Characterization of $(CH_2OH)_2$ in High Mass Young Stellar Object NGC 7538 IRS 1

Jodi Balfe
Supervised by Prof. Karin I. Öberg
Astronomy 98, Fall 2013
Harvard College

11 December 2013

Abstract

The utility of complex organic molecules as probes of star formation and astrophysical environments depends on understanding of complex chemistry in protostellar environments. Existing protostellar models of hot core chemistry predict very low abundances of ethylene glycol, $(CH_2OH)_2$, a very large and hydrogen rich complex molecule, in star-forming regions. However, recent observations of Comet Hale-Bopp and Central Molecular Zone (CMZ) clouds have suggested that $(CH_2OH)_2$ can be produced in higher quantities than previously assumed. However, since these sources have extreme environmental conditions that complicate analysis of the chemistry, we cannot directly compare these results to current model predictions. As such, we examine 16 GHz of spectra from high mass young stellar object (YSO) NGC 7531 IRS 1 from the IRAM 30m telescope to determine the molecular abundance and excitation temperature of $(CH_2OH)_2$. Since NGC 7538 IRS 1 is an archetypal hot core, we can directly test current model predictions and benchmark the relative importance of different stages of star formation region complex chemistry. The column density and excitation temperature is calculated in two ways: by qualitative fitting of the most promising line candidates and via the rotation diagram method. We find that current models strongly underestimate $(CH_2OH)_2$, providing the first indication that $(CH_2OH)_2$ is much more common in star-forming regions. Since laboratory experiments have linked $(CH_2OH)_2$ to warm ice photochemistry, the results imply that this stage in star formation and the formation pathways of very complex molecules such as $(CH_2OH)_2$ are of greater relative chemical importance than previously assumed.

1 Astrochemistry Introduction

Astronomical molecules are found in variety of sources, and in addition to being of interest themselves due to the unique nature of space chemistry, are also probes of the physical conditions and histories of their host environments. High-resolution rotational spectra yields information on the density, temperature, and dynamics of surrounding molecular gas, while vibrational spectra reveal the polar or nonpolar character of the reaction surfaces. Since chemical composition is highly dependent on physical environmental conditions such as temperature and density, calculations of molecular abundances can constrain formation pathways and yield information on the density-temperature-radiation history of an astrophysical object. Moreover, chemical composition is also time-dependent, so the molecular inventory of a region can reveal information on the history and evolution of the environment.

1.1 Complex Organic Interstellar Molecules

Over 150 different molecular species have been detected in the interstellar medium, primarily through rotational emission spectra obtained by both ground and space millimeter wave telescopes. While most of the astronomical molecules are ubiquitous on Earth, some are quite rare. For instance, both positively (e.g HCO^+) and negatively (e.g. C_4H^-) charged molecular ions; unpaired electron radicals (e.g. C_6H); and isomer pairs have been detected. Additionally, isotopologues, or molecules containing unusual isotopes (e.g. deuterium, ^{13}C, ^{15}N, ^{17}O, ^{18}O, and ^{34}S) have been identified at higher abundances than expected from cosmic isotope ratios (Herbst and van Dishoeck 2009).

Many of these astronomical molecules contain the heavy element carbon and are thus of an organic nature. Of the larger (more than 6 atoms) molecules, all of the detected species contain carbon. For astrophysical purposes, these molecules can be considered complex organic interstellar molecules. The detected complex organic molecules can be designated either as unsaturated or saturated on the basis of hydrogen richness. Unsaturated molecules are relatively hydrogen-poor, and include carbon clusters (C_n; $n = 2, 3, 5$), radicals of the form C_nH (where $n = 2 - 8$), and cyanopolyynes (HC_nN; $n = 3, 5, 7, 9, 11$). Meanwhile, saturated molecules are hydrogen-rich, and fully saturated molecules only possess single chemical bonds with hydrogen through valence electrons. There are five primary sources of complex molecules as detected in our galaxy: cold interstellar cores, circumstellar envelopes, hot cores, lukewarm cores, and outflows. Tables $1 - 4$ list the detected complex interstellar organic molecules by molecular family, in addition to their sources (cc denotes cold core, hc denotes hot core, of denotes outflow, and env denotes circumstellar envelope). Hot cores exhibit the most diverse and complex molecular inventories relative to other sources, including saturated molecules such as methanol (CH_3OH), dimethyl ether (CH_3OCH_3), and propionitrile (CH_3CH_2CN) at relatively high

Species	Name	Source
C_2H_4	Ethene	env
HC_4H	Butadiyne	env
H_2C_4	Butatrienylidene	env, cc, lc
C_5H	Pentadiynyl	env, cc
CH_3C_2H	Propyne	cc, lc
C_6H	Hexatriynyl	env, cc, lc
C_6H^-	Hexatriynyl ion	env, cc, lc
H_2C_6	Heapentaenylidene	env, cc, lc
HC_6H	Triacetylene	env
C_7H	Heptatyinyl	env, cc
CH_3CH_4H	Methyldiacetylene	cc
CH_3CHCH_2	Propylene	cc
C_8H	Octatetraynyl	env, cc
C_8H^-	Octatetraynyl ion	env, cc
CH_3C_6H	Methyltriacetylene	cc
C_6H_6	Benzene	env

Table 1: Hydrocarbons

fractional abundances on the order of 10^{-7} (Herbst and van Dischoeck 2009). Cold cores are the source of many exotic complex molecular species, including ions, radicals, isotopologues, and unsaturated complex organic molecules; however, in both collapsing (pre-stellar) and non-collapsing cold cores, the molecular complexity and diversity is much less than that of hot cores.

Although nearly all of the complex molecules have been detected in the gas phase, there is strong evidence to suggest that these molecules can be formed in the ice mantles of interstellar grains via solid phase chemistry. However, absorption spectra from solid state molecules are too broad to be assigned to species with necessary certainty, whereas gas phase spectral features are sufficiently narrow. Absorption features arising from the complex species in grain mantles are too weak, so only small molecules have been explicitly identified in ice grain mantles, the most common being water (fractional abundance with respect to H_2 of 10^{-4}). The only complex molecule detected unambiguously is solid methanol CH_3OH with a fractional abundance with respect to H_2 of up to 3×10^{-5}. However, significant fractional abundances for other complex organics have been derived in CH_3OH ice laboratory experiments, demonstrating that CH_3OH ice photochemistry is efficient enough to explain observed abundances of complex organics around protostars (Öberg et al. 2009).

1.2 Chemical Processes of Star Forming Regions

In this section, we provide a general overview of the types of reactions that are significant in star forming regions. Star formation occurs in interstellar dense molecular clouds with temperatures $\lesssim 300$ K and densities $\gtrsim 10^3$ cm^{-3}. The majority of gas present is H_2, followed by He at an abundance of 0.20 relative to the H_2 and then by CO at a relative abundance of $\approx 2 \times 10^{-4}$. The remaining oxygen is believed to be contained primarily in H_2O, O, and O_2 (van der Tak 2005).

Species	Name	Source
CH_3OH	Methanol	cc, hc, gc, of
HC_2CHO	Propynal	hc, gc
$c-C_3H_2O$	Cyclopropenone	gc
CH_3CHO	Acetaldehyde	cc, hc, gc
C_2H_3OH	Vinyl alcohol	hc
$c=CH_2OCH_2$	Ethylene oxide	hc, gc
$HCOOCH_3$	Methyl formate	hc, gc, of
CH_3COOCH_3	Acetic acide	hc, gc
$HOCH_2CHO$	Glycoaldehyde	hc, gc
C_2H_3CHO	Propenal	hc, gc
C_2H_5OH	Ethanol	hc, of
CH_3OCH_3	Methyl ether	hc, gc
CH_3COCH_3	Acetone	hc
$HOCH_2CH_2OH$	Ethylene glycol	hc, gc
C_2H_5CHO	Propanal	hc, gc
$HCOOC_2H_5$	Ethyl formate	hc

Table 2: O-Containing

Species	Name	Source
CH_3CH	Acetonitrile	cc, hc, of
CH_3NC	Methylisocyanide	hc
CH_2CNH	Keteneimine	hc
HC_3NH^+	Prot. cyanoacetylene	cc
C_5N	Cyanobutadiynyl	circ
CH_3NH_2	Methylamine	hc, gc
C_2H_3CN	Vinylcyanide	cc, hc
HC_5N	Cyanodiacetylene	circ, cc
CH_3C_3N	Methylcyanoacetylene	cc
CH_2CCHCN	Cyanoallene	cc
NH_2CH_2CN	Aminoacetonitrile	hc
HC_7N	Cyanotriacetylene	circ, cc
C_2H_5CN	Propionitrile	hc
CH_3C_5N	Methylcyanodiacetylene	cc
HC_9N	Cyanotetraacetylene	circ, cc
C_3H_7CN	N-propyl cyanide	hc
$HC_{11}N$	Cyanopentaacetylene	circ, cc

Table 3: N-Containing

Species	Name	Source
CH_3SH	Methyl mercaptan	hc
NH_2CHO	Formamide	hc
CH_3CONH_2	Acetamide	hc, gc

Table 4: N & O, S-Containing

In addition to gaseous species, star forming regions contain solid dust grains with of radius $\approx 0.1\mu$m at an abundance of $\approx 1\%$ by mass or 3×10^{-12} by number, as schematically represented in Figure 1. At the center of these grains is a *grain core* of carbon and silicate, covered with layers of volatile material referred to as the *ice mantle*. As more atoms and molecules collide with the grain surface, the incoming material (adsorbates) accrete onto the ice mantle. As a result, the composition of the ice mantle depends more on the chemical composition of the surrounding gas rather than the grain temperature itself. Dust grains are identified through the broad absorption features that arise in mid-infrared spectra of molecular clouds resulting from the ice mantle, which is predominantly H_2O (abundance of $\sim 10^{-4}$). The calculated solid state abundances of complex organic molecules is significantly higher than can be produced solely via gas phase reactions, strongly suggesting the occurrence of grain surface reactions (van der Tak 2005). In the following sections, we describe the types of gas phase and solid state reactions that can occur in star-forming regions.

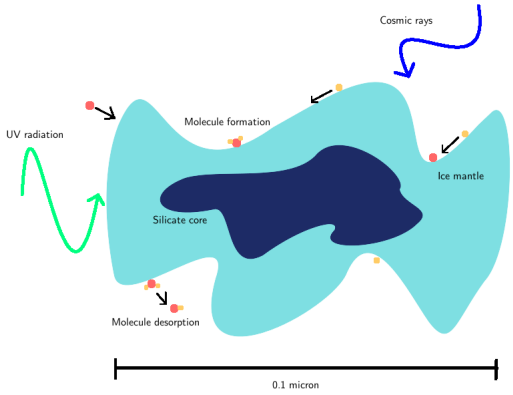


Figure 1: *Schematic representation of an interstellar dust grain.* A silicate grain core is covered with an icy mantle as atoms collide and accrete on the grain surface.

1.2.1 Gas Phase Processes

The majority of astronomical molecules have been detected as products of gaseous reactions, which can involve not only molecules, ions and atoms, but also electrons, cosmic rays, and ultraviolet photons in the interstellar medium (van der Tak 2005). However, in dense clouds (observationally defined as $A_V \gtrsim 3$) the primary chemical processes and their rate coefficients, denoted by k, are as follows:

- *Ion-molecule reactions* ($k \sim 10^{-9}$ cm^3 s^{-1}) : Since these reactions do not involve high activation barriers, they dominate the chemistry of cold dark clouds with very little kinetic energy ($T \approx 10$ K). These reactions tend to have unsaturated products with double or multiple bonds.

- *Dissociative recombination reactions* ($k \sim 10^{-6}$ cm^3 s^{-1}): While these reactions have relatively high rate coefficients, this is primarily due to Coulomb attraction and the actual reaction rate, also dependent on species density, is low due to the low electron abundance ($\sim 10^{-8}$) in dense clouds. Consequently,

dissociative reaction reactions are much less frequent, particularly relative to exothermic H_2 or CO (high density) reactions.

- *Neutral-neutral reactions* ($k \sim 10^{-12}$ cm^3 s$^{-1} - 10^{-10}$ cm^3 s^{-1}): In addition to having low rate coefficients, neutral-neutral reactions often require overcoming substantial activation barriers (~ 0.1 eV). Consequently, neutral-neutral reactions are more present in hot core chemistry rather than the processes of cold clouds.

In the initial phase of star formation, occurring in cold, dense clouds, the absence of kinetic energy (≈ 10 K) results in very slow speeds. As the interstellar gas and dust condense into a cold core, gas phase chemistry proceeds, and as ice mantles form, solid state chemistry is initiated. The low kinetic energy of cold cores ($T_{gas} \approx 10$ K) requires that reactions must be exothermic and have no potential barriers in order to proceed (Herbst and Klemperer 1973). This is in contrast with gas phase reaction conditions on Earth, where not only are temperatures relatively high, but also atmospheric densities are orders of magnitude greater so reactions between three or more species at a time can occur. In three-body reactions, the third body can absorb excess bond energy. However, in cold cores collisions are relatively infrequent occurrences, and only two-body reactions are significant. Consequently, ion-neutral reactions dominate gas phase cold core chemistry. The relatively high abundance of positive ions is enabled by cosmic ray bombardment. For instance, letting γ_{CR} denote a cosmic ray photon, the simplest polyatomic ion H_3^+ can be formed as follows:

$$H_2 + \gamma_{CR} \rightarrow H_2^+ + e^- + \gamma_{CR},$$

$$H_2^+ + H_2 \rightarrow H_3^+ + H.$$

Several chains of ion-neutral reactions begin with H_3^+, and many of the subsequent products are precursors for complex organic molecules. Since most hydrogenation reactions involve H-atom transfer to a positive ion (X^+) of the form $X^+ + H_2 \rightarrow XH^+ + H$, the endothermic nature of this process prevents it from occurring in cold cores. Consequently, the complex icons formed at these low temperatures tend to be unsaturated. For hydrocarbon ions $C_nH_m^+$, for $n > 3$, H-atom transfer reactions cannot produce hydrocarbon ions with $m > 2$, resulting in only unsaturated ions. Furthermore, subsequent dissociative recombination processes can dismantle hydrocarbon ions in a reaction of the form $C_nH_2^+ + e^- \rightarrow C_nH + H$, resulting in neutral molecules with fewer hydrogen atoms than the complex ion reactant (McEwan et al. 1999).

1.2.2 Solid State Processes

In star-forming regions, solid state processes primary involve dust grains and can enable reactions that would not occur otherwise in the gas phase, particularly in cold, dense cores. In particular, in $T \lesssim 10$ K environments, the evaporation time of an atomic species is greater than the time necessary to traverse the

entire grain surface by means of thermal hopping or quantum tunneling. Consequently, the grain surface enables reactions to occur that would not otherwise be possible in such low kinetic energy environments, especially neutral-neutral reactions. Moreover, the products of grain surface reactions are predominantly saturated species dominated by single bonds, as opposed to the products of gas phase ion-molecule reactions. The primary solid state processes are as follows:

- *Hydrogenation*: In regions of densities less than 10^4 cm^{-3}, H is more abundant than O and hydrogenation is the dominant process. In particular, O reacts to produce H_2O, N to produce NH_3, C to produce CH_4, S to produce H_2S, and CO to produce H_2CO and CH_3OH. Most ices are dominated by H_2O and are referred to as polar ices, where those dominated by CO products are referred to as apolar ices. Spectral analysis of cold cloud ice mantles reveals the prevalence of these hydrogenation products, and thus the dominance of the hydrogenation process (Gibb et al. 2000).

- *Oxygenation*: In regions of higher densities, O is more abundant than H and oxygenation becomes a more significant process, as evidenced by large abundances of CO_2 ices (Garrod and Pauly, 2011).

- *Deuteration*: While deuterated ice species have not been observed yet, the abundance ratio of D to H is several orders of magnitude in the solid state compared to the gas phase elemental ratio. Therefore, grain surface chemistry may also include deuteration, as evidenced by the deuterium enrichment of evaporated H_2CO and CH_3OH (van der Tak 2005).

At cold region temperatures, sublimation can occur effectively for very light adsorbates (e.g. H, H_2, He) that can be held to the dust grain core by weak bonds. Accretion is achieved via van der Waals forces (physisorption) and valence bonding (chemisorption). Since chemisorption has a potential barrier, accretion is considered to be predominantly physisorption at cold temperatures. Grain surface reactions occur when reactive adsorbates are contained in the same potential minimum, allowing the adsorbates to combine as the grain absorbs the excess energy. The accretion of adsorbates allows for an ice mantle to form from grain surface chemistry through three primary mechanisms: the Langmuir-Hinshelwood mechanism, the Eley-Rideal mechanism, and the hot atom mechanism. These three mechanisms enable the adsorbates to reach a common potential minimum before evaporation, as schematically visualized in Figure 2. In the *Langmuri-Hinshelwood mechanism*, diffusion across the grain surface is occurs by tunneling or thermal hopping over the potential barrier between binding sites. When reactions occur via the *Eley-Rideal mechanism*, a gas-phase species lands directly on top of an adsorbate and a reaction is initiated. The *hot atom mechanism* involves a gas-phase species landing on the grain and scanning the surface to collide with an adsorbate before thermalization. The Langmuir-Hinshelwood mechanism is considered to be the predominant means by which surface reactions occur (Herbst and van Dischoeck 2009). Since atomic hydrogen adsorbates diffuse quickly on grain surfaces, atomic hydrogen is a dominant reactant and hydrogenation reactions are relatively efficient

on grain surfaces. Consequently, the products of diffusive chemistry are mainly saturated species, in contrast with the cold core gas phase reactions. Water ice, the dominant ice species, is produced primarily by the hydrogenation of O atoms that land on the surface, and hydrogenation reactions also form NH_3 from N and CH_4 from C. Moreover, when CO accretes onto the grain surface, sequential hydrogenation reactions can convert it to methanol, the precursor to other complex organics:

$$CO \rightarrow HCO \rightarrow H_2CO \rightarrow H_2COH \rightarrow CH_3OH \tag{1}$$

Although more complex, larger molecules can diffuse less efficiently, there is a possibility that ethanol and acetaldehyde can also be formed from CO in low temperature environments on grain surfaces:

$$CO \rightarrow HCO \rightarrow HCCO \rightarrow CH_2CO \rightarrow CH_3CO \rightarrow CH_3CHO \rightarrow C_2H_5OH. \tag{2}$$

Thus, dust grain surfaces host reactions that would otherwise be impossible in low temperature cold cores, resulting in the formation of especially many hydrogen-rich species and complex molecules, including methanol (Bisschop et al. 2007).

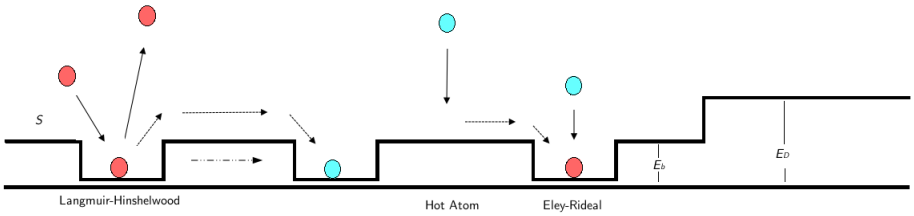


Figure 2: *Three mechanisms of grain surface chemistry.* The Langmuir-Hinshelwood mechanism, hot-atom mechanism, and Eley-Rideal mechanism are the three known means by which adsorbates reach a common potential minimum, allowing a reaction to occur, on an interstellar grain surface.

1.3 High Mass Young Stellar Objects and Complex Interstellar Molecules

The stages of formation of high mass young stellar objects (YSOs) occur in highly obscure regions in which the young stellar object is surrounded by envelopes of cold, lukewarm, and hot gas and dust. Due to the presence of large surrounding regions of gas and dust, high mass YSOs have not been spatially resolved with millimeter interferometers, hence have not been as extensively studied as low mass YSOs. As such, the formation of more massive YSOs is not as well understood. Since the central star is so deeply embedded, observations of trace species frequently provide the best constraints on the early formation stages of high mass YSOs (van Dischoeck and Blake 1998). However, the usefulness of dust and gas phase molecular probes critically hinges on understanding of the chemical behavior of these molecules. In particular, complex organic

molecules are frequently used as molecular probes for the hot core stage of high mass YSO formation due to their strong emission from the dense hot core region surrounding the protostar. Thus, it is necessary to understand the chemical behavior of these complex organic species in order to meaningfully utilize them to understand the formation of high mass YSOs.

1.3.1 Chemistry of Hot Cores

As mentioned previously, hot cores possess the most diverse and complex molecular inventories, allowing various complex molecules to be employed as probes. The chemistry of hot cores can be divided into three phases following star evolution, as schematically represented in Figure 3. In the initial phase, also known as the *cold phase* (< 25 K), the cold core ages and the protostellar envelope starts to collapse due to its own gravity or external pressure. While temperatures are still very low, molecules and atoms still undergo collisions and adsorbates accrete onto grain surfaces, forming icy mantles. In the second phase, known as the *warm-up phase* (> 25 K), the luminosity of the young star increase due to both accretion of gas and dust onto the protostar and nuclear fusion. The inner envelope of the protostar is heated up to $100 - 300$ K from the cold core temperature of < 25 K and the ice mantles start to sublimate as material flows inward towards the protostar. The third phase, known as as the *hot-core phase* (> 100 K) occurs when the temperature has stabilized and the ice mantles completely evaporate, releasing grain-surface products to further evolve through gas phase chemistry as the hot core condenses. In the hot-core phase, chemistry is dominated by high temperature gas phase reactions since the ice mantles have evaporated (Herbst and van Dischoek 2009). In the following section, we examine the phases and their chemical processes in more detail.

1.3.2 Zeroth-Generation Species

The cold phase is characterized by cold gas chemistry, forming unsaturated molecules, and hydrogenation grain reactions resulting the accretion of adsorbates onto cold dust grain surfaces. The products of cold phase reactions are known as *zeroth-generation species*. The grain surface enables simple ices to be formed in addition to more complex and saturated molecules that would not otherwise be produced at such low temperatures (Herbst and van Dischoek 2009). In particular, CH_3OH, the key precursor to many more complex molecules, can be produced from the grain surface hydrogenation of CO (van der Tak 2005). Cosmic rays can heat grains and cause local evaporation of the ice mantle, releasing the products of solid state reactions to be subsequent reactants in gas phase processes, reflecting the interaction between chemistry in the solid state and gas phase during the cold phase.

1.3.3 First-Generation Species

During the warm-up phase, inflowing material passively heats the protostar from cold core temperatures ($\sim$ 10 K) to $100 - 300 \sim 10$ K as a YSO develops (van der Tak 2005). Due to this temperature transition, the warm-up phase chemistry is characterized by significant changes in grain surface reactions and efficient complex ice chemistry. The products of warm-up phase chemistry are referred to as *first-generation species*, and are dominated by products of photodissociation reactions of zeroth-generation ices combined with radial diffusion and recombination in icy mantles. Specifically, as the temperature exceeds $\sim$ 20 K, the evaporation time of H atoms significantly decreases and H atoms are no longer the dominant reactants on grain surfaces. Consequently, species such as CO, CH_4, and N_2 can sublimate at higher temperatures, and reactions involving heavier molecules that can now diffuse more quickly compete with hydrogenation reactions. While these larger species are generally too stable to be reactive, cosmic-ray UV photons can drive active photodissociation in which reactive neutral species or radicals are created from zeroth-generation precursors (e.g. H_2O, H_2CO, CH_3OH, CH_4) and allowed to diffuse int he mantle. As a result, larger complex organic species are formed in the warm-up phase. For instance, warm-up phase reactions forming methyl formate ($HCOOCH_3$), dimethyl ether (CH_3OCH_3), and formic acid (HCOOH) have been proposed (Garrod and Herbst 2006):

$$HCO + CH_3O \rightarrow HCOOCH_3,$$

$$CH_3 + CH_3O \rightarrow CH_3OCH_3,$$

$$HCO + OH \rightarrow HCOOH.$$

1.3.4 Second-Generation Species

The hot-core phase is characterized by a complete sublimation of ice mantles on dust grains, resulting in a dominance of hot gas-phase chemistry. The high kinetic energy in the hot core allows many more reactions, including endothermic reactions and reactions with potential barriers, to occur. In particular, both ion-neutral and neutral-neutral reactions can occur. The products of hot gas phase chemistry are referred to as *second-generation species*. In the hot-core phase, the ices evaporate on short time scales, releasing the zeroth and first generation complex molecular species into the gas phase and creating second generation products through high density hot gas phase chemistry.

1.3.5 Previous Observational, Theoretical, and Experimental Work

The earliest studies of hot cores demonstrated the diversity and complexity of their molecular inventories. Observations of hot cores Orion-KL and Sgr B2(N) resulted in the identification of more complex astronomical molecules than previously detected in cold cores. Observational surveys shifted starting in 2000 to

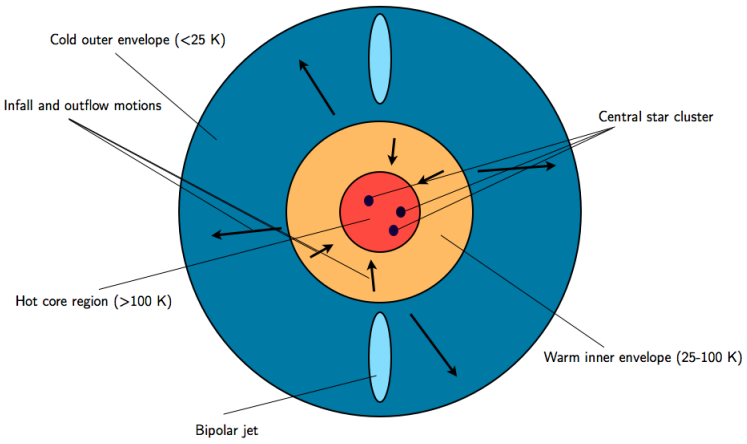


Figure 3: *Chemistry of star-forming regions.* In a high mass YSO, the central protostar heats a large surrounding envelope. Cold phase chemistry, dominated by grain surface hydrogenation and cold gas phase reactions occurs in the outer envelope, leading to the formation of zeroth generation molecules. First generation molecules are formed in the warm up phase as released radicals can diffuse and recombine and photodissociation occurs, enabling very efficient complex ice chemistry. Through non-thermal desorption, zeroth and first generation products can be released into the gas phase and allowed to react. In the hot core, the icy mantles are sublimated, releasing the zeroth and first generation molecules to evolve further via hot gas phase reactions into second generation molecules. Meanwhile, the infall and outflow motions enable further chemical mixing, and the bipolar jet provides UV radiation to drive photodissociation reactions.

targets with narrower lines (e.g. G 327.3-0.6 and NGC 6334) to reduce line assignment ambiguities (Schilke et al. 2006). Moreover, examining high mass YSOs at small frequency windows (e.g. 230 and 345 GHz surveys of Sgr B2 (N), (M), and (NW); NGC 6334 I; and W3 IRS4, IRS 5 and H_2O) has enabled the comparison of chemistry within hot cores, revealing significant diversity in molecular inventories (Helmich and van Dischoeck (1997)). For instance, while Sgr B2(N), NGC 6334 I, and W3 H_2O contain significant complex organic spectral abundances linked to second-generation chemistry, the same molecular presence is not found in other high mass YSOs, despite their similar physical characteristics. For instance, the spectra of Sgr B2(M) and W 3 IRS5 are dominated by SO and SO_2 lines, whereas the spectrum of W 3 IRS 4 displays strong radical lines from CN and C_2H. Therefore, although hot cores may share physical conditions and location, their detected chemical abundances can significantly differ. If deeper surveys reveal closer abundance values and molecular presences, then perhaps the chemistry between hot cores is not changing but rather the varying sizes of the hot core region is affecting the beam fill factor, resulting in different measured molecules.

Specific searches for individual molecules or molecule families have been carried out for a number of complex organic molecules, providing demonstrating the efficiency of the different stages of complex chemistry

in star-forming regions. Bisschop et al. (2007) examined the chemical origins of a set of complex organic molecules in 7 high mass YSOs and classified complex organic molecules as either hot or cold based on their excitation temperature, suggesting two regions of complex organic molecule formation. The detected high column densities of complex organic molecules could not have been produced solely by pure gas-phase processes, but must have been the product of successive hydrogenation and oxygenation grain surface reactions. In particular, the abundances of hot complex organic molecules are correlated, which is most likely due to shared chemical history as first-generation species formed via solid-state chemistry. The high detected abundances are unable to be explained by pure gas phase molecules, particularly for saturated molecules. Meanwhile, second generation gas phase species are expected to peak in abundance independently and thus cannot support the observed correlations. Thus, most complex organic molecules, such as H_2CO, CH_3OH, C_2H_5OH, $HCOOCH_3$, and CH_3OCH_3, are most likely first generation species formed in solid state reactions on grain mantles.

Larger complex organic molecules, including both aldehydes and reduced alcohols, were later detected and studied by Requena-Torres et al. (2007) in molecular clouds MC G-0.11-0.08 and MC G-0.02-0.07 in the Sgr A* complex and MC G +0.693-0.03 in the Sgr B2 complex. While complex organic molecules have typically been associated with hot cores, Requena-Torres et al. found very high abundances for all observed complex organic molecules in the CMZ, indicating that these complex organic molecules were formed on grain mantles to reach such large quantities. Moreover, the study found similar relative abundances between all three clouds, despite the sources being located in very different environments, implying common chemical processes involving a shared grain mantle composition and within the CMZ. For both Galactic disk hot cores and the CMZ, observations show that complex organic chemistry favors the reduced alcohol over the corresponding aldehyde, elucidating the largely unknown formation processes for complex organic molecules.

The efficiency of grain surface chemistry in producing large complex organic molecules in protostars is quantitatively supported in the two-phase (Garrod and Weaver 2008) and three-phase (Garrod 2013) chemical network models of hot core chemistry. The two-phase model assumes homogeneity in chemical behavior on the surface and within the ice, so surface kinetics were used to simulate all dust-grain chemical processes on the surface of the grain and within the bulk ice, in addition to gas-phase reactions. After experimental results demonstrated that sub-surface processes could occur via UV radiation (Öberg et al. 2009), allowing complex molecules to further evolve even in very thin ices ($\sim$ 10 monolayers), a three-phase model was developed. In contrast to two-phase models, the three-phase model simulates bulk ice chemistry, ice surface chemistry, and gas phase chemistry. The chemical networks utilize assume grain surface radical-radical formation mechanisms ans destruction mechanisms, including cosmic UV ray photodissociation (both on the

grain surface and in the gas phase), gas-phase ion-molecule reactions, electronic recombinations of products, the accretion of gas phase species onto grain surfaces, and both thermal and non-thermal desorption from the grains. The physical model utilizes a collapsing envelope (cold core) phase followed by a static-warm u phase in which the core is heated from 8-400 K. Three warm-up speeds (designated fast, medium, slow) are tested, and the abundances of warm-up phase species are simulated. In particular, the models show that the gradual warm up of of the hot core enables strongly bound "heavy" radicals to become mobile on the grain surface, allowing them to recombine into complex molecules. As the models progress and the ice is sublimated, the resulting abundances are comparable to the observed abundance ratios of complex organic molecules detected towards hot cores. The coupled two-phase gas phase, grain surface model and the three phase gas phase, grain surface, and bulk ice model both fully simulate the formation of molecules reaching a high degree of complexity, supporting grain chemistry as the formation mechanism enabling the observed large abundances of complex organic molecules in hot cores. In particular, CH_3OH is shown to be a critical reactant of known zeroth-generation molecules contributing to the production of first-generation complex molecules.

In addition to observational and theoretical evidence, the efficiency of warm-up stage photochemistry chemistry in producing complex organics at large detected abundances is supported by laboratory experiments. Öberg et al. (2009) examined the efficiency of photochemistry in CH_3OH ices with a combination of experiments and kinetic modeling to determine the CH_3OH photodesorption yield, the CH_3OH dissociation branching ratios upon UV radiation, and diffusion barriers of formed radicals and reaction barriers in the formation of more complex molecules. The CH_3OH ices were subject to astrophysical conditions, and first generation complex molecules were identified and their abundances were determined. These laboratory studies found that warm ice photochemistry is efficient enough to enable a significant fraction of the CH_3OH ice to be converted into more complex molecules. The abundances of oxygen-rich complex species in the ice experiments simulating prestellar and protostellar conditions were comparable to detected abundances in astrophysical sources.

2 Motivation and Previous Work

Gas phase processes were long believed to dominate the formation of complex organic molecules in star forming regions. However, recent observational, experimental and theoretical evidence has demonstrated that gas phase chemistry is not efficient enough to produce molecules of a such a high degree of complexity at detected high abundances. Consequently, as discussed in the previous section, complex organic molecules are now propose to form in the icy mantles surrounding interstellar dust grains, which enable solid state reactions to proceed. Meanwhile, the relative importances of the cold, warm-up, and hot phases of complex molecule formation in star forming regions are not as well-understood, especially since many complex molecules can be produced through both first and second generation chemistry.

2.1 Ethylene Glycol ($(CH_2OH)_2$) and Warm Ice Photodissociation Chemistry

Ethylene glycol, ($(CH_2OH)_2$), is one of the largest hydrogen-rich (relative to carbon chains or polycyclic aromatic hydrocarbons), was first detected in the interstellar medium towards a CMZ cloud, Sgr B2(N-LMH), yet only at low abundance of 0.1% with respect to CH_3OH. Consequently, $(CH_2OH)_2$ has been assumed to be present in low abundances in all star forming regions, even beyond the CMZ, with other complex molecules favored. Furthermore, the two- and three-phase protostellar models of the warm-up stage in hot cores as described previously (Garrod and Weaver 2008, Garrod 2013) predict $(CH_2OH)_2$ abundances relative to CH_3OH on the order of $0.0001 - 0.01\%$. As such, the abundance of $(CH_2OH)_2$ has always been assumed to be present in low abundances. However, recent experimental and observation evidence suggest that $(CH_2OH)_2$ should be produced in larger, more significant quantities via low-temperature chemistry. Since $(CH_2OH)_2$ is a molecule of relatively high complexity, understanding its abundance probes the conditions and mechanisms for very complex molecule formation in star-forming regions.

2.1.1 Comet Hale-Bopp

Crovisier et al. (2004) detected 8 lines of $(CH_2OH)_2$ while observing Comet Hale-Bopp with the IRAM 30m telescope and found that $(CH_2OH)_2$ is one of the most abundance CH) molecules in the comet. The abundance of $(CH_2OH)_2$ was $10 - 20\%$ that of CH_3OH, and at least 5 times more abundant than C_2H_6O. Unlike in the case of Sgr B2(N-LMH), $(CH_2OH)_2$ is of highly comparable abundance to other complex molecules in the comet. Moreover, the complex molecule abundances for species besides $(CH_2OH)_2$ are comparable to those found in hot cores, unlike Sgr B2(N-LMH). Crovisier et al. propose that $(CH_2OH)_2$ may not have been been as relatively abundant in Sgr B2(N-LMH) due to the large spatial scale of glycoaldehyde, the likely precursor for $(CH_2OH)_2$, which may have prevented the grain sublimation necessary for hydrogenation reactions.

2.1.2 Central Molecular Zone (CMZ) Clouds

Following the detection of $(CH_2OH)_2$ in relatively high quantities in Comet Hale-Bopp, additional evidence implies that $(CH_2OH)_2$ is much more common than previously assumed. Requena-Torres et al. (2007) detected $(CH_2OH)_2$ at significant abundances in the CMZ ($0.4 - 2.2 \times 10^{-8}$ with respect to H_2 and $1.2 - 2.4 \times 10^{-2}$ with respect to CH_3OH), which hosts physical conditions very different from those in the hot core. Despite the differences in physical conditions between hot cores and Galactic Center clouds, the abundances of other detected complex organic molecules (C_2H_5OH, CH_3OCH_3, $HCOOCH_3$, HCOOH) relative to CH_3OH are similar, implying a shared grain mantle composition in the hot cores of the Galactic disk and the CMZ. The abundance similarities for other complex organic molecules relative to methanol would suggest that CH_3OH is present at higher abundances than previously detected in hot cores. Moreover, ice chemistry laboratory experiments yield an abundance of $(CH_2OH)_2$ relative to CH_3OH of 0.4, the same order of magnitude as the Comet Hale-Bopp abundance.

2.1.3 Laboratory Experiments

In addition, the laboratory experiments discussed in the previous section involving the UV-radiation of CH_3OH ices demonstrate that $(CH_2OH)_2$ can be formed in relatively high quantities in warm-up phase temperatures. When a CH_3OH:CO 1:10 ice mixture was subjected to astrophysical conditions, the abundance of (CH_2OH) relative to CH_3OH was ~ 0.01, and with optimal conditions of pure CH_3OH ice, the abundance of $(CH_2OH)_2$ reached 0.4 relative to CH_3OH ice. Both pure CH_3OH and CH_3OH:CO ice mixtures are possible since CH_3OH is proposed to form from CO ice. The proposed reaction scheme based on laboratory experiments for $(CH_2OH)_2$ is presented in Figure 4.

Therefore, recent observations of Comet Hale-Bopp and CMZ clouds in addition to laboratory experiments realize much higher $(CH_2OH)_2$ abundances than predicted by current hot core models. However, both Comet-Hale Bopp and CMZ clouds are relatively extreme sources with more volatile environments which complicate the interpretation of the chemistry. Therefore, the aim of this study is to determine the abundance of $(CH_2OH)_2$ in a characteristic, typical hot core to provide directly comparable results to protostellar model predictions. Furthermore, since laboratory experiments linked $(CH_2OH)_2$ to the warm-up stage, analysis of a hot core will provide insight into the relative importance of warm ice photochemistry and formation pathways of very complex molecules. In turn, understanding the chemistry of hot cores, specifically the relative importance of different formation and destruction stages, allows very complex molecules such as $(CH_2OH)_2$ to be used as molecular probes of high mass YSO formation.

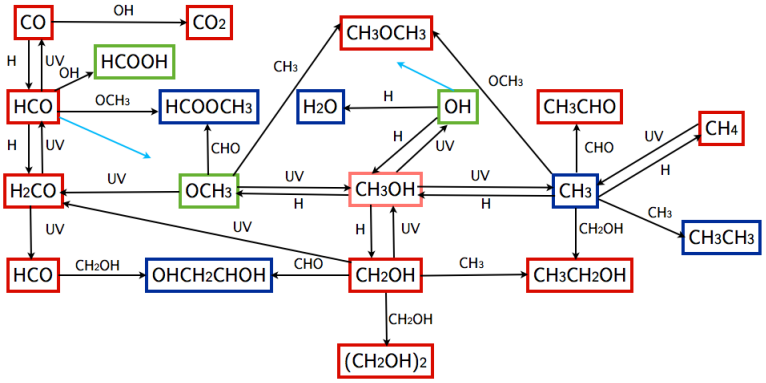


Figure 4: *Proposed reaction scheme.* Laboratory experiments involving UV-radiation have shown that $(CH_2OH)_2$ is formed from CH_3OH ices through warm ice photochemistry at high abundances. The reaction scheme above proposed by Öberg et al. (2009) leads to various complex products form CH_3OH ice, outlined in the pink box, reflecting how CH_3OH is the key zeroth generation precursor for more complex molecules. The products of pure CH_3OH ice reactions are shown in red, and the products of either CO:CH_3OH or CH_4:CH_3OH ice mixtures are shown in blue. The products whose formation could only be constrained by upper limits are shown in green, and cyan arrows show radicals that can be part of additional reactions. All reactions are reversible.

2.2 NGC 7538 IRS 1

In light of the evidence suggesting higher abundances of $(CH_2OH)_2$ in star-forming regions, this paper aims to estimate the $(CH_2OH)_2$ abundance in NGC 7538 IRS 1, a high mass young stellar object (YSO). Unlike the CMZ clouds and Comet Hale-Bopp, NGC 7538 IRS 1 is an archtypal hot core. Hence, analysis of this massive YSO will provide us with chemical information directly comparable to existing protostellar models. The large hypercompact ($\sim 4'$) HII cloud region NGC 7538, located at 2.65[0.12] kpc, hosts several massive young stellar objects, of which NGC 7538 IRS 1 is the brightest (Moscadelli et al. 2009). Analysis of radio and infrared data suggest that NGC 7538 IRS 1 is an O6/7 star with a luminosity on the order of 10^5 $L_\odot$ (Akabane and Kuno 2005). NGC 7538 IRS 1 was first resolved with the Very Large Array (VLA) at 14.9 GHz, and later found to have a compact ($\sim 0.2''$) bipolar north-south core with extended free emission lobes within the center $2''$ region (Campbell 1984; Gaume et al. 1995). The source has been previously found to host a prominent molecular outflow in CO in the north-southwest direction, photodissociation fronts, and very broad hydrogen recombination lines indicating substantial mass motion of ionized gas (Minier et al. 2000; Keto et al. 2000). In addition, a northwest-southeast linear structure of methanol masers with a velocity gradient of ~ 0.02 km s^{-1} AU^{-1} was found by Minier et al. (1998, 2000). Moreover, the strong hot core emission from NGC 7548 IRS 1 has allowed very large abundances of many complex molecules to be determined (Bisschop et al. 2007).

In this paper, we analyze spectral observations from high mass YSO NGC 7538 IRS 1. In Section 3, we review the observational settings and processing of deep 1mm spectral data of 16 GHz bandwidth, then present the identification of $(CH_2OH)_2$ line candidates and analyze the effects of line blending on competing assignments. In Section 4, we obtain first estimates of the column density and excitation temperature of $(CH_2OH)_2$ in the source by simulating uncontaminated spectra by fitting the strongest transitions, and obtain a qualitative means of visualizing the effects of line blending. To obtain more quantitative results with formal uncertainties, we construct a rotation diagram from Gaussian fits to the spectral profile in order to obtain second estimates of the column density and excitation temperature in Section 5. A quantitative and visual comparison between the two parameter sets is presented in Section 6. In Section 7, we compare our results to $(CH_2OH)_2$ abundances as observed in other astrophysical sources, generated in laboratory experiments, and most critically, to current protostellar model predictions. Additionally, the abundances and excitation temperatures of other complex molecules previously studied in NGC 7538 IRS 1 are compared to the results of this study. Section 8 presents concluding remarks.

3 Observations

3.1 Data Sample and IRAM Settings

NGC 7538 IRS 1 was observed with the IRAM 30m Telescope on February 19-20, 2012. The data was collected in four sidebands with frequency ranges given in Table 5 at a spectra resolution of $\sim$ 0.2 km s^{-1} with a sideband rejection of -15dB (Carter et al. 2012). The IRAM 30m Telescope has a beam size (θ_{beam} given in arcsec) of $10.5''$ in collected frequency ranges. The front-end was EMIR 230 GHz facility receiver, and FTS was used as the back-end, with a channel width of 0.0002 GHz. The pointing position was $\alpha = 23:14:01:60$, $\delta = 61:27:20.4$, and pointing was checked periodically every $1-2$ hours and always remained within $2-3''$. and corrections were always on the order of $0.1-0.7''$. Focus was checked every 4 hours and corrections were minor (on the order of 0.2-0.4, except for 0.7 once). The sky, atmosphere and instrument backgrounds were subtracted automatically by the telescope, which performs both position and wobbler switching to alternatively observe positions +220" and -220" away from the source. The total integration time was $\sim$ 4 hours.

Sideband	**Frequency Range (GHz)**
Lower outer	223.32323 - 227.37251
Lower inner	227.05256 - 231.10184
Upper inner	239.00042 - 243.04969
Upper outer	242.72975 - 246.77902

Table 5: *Observing Frequency Ranges.* IRAM 30m observations were taken in 4 sidebands to cover 16 GHz.

3.2 Data Processing

The initial spectra were reduced using `GILDAS` reduction software in `IDL`. A baseline was fitted to each 4 GHz interval, resulting in 4 windows. The data was acquired with intensity and frequency units that required correction. The initial spectral data was collected with two reference frequencies for the upper (241.3 GHz) and lower (228.8 GHz) sidebands and output in terms of velocity offset (km/s) from the reference frequency. For reference frequency f_0 and velocity offset $\Delta v = v_{source} - v$, the frequency was set to a GHz scale, f, by first applying a Doppler velocity correction using the known systemic velocity of the source (-57 km/s) to the offset, then applying the offset to the reference frequency:

$$f = f_0 \left(1 - \frac{\Delta v}{c}\right)$$

The intensity observations were the converted from original antenna temperature, T_a, to main beam temperature units T_{MB}, (both in units of K) using $F_{eff}/B_{eff} = 1.533$ for the lower sideband and $F_{eff}/B_{eff} = 1.607$ for the upper sideband.

3.3 Data Analysis

3.3.1 Preliminary Identification of $(CH_2OH)_2$

The 16 GHz of spectral data includes a variety of species, including a diverse complex organic inventory. Since $(CH_2OH)_2$ is relatively less abundant and susceptible to blending with stronger emission lines, the $(CH_2OH)_2$ lines are weaker and relatively difficult to identify. To determine $(CH_2OH)_2$ the spectral lines, we start with The *Splatalogue* Database for Astronomical Spectroscopy (`http://www.splatalogue.net`), which includes data from the JPL Catalog, the Cologne Database for Molecular Spectroscopy (CDMS), the Lovas/NIST database. The $(CH_2OH)_2$ line candidates were selected by comparing the observed spectral frequencies with the database $(CH_2OH)_2$ line frequencies, and $(CH_2OH)_2$ line candidates were identified as frequency coincidences within ~ 1 MHz. We were able to identify 258 line candidates corresponding to database $(CH_2OH)_2$ line entries.

3.3.2 Competing Assignments

Although 258 lines were associated with $(CH_2OH)_2$, most of the lines are weak or difficult to discern from neighboring stronger lines. Line mixing is particularly problematic since the source contains other complex molecules and methanol products that have lines in the vicinity of $(CH_2OH)_2$ assignments, many of which have similar energies and line intensities. To identify the most promising $(CH_2OH)_2$ candidates, we first isolated the 22 strongest $(CH_2OH)_2$ transitions, then used the database to identify all possible line assignments across all molecular species within ± 6 MHz of the $(CH_2OH)_2$ line frequency. Next, we took these other possible lines in the frequency vicinity of the $(CH_2OH)_2$ line and isolated those that would have comparable spectral profiles with similar energies and intensities that would be the most likely contamination candidates. In particular, related complex molecules (*e.g.* methanol, ethanol, methyl formate, dimethyl ether, ketene, acetaldehyde, formamide) were considered as competing assignments. Out of the 22 strongest lines, only 11 did not have significant competing assignments and appeared to be solely due to $(CH_2OH)_2$. Figure 5 displays the 241-242 GHz window, with the possible $(CH_2OH)_2$ transitions identified to illustrate the density of line candidates. The entire set of spectral data can be found in the Appendix.

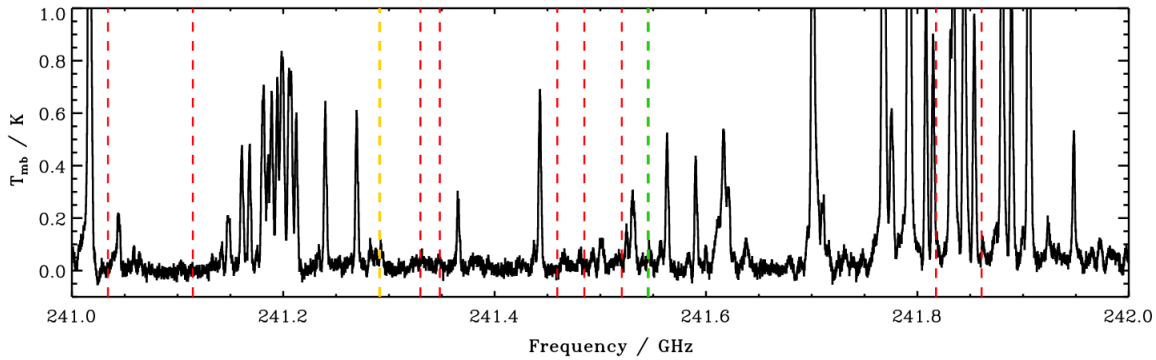


Figure 5: *241-242 GHz spectral data.* The $(CH_2OH)_2$ line candidates are plotted in red in this 1 GHz window. One of the 11 strongest lines without an identified competing assignment is within this frequency window and plotted in yellow. One of the 11 likely contaminated transitions out of the 22 strongest features is plotted in green. In this case, the competing assignment is methyl formate ($C_2H_4O_2$), another CH_3OH product.

4 Simulated $(CH_2OH)_2$ Spectra

Simulating the spectral profiles for the listed $(CH_2OH)_2$ lines within the frequency ranges of interest allow confirmation of the $(CH_2OH)_2$ line candidates and understanding of overall contamination and line blending effects. Moreover, the parameters used to construct the model spectra provide physical insight and parameter estimates. In particular, integrated intensities, $(CH_2OH)_2$ column density ($N_{(CH_2OH)_2}$), and excitation temperature (T_{ex}) were estimated by generating simulating $(CH_2OH)_2$ spectra to fit the 11 cleanest line candidates identified in our preliminary analysis.

4.1 Input Parameters and Modeling

The observational parameters of telescope beam size ($\theta_{beam} \simeq 10.5''$), input emission area ($\theta_{source} \simeq 1''$), and channel width (0.0002 GHz) were inputted in addition to starting estimates for the excitation temperature, $(CH_2OH)_2$ column density, and line widths to generate a model spectral profile for the $(CH_2OH)_2$ database lines within the data range. Starting estimates for the physical parameters ($N_{(CH_2OH)_2}$ and T_{ex}) were obtained from literature (Bisschop et al. 2007) for comparable molecules, then varied when compared to observations to obtain more precise estimates for $(CH_2OH)_2$. The parameters were varied until the difference in intensity between the model and the actual data for the 11 of the 22 strongest lines was minimized. While both an increase in T_{ex} and $N_{(CH_2OH)_2}$ would produce stronger lines, T_{ex} affects the distribution of the lines whereas $N_{(CH_2OH)_2}$ determines the height of the spectral feature. Moreover, higher energy lines are relatively more affected by higher T_{ex} than lower energy lines. Thus, we are able to simultaneously constrain T_{ex} and $N_{(CH2OH)_2}$ in qualitatively optimizing our model spectra. The resulting parameter values are $T_{ex} = 198$ K and $N_{(CH_2OH)_2} = 2.2 \times 10^{15}$ cm^{-2} (using literature values, abundance relative to H_2 of 1.05×10^{-8}). The

main beam temperature of the source was calculated as follows:

$$T_{MB,source} = \frac{N_{(CH_2OH)_2} \nu S_{ij}\mu^2}{\beta Q_{rot} e^{(E_u/T_{ex})}} \tag{3}$$

where ν denotes line frequency, $S_{ij}\mu^2$ a measure of line intensity, β a consolidation of constants $\sim 1.67 \times 10^{14}$, E_u the energy in K, and Q_{rot} the rotation partition function at T_{ex}, calculated by interpolating given Q_{rot} values across temperatures ranging from 9.375 to 300 K. The main beam temperature, T_{MB}, was then calculated by multiplying the source main beam temperature by a correction factor:

$$T_{MB} = \left(\frac{\theta_{source}}{\theta_{betam}}\right) T_{MB,source}. \tag{4}$$

The main beam temperature was then applied to a Gaussian intensity distribution centered at each Doppler corrected line frequency.

4.2 Physical Parameter Estimates from Simulated Spectra

In summary, a model spectra was generated for $(CH_2OH)_2$ lines listed in the database within the frequency ranges of the collected data. The model spectra itself for the same 241-242 GHz window can be found in the first panel of Figure 6, and overplotted in the second panel. The intensity range is restricted to $(CH_2OH)_2$ values in the third panel to see the modeled spectra overplotted more clearly. The complete set of overplotted model spectra can be found in the Appendix. The model spectra was then compared to the actual spectra to validate f $(CH_2OH)_2$ line candidates. Furthermore, the calculated integrated intensities are solely due to $(CH_2OH)_2$, so generation of a model spectra allows visualization of how all the lines are contaminated (beyond the 22 strongest $(CH_2OH)_2$ lines for which we explicitly identified competing assignments) and how the data line widths and intensities are affected by line mixing. While the line candidates appear consistent, we do see significant line blending across all 258 transitions. We also have first estimates of $(CH_2OH)_2$ column density and excitation temperature. In the following sections, we calculate $N_{(CH_2OH)_2}$ and T_{ex} again as a more quantitative method check, but these physical parameter estimates yield a $(CH_2OH)_2$ abundance very comparable to other complex molecules, notably methanol and methanol products (Bisschop et al. 2007).

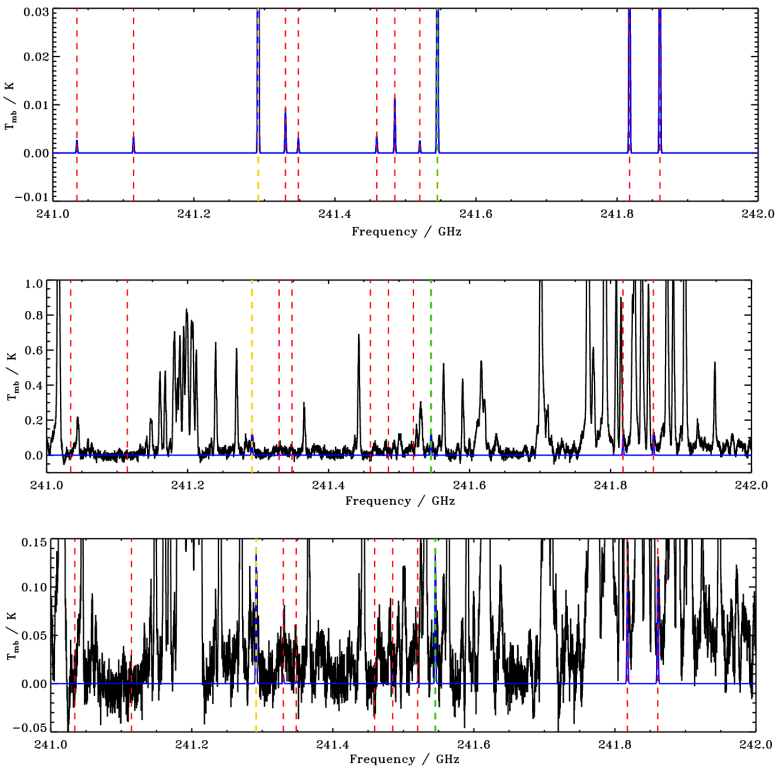


Figure 6: *Model Spectra for* $(CH_2OH)_2$. The simulated spectra is shown in blue, and then overplotted onto the spectral data. The model spectra pvodies a visual way of checking all 258 line candidates for mixing effects and provides physical parameter estimates of column density and excitation temperature.

5 Rotation Diagram Analysis

The rotation diagram method provides an alternative method of deriving physical properties (e.g. temperatures and densities) and molecular abundances from spectral lines. A *rotation diagram* is a plot of molecular column density per energy level statistical weight as a function of energy above the ground state. In this section, we utilize the Gaussian parameters fitted to spectral line profiles computed previously to construct a rotation diagram for $(CH_2OH)_2$ and to derive another estimate of the excitation temperature and molecular column density.

5.1 Derivation of Rotation Diagram

The rotation diagram relates spectral features to physical characteristics, and in this section we describe how these physical parameters can be derived, following the method of Goldsmith and Langer (1998).

5.1.1 Line Emission and Main Beam Temperature

Assuming that a single large bright body is in the main antenna beam, the main beam temperature, T_{MB} produced at frequency ν by a source having brightness distribution $B_\nu(\theta, \phi)$ is given by:

$$T_{MB} = (A_e/2k) \int_{4\pi} B_\nu (\theta, \phi) P_n (\theta, \phi) \, d\Omega \tag{5}$$

where $P_n (\theta, \phi)$ is the normalized power pattern of the antenna, A_e is the effective temperature, k is the Boltzmann constant, and Ω is the solid angle subtended by the antenna. This relation describes the relative response of the antenna to signals from various directions normalized on the direction of maximum response. We can assume that the source is sufficiently distant such that the solid angle subtended by a point-like source is much smaller than the antenna solid angle, $\Delta\Omega_A$, allowing us to set P_n equal to unity over the solid angle of the source. Furthermore, assuming that the source is homogenous, we can take B_ν to be constant. Therefore, the integral simplifies to the source solid angle, $\Delta\Omega_S$ and Equation 5 can be rewritten as:

$$T_{MB} = (A_e/2k) (\Delta\Omega_A) \left(\frac{\Delta\Omega_S}{\Delta\Omega_A} \right) B_\nu. \tag{6}$$

where $\left(\frac{\Delta\Omega_S}{\Delta\Omega_A} \right)$ is the beam dilution factor. For a very small source, the beam dilution factor is very small, but assuming an extended, uniform source, the beam dilution factor approaches unity. From the antenna theorem (Kraus, 1982), assuming that the source radiates isotropically, we can relate the source solid angle and effective area as $A_e \Delta\Omega_A = \lambda^2$, allowing us to rewrite Equation 6 as:

$$T_{MB} = \left(\lambda^2/2k\right) \left(\frac{\Delta\Omega_S}{\Delta\Omega_A} \right) B_\nu. \tag{7}$$

The brightness produced by a source at excitation temperature T_{ex} at optical depth τ is given by:

$$B_\nu = \frac{h\nu/k}{e^{h\nu/kT_{ex}} - 1} \left(1 - e^{-\tau}\right)$$

where ν is the transition frequency. In considering pure rotational transitions, the excitation temperature can also be referred to as the rotational temperature, To simplify later calculations, we rewrite this as:

$$B_\nu = \frac{h\nu/k}{e^{h\nu/kT_{ex}} - 1} \left(\frac{1 - e^{-\tau}}{\tau} \right) \tau. \tag{8}$$

The optical depth of the transition, τ, can be written in terms of column density, N_u, in the upper state u as follows:

$$\tau = \frac{h}{\Delta v} N_u B_{ul} \left(e^{h\nu/kT_{ex}} - 1 \right), \tag{9}$$

where B_{ul} is the Einstein B-coefficient for the transition and Δv is the full width at half-maximum line width (velocity units). Now we can substitute Equations 8 and 9 in Equation 7 to obtain the main beam temperature as a function of column density:

$$T_{MB} = \frac{hc^3 N_u A_{ul}}{8\pi k \nu^2 \Delta v} \left(\frac{\Delta\Omega_S}{\Delta\Omega_A}\right) \left(\frac{1 - e^{-\tau}}{\tau}\right) \tag{10}$$

where A_{ul} is the Einstein A-coefficient for spontaneous emission equivalent to $A_{ul} = 8\pi h\nu^3 B_{ul}/c^3$. Importantly, Equation 10 does not explicitly include the source excitation temperature, T_{ex}, but only observed quantity T_{MB}. Let W denote the integrated line intensity, which we can write as:

$$\begin{aligned} W &= \int T_{MB} dv \\ &= T_{MB} \Delta v. \end{aligned}$$

By inverting Equation 10 and substituting in W, we can obtain an expression for column density of the upper level:

$$N_u = \frac{8\pi k \nu^2 W}{hc^3 A_{ul}} \left(\frac{\Delta\Omega_A}{\Delta\Omega_S}\right) \left(\frac{\tau}{1 - e^{-\tau}}\right). \tag{11}$$

If we assume that *(i.)* the source is homogenous and fills the telescope beam and *(ii.)* the lines are optically thin, Equation 11 simplifies to:

$$N_u = \frac{8\pi k \nu^2 W}{hc^3 A_{ul}}. \tag{12}$$

We evaluate the assumption that the lines are optically thin by examining CH_3OH, which is present in much greater quantity than $(CH_2OH)_2$ with much stronger emission features. Since no isotopologues were detected for CH_3OH, the CH_3OH transitions are most likely not optically thick; furthermore, since $(CH_2OH)_2$ has a much wider partition function, the $(CH_2OH)_2$ are at most only slightly optically thick. Let us define a energy state constant γ_u that consolidates observable quantities and physical constants as follows:

$$\gamma_u = 8\pi k \nu^2 W / hc^3 A_{ul}$$

We can now express N_u in terms of γ_u as:

$$N_u = \gamma_u W. \tag{13}$$

5.1.2 Upper Level and Total Column Densities

Now let us assume that the the system is in *local thermal equilibrium* (LTE), which implies that all level populations can be characterized by a single excitation temperature, T_{ex}. It has been shown that for a single massive source filling the beam, assuming a loss-free antenna, the main beam temperature is equal to the

object's temperature. The total column density, N_{tot}, can be related to the column density of the upper level, N_u by:

$$\frac{N_u}{g_u} = \left(\frac{N_{tot}}{Q_{rot}}\right) e^{-E_u/kT_{ex}} = \frac{\gamma_u W}{g_u} \tag{14}$$

where g_u is the statistical weight of level u, $Q_{rot} = Q_{rot}(T_{ex})$ is the rotational partition function, and E_u is the energy of the upper level. Therefore, in LTE the total column density from the column density of a single transition can be calculated for a known T_{ex}.

5.1.3 Rotation Diagram Under Local Thermodynamic Equillibrium

Assuming LTE allows us to obtain the total column density directly from the upper column density. We can now rearrange Equations 13 and 14, to obtain a linear relationship:

$$\ln\left(\frac{\gamma_u W}{g_u}\right) = \ln(N_{tot}) - \ln(Q_{rot}) - \frac{E_u}{kT_{ex}}. \tag{15}$$

Thus, in a logarithmic plot of the left (observable) side of Equation 15 as a function of E_u/k provides a diagram with an intercept of N_{tot}/Q_{rot} and a slope of $-1/T_{ex}$, yielding a visual method to analyze spectra of complex molecules to derive physical quantities and properties. Moreover, the accuracy of a linear fit to the rotation diagram provides insight on the appropriateness of the assumptions, namely optical thinness and local thermal equilibrium.

5.2 Gaussian Fitting Spectral Profiles

We then applied the method described above to the $(CH_2OH)_2$ data. The spectral lines were fitted with Gaussian profiles in velocity space using `IDL` routine `gaussfit`, allowing the full width half maximum and peak intensity to be calculated from Gaussian fit parameters. Then, the integrated intensities were calculated, yielding values for W as described above. Using observational parameters, we can then construct a rotation diagram for $(CH_2OH)_2$.

5.2.1 Gaussian Parameters

In order to provide an additional model comparison to the spectral data and alternative physical parameter estimates, Gaussian profiles were interactively fit to each of the 258 $(CH_2OH)_2$ line candidates. The Gaussian fits allow us to further examine the observed $(CH_2OH)_2$ lines for contamination by comparing the Gaussian height and width to the line profile. A small window of 8 MHz before and after the observed line frequency was isolated and a 5-term Gaussian with local linear baseline was fit to this 16 MHz region. The method of Gaussian fitting allow us to quantitatively model the line profiles and obtain formal uncertainties for the Gaussian fit parameters (Gaussian height, center, and standard deviation) in addition to the error on

the flux determined by the rms on the lower and upper sidebands. All 258 of the observed $(CH_2OH)_2$ lines were fit with Gaussian profiles and parameter estimates for height at peak intensity, center frequency, and line width were obtained (along with errors). As expected, most of the lines are blended with neighboring lines or overwhelmed by stronger emission features in the vicinity of the center frequency in the 16 MHz region. Moreover, 72 of the 258 lines are part of a multiplet and were incorrectly fitted as as a single broad, strong emission. The resolution of multiplets is discussed in the following section when integrated intensities are calculated from the Gaussian parameters.

5.2.2 Selecting $(CH_2OH)_2$ from Contaminated Lines and Multiplets

Using the parameters derived from fitting the spectral lines with Gaussian profiles, we calculated the integrated intensity, W, and subsequently the upper column density N_u. Interpolating the rotation partition function, Q_{rot}, we then derived the total column density, N, and excitation temperature, T_{ex} using all 258 $(CH_2OH)_2$ lines. However, most of the lines are subject to contamination from competing spectral features, leading to an overestimation of the integrated intensity. Moreover, 72 of the 258 lines are part of multiplets that are fit by single Gaussian profiles. In order to address this, we compared the integrated intensities calculated from the Gaussian parameters to the integrated intensities generated by the simulated spectra in the previous section. We remove potentially blended lines in which the integrated intensity calculated from the Gaussian fit exceeds the simulated spectra integrated intensity by a factor of 3. Additionally, in crowded regions, one Gaussian was incorrectly fit to a multiplet, leading to an overestimate of the line intensity and width. We examine the multiplet cases by averaging the parameter values over all lines included to create a new line entry, then compare the new integrated intensity to the corresponding simulated spectra integrated intensity value and include it if it is within our criteria. Since the simulated spectra are generated solely from $(CH_2OH)_2$ emission, Gaussian intensities exceeding our threshold factor of 3 are likely contaminated with other transitions, so including them in our rotation diagram analysis would lead to incorrect estimates of our physical parameters.

5.3 Excitation Temperature and Column Density for $(CH_2OH)_2$

After filtering the 258 lines, we construct a new rotation diagram with the best $(CH_2OH)_2$ line candidates, as shown in Figure 7.

We obtain an excitation temperature of $T_{ex} = 190.71$ K and a total column density of $N_{(CH_2OH)_2} = 1.09 \times 10^{15}$ cm^{-2}. Sources of error in our estimated excitation temperature and column density values could include the effects of a finite optical depth, lingering contamination effects, or non-LTE excitation. Further details on

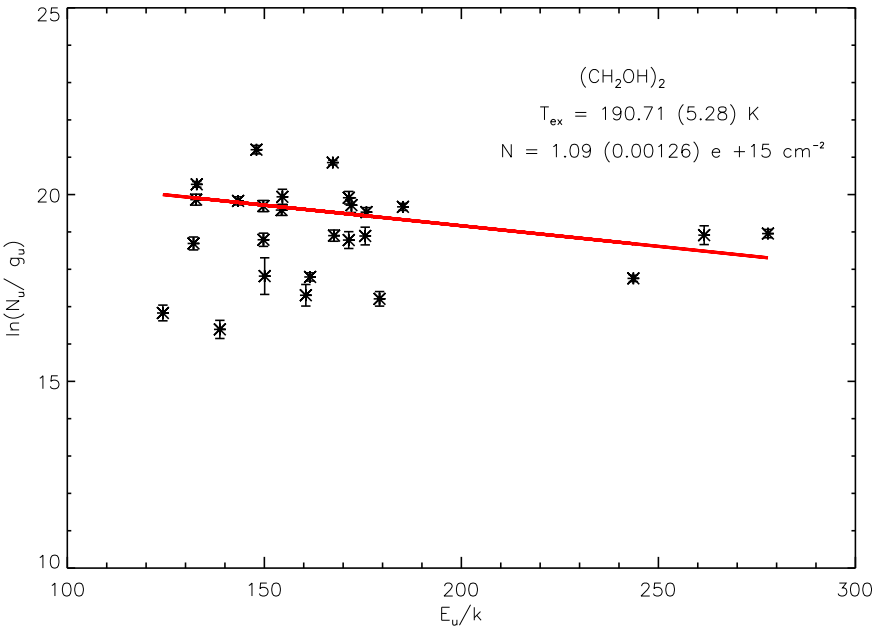


Figure 7: *$(CH_2OH)_2$ rotation diagram.* Only lines with integrated intensities sufficiently close to the integrated intensities from the simulated spectra are included to minimize the contamination from blended lines. Moreover, multiplets are extracted and an averaged line entry is created since multiplets are likely to be incorrectly fitted due to their proximity in frequency.

the assumptions of the rotation diagram method and correction effects can be found in Appendix 2. Lingering contamination effects from inextricably blended lines could exist, but the overall intensity is still within a reasonable factor of the non-contaminated simulated spectra, so this should not cause a substantial error. Nonetheless, the rotation diagram method does yield the average excitation temperature from which most molecular emission arises, so it still is a meaningful parameter. If the spectral lines are sub-thermally excited because the density is below the critical density, the derived excitation temperature given above is a lower limit of the kinetic temperature. Small-scale non-LTE conditions are could be contributing to the error in the lines with larger uncertainties. However, since we are analyzing hot core emission in which collisions dominate radiative excitation in the dense, hot central protostar, we can assume that energy is transferred primarily through kinetic interactions and LTE is a reasonable and common assumption.Meanwhile, since we can fit both weak and strong transitions reasonably with a single line, our rotation diagram is consistent with general optical thinness. If some lines are marginally optically thick, particularly those with larger uncertainties and scatter, the column density is underestimated and the excitation temperature is overestimated. However, by checking the integrated intensity with the simulated integrated intensity, the most optically thick lines are eliminated and excluded from the fit. With these limitations, the rotation diagram yields an additional useful column density and excitation temperature estimate that for $(CH_2OH)_2$ that more quantitatively and comprehensively fits all $(CH_2OH)_2$ line candidates.

6 Comparison of Spectral Models

The simulated spectra and the rotation diagram allow us to obtain two estimates each for the column density, $N_{(CH_2OH)_2}$, and excitation temperature, T_{ex}, as listed in Table 6:

Model	$N_{(CH_2OH)_2}$	T_{ex}
Simulated spectra	2.20×10^{15} cm^{-2}	198 K
Rotation diagram	1.09×10^{15} cm^{-2}	190 K

Table 6: *Summary of parameter values.* The column density and excitation temperature estimates of both methods (simulating spectra and constructing a rotation diagram) are given.

6.1 Simulating Spectra from Rotation Diagram Parameters

To understand how the difference in parameter estimate values affects the spectral profile of $(CH_2OH)_2$, we simulated an additional model spectra using the estimates of column density and excitation temperature given by the rotation diagram. In generating this simulated spectra, instead of inputting starting parameters from literature and varying them to obtain a reasonable fit to the data, we directly input $N_{(CH_2OH)_2} = 1.09 \times 10^{15}$ cm^{-2} and $T_{ex} = 190$ K to simulate a new a spectra in the same manner as done previously. These two models are then overplotted onto the data, and we see that the model spectra from the rotation diagram is characterized by lower intensities than the previously simulated spectra. The complete set of overplotted spectra for both the initial simulated spectra and rotation diagram model can be found in the Appendix.Figure 8 show the original simulated spectra and the rotation diagram generated simulated spectra overplotted onto the 241-242 GHz spectra window, with the second panel restricted to a smaller intensity range to see the differences in models clearly. This trend is most likely due to the fact that the previously simulated spectra was generated by fitting 11 of the 22 strongest lines, but the lines used to construct the rotation diagram are less likely to be blended with features from other species or in multiplets since they were subject to stricter selection criteria. Moreover, we see that the model generated from the rotation diagram parameters seem to fit the spectral data more closely, so we can conclude that the values of $N_{(CH_2OH)_2} = 1.09 \times 10^{15}$ cm^{-2} and $T_{ex} = 190$ K are more accurate.

6.2 Examining Effects of Parameter Estimate Differences

To obtain an alternative visual comparison of the effects of the physical parameter differences, the initial simulated spectra, the fitted Gaussian, and the spectra simulated from rotation diagram parameters are overplotted onto the profiles of the $(CH_2OH)_2$ lines used to generate the rotation diagram in Figure 9.

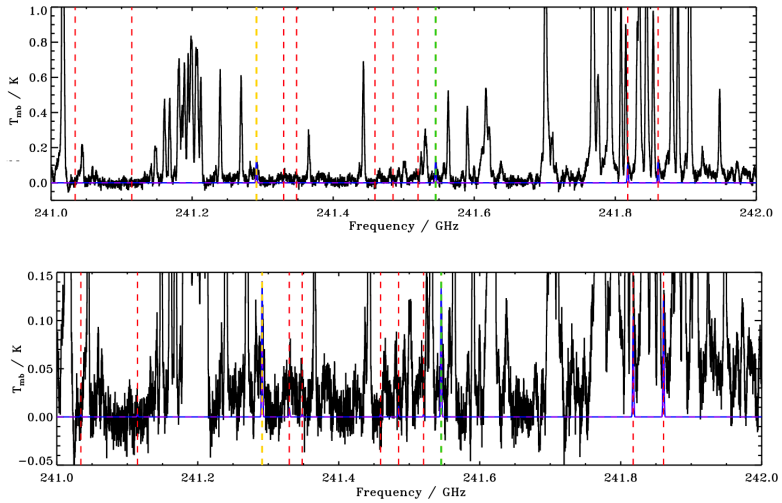


Figure 8: *Comparing spectral models.* The initial simulated spectra is again overplotted in blue, and the model spectra generated from the rotation diagram parameters is overplotted in the purple dashed line. The difference between the predicted column densities is possibly due to lingering contamination causing an overestimate in the original simulated spectra.

7 Analysis and Discussion

7.1 Comparison of $(CH_2OH)_2$ Abundances

The utility of the analysis strategy employed in this study and presented above is in the capability to test model predictions and benchmark currently existing theoretical models and acquire an understanding of the relative importance of formation and destruction stages of hot core complex molecules. As mentioned previously, NGC 7538 IRS 1 is a known massive YSO exemplifying strong hot core emission rich in complex organic molecules. As discussed in Section 2, $(CH_2OH)_2$ has been observed in astrophysical sources (*e.g.* Comet Hale-Bopp, CMZ), simulated in two-phase and three-phase kinetic models, and detected in laboratory experiments. However, observations of NGC 7538 IRS 1, unlike those of more extreme sources Comet Hale-Bopp and CMZ clouds, allow us to directly test existing protostellar models of warm-up phase photochemistry on an archetypal massive YSO. Table 7 lists the detected abundances of $(CH_2OH)_2$ and the abundances calculated from the simulated spectra and the rotation diagram.

We normalize abundances to CH_3OH since it is considered the critical zeroth-generation precursor to more complex molecules, including $(CH_2OH)_2$, and since it is readily detectable due to strong emission features. First, we note that that the abundances of $(CH_2OH)_2$ are most comparable to those found in the CMZ molecular clouds studied by Requena-Torres et al. (2008). We previously proposed that since the abundances of other complex molecules were comparable between the CMZ and hot cores, the abundances of $(CH_2OH)_2$ should be similar between the two types of sources, which seems to be supported here. Second,

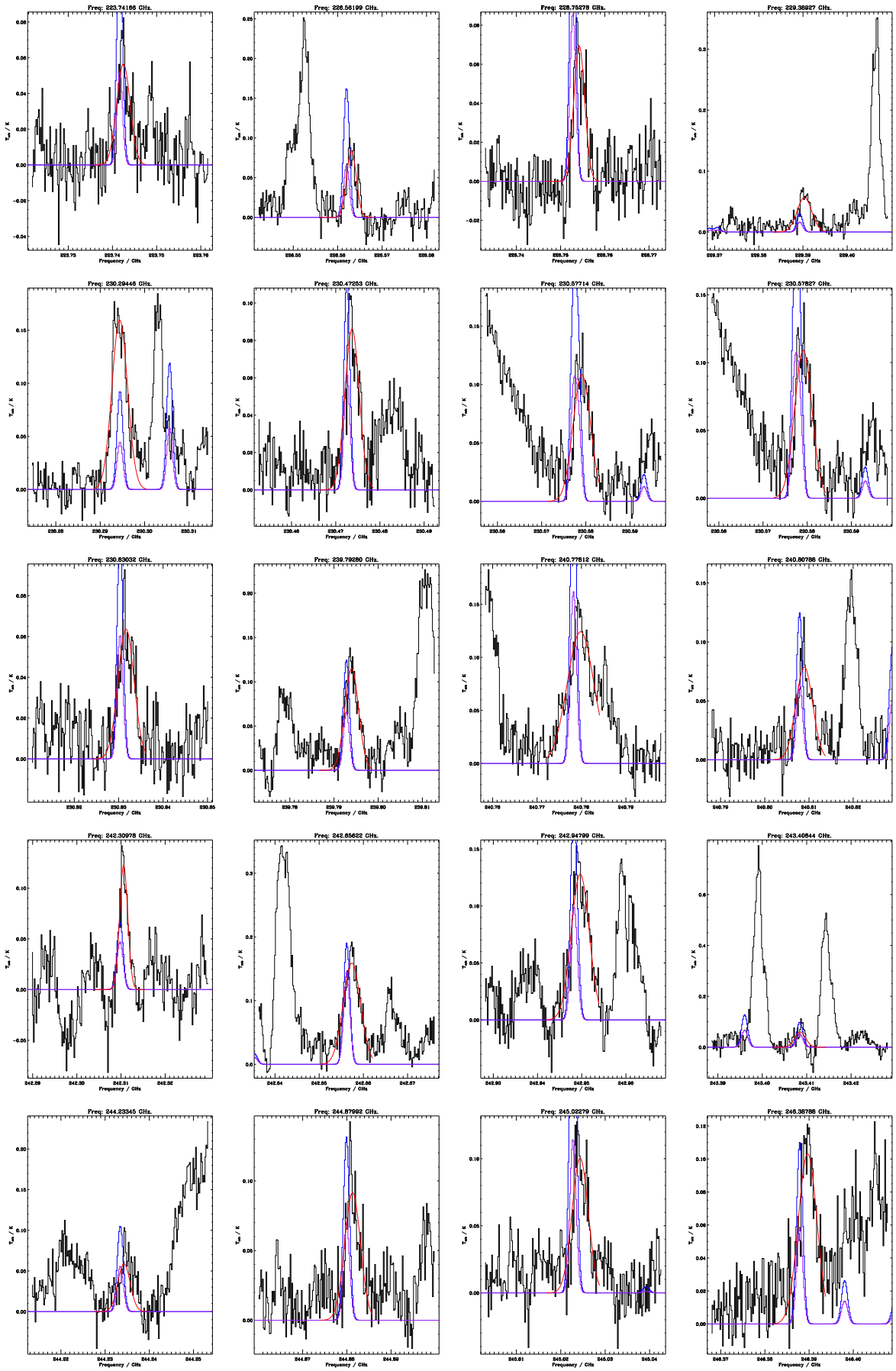


Figure 9: *Rotation Diagram Lines.* Gaussian fits (red), original model spectra (blue) and rotation diagram model (purple).

	$N((CH_2OH)_2)/N(CH_3OH)$
Comet Hale-Bopp[1]	0.104
MC G-0.02[2]	0.012
MC G-0.11[2]	0.02
MC G+0.693[2]	0.024
Experimental[3]	0.4
Warm-Up Model (Fast, T_{ex} = 206 K)[4]	3.4E-4
Warm-Up Model (Medium, T_{ex} = 198 K)[4]	1.2E-4
Warm-Up Model (Slow, T_{ex} = 158 K)[4]	1.1E-6
[a]$(CH_2OH)_2$	0.02
[b]$(CH_2OH)_2$	0.01

[1]Crovisier et al. (2004).
[2]Requena-Torres et al. (2008).
[3]Öberg et al. (2009).
[4]Garrod (2013).
[5][a] Parameter values from initial simulated spectra. [b] Parameter values from rotation diagram.

Table 7: *Abundances of $(CH_2OH)_2$ relative to CH_3OH.* We see that the abundances derived from the simulated spectra and rotation diagram are comparable to detected and experimental values. The abundances exceed those derived in the current protostellar warm-up models, indicating greater importance of formation mechanisms or special ice conditions leading to the overproduction of $(CH_2OH)_2$ over other molecules.

while the abundances found in this study of $(CH_2OH)_2$ are significantly greater than those found in Sgr B2(N) (Hollis et al. 2002), they are less that those detected in Comet Hale-Bopp. This could be because the source is not in a peak formation stage, or gas phase reactions could be forming and destroying the $(CH_2OH)_2$ at different rates. Additionally, since laboratory experiments generate very high relative abundances of $(CH_2OH)_2$ under the optimal conditions of pure CH_3OH ice and photodissociation activity, this could imply there is CO ice or less radiative activity in NGC 7538 IRS 1 relative to Comet Hale-Bopp, decreasing the formation of $(CH_2OH)_2$. Thus, laboratory experiments suggest that the increase in $(CH_2OH)_2$ abundance between massive YSO NGC 7538 IRS 1 and Comet Hale-Bopp could be due to the presence of more mixed CO:CH_3OH ice and less photodissociation activity in NGC 7538 IRS 1.

Most importantly, the existing protostellar warm-up phase models severely underestimate the abundance of (CH_2OH_2) in NGC 7538 IRS 1. In particular, the medium stage warm-up model has an excitation temperature equal to that of the simulated spectra but an abundance 2 orders of magnitude lower. While the high quantities of $(CH_2OH)_2$ in Comet Hale-Bopp and the CMZ clouds could possibly be attributed to unique chemistry from their more volatile and extreme environments, the results from NGC 7538 IRS 1 are directly relevant to existing protostellar models as NGC 7538 IRS 1 is representative of hot cores. The significant discrepancy between the abundance of $(CH_2OH)_2$, a one of the largest hydrogen rich astronomical complex molecules, in NGC 7538 IRS 1 and current protostellar model predictions imply that the formation pathways of very complex molecule formation may be underestimated. Specifically, since $(CH_2OH)_2$ has been shown to be produced at high quantities from CH_3OH ice via warm ice photochemistry, these results suggest that the warm-up phase is of greater relative importance in hot core stage of massive YSO formation than currently assumed in existing protostellar models.

7.2 Complex Molecules in NGC 7538 IRS 1

NGC 7538 IRS 1 is known to possess a rich and diverse complex molecular inventory as studied through its hot core emissions. In addition to comparing the model predictions to previous astrophysical observations and models, we can also compare the calculated abundances and excitation temperatures of $(CH_2OH)_2$ with those of other complex molecules detected in NGC 7538 IRS 1, as summarized in Table 8.

Species	N[cm^{-2}]	T_{ex} [K]	$N(X)/N(H_2)^2$	$N(X)/N(CH_3OH)$
H_2CO	2.5E+16	87	1.2E-7	0.21
CH_3OH	1.2E+17	156 ± 10	5.7E-7	1
C_2H_5OH	5.7E+17	164 ± 17	2.5E-8	0.04
$HNCO$	2.3E+15	278 ± 88	1.1E-8	0.02
NH_2CHO	5.7E+14	164 ± 28	2.6E-9	0.005
CH_3CN	8.2E+15	218	3.9E-8	0.07
C_2H_5CN	9.2E+14	96	4.4E-9	0.008
$HCOOCH_3$	1.4E+16	134 ± 8	6.7E-8	0.12
CH_3OCH_3	1.6E+16	130	7.8E-8	0.13
$^a(CH_2OH)_2$	2.2E+15	198	1.1E-8	0.02
$^b(CH_2OH)_2$	1.1 E+15	190 ± 7	5.5E-9	0.01

Table 8: *Hot core molecules in NGC 7538 IRS 1.* We see that $(CH_2OH)_2$ is of comparable column density to other complex molecules previously detected in NGC 7538 IRS 1, including less complex molecules.
[1] ^{a} Parameter values from initial simulated spectra. ^{b} Parameter values from rotation diagram.
[2] Molecular abundances calculated with literature value of N(H_2) = 2.1×10^{23} (Bisschop et al. 2007).

We see that $(CH_2OH)_2$ is present in very relevant quantities to other complex molecules in the source, notably with abundances comparable to ethanol. The derived abundance $(CH_2OH)_2$ is comparable to that of less complex organic molecules in the source, indicating that formation pathways could be more favorable towards complexity than previously anticipated. The excitation temperature of $(CH_2OH)_2$ is close to that of CH_3OH, suggesting that $(CH_2OH)_2$ could be the product of CH_3OH ice reactions, as demonstrated previously in laboratory experiments. In general, the excitation temperature is consistent with that of other complex molecules in the source, suggesting that $(CH_2OH)_2$ should be produced in the same protostellar environment. In summary, the comparability of the excitation temperature and abundance of $(CH_2OH)_2$ to other organic species, even those of lesser complexity, also implies that warm photodissociation ice chemistry may have been previously underestimated in relevance to cold phase and hot phase chemistry in this high mass YSO.

8 Conclusion

Ethylene glycol, $(CH_2OH)_2$, a very large, hydrogen rich complex organic molecule, has previously been assumed to be of relatively low abundance in star forming regions. Since $(CH_2OH)_2$ is of high complexity, its abundance constraints formation pathways of very complex molecules. Moreover, complex organic molecules such as $(CH_2OH)_2$ have the potential to probe the hot core stage of formation of high mass YSOs, which is less understood relative to low mass YSOs. The utility of complex organic molecule probes depends on the understanding of the chemistry of these molecules in star-forming regions.

However, recent observational and experimental studies, notably of CMZ clouds and Comet Hale-Bopp, suggest that $(CH_2OH)_2$ is present in higher quantities. Additionally, laboratory experiments have linked high production of $(CH_2OH)_2$ to CH_3OH ice photochemistry in the warm-up phase of hot core chemistry, suggesting that warm ice photochemistry and formation pathways for very complex molecules are more important that previously considered. Nonetheless, since CMZ clouds and Comet Hale-Bopp are relatively volatile, extreme environments with complicated chemistry, the results are not directly comparable to predictions from existing protostellar warm-up phase models. As such, we analyze hot core emission from archetypal high mass YSO NGC 7538 IRS 1. After identifying 258 line candidates in a 16 GHz spectral window, generated two model spectra: one fitting the most promising strong transitions and one generated by parameters fit by the rotation diagram method. To construct the rotation diagram, we fit Gaussians to all of the $(CH_2OH)_2$ line profiles and used the Gaussian parameters to obtain integrated intensities and line widths, which we could then use to calculate excitation temperatures and column densities. As a result, the simulated spectra yielded a column density estimate of $N = 2.2 \times 10^{15}$ cm^{-2} and an excitation temperature estimate of $T_{ex} = 198$ K, and the rotation diagram generated $N = 1.09 \times 10^{15}$ cm^{-2} and $T_{ex} = 190$ K, or $1 - 2\%$ relative to CH_3OH. We find that the abundances predicted by current protostellar models strongly underestimate the quantities determined in this study of a characteristic hot core. Thus, we have found the first indication that $(CH_2OH)_2$ is much more prevalent around YSOs than previously assumed, in turn suggesting that warm ice photochemistry and the formation of very complex molecules is of greater relative importance than is considered in existing models of star formation chemistry.

9 References

Bisschop, S.E., Jørgensen, J.K., van Dischoek, E.F., de Wachter E.B.M.2007. *Astron. Astrophys.* 465: 913-927.

Black, G.A., Sutton, E.C., MAsson, C.R., and Phillips, T.G. 1987, *Ap.J.*, 315, 621.

Carr, J.S., Nagita, J.R. 2008. *Science* 319:1504.

Crovisier, J., Bockelee-Morvan, D., Biver, N., Colom, P., Despois, D., Lis, D.C. *Astron. Astrphys.* 418:L35-L38.

Garrod R.T., Weaver S. 2008. *Ap. J.* 682: 283-302.

Garrod, R.T., Pauly, T. 2011. *Ap. J.* 735:15.

Garrod R.T. 2013. *Ap. J.* 765: 60.

Goldsmith, P.F., Langer, W.D. 1998. *Ap. J.* 517: 209-225.

Gibb E.L., Whittet, D.C.B., Boogert A.C.A. Tielens, A.G.G.M. 2004. *Ap. J. Suppl.* 151:35.

Helmich, F.P., van Dischoek, E.F. 1997. *Astron. Astrophys. Suppl.* 124:205.

Herbst, E., Klemperer W. 1973. *Ap. J* 185:505.

Herbst, E., van Dischoek E.F. 2009. *Annu. Rev. Astron. Astrophys.* 47:427-480.

Hollis, J.M., Lovas, F.J., Jewell, P.R., Coudert, L.H. 2002. *Ap. J.* 571:L59.

Ikeda, M., Onishi, M., Nummelin, A., Dickens, J.E., Bergman, P., et al. 2001. *Ap. J.* 611:605.

Öberg, K.I., Garrod, R.T., van Dischoek E.F., Linnartz, H. 2009. *Astron. Astrophys.* 504:891-913.

Öberg, K.I., et al. 2013. *arXiv: 1305.315v1* preprint.

Remijan, A.J., Snyder, L.E., Friedel, D.N., Liu, S.Y., Shah, R.Y. 2003. *Ap. J.* 590:314.

Remijan, A.J., Wyrowski, F., Friedel D.N., Meier D.S., Snyder L.E. 2005b. *Ap. J.* 626:233.

Requena-Torres, M.A., Marcelino, N., Jiminez-Serra, I., Martin-Pintado, J., Martin, S., Mauersberger, R. 2007. *Ap. J. Lett.* 655:L37.

Moscadelli, L., Reid, M.J., Menten, K.M., et al. 2009, *Ap.J.* , 693, 406.

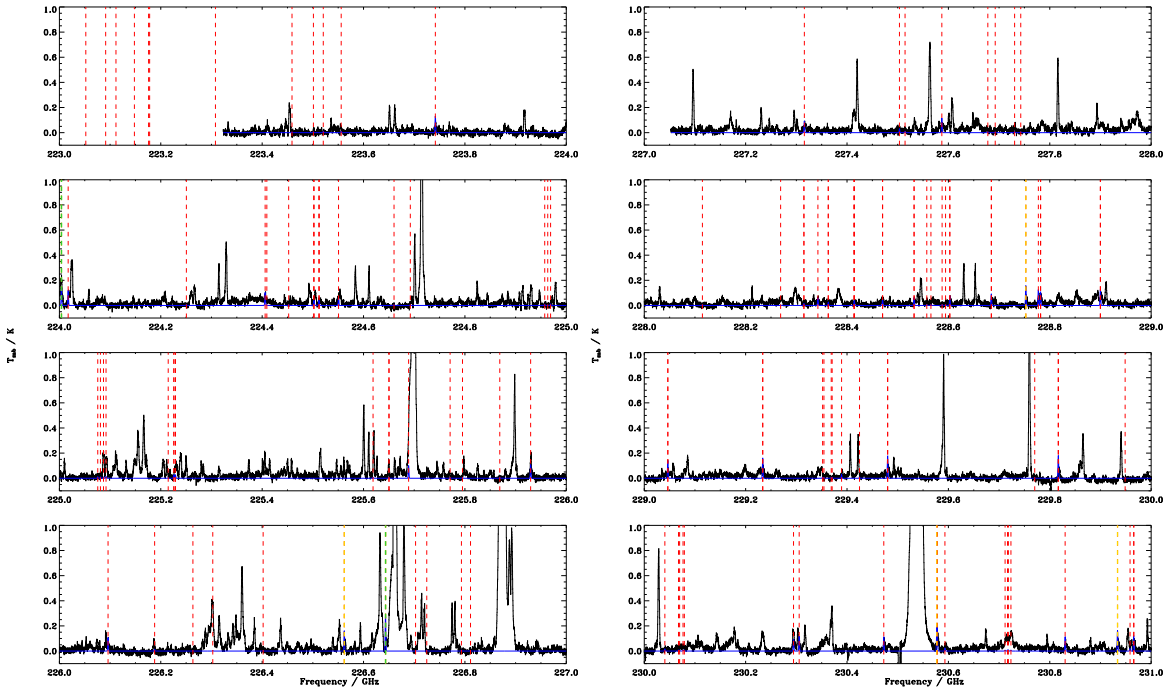
Requena-Torres, M.A., Martin-Pintado, J., Martin, S., Morris, M.R. 2008. *Ap. J.* 672:352-360.

Smith, I.W.M., Herbst, E., Chang, Q. 2004. *MNRAS* 350:323.

Turner, BE. 1998. *Ap. J.* 501:731.

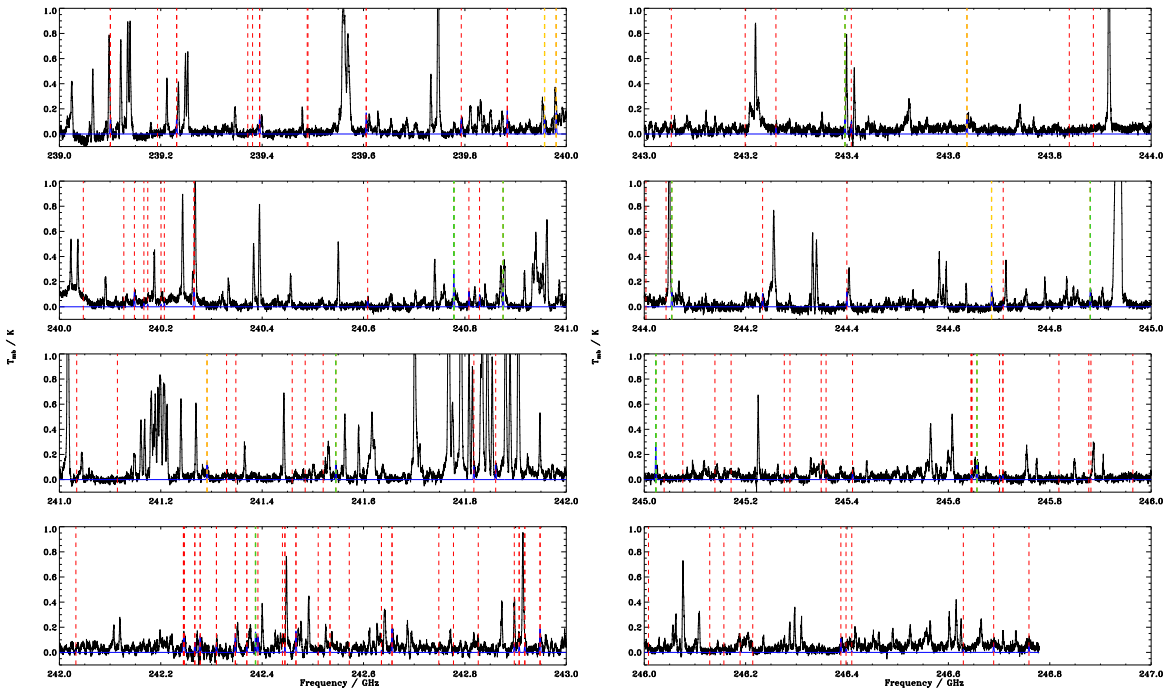
van der Tak, F.F.S. 2005. *Proceedings IAU Symp.* 227.

Appendix I: Complete Spectral Data and Models



(a) Lower Outer Sideband (223.32 - 227.37 GHz)

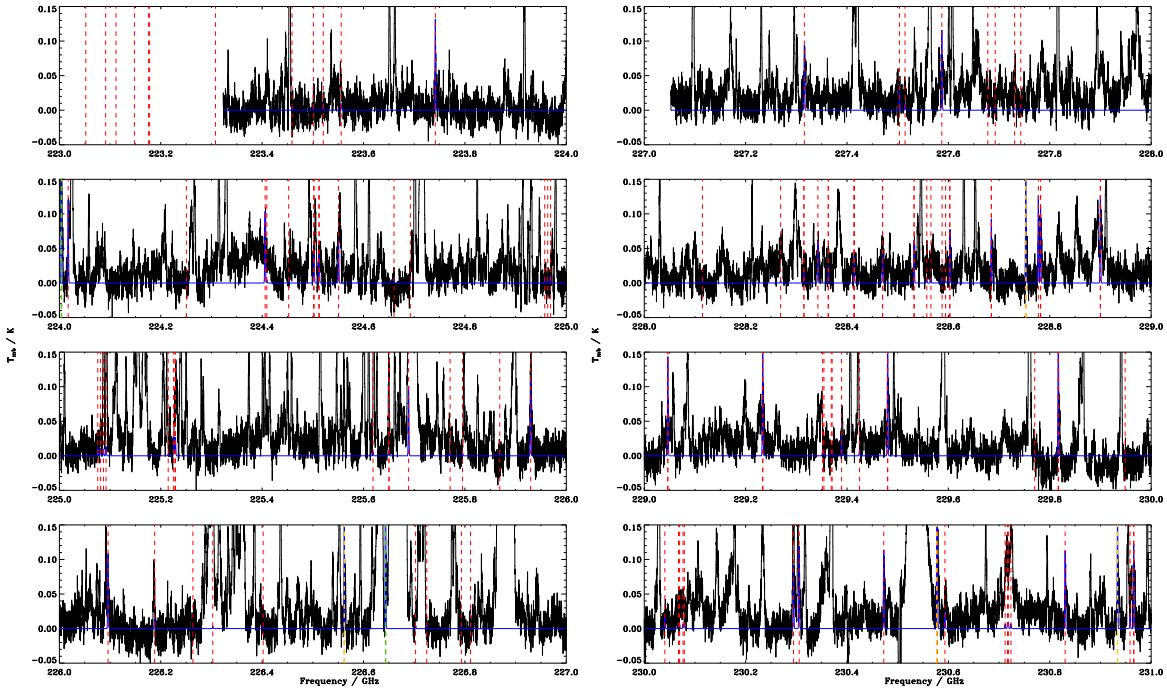
(b) Lower Inner Sideband(227.05 - 231.10 GHz)



(c) Upper Inner Sideband (239.00 - 243.05 GHz)

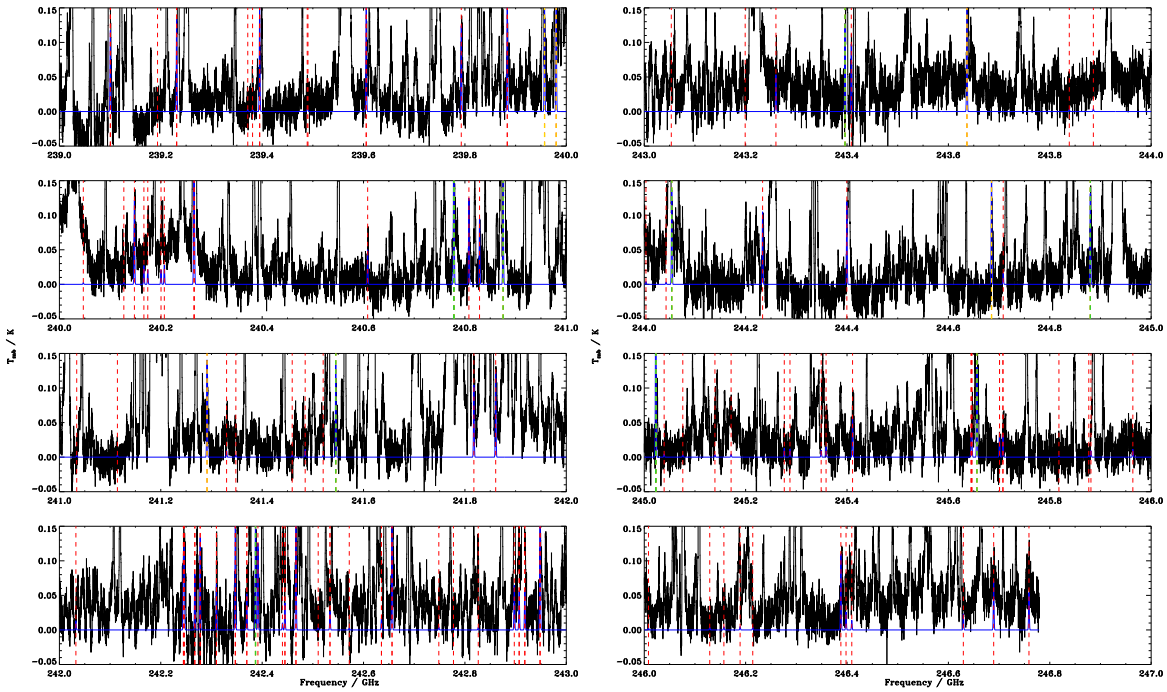
(d) Upper Outer Sideband (242.73 - 246.78 GHz)

Figure 10: *Spectra from NGC 7538 IRS 1.* The 258 $(CH_2OH)_2$ line candidates are plotted in red across all 16 GHz of data. The 11 strongest lines without competing assignments are plotted in yellow, and the 11 likely contaminated lines are plotted in green. Moreover the simulated spectra is overplotted in blue. The simulated spectra provides a visual way of checking all 258 lines for mixing effects and confirms the assignment of $(CH_2OH)_2$ lines.



(a) Lower Outer Sideband (223.32 - 227.37 GHz)

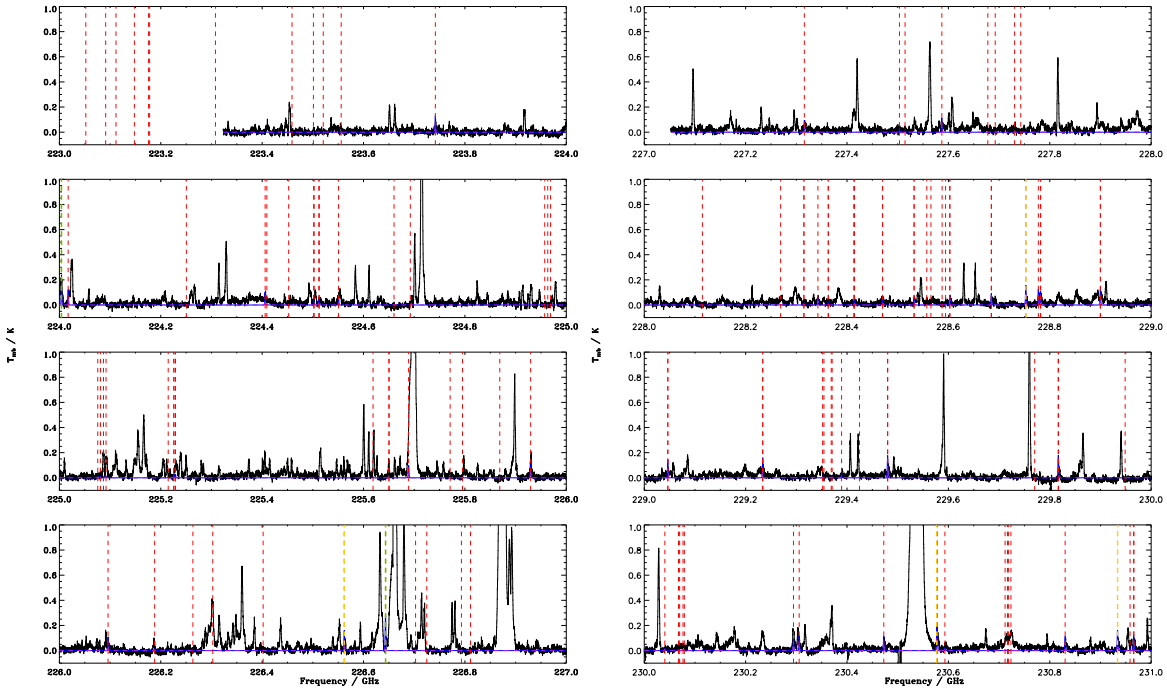
(b) Lower Inner Sideband(227.05 - 231.10 GHz)



(c) Upper Inner Sideband (239.00 - 243.05 GHz)

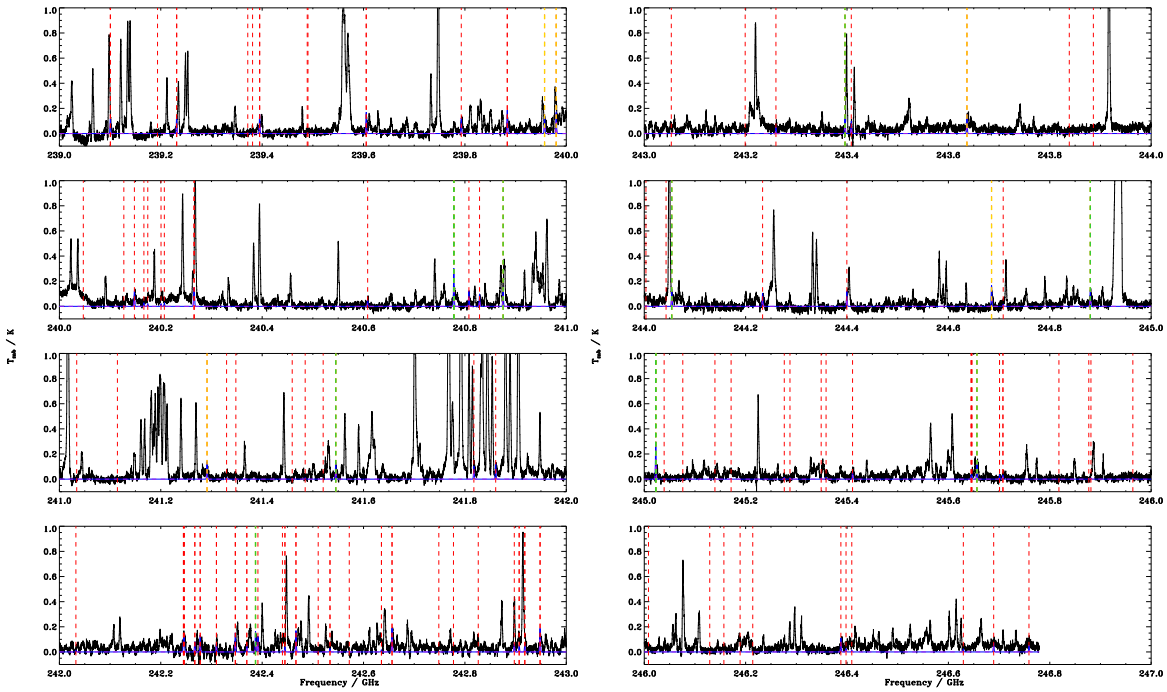
(d) Upper Outer Sideband (242.73 - 246.78 GHz)

Figure 11: *Spectra from NGC 7538 IRS 1.* The intensity y-range of the spectra is restricted to $(CH_2OH)_2$ lines to see the difference between the simulated and actual spectra more clearly. We see that contamination is a significant problem for many of lines in which the simulated spectral line intensity is less than the actual line intensity and the simulated spectral line width is narrower than the actual line width.



(a) Lower Outer Sideband (223.32 - 227.37 GHz)

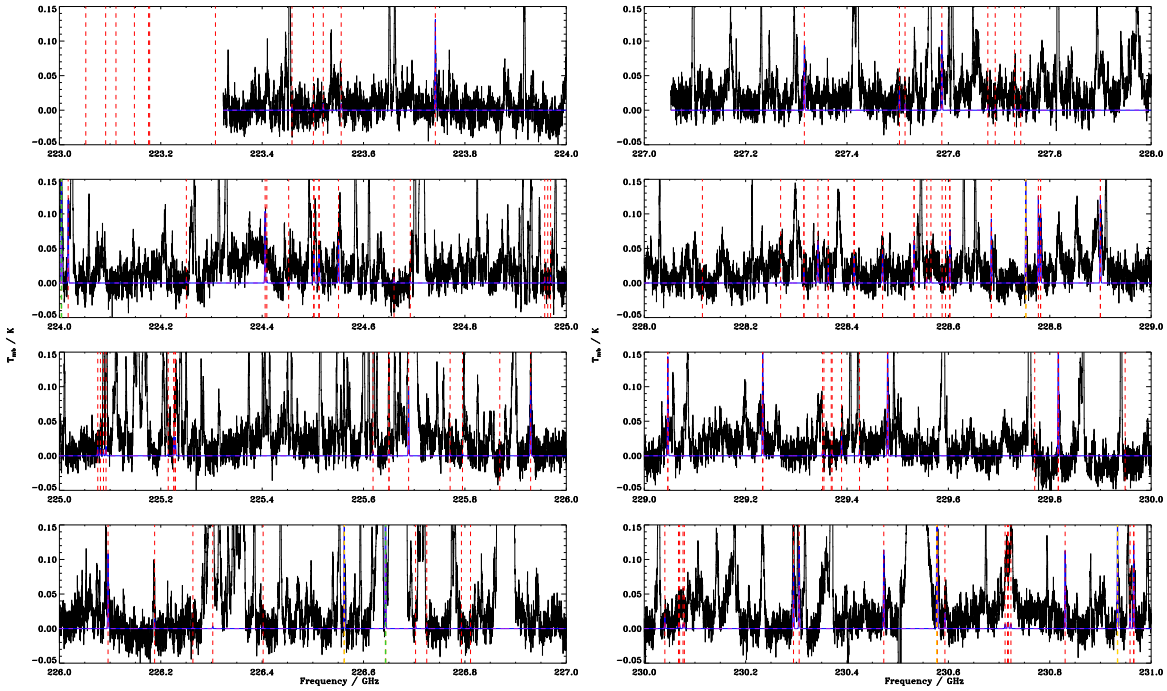
(b) Lower Inner Sideband(227.05 - 231.10 GHz)



(c) Upper Inner Sideband (239.00 - 243.05 GHz)

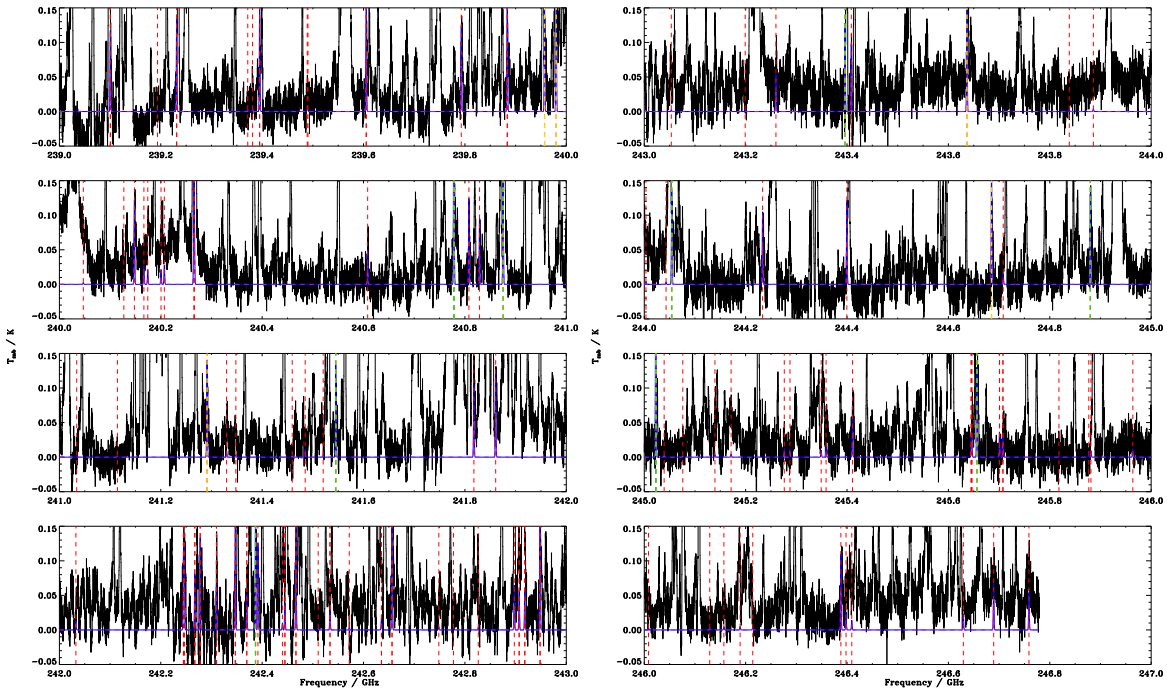
(d) Upper Outer Sideband (242.73 - 246.78 GHz)

Figure 12: *Comparing Model Spectra.* The original simulated spectra is plotted in blue, and the model spectra generated by the rotation diagram physical parameters is plotted by the purple dashed line. Since the predicted column density of the rotation diagram is around 50% of the simulated spectra, the overall intensities predicted by the rotation diagram model are weaker. However, the models seem to reasonably coincide and have similar line widths. The column density difference is possibly due to lingering contamination causing an overestimate in the original simulated spectra.



(a) Lower Outer Sideband (223.32 - 227.37 GHz)

(b) Lower Inner Sideband(227.05 - 231.10 GHz)



(c) Upper Inner Sideband (239.00 - 243.05 GHz)

(d) Upper Outer Sideband (242.73 - 246.78 GHz)

Figure 13: *Comparing Model Spectra.* We again restrict the y-range of the spectra to see the difference between the models and actual spectra more clearly.

Appendix II: Assumptions of the Rotation Diagram Method

While the rotation diagram method described above is frequently used, the conclusions hinge on several assumptions may not always hod in actual systems. In this section, we review the critical assumptions and uncertainties involved with rotation diagrams.

Finite Optical Depth

The lines of many complex organic molecules, notably CH_3OH and CH_3CN, are sometimes optically thick, especially if they can be sub-thermally excited. Under the assumption of optically thin lines, this will lead to an overestimate of T_{ex} and an underestimate of N_{tot}. To address a possible optical thick emission, we could introduce the optical depth correction factor C_τ as:

$$C_\tau = \frac{\tau}{1 - e^{-\tau}}.$$

Referring back to Equation 11, it follows that for optically thick lines:

$$N_u = \gamma_u W C_\tau. \tag{16}$$

This results in an added factor in Equation 15:

$$\ln\left(\frac{\gamma_u W}{g_u}\right) = \ln(N_{tot}) - \ln C_\tau - \ln(Q_{rot}) - \frac{E_u}{kT_{ex}}. \tag{17}$$

The optical depth correction factor is greater than or equal to unity and describes how the column density of the upper state from observed integrated line intensity would be artificially small due to a finite optical depth.

Non-Local Thermodynamic Equilibrium Excitation

In non-LTE excitation, the excitation temperature may differ the populations of different energy levels. Moreover, when the hydrogen density is insufficient to thermalize some or all of the transitions, LTE would not accurately describe energy level populations. We can use the Boltzmann distribution to define the excitation temperature, T_{ex} between two levels i and j, by the relative column densities N_i and N_j and relative energies E_i and E_j:

$$\frac{N_j}{N_i} = \frac{g_j}{g_i} \exp\left[-(E_j - E_i)/kT_{ex}\right],$$

where g_j and g_i are statistical weights. Non-LTE excitation significantly complicates analysis, particularly through dependence of the excitation temperature on energy state.

Moreover, in complete thermalization, the excitation temperature should equal the kinetic temperature, but it has been found that higher frequency transitions can be sub thermally excited. For a two-level system, the critical density for thermalization, n_{cr} is given by $n_{cr} = A_{ul}/q_{ul}$ where q_{ul} is the rate coefficient for collisional de-excitation. Since the Einstein A-coefficient is proportional to $\mu^2\nu^3$, where μ is the permanent dipole moment, and q_{ul} does not vary significantly over energy levels, the critical density increases strongly with greater transition frequency. For instance, Goldsmith and Langer (1998) examined this for HC_3N: the critical density of HC_3N at the $J = 10 - 9$ transition at 91 GHz is 10^5 cm^{-3} ($E_u = 24$ K) whereas the critical density of the $J = 27 - 26$ transition at 245 GHz is 2×10^6 cm^{-3} ($E_u = 165$ K). In turn, since the rotational energy level separation is proportional to rotational quantum number J, higher frequency, higher J transitions require higher critical densities and thus temperatures to be thermalized. This was evidenced by Johnstone et al. (2003) for CH_3OH, which had an excitation temperature ranging from $25 - 80$ K in an environment with a kinetic temperature of 100 K. As a result, higher transition frequency transitions exhibit subthermal excitation, and observed low excitation temperatures found in cold regions (eg. outflows, cold cores) may not necessarily imply low kinetic temperature, but instead low densities.

Beam Adjustments and Heterogeneity in the Source

In many cases, the assumption that the source fills the beam does not hold, particularly for hot core observations. Additionally, many molecular clouds are known to be very "clumpy" on all scales with significant chemical and physical inhomogeneities. When analyzing a heterogenous source, the beam temperatures are reduced by the beam dilution factor, leading to detection of weak emission with the beam temperature much less than the kinetic temperature, leading us to believe that the emission is optically thin. As hot cores are typically only $1''$ on the sky and single dish beams are on the order of $15'' - 30''$, the beam dilution factor can be on the order of 100 and significantly affect optical depth estimates. Assuming LTE and a derivable optical correction factor or optical thinness, we can accommodate this error by introducing the the beam dilution factor ($\Delta\Omega_S/\Delta\Omega_A$) must be introduced in to Equation 17 to include a constant offset due to beam dilution:

$$\ln\left(\frac{\gamma_u W}{g_u}\right) = \ln(N_{tot}) - \ln\left(\frac{\Delta\Omega_S}{\Delta\Omega_A}\right) - \ln C_\tau - \ln(Q_{rot}) - \frac{E_u}{kT_{ex}}. \quad (18)$$