 mm–1 m)	$\sim 780''/\nu$ [GHz]; ∼ 7.8″ at 100 GHz	down to 0.02 kHz (VEGAS)	up to 4–6 GHz
IRAM 30-m[d]	Single dish (30 m)	Pico Veleta, Spain (2850 m)	80–370 GHz (0.8–3.7 mm)	$2460''/\nu$ [GHz]; 6.6–31″	50/195 kHz (FTS)	up to 32 GHz (FTS)
Yebes 40-m[g]	Single dish (40 m)	Yebes, Castilla–La Mancha, Spain (∼980 m)	31.5—-50 & 72—-90.5 GHz (3.3–9.5 mm)	$1763''/\nu$ [GHz]; ∼20–56″	38 kHz	up to 18.5 GHz
JCMT[f]	Single dish (15 m)	Maunakea, Hawaii, USA (4092 m)	325–375 GHz (0.8–0.92 mm; HARP)	∼ 14″ at 345 GHz	30.5 kHz to ∼1 MHz	up to 1.8 GHz (ACSIS)

The full name and the link for technical information (up to Sep 2025) of each facility are listed below:
[a] Atacama Large Millimeter/submillimeter Array (`https://almascience.eso.org/documents-and-tools`)
[b] NOrthern Extended Millimeter Array (`https://www.iram.fr/IRAMFR/GILDAS/doc/html/noema-intro-html/node1.html`)
[c] SubMillimeter Array (`http://sma1.sma.hawaii.edu/status.html#arrayconf`)
[d] Institut de Radioastronomie Millimétrique 30-meter telescope (`https://www.iram.fr/GENERAL/calls/w25/30mCapabilities.pdf`)
[e] Green Bank Telescope (`https://www.gb.nrao.edu/scienceDocs/GBTpg.pdf`)
[f] James Clerk Maxwell Telescope (`https://www.eaobservatory.org/jcmt/instrumentation/`)
[g] `https://rt40m.oan.es/rt40m_en.php` and Tercero et al. (2021)

lution and lower line sensitivity than ALMA due to shorter maximum baselines and smaller collecting area, but they also provide important complementary capabilities in observing northern sources that are unreachable by ALMA and offering wider instantaneous bandwidths for line surveys. Single-dish telescopes, though with much lower angular resolution, are still actively used for COM observations. Their high spectral resolution (kHz-level), wide instantaneous bandwidth, large recoverable scale, and decent sensitivity make them more favorable than ALMA for surveys of narrow lines and extended emission toward prestellar cores (Sect. 1.3.3.3). They also remain indispensable in mapping large-scale chemistry and preselecting promising regions for follow-up observations with interferometer arrays.

1.3.3.3 Line surveys

To better identify and quantify the abundances of gas-phase COMs, it is preferable to observe a wide frequency range to maximize line coverage; therefore, such observations are also referred to as "line surveys". With single dishes and early interferometer arrays, (sub)millimeter line surveys toward line-rich YSOs such as Orion KL and Sgr B2 had already revealed dozens of lines per GHz (e.g., Fig. 1.2), many of which are attributable to gas-phase COMs emitting from hot cores (see review by Herbst & van Dishoeck 2009). Long carbon chains were also discovered in cold prestellar cores, represented by the Taurus Molecular Cloud 1 (TMC-1). However, many of these observations suffered from beam dilution, in which case the molecular emission is not spatially resolved and the actual source size of each emitting molecule is unknown, causing uncertainties in the derived column densities and relative ratios.

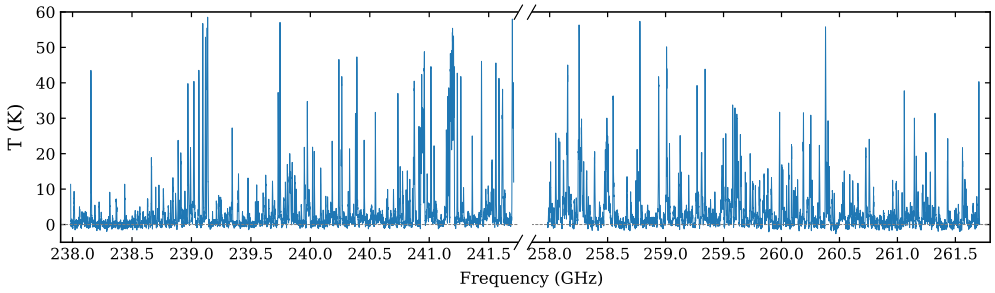

Figure 1.2: ALMA spectrum of G19.88-0.53, one of the high-mass hot cores studied in Chapters 2 & 3. Most of the displayed lines are emitted by gas-phase COMs.

The high sensitivity and angular resolution of ALMA significantly improve the depth, accuracy, and efficiency of line surveys. In early cycles, several dedicated line surveys were conducted for well-known hot corino/cores: the ALMA Protostellar Interferometric Line Survey (PILS) for the Class 0 binary system IRAS 16293-2422 (Jørgensen et al. 2016), the Exploring Molecular Complexity with ALMA (EMoCA[10] for Sgr B2(N), a massive star-forming region located in the central molecular zone (Belloche et al. 2016), and the G31.41+0.31 Unbiased ALMA sPectral Observational

[10]EMoCA was later upgraded to Re-exploring Molecular Complexity with ALMA (ReMoCA) with a higher sensitivity and angular resolution (Belloche et al. 2019).

Survey (GUAPOS) for the massive hot core G31.41+0.31 (Mininni et al. 2020). Covering a continuous ∼30 GHz bandwidth around 345 GHz (0.87 mm; ALMA Band 7) or 100 GHz (3 mm; ALMA Band 3), these programs delivered the most complete COM gas inventories for the three prototypical hot corino/cores. Many new species were discovered, including but not limited to chloromethane (CH_3Cl; Fayolle et al. 2017), urea (NH_2CONH_2; Belloche et al. 2019), propyl cyanide (C_3H_7CN; Belloche et al. 2014), propanol (C_3H_7OH; Belloche et al. 2022), and a number of deuterated COMs (e.g., Coutens et al. 2016; Manigand et al. 2019; Richard et al. 2021; Ilyushin et al. 2022; Drozdovskaya et al. 2022; Ferrer Asensio et al. 2023). The rich detection lists of these dedicated line surveys enable detailed investigations of isotopologs, isomers, and chemical similarities and differentiations between different species and different cores (e.g., Belloche et al. 2016; Calcutt et al. 2018; Mininni et al. 2023; Busch et al. 2024; Belloche et al. 2025). This is only available with sufficient depth and frequency coverage, but such observations are usually very expensive or even unrealistic with mm facilities other than ALMA.

Thanks to the high sensitivity of ALMA, deep and unbiased line surveys (i.e., those with low noise levels and large bandwidths) have become much more affordable, promoting the transition from case studies to large-sample surveys. Instead of staring at individual sources, the ubiquity of COMs has been examined from a broader perspective by exploring statistics of larger samples. For low-mass sources, the Perseus ALMA Chemistry Survey (PEACHES) targeted 50 embedded (Class 0/I) protostars and reported a detection rate of ≤58 % for COM emission (Yang et al. 2021). A sibling program of PEACHES, the ORion ALMA New GEneration Survey (ORANGES), detected a smaller fraction (∼26 %) of CH_3OH emission in 5 out of 19 targeted sources in the Orion Molecular Complex (OMC), which is a more illuminated environment than the Perseus (low-mass) star-forming region (Bouvier et al. 2022). A similar low detection rate for CH_3OH in OMC was found by another program, the ALMA Survey of Orion Planck Galactic Cold Clumps (ALMASOP), where only 11 out of 56 (∼20 %) Class 0/I sources exhibit CH_3OH emission (Hsu et al. 2022). Although these statistics hint at the non-ubiquity of COMs in star-forming regions, the non-detection of COM emission does not necessarily mean the absence of COMs; the interpretation of non-detection can be affected by factors such as detection sensitivity, disk shadowing, and dust attenuation (e.g., van Gelder et al. 2022; Nazari et al. 2022b).

In the high-mass regime, a substantial amount of ALMA data useful for COM studies has been obtained by a number of large-sample surveys, albeit not primarily intended for chemical purposes. These surveys include two ALMA Large Programs: the ALMA-IMF program (where IMF refers to initial mass function; Motte et al. 2018) targeting 15 HMSF regions in the Galactic plane, and the ALMA evolutionary study of high-mass protocluster formation in the GALaxy (ALMAGAL; Molinari et al. 2025) targeting over 1000 intermediate/high-mass dense clumps in the Galaxy. Another two ALMA programs, ALMA Three-millimeter Observations of Massive Star-forming regions (ATOMS; Liu et al. 2020) and Querying Underlying mechanisms of massive star formation with ALMA-Resolved gas Kinematics and Structures (QUARKS; Liu et al. 2024), observed the same sample of 146 HMSF regions at different bands (ALMA Bands 3 and 6). *Chapters 2 and 3* of this thesis also introduce a chemically oriented program, the Complex Chemistry of hot Cores with ALMA (CoCCoA), which exclusively focuses on line-rich hot cores with customized spectral setups for COM lines.

These large-sample surveys of HMSF regions not only increased the sample size of hot cores, but also enabled systematic and comparative studies of the abundances and distributions of COMs (e.g., Qin et al. 2022; Nazari et al. 2022a; Baek et al. 2022; Bonfand et al. 2024; Li et al. 2024, 2025; *Chapters 2 and 3* of this thesis). In particular, a similar trend of column density ratios between CH_3OH and some other COMs (especially CH_3OCH_3 and CH_3OCHO) is found for sources with luminosities across multiple orders of magnitude, shedding light on the histories and mechanisms of COM formation (e.g., Coletta et al. 2020; Nazari et al. 2022a; Chapter 2).

The investigation of gas-phase COMs is not limited to protostellar sources. Intensive line surveys have also been carried out for prestellar cores and protoplanetary disks (Class II objects). Numerous first detections of hydrocarbons (including long carbon chains, aromatics, and ions) were made toward the famous prestellar cloud TMC-1 by single dishes (e.g., Cernicharo et al. 2021; Wenzel et al. 2024, 2025). Two representative programs are the Q-band Ultrasensitive Inspection Journey to the Obscure TMC-1 Environment (QUIJOTE, with Yebes 40-m and IRAM 30-m radio telescopes; Cernicharo et al. 2022) and the GBT Observations of TMC-1: Hunting for Aromatic Molecules (GOTHAM; McGuire et al. 2020). These single-dish surveys excelled in detecting narrow lines of large molecules at low frequencies with their wide bandwidths ($\lesssim$50 GHz) and high spectral resolutions (kHz-level, $\lesssim$0.1 km s^{-1}).

For protoplanetary disks, it is already difficult to detect the most abundant COMs such as CH_3OH and CH_3CN, as these detections usually require a long integration time to achieve sufficient noise levels and angular resolutions. Early COM detections were reported for several disks that are either nearby (TW Hya; Walsh et al. 2016) or warm due to outbursts (V883 Ori; Lee et al. 2019b) or surrounding Herbig Ae/Be stars (MWC 480, HD 100546, and Oph IRS 48; Öberg et al. 2015; Booth et al. 2021; van der Marel et al. 2021; Brunken et al. 2022). In recent ALMA Cycles, follow-up line surveys have been conducted and more COMs in more disks are found (e.g., Booth et al. 2023, 2024a,b, 2025). The presence of COMs in protoplanetary disks, which is suggested to originate from sublimation of ice mantles on dust grains, provides evidence for ice inheritance across different star formation stages. Whether these observed COMs were inherited from earlier stages or formed in situ depend on many factors (e.g., thermal history and disk geometry), and recent studies suggest that both scenarios can play a role in shaping the disk chemistry (Öberg et al. 2023).

1.3.4 Observations of solid-phase COMs

1.3.4.1 Vibrational bands of interstellar ices

In dense star-forming regions, ices on dust grains serve as important reservoirs for interstellar molecules, where more than 50% of the heavy elements can be locked up. In the solid phase, molecules cannot rotate freely and emit rotational lines at (sub)millimeter wavelengths like those in the gas phase; instead, they can only emit or absorb energy through their vibrational transitions which fall in the near–to mid-infrared (∼2–20 µm). Because the dust extinction is lower in the infrared than in the optical, infrared light from a background star or a protostar can pass through dense cores and be absorbed by ices along the line of sight, making the IR vibrational absorptions a clean probe of interstellar ices.

For a free molecule, the total number of fundamental vibrational modes depends

on the number of atoms (N) and its structure; non-linear molecules have $3N-6$ and linear molecules have $3N-5$ fundamental modes, where $3N$ is the total number of mechanical degrees of freedom (DoF), and the subtracted degrees are for translations (i.e., moving as a whole in three directions) and rotations (around three or two axes, depending on the linearity). For symmetric molecules (e.g., H_2O and CO_2), some vibrational modes are degenerate with each other. In reality, molecules are not perfect harmonic oscillators, and non-fundamental modes such as overtones and combinations may appear due to anharmonicity (of which the observations have been reported by, e.g., Carr 1989; Keane et al. 2001; GRAVITY Collaboration et al. 2020; Brunken et al. 2024b). In ice mantles, interactions with environments can also produce additional features such as the H_2O libration mode at $\sim$13 µm and the O–H dangling mode of porous H_2O ice at $\sim$2.7 µm (Ehrenfreund et al. 1996; Noble et al. 2024).

With a larger number of atoms, COMs can have dozens of fundamental vibrational modes, but not all of them are detectable in IR observations. In principle, the intrinsic band strength A scales with $(\mathrm{d}\mu/\mathrm{d}Q)^2$, the square of the change in dipole along the mode, where μ and Q refer to the dipole moment and the vibrational coordinate, respectively. That is, modes that cause more charge movements are stronger (e.g., stretches of X–H, C–O, and C=O bonds). There are many strong COM bands, mostly CH_3 deformation and C–O stretching modes, sitting within the 6.8–8.8 µm range, which is also known as the **COM ice fingerprint range**. This range will be the focus of the COM ice studies presented in Chapters 4 and 5.

Unlike narrow gas-phase lines of rotational transitions, vibrational absorption features of ices are much broader, hence called "vibrational bands". The broadening effect is induced by interactions with neighboring molecules that shift the original vibrational transition by different amounts of energy. As a result, the profile of a vibrational band (including peak positions, widths, and shapes) can vary significantly with environments, with the two dominant factors being matrix polarity (polar/H_2O-rich or apolar/CO-rich) and temperature (amorphous/low-T or crystalline/high-T), as shown in Fig. 1.3. The diversity of band profiles was already observed in early detections of H_2O ices toward different sources (e.g., Merrill et al. 1976) and revealed the fact that interstellar ices are present in different mixtures rather than in pure form.

1.3.4.2 Laboratory IR spectroscopy

To infer ice compositions from observed vibrational bands, IR spectra of different ices are measured in the laboratory and compared with observations (see e.g., reviews by Allodi et al. 2013; Linnartz et al. 2015; Cuppen et al. 2024). Astrochemical experimentalists grow thick (10^1–10^3 monolayers) interstellar ice analogs by vapor-depositing gases onto cryogenic substrates ($\gtrsim$10 K) in ultrahigh vacuum ($\lesssim 10^{-9}$ mbar). Transmission or reflection IR spectra of deposited ices are then recorded using Fourier-transform infrared spectroscopy (FTIR) or reflection adsorption infrared spectroscopy (RAIRS), respectively. Parameters such as composition, thickness, temperature, and radiation processing are varied per experiment to investigate different topics. In practice, transmission spectroscopy is the default choice for characterizing IR features of ices (which is also the case relevant to this thesis), and reflection spectroscopy is preferred for studying surface chemistry. Three important properties can be collected from transmission FTIR for each ice species or mixture: (1) peak wavelengths (λ_{peak}) and shapes (including FWHM) of vibrational bands, (2) intrinsic band strengths A,

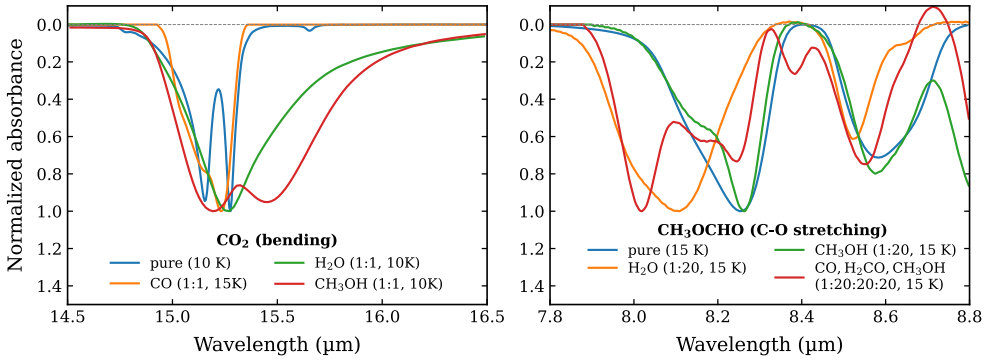

Figure 1.3: Laboratory-measured IR spectra of pure and mixed ices of CO_2 (left) and CH_3OCHO (right). The mixing constitute(s), mixing ratio, and temperature are indicated in the legend. In the right panel, some features do not belong to CH_3OCHO but the mixing constitutes (e.g., CH_3OH and H_2CO have a band at ∼8.9 μm and ∼8.0 μm, respectively).

and (3) optical constants n and k serving as required inputs for radiative transfer modeling. In this thesis, $\lambda_{\rm peak}$ and FWHM are used to identify ice compositions from the observed IR bands, and A is used to calculate ice column densities ($N_{\rm ice}$) of detected species. The relation between A and $N_{\rm ice}$ is given by

$$N_{\rm ice} = \frac{1}{A}\int_{\tilde{\nu}_1}^{\tilde{\nu}_2} \tau(\tilde{\nu})\,{\rm d}\tilde{\nu} = \frac{\ln 10}{A}\int_{\tilde{\nu}_1}^{\tilde{\nu}_2} Abs(\tilde{\nu})\,{\rm d}\tilde{\nu}, \tag{1.3}$$

where $\tilde{\nu}$ is the wavenumber, $\tilde{\nu}_1$ and $\tilde{\nu}_2$ indicate the range of the targeted vibrational band, τ is the optical depth, and Abs is the absorbance.

Driven by the demand for analyzing observational data, great efforts have been made in laboratories since the detection of the first interstellar ice band at 3.1 μm (Gillett & Forrest 1973), which was later attributed to amorphous (instead of crystalline) H_2O ice by Léger et al. (1979). This identification highlighted the need for laboratory spectra measured under astrophysical (i.e., low-temperature and low-pressure) conditions. More systematic experiments measured IR spectra and optical constants for various pure and mixed ices, while showing how temperature, dilution, and mixture composition can alter the IR features of these ices (e.g., Hagen et al. 1983; Hudgins et al. 1993). In parallel, band strengths of common interstellar ices were revisited and calibrated to improve the accuracy of ice column density determinations from observations (e.g., Gerakines et al. 1995; Mastrapa et al. 2009; Bouilloud et al. 2015). As IR spectroscopy observations developed (especially with the advent of space-based telescopes, see Sect. 1.3.4.3), ice experiments were also advanced to include warm-up and energetic irradiation, in purpose of studying the evolution of interstellar ice analogs and characterizing the processing products such as ions and more complex species (e.g., Allamandola et al. 1988; Ehrenfreund et al. 1997; Öberg et al. 2009b).

Before the launch of JWST, the laboratory group at Leiden conducted systematic IR spectral measurements for various COM ices under different mixing conditions and temperatures. The selected COMs are those that are abundantly detected in the gas phase, including CH_3CHO, C_2H_5OH, CH_3OCH_3, CH_3OCHO, CH_3COCH_3, CH_3CN, CH_3NH_2, and NH_2CHO (Terwisscha van Scheltinga et al. 2018, 2021; Rachid

et al. 2020, 2021, 2022; Slavicinska et al. 2023), of which the IR spectra are publicly available in the Leiden Ice Database for Astrochemistry (LIDA[11]; Rocha et al. 2022). IR spectra of other COMs (pure or mixed with H_2O) such as CH_2OHCHO, $(CH_2OH)_2$, CH_3COOH, C_3H_7OH and hydrocarbons with ≥ 6 atoms are provided by the Cosmic Ice Laboratory of NASA (Hudson et al. 2014; Hudson & Gerakines 2019). These laboratory measurements provide an indispensable toolkit for observers to identify interstellar ices, derive reliable column densities, and infer their thermal and chemical processing histories from observations.

1.3.4.3 Observations in the pre-JWST era

It is not trivial to observe at IR wavelengths due to the absorption and contamination of the Earth's atmosphere. IR facilities have to be built on high mountains or carried on aircraft to minimize the influence of water vapors and thermal emission of the atmosphere itself. The ideal way is to observe from space, which was finally realized in the 1980s by IRAS (Neugebauer et al. 1984).

Ice mantles were suggested to be present on dust grains about 80 years ago (van de Hulst 1946; Oort & van de Hulst 1946), but early attempts of searching for the 3.1 µm H_2O ice band turned out vain toward the diffuse ISM (Danielson et al. 1965; Knacke et al. 1969). It was not until the 1970s that the first detection of this band was made toward embedded protostars in the Orion BN-KL complex (Gillett & Forrest 1973), followed by more detections toward YSOs associated with dense molecular clouds (Merrill et al. 1976; Willner et al. 1982). In addition to H_2O, other key ice species were identified subsequently via their strong vibrational bands, including CO at 4.67 µm (Soifer et al. 1979; Lacy et al. 1984), XCN (OCN^-) at 4.62 µm (Geballe 1986; Grim & Greenberg 1987), NH_3 at 2.97 µm (Knacke et al. 1982; Leger et al. 1983), CO_2 at 15.2 µm (D'Hendecourt & Jourdain de Muizon 1989) and 4.27 µm (de Graauw et al. 1996; more convincing multi-band detections), CH_3OH at 3.53 µm (Grim et al. 1991), OCS at 4.9 µm (Palumbo et al. 1995), H_2CO at 3.47 µm (Schutte et al. 1996), and CH_4 at 7.67 µm (Boogert et al. 1996).

These early detections were made toward massive protostars and mostly in the near-IR with ground-based and airborne IR telescopes, except that mid-IR detections of CO_2 and CH_4 were realized by the ISO (Kessler et al. 1996). The Short Wavelength Spectrometer (SWS) of ISO provided continuous observations between 2.4 and 45 µm band at moderate spectral resolution ($\lambda/\Delta\lambda \sim$1500–2000), delivering the first near-complete inventories of common ices and the complex absorption structure over the entire 2.5–20 µm range (e.g., Gibb et al. 2004). In parallel, near-IR spectrometers with higher spectral resolution ($\lambda/\Delta\lambda$ up to a few 10^4) started to serve on ground-based optical telescopes such as the Very Large Telescope (VLT), the Keck Observatory (McLean et al. 1998), and the Subaru 8.2 m Telescope. These observations, in synergy with models and experiments, gradually established an analytical framework that links band shapes to thermal and chemical processing of interstellar ices (e.g., Pontoppidan et al. 2003; van Broekhuizen et al. 2005).

As time moved to the 21st century, the launch of the *Spitzer* Space Telescope was a major step toward large-sample IR studies of YSOs, mostly owing to its high sensitivity enabled by the cryogenic environment. In particular, the Cores to Disks (c2d) Legacy

[11] `https://icedb.strw.leidenuniv.nl`

Program investigated more than one thousand YSOs in five nearby low-mass star-forming regions, with high-SNR spectra obtained for hundreds of YSOs between 5 and 38 µm (Evans et al. 2003, 2009). The c2d survey built the first large, homogeneous ice census toward low-mass YSOs, providing valuable statistics for common interstellar ices (H_2O, CO, CO_2, NH_3, CH_4, CH_3OH) and the 5–8 µm complex (Boogert et al. 2008; Pontoppidan et al. 2008; Öberg et al. 2008; Bottinelli et al. 2010; Öberg et al. 2011).

IR observations of the universe were continued with ground-based and airborne facilities; new space telescopes such as AKARI (Murakami et al. 2007), the *Herschel* Space Observatory (Pilbratt et al. 2010), and the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) were launched, but they did not provide higher-quality mid-IR spectroscopic observations after the decommissioning of *Spitzer*. The explored ice inventory thus did not extend far beyond simple species, with CH_3OH being the only COM that had been firmly detected in the solid phase before the advent of JWST. Nevertheless, substantial achievements have been made in understanding the icy universe in the pre-JWST era (see review by Boogert et al. 2015).

1.3.4.4 Findings and challenges of JWST observations

The launch of JWST on the Christmas Day of 2021 marked the beginning of a renaissance of IR astronomy. JWST is a 6.5-meter, cryogenic space telescope orbiting the Sun at the second Lagrange point (L2). With comprehensive improvements in sensitivity (μJy–mJy), spatial resolution (sub-arcsec), and resolving power (R up to 4000, wavelength-dependent), JWST offers unprecedented high-quality observations between 0.6 and 28 µm. In particular, the Near Infrared Spectrograph (NIRSpec) and the Medium Resolution Spectroscopy (MRS) mode of the Mid-Infrared Instrument (MIRI) provide integrated field unit (IFU) spectroscopic observations at sub-arcsec spatial resolutions and moderate resolving powers (1000–4000). In comparison, the resolving power of the InfraRed Spectrograph (IRS) of *Spitzer* is several times to order-of-magnitude lower (R ∼60–120 for ∼5–40 µm and $R = 600$ for ∼9.9–37 µm).

The advanced capabilities of JWST's mid-IR spectroscopy open a new window on the study of interstellar ices (see Fig. 1.4 for a typical IR spectrum of protostellar sources). In addition to the key vibrational bands of common ices (H_2O, CO, CO_2, NH_3, CH_4, CH_3OH, etc.), JWST is also capable to detect weak features of minor isotopologs of simple species and COMs larger than CH_3OH in the solid phase. The Early Release Science (ERS) program Ice Age toward background stars behind the Chamaeleon I cloud reveals a rich inventory of interstellar ices (including ^{13}C isotopologs and ions), providing an elemental budget along prestellar sightlines for the first time (McClure et al. 2023). The decent sensitivity and spectral resolution of JWST also enable systematic measurements of solid-phase $^{12}C/^{13}C$ ratios in solar-mass protostars (Brunken et al. 2024c) and secure detection of HDO ices in multiple YSOs (Slavicinska et al. 2024, 2025b).

With JWST, the fingerprint range of COM ices (6.8–8.8 µm) can be observed in greater detail. The COMs ORigin Investigated by the Next-generation Observatory in Space (CORINOS) program detects the 7.24 and 7.4 µm bands with high SNR toward the Class 0 source IRAS 15398-3359 (Yang et al. 2022). These two bands were originally noticed in the spectrum of a high-mass YSO (W33A) and studied by Schutte et al. (1999), who attributed them to $HCOO^-$, HCOOH, and CH_3CHO, but

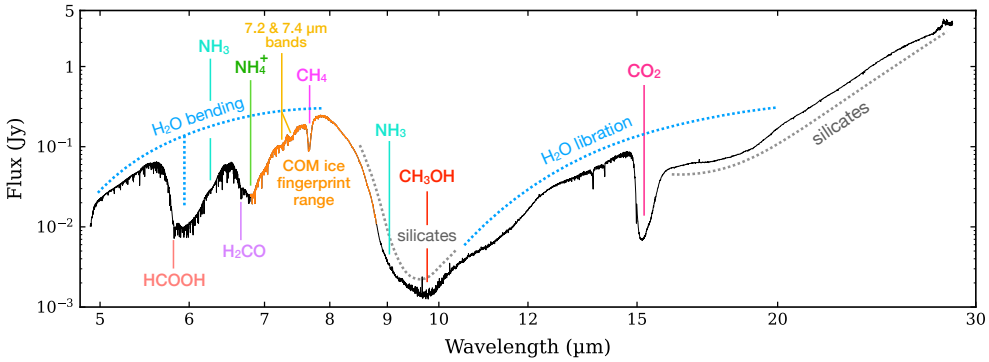

Figure 1.4: JWST/MIRI-MRS spectrum of B1-c, a low-mass protostar studied in Chapter 4. The major ice bands and the COM ice fingerprint range are labeled in different colors.

without claiming a firm detection. Yang et al. (2022) revisited the attribution of the aforementioned species plus C_2H_5OH and CH_3OCHO by comparing the peak wavelengths of the absorption bands at ∼7.24, 7.4, and 11 μm between observations and laboratory spectra. However, no firm identification was claimed until dedicated fitting was performed by Rocha et al. (2024) for one low-mass and one high-mass protostars (NGC 1333 IRAS 2A and IRAS 23385+6053), where laboratory spectra of candidate ices including CH_4, $HCOO^-$, HCOOH, and several two-carbon O-COMs were fitted to the observed COM ice fingerprint range. In *Chapter 4*, similar fitting is performed for another low-mass protostar B1-c; although with different fitting strategies, the identification of O-COMs larger than CH_3OH remains robust in both studies.

The exciting fact that quantitative analysis of COM ices is now available with JWST comes with challenges and caveats. One of the challenges lies in the isolation of weak COM ice bands from the entire spectrum. In Rocha et al. (2024) and Chapter 4, this process is divided into several steps, of which the performance is subjective and case-dependent. Although Chapter 5 tries to optimize and standardize these steps for a larger sample, it is impossible to rule out subjectivity completely. As a consequence, the isolated COM fingerprint range can be different when analyzed by different people/groups, and these differences may be propagated and amplified in the results and conclusions. Therefore, it is important for COM ice studies to elaborate on the methodology employed and provide conservative uncertainties for key quantities such as optical depths and column densities. Another challenge concerns the quantitative fitting of COM ices using laboratory spectra. With current laboratory databases, although significant development has been made in the past decade (Sect. 1.3.4.2), it is still difficult to obtain unambiguous decompositions for observed absorption features, especially those in the COM ice fingerprint range, because of the intrinsic complexity and degeneracy lying in the vibrational bands of mixed ices (Sect. 1.3.4.1). A specific example is given in Chapter 4, where different laboratory spectra are compared to observations, showing the non-uniqueness of COM ice fitting; that is, plausible fitting can be achieved by laboratory spectra with a range of mixing conditions and temperatures. Overall, the quantitative study of COM ices is still an emerging area in astrochemistry that requires further development and investigation.

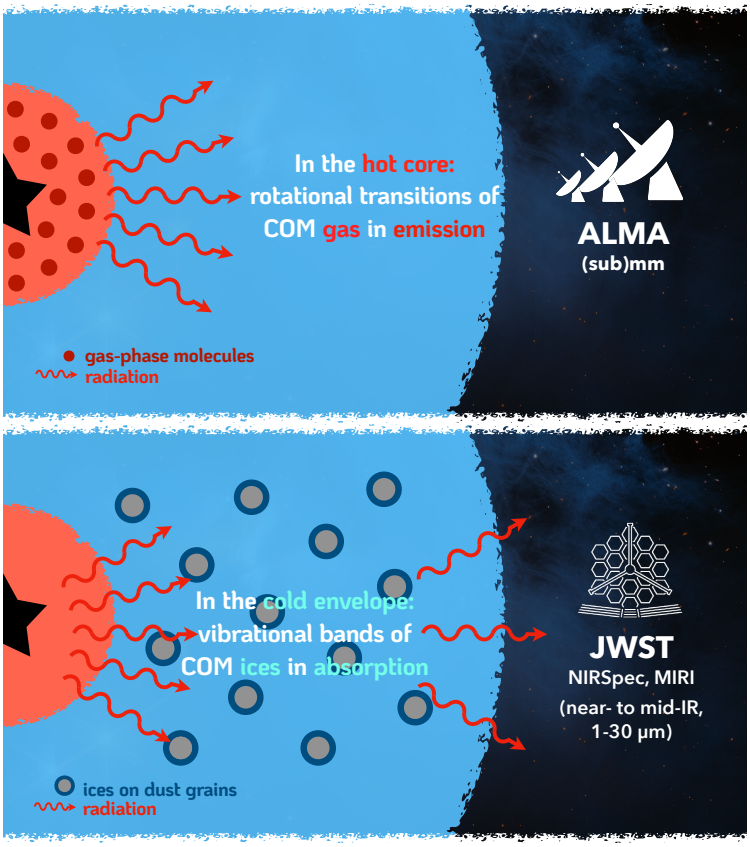

Figure 1.5: Cartoon of ALMA and JWST tracing the gas and ice reservoirs of COMs.

1.4 This thesis

This thesis presents a series of observational studies of COMs in the gas and solid phases using ALMA and JWST, respectively. The highlight as well as the ultimate goal is to make direct comparisons between the gas and ice reservoirs of COMs, which have only become feasible with the combination of ALMA and JWST observations (Fig. 1.5). Thanks to the successful operation of JWST, we are now able to dive deeper into the icy world of interstellar molecules and link the origin of COM formation with the subsequent gas-phase evolution during early stages of star formation.

1.4.1 Overview of chapters

This thesis consists of equal numbers of chapters relevant to ALMA and JWST, well reflecting its topic of combining these two powerful facilities in COM studies. The first two chapters (*2 and 3*) present systematic studies of gas-phase O-COMs in 12–14 high-mass hot cores observed by the CoCCoA program (PI: Brett A. McGuire). The following two chapters (*4 and 5*) shift the focus to searching for COM ice fingerprints in the MIRI-MRS spectra of (mostly) low-mass protostars in the JWST Observations of

Young protoStar (JOYS) program (PI: Ewine van Dishoeck) and several collaborating programs. In particular, *Chapter 4* directly engages with the topic of this thesis by making the first gas-to-ice comparisons of COMs in the *same* sources. The summary of each chapter is provided below:

Chapter 2 investigates six two-carbon O-COMs: acetaldehyde (CH_3CHO), ethanol (C_2H_5OH), dimethyl ether (CH_3OCH_3), methyl formate (CH_3OCHO), glycolaldehyde (CH_2OHCHO), and ethylene glycol ($(CH_2OH)_2$) in the gas phase toward 14 high-mass hot cores observed by the CoCCoA program, motivated by the scarcity of ALMA-based large-sample studies of O-COMs in high-mass protostars. The column density ratios between the six O-COMs and methanol (CH_3OH), the simplest and most abundant COM, are measured and compared to literature values of five low-mass protostars. No clear trend of O-COM/CH_3OH ratios as a function of stellar luminosity is observed between the low- and high-mass samples, suggesting common conditions of COM formation which is most likely to occur on icy dust grains in the cold prestellar phase. The observed gas-phase reservoirs then originate from the following ice sublimation as the protostar warms up the envelope. Among the studied O-COMs, CH_3OCH_3 and CH_3OCHO exhibit the highest and most consistent ratios with respect to CH_3OH. In contrast, the other four O-COMs display greater scatters in their ratios (typically around one order of magnitude), which may result from gas-phase chemistry or differences in emitting areas. Comparisons with astrochemical experiments and simulations show that CH_3OCH_3 and CH_3OCHO tend to be underproduced by experiments, whereas C_2H_5OH and CH_2OHCHO are overproduced by models.

Chapter 4 extends the systematic observational study from two-carbon O-COMs to two three-carbon ones, acetone (CH_3COCH_3) and propanal (C_2H_5CHO), which lack large-sample measurements of their gas-phase abundances. Physical properties of CH_3COCH_3, C_2H_5CHO, and two potentially chemically relevant species, ketene (CH_2CO) and propyne (CH_3CCH), are measured toward 12 out of the 14 CoCCoA sources studied in Chapter 2. Morphologies and kinematics indicate that CH_3COCH_3, C_2H_5CHO, and CH_2CO arise from compact hot cores like other O-COMs, whereas CH_3CCH is more likely to trace extended outflows. The statistics of gas-phase column density ratios between eight O-COMs (six studied in Chapter 2 and two in this chapter) and CH_3OH show that CH_3COCH_3 has quite high gas-phase abundances comparable to some two-carbon O-COMs, whereas aldehydes (those with –CHO group) are systematically less abundant than alcohols, ethers, esters, and ketones with the same amount of carbon atoms. Possible formation routes of acetone are searched in the literature and examined by comparing the observed column density ratios between CH_3COCH_3 and its chemical neighbors with modeling results. All discussed pathways are plausible from an observational point of view, but more evidence (e.g., from theoretical calculations and experiments) is needed to determine which pathway(s) is more likely to dominate.

Chapter 4 presents the first gas-to-ice comparative study of COMs in the same sources—two chemically-rich low-mass protostars, NGC 1333 IRAS 2A and B1-c. Column densities and temperatures of O-COMs in the gas and solid phases are measured from ALMA and JWST/MIRI-MRS spectra, respectively. Dedicated decomposition of the COM ice fingerprint range (6.8–8.8 μm) in the JWST spectrum of B1-c with laboratory spectra of ice mixtures suggests that most O-COM ices are likely present

in a H_2O- or CH_3OH-rich environment. Temperature, however, is difficult to distinguish, as COM ices usually have similar band profiles at low temperatures below their crystallization points. Four O-COMs were selected to make gas-to-ice comparisons in their column density ratios with respect to CH_3OH, and two cases are observed: CH_3OCH_3 and CH_3OCHO have similar ratios in both phases, while CH_3CHO and C_2H_5OH have higher ice ratios by 1–2 orders of magnitude. This implies that both inheritance and gas-phase reprocessing can play a role during the evolution of COMs.

Chapter 5 expands the investigation of ice features from the case study of two low-mass protostars (*Chapter 4*) to a systematic study of more than 20 (mostly low-mass) sources. An interactive workflow for isolating the COM ice fingerprint range is designed for large-sample analysis. As a preparatory step for isolating the weak COM ice bands, broad absorption bands of silicates (at ∼9.8 and 18 µm) and H_2O (at ∼13 µm) are fit. A strong correlation is found between the silicate optical depth and the H_2O ice column densities, confirming their roles of probing dust and ice reservoirs in protostellar envelopes. Regarding the COM ice fingerprint range, absorption bands at ∼7.02, 7.24, 7.40, and 7.67 µm (CH_4 band) are detected in most of the Class 0 and 0/I borderline sources; Class I sources generally have fewer dust and ices probed. Band properties including peak wavelength, FWHM, and integrated area of the four aforementioned bands are measured and compared with the same values of laboratory spectra of attributable species. Comparisons show that the co-occurrence of the 7.24 and the 7.42 µm bands can reasonably be attributed to $HCOO^-$ mixed with H_2O, but other COMs such as C_2H_5OH and CH_3CHO are also likely to contribute, especially to the 7.24 µm band. The 7.67 µm band is solely contributed by CH_4 ice mixed with H_2O and possibly other common species like CO_2 and CH_3OH. This first systematic study of the COM ice fingerprint range provides more observational evidence for solid-phase formation of salts and COMs during early stages of star formation.

Taken together, the four chapters presented in this thesis aim to provide a coherent picture of the origin and early evolution of COMs in star-forming regions. By combining systematic observations of gas-phase COMs in both low- and high-mass protostars with the first direct comparisons between gas and ice in the same sources, this thesis not only establishes an observational framework that links the gas and ice reservoirs of COMs, but also deepens our understanding of their origins in the cold prestellar stage and their subsequent release and reprocessing during protostellar evolution.

1.4.2 Outlook

Looking ahead, the study of COMs in both gas and ice will be continuously advanced by the high-quality, large-sample data obtained from ALMA and JWST observations. There are several approved but yet unpublished programs that focus specifically on COM chemistry in protostars. The ALMA Cycle 9 Large Program Complex Organic Molecules in Protostars with ALMA Spectral Surveys (COMPASS; PI: Jes Jørgensen) will extend the 30-GHz-wide line survey of IRAS 16293 to a sample of 11 low-mass protostars in different star-forming regions and at different evolutionary stages. IR spectra of the same sample will also be obtained by a complementary JWST program. The Cycle 3 JWST Large Program High angular resolution observations of stellar

Emergence in Filamentary Environments (HEFE; PI: Tom Megeath) will deliver IFU spectra in the full 2.9–28 µm range for 13 Class 0 protostars in the OMC2/3 region. Recently, a JWST Cycle 4 program proposed by the author has been approved, in which 15 high-mass protostars with rich detections of gas-phase COM will be observed by MIRI-MRS in search of COM ice fingerprints. Combined with the existing ALMA data for the same sources, the gas-to-ice comparisons presented in Chapter 4 will hopefully be extended to a larger sample and a different mass regime.

As observers look farther and deeper into the universe, more sophisticated models and experiments have also been developed by modelers and experimentalists. The evaluation by models and experiments not only deepens the interpretation of existing data, but also provides guidance for new observations. In particular, the laboratory-measured mm and IR spectra are essential for spectral fitting and directly affect the observational results. Although this thesis focuses exclusively on observational studies, a substantial part of the discussion is devoted to the comparisons with simulation and experimental results. The collaborations among observers, modelers, and experimentalists set the foundation of astrochemistry, and will only become tighter in the future.

1