Probing the Conditions for the H I-to-H₂ Transition in the Interstellar Medium

Gyueun Park[1,2] , Min-Young Lee[1,2] , Shmuel Bialy[3] , Blakesley Burkhart[4,5] , J. R. Dawson[6,7] , Carl Heiles[8] ,
Di Li[9,10,11] , Claire Murray[12,13] , Hiep Nguyen[14] , Anita Hafner[7] , Daniel R. Rybarczyk[15] , and Snežana Stanimirović[15]

[1] Korea Astronomy and Space Science Institute, 776 Daedeok-daero, Daejeon 34055, Republic of Korea; mlee@kasi.re.kr
[2] Department of Astronomy and Space Science, University of Science and Technology, 217 Gajeong-ro, Daejeon 34113, Republic of Korea
[3] Department of Physics, Technion—Israel Institute of Technology, Haifa, 3200003, Israel
[4] Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA
[5] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[6] School of Mathematical and Physical Sciences and Macquarie University Astrophysics and Space Technologies Research Centre, Macquarie University, NSW 2109, Australia
[7] Australia Telescope National Facility, CSIRO Space & Astronomy, PO Box 76, Epping, NSW 1710, Australia
[8] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[9] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[10] NAOC-UKZN Computational Astrophysics Centre, University of KwaZulu-Natal, Durban 4000, South Africa
[11] Research Center for Intelligent Computing, Zhejiang Laboratory, Hangzhou 311100, People's Republic of China
[12] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[13] Department of Physics & Astronomy, Johns Hopkins University, MD 21218, USA
[14] Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia
[15] Department of Astronomy, University of Wisconsin, Madison, WI 53706-15821, USA

Abstract

We investigate the conditions for the H I-to-H₂ transition in the solar neighborhood by analyzing H I emission and absorption measurements toward 58 Galactic lines of sight (LOSs) along with ^{12}CO(1–0) (CO) and dust data. Based on the accurate column densities of the cold and warm neutral medium (CNM and WNM), we first perform a decomposition of gas into atomic and molecular phases, and show that the observed LOSs are mostly H I-dominated. In addition, we find that the CO-dark H₂, not the optically thick H I, is a major ingredient of the dark gas in the solar neighborhood. To examine the conditions for the formation of CO-bright molecular gas, we analyze the kinematic association between H I and CO, and find that the CNM is kinematically more closely associated with CO than the WNM. When CNM components within CO line widths are isolated, we find the following characteristics: spin temperature < 200 K, peak optical depth > 0.1, CNM fraction of ~ 0.6, and V-band dust extinction > 0.5 mag. These results suggest that CO-bright molecular gas preferentially forms in environments with high column densities where the CNM becomes colder and more abundant. Finally, we confront the observed CNM properties with the steady-state H₂ formation model of Sternberg et al. and infer that the CNM must be clumpy with a small volume filling factor. Another possibility would be that missing processes in the model, such as cosmic-rays and gas dynamics, play an important role in the H I-to-H₂ transition.

1. Introduction

As the most abundant molecule in the universe, molecular hydrogen (H₂) plays a key role in the heating and cooling of the interstellar medium (ISM), as well as in the formation of other heavier molecules (e.g., Sternberg & Dalgarno 1995; Hollenbach & Tielens 1997). In addition, H₂ is an essential ingredient for star formation, as extensively shown by Galactic and extragalactic observations (e.g., Kennicutt & Evans 2012). Considering this significance of H₂ in astrophysics, it is of critical importance to understand how H₂ forms out of the surrounding diffuse atomic (H I) gas.

Observationally, the H I-to-H₂ transition has been directly examined through ultraviolet (UV) absorption measurements toward early-type stars or active galactic nuclei (e.g., Savage et al. 1977; Rachford et al. 2002; Shull et al. 2021). These measurements in the Lyman α (1216 Å) and Lyman–Werner (LW; 912–1108 Å) bands probe diffuse to translucent gas with the color excess $E(B - V)$ of ~ 0.01–1.0 mag, and were analyzed to derive H I and H₂ column densities ($N(\mathrm{H\,I})$ and $N(\mathrm{H_2})$). The molecular fraction, $f(\mathrm{H_2}) = 2N(\mathrm{H_2})/[N(\mathrm{H\,I}) + 2N(\mathrm{H_2})]$, was then found to increase from very low ($\lesssim 0.01$) to high values ($\gg 0.1$) at the total hydrogen column density $N(\mathrm{H}) = N(\mathrm{H\,I}) + 2N(\mathrm{H_2})$ of $\sim 10^{21}$ cm^{-2} or $E(B - V)$ of ~ 0.1 mag, indicating a sharp conversion from H I to H₂.

In addition, the H I-to-H₂ transition has been indirectly inferred from the flattening of the H I column density with respect to other dense gas tracers. For example, Barriault et al. (2010) compared H I and OH emission in infrared (IR) cirrus clouds and showed that the OH column density increases with the H I column density up to $N(\mathrm{OH}) \sim 0.3 \times 10^{14}$ cm^{-2}. At higher OH column densities, the H I column density saturates to $\sim 5 \times 10^{20}$ cm^{-2}, which implies the presence of molecular gas not traced by H I emission. Similarly, IR studies of diffuse clouds found a positive deviation from the linear relation between the H I column density and IR emission (e.g., Reach et al. 1994; Douglas & Taylor 2007). The observed excess in

IR emission indicates that a substantial amount of H_2 exists beyond the threshold H I column density of $\sim 5 \times 10^{20}\,\mathrm{cm}^{-2}$.

Theoretically, the H I-to-H_2 transition has been explored as one of the key processes in photodissociation regions (PDRs; e.g., van Dishoeck & Black 1986; Draine & Bertoldi 1996; Browning et al. 2003; Goldsmith et al. 2007; Liszt 2007). In interstellar space, molecular-dominated regions are found in dense regions where gas and dust grains provide sufficient shielding against dissociating UV radiation. These molecular regions are bound by PDRs, where the gas is primarily neutral. The structure of PDRs has been solved numerically and analytically, and recent analytical models (Krumholz et al. 2009; Sternberg et al. 2014; Bialy & Sternberg 2016) predict that the minimum H I column density to shield H_2 from photodissociation depends on ISM conditions (e.g., $N(\mathrm{H\,I}) \sim 10^{21}\,\mathrm{cm}^{-2}$ for solar metallicity). Once this minimum H I column density is accumulated, all excess H I is converted into H_2, resulting in the uniform H I distribution.

While the observed threshold H I column density of $\sim(0.5–1) \times 10^{21}\,\mathrm{cm}^{-2}$ is consistent with what the analytical H_2 formation models predict for H I shielding layers, the previous observational studies could not provide insights into what H I conditions aside from the minimum column density are required for H_2 formation because they did not distinguish between different H I phases. The distinct velocity structures between H I emission and absorption spectral pairs have been interpreted as the presence of H I gas with a range of temperatures and densities (e.g., Radhakrishnan et al. 1972), and theoretical models of neutral atomic gas have indeed suggested that two H I phases can coexist over the range of thermal pressures $P/k_B \sim 10^3–10^4\,\mathrm{cm}^{-3}\,\mathrm{K}$ ($k_B =$ Boltzmann constant): cold neutral medium (CNM) and warm neutral medium (WNM) with densities and temperatures of $(n, T) \sim (5–120\,\mathrm{cm}^{-3}, 40–180\,\mathrm{K})$ and $(0.04–1\,\mathrm{cm}^{-3}, 7000–8000\,\mathrm{K})$ (e.g., Wolfire et al. 1995, 2003; Bialy & Sternberg 2019). In addition to these stable phases, the thermally unstable medium (UNM) with intermediate densities and temperatures has been commonly observed (e.g., Murray et al. 2015, 2018b). As for the formation of molecular gas, the denser and colder CNM is expected to be crucial (e.g., H_2 formation $\propto$ H I density), but the impact of the different H I phases on the H I-to-H_2 transition has been largely unexplored mainly because of a lack of observational constraints.

In this paper, we examine how the different H I phases are related to the H I-to-H_2 transition by analyzing H I emission and absorption spectra along with $^{12}\mathrm{CO}(J = 1 \rightarrow 0)$ (CO(1–0) hereafter) data toward 58 lines of sight (LOSs) at Galactic latitudes $b < -5°$. These data have been obtained as part of the Galactic Neutral Opacity and Molecular Excitation Survey (GNOMES) collaboration, whose primary science goal is to understand the properties of atomic and molecular gas in and around molecular clouds. So far, the H I and OH data were presented in Stanimirović et al. (2014), Nguyen et al. (2019), and Petzler et al. (2023), and we make use of the derived H I properties, such as the optical depth (τ_{CNM}) and spin temperature (T_s) of the CNM and the column densities of the CNM and WNM (N_{CNM} and N_{WNM}), to explore what conditions are required for the formation of CO-bright molecular gas. The observed H I properties are also compared to the analytical model of Sternberg et al. (2014) (S14 hereafter) to test if H_2 formation in steady state is indeed valid for solar neighborhood conditions.

This paper is organized as follows. In Section 2, we summarize two of the most relevant studies, Nguyen et al. (2019) and S14, to provide background information. In Sections 3 and 4, we present the H I, CO, and dust data for our analyses and investigate the environmental conditions of the observed GNOMES LOSs. In Section 5, we describe the results from the CO observations and decompose the gas along each LOS into different atomic and molecular gas phases. The observed H I and CO properties are compared to each other, as well as to the prediction from the S14 model, to provide observational and theoretical perspectives of the conditions for the formation of CO-bright molecular gas (Sections 6 and 7). Finally, our results are discussed and summarized in Sections 8 and 9.

2. Background

In this section, we summarize the recent observational and theoretical studies that are most relevant to our work.

### 2.1. CNM and WNM in and around Molecular Clouds

As part of the GNOMES collaboration, Nguyen et al. (2019) analyzed Arecibo H I emission and absorption spectra toward 77 continuum sources located behind Perseus, Taurus, California, Rosette, NGC 2264, and Mon OB1. For their analyses, the authors divided the observed LOSs into the following three environments: (1) 22 LOSs at $b > 5°$ tracing the diffuse medium (diffuse); (2) 20 LOSs at $|b| < 5°$ penetrating the dense Galactic Plane with likely strong UV radiation field (Plane); and (3) 35 LOSs at $b < -5°$ probing the surroundings of local molecular clouds including Taurus and Perseus (Perseus). The H I spectra along these LOSs were examined via the Gaussian decomposition method of Heiles & Troland (2003a) to estimate the physical properties of H I, such as the optical depth, spin temperature, and column density of the CNM and WNM (see Section 3.1 for details on the observations and analysis methods).

Strong H I absorption was detected toward all the observed LOSs, and a total of 349 CNM and 327 WNM components were identified. For the identified CNM components, the peak optical depth ranges from ~ 0.01 to ~ 16.2 with a median of ~ 0.4, and the spin temperature varies from $\sim 10\,\mathrm{K}$ to $\sim 480\,\mathrm{K}$ with the distribution peak at $\sim 50\,\mathrm{K}$. Interestingly, these individual properties are comparable between the three environments and agree with the results from previous measurements of random LOSs (e.g., Heiles & Troland 2003b; Murray et al. 2015, 2018b), which implies that the CNM has universal properties throughout the Galaxy. Meanwhile, the CNM fraction, which is defined as the ratio of the CNM to total H I column density, is systematically higher in molecular cloud environments (median fractions of 0.43 and 0.37 for the Plane and Perseus LOSs versus 0.16 for the diffuse LOSs), suggesting a close association between the abundance of the CNM and the formation of molecular gas.

2.2. Theoretical Modeling of H_2 Formation in the Steady-state Medium

S14 developed an analytical model of the H I-to-H_2 transition in a one-dimensional plane-parallel slab of gas and dust and provided the following expression of the total H I column density for two-sided isotropic UV radiation:

$$N(\mathrm{H\,I})\,(\mathrm{cm}^{-2}) = \frac{8.4 \times 10^{20}}{\tilde{\sigma}_g} \ln\left(\frac{\alpha G}{3.2} + 1\right) \qquad (1)$$
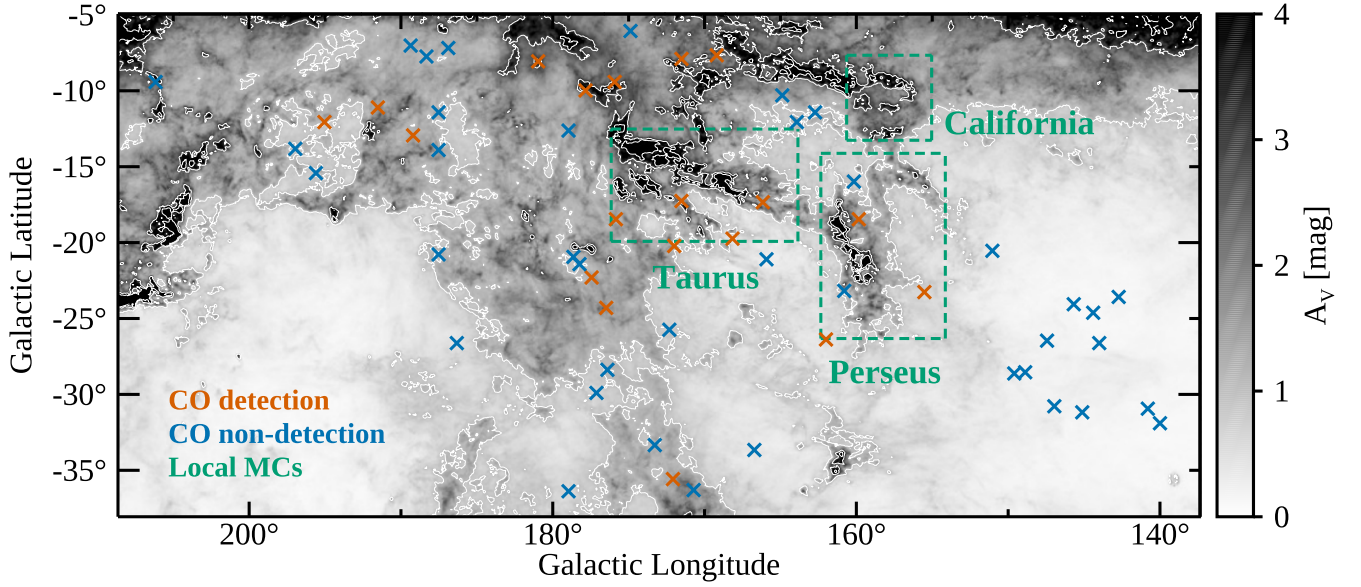
Figure 1. GNOMES LOSs at $b < -5°$ overlaid on the Planck A_V image (Section 4.1; 1, 3, and 5 mag are shown as white contours). Among the total 58 LOSs, the 19 LOSs where CO(1–0) emission is clearly detected are shown as tan crosses. The remaining 39 LOSs without CO detection are indicated as blue crosses. The green boxes represent the approximate extents of local molecular clouds (top to bottom: California, Taurus, and Perseus; Lee et al. 2018).

where $\tilde{\sigma}_g$ is the dust absorption cross-section per hydrogen nucleus in the LW band (σ_g) normalized to the canonical solar metallicity value of 1.9×10^{-21} cm^2.

The dimensionless parameter α in Equation (1) is the ratio of the unattenuated H_2 photodissociation rate to the H_2 formation rate, which can be expressed as

$$\alpha = \frac{D_0}{Rn}$$
$$= 1.9 \times 10^4 \left(\frac{I_{UV}}{\tilde{\sigma}_g}\right)\left(\frac{100 \text{ cm}^{-3}}{n}\right), \quad (2)$$

where D_0 is the free-space H_2 photodissociation rate, R is the rate coefficient for H_2 formation on dust grains, $n = n_1 + 2n_2$ is the total gas number density, n_1 is the H I number density, n_2 is the H_2 number density, and I_{UV} is the strength of UV radiation relative to the Draine field (Draine 1978; Bialy 2020). Meanwhile, the other dimensionless parameter G can be interpreted as the average H_2 self-shielding factor. Here we employ the expression derived by Bialy & Sternberg (2016), which uses a more accurate fitting function for the H_2 dissociation bandwidth and reads as

$$G = 3 \times 10^{-5} \tilde{\sigma}_g \left(\frac{9.9}{1 + 8.9\tilde{\sigma}_g}\right)^{0.37}. \quad (3)$$

By combining Equations (2) and (3), αG can be written as[16]

$$\alpha G = 0.59 \, I_{UV} \left(\frac{100 \text{ cm}^{-3}}{n}\right)\left(\frac{9.9}{1 + 8.9\tilde{\sigma}_g}\right)^{0.37} \quad (4)$$

---

[16] Equation (4) is taken from Bialy & Sternberg (2016), who examined the H I and H_2 density profiles of optically thick interstellar clouds based on the S14 model. While this expression was originally derived for beamed UV radiation, it is also applicable for isotropic UV radiation, considering that α is equal for beamed and isotropic UV radiation fields with the same strength and G is independent of the UV field geometry (S14).

and has the physical meaning of the ratio of the effective H_2 photodissociation rate (accounting for UV shielding) to the H_2 formation rate. For realistic ISM conditions, αG can range from large to small values. For example, when αG is small ($\ll 1$; weak-field limit), H_2 self-shielding primarily protects H_2 from dissociating UV photons, and the H I-to-H_2 transition is gradual. In other words, most of the H I column density is built up beyond the transition point where the gas is mainly molecular. On the contrary, when αG is large ($\gg 1$; strong-field limit), dust absorption becomes important, resulting in a sharp H I-to-H_2 transition due to the exponential reduction of UV radiation with cloud column density. In this case, the H I column density is built up in the outer layer of the gas slab prior to the transition point. We refer the interested reader to S14 and Bialy & Sternberg (2016) for further details of the model and the parameters.

3. Data

3.1. GNOMES: H I and OH

In this study, we make use of the H I (1.4204 GHz) and OH (1.6654 and 1.6673 GHz) emission/absorption spectra from Stanimirović et al. (2014) and Nguyen et al. (2019). These spectra were obtained with the 305 m Arecibo telescope (providing angular and velocity resolutions of 3.5′ and 0.16 km s^{-1}) toward 100 extragalactic continuum sources that were selected from the NRAO VLA Sky Survey (NVSS; Condon et al. 1998) with 1.4 GHz flux densities $S_{1.4} \gtrsim 0.6$ Jy. Among the observed sources, 58 at $b < -5°$ probing the surroundings of the Perseus, Taurus, and California molecular clouds were considered for our study (Figure 1 and Table 1).

The methodology of the observations and data reduction in Stanimirović et al. (2014) and Nguyen et al. (2019) is essentially based on Heiles & Troland (2003a), and we provide here a summary of the H I data. For each source, one on-source and 16 off-source measurements were made to obtain optical depth (τ_{CNM}) and expected emission (T_{exp}) spectra. The

Table 1
58 LOSs in Our Study

Source	R.A. (J2000) (hh:mm:ss)	Decl. (J2000) (dd:mm:ss)	l (°)	b (°)	$S_{1.4}$ (Jy)	T_{sky} (K)	CO(1–0)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
J034053+073525 (4C+07.13)	03:40:53.73	07:35:25.40	178.87	−36.27	1.0	4.07	
J032153+122114 (PKS0319+12)	03:21:53.11	12:21:14.00	170.59	−36.24	1.9	4.51	
J032723+120835 (4C+11.15)	03:27:23.11	12:08:35.80	171.98	−35.48	1.2	4.17	✓
J031857+162833 (4C+16.09)	03:18:57.77	16:28:33.10	166.64	−33.60	8.0	6.93	
J033626+130233 (3C090)	03:36:26.56	13:02:33.20	173.15	−33.29	2.0	4.67	
J015712+285138 (NV0157+28)	01:57:12.85	28:51:38.49	139.90	−31.83	1.4	2.78	
J021701+280458 (4C+27.07)	02:17:01.89	28:04:59.12	145.01	−31.09	1.0	2.79	
J020136+293340 (4C+29.05)	02:01:35.91	29:33:44.18	140.72	−30.88	1.2	2.79	
J022412+275011 (3C067)	02:24:12.31	27:50:11.69	146.82	−30.70	3.0	2.79	
J035613+130535	03:56:13.81	13:05:35.80	177.02	−29.78	0.9	4.14	
J023752+284809 (4C+28.07)	02:37:52.42	28:48:09.16	149.47	−28.53	2.2	2.79	
J023535+290857 (4C+28.06)	02:35:35.41	29:08:57.73	148.78	−28.44	1.3	2.79	
J035900+143622 (3C096)	03:59:00.91	14:36:22.50	176.27	−28.26	1.2	4.37	
J022048+324106 (5C06.237)	02:20:48.06	32:41:06.64	143.88	−26.53	0.9	2.79	
J042725+085330 (4C+08.15)	04:27:25.05	08:53:30.30	186.21	−26.51	0.9	4.08	
J023423+313418 (3C068.2)	02:34:23.87	31:34:17.62	147.33	−26.38	1.0	2.79	
J032504+244445 (4C+24.06)	03:25:04.35	24:44:45.60	161.92	−26.26	0.8	4.13	✓
J035633+190034 (4C+18.11)	03:56:33.46	19:00:34.60	172.23	−25.66	1.1	4.15	
J022610+342130 (4C+34.07)	02:26:10.34	34:21:30.45	144.31	−24.55	2.9	2.79	
J041140+171405 (4C+17.23)	04:11:40.77	17:14:05.10	176.36	−24.24	1.0	4.26	✓
J023228+342405 (NV0232+34)	02:32:28.72	34:24:06.08	145.60	−23.98	2.6	2.79	
J022105+355613 (B20218+35)	02:21:05.48	35:56:13.91	142.60	−23.49	1.7	2.79	
J031135+304320 (4C+30.04)	03:11:35.19	30:43:20.62	155.40	−23.17	1.0	2.79	✓
J032957+275615 (B20326+27)	03:29:57.69	27:56:15.64	160.70	−23.07	1.3	2.79	
J042022+175355 (3C114)	04:20:22.17	17:53:55.20	177.30	−22.24	1.1	4.23	✓
J042524+175525 (4C+17.25)	04:25:24.43	17:55:25.30	178.11	−21.31	0.9	4.16	
J035204+262418 (4C+26.12)	03:52:04.36	26:24:18.11	165.82	−21.06	1.4	2.78	
J042756+175242 (4C+17.26)	04:27:56.98	17:52:42.80	178.56	−20.88	1.0	4.22	
J044907+112128 (PKS0446+11)	04:49:07.65	11:21:28.20	187.43	−20.74	0.9	4.16	
J030142+351219 (4C+34.09)	03:01:42.38	35:12:20.84	150.94	−20.49	1.9	2.79	
J041243+230506 (3C108)	04:12:43.69	23:05:05.53	171.87	−20.12	1.5	2.79	✓
J040305+260001 (B20400+25)	04:03:05.61	26:00:01.61	168.03	−19.65	0.9	2.79	✓
J034008+320901 (3C092)	03:40:08.54	32:09:01.30	159.74	−18.41	1.6	3.95	✓
J042846+213331 (4C+21.17)	04:28:46.64	21:33:31.40	175.70	−18.36	1.3	4.35	✓
J040442+290215 (4C+28.11)	04:04:42.82	29:02:15.90	166.06	−17.22	1.0	3.69	✓
J042049+252627 (4C+25.14)	04:20:49.30	25:26:27.63	171.37	−17.16	1.0	2.79	✓
J034846+335315 (3C093.1)	03:48:46.93	33:53:15.41	160.04	−15.91	2.4	2.80	
J052424+074957 (4C+07.16)	05:24:24.04	07:49:57.10	195.51	−15.35	0.8	4.25	
J051240+151723 (PKS0509+152)	05:12:40.99	15:17:23.80	187.41	−13.79	1.0	4.11	
J053239+073243	05:32:39.01	07:32:43.50	196.84	−13.74	2.7	4.96	
J051930+142829 (4C+14.14)	05:19:30.95	14:28:29.00	189.04	−12.85	0.9	4.15	✓
J045643+224922 (3C132)	04:56:43.08	22:49:22.27	178.86	−12.52	3.4	2.80	
J053450+100430 (4C+09.21)	05:34:50.82	10:04:30.30	194.89	−11.98	1.1	4.62	✓
J041437+341851 (B20411+34)	04:14:37.28	34:18:51.31	163.80	−11.98	1.9	2.79	
J041236+353543 (4C+35.07)	04:12:36.28	35:35:43.20	162.58	−11.36	0.9	3.93	
J052109+163822 (3C138)	05:21:09.93	16:38:22.20	187.41	−11.34	8.6	7.59	
J053056+133155 (PKS0528+134)	05:30:56.44	13:31:55.30	191.37	−11.01	1.6	4.64	✓
J042353+345144 (3C115)	04:23:53.25	34:51:44.80	164.76	−10.24	1.3	3.88	
J050258+251624 (3C133)	05:02:58.51	25:16:25.16	177.73	−9.91	5.8	2.80	✓
J060536+014512 (4C+01.17)	06:05:36.56	01:45:12.70	206.08	−9.37	0.6	4.07	
J045956+270602 (4C+27.14)	04:59:56.09	27:06:02.90	175.83	−9.36	0.9	3.90	✓
J051740+235110 (4C+23.14)	05:17:40.81	23:51:10.20	180.86	−8.01	1.0	4.32	✓
J045323+312924 (3C131)	04:53:23.34	31:29:24.20	171.44	−7.80	2.9	4.04	✓
J053557+175600 (4C+17.33)	05:35:57.42	17:56:00.70	188.22	−7.67	0.8	4.23	
J044708+332747 (4C+33.10)	04:47:08.90	33:27:46.85	169.05	−7.57	1.2	2.80	✓
J053444+192721 (PKS0531+19)	05:34:44.51	19:27:21.70	186.76	−7.11	7.0	6.48	
J054046+172839 (4C+17.34)	05:40:46.05	17:28:39.20	189.21	−6.93	1.5	4.50	
J050929+295755 (4C+29.16)	05:09:29.51	29:57:55.80	174.77	−5.97	1.1	4.03	

Note. (1) Source name; (2 and 3) right ascension (R.A.) and decl. (Decl.) coordinates; (4 and 5) Galactic coordinates; (6) flux density at 1.4 GHz; (7) diffuse background radio continuum emission; and (8) detection of CO(1–0) emission. Here, columns (1)–(7) are from Stanimirović et al. (2014) and Nguyen et al. (2019).

expected emission spectrum is the one that would be observed at the source position if the source were turned off and was derived by approximating the off-source spectra as a second-order Taylor expansion of the expected emission spectrum. This approximation was used to consider spatial variations in H I emission, and the derivatives were used to estimate the uncertainty spectrum of expected emission. The median 1σ uncertainties in the measured optical depth ($\sigma_{e^{-\tau}}$) and expected emission ($\sigma_{T_{exp}}$) at a velocity resolution of 0.16 km s^{-1} are 0.02 and 0.36 K, respectively.

The obtained H I absorption and emission spectra were analyzed through the Gaussian decomposition method of Heiles & Troland (2003a). This method simultaneously fits the absorption and emission spectra with individual Gaussian components under the assumption that the CNM is detected in both absorption and emission, while the WNM contributes to the emission spectrum only. In the fitting process, all possible permutations of the CNM components are considered to find the best-fit model with a minimum chi-square value. In addition, the fitting takes into account the possibility that a certain fraction of the WNM (F) could be located in front of the CNM by assuming three cases $F = 0$, 0.5, and 1 (e.g., $F = 1$ means that the WNM is not absorbed by the CNM at all). The final parameters from the fitting process include the velocities, widths, spin temperatures, peak optical depths, and H I column densities of individual Gaussian components,[17] and we refer the interested reader to Section 3 of Stanimirović et al. (2014) for further details of the fitting procedure. For our analyses, we mostly used the derived H I properties and utilized the OH spectra only to separate LOSs with molecular gas (Section 4.1).

Finally, we note that 3C092, 3C131, and 4C+27.14 were observed in both Stanimirović et al. (2014) and Nguyen et al. (2019). These observations are essentially consistent within uncertainties, and we used the spectra from Nguyen et al. (2019) for our analyses because they have better sensitivities.

3.2. CO

Single-pointing observations of the CO(1–0) transition at 115.2712 GHz were carried out toward the 58 GNOMES LOSs at $b < -5°$ using the 13.7 m telescopes at the Taeduk Radio Astronomy Observatory (TRAO) and Purple Mountain Observatory (PMO). The TRAO observations were made in March and December 2020 and in February 2021, while the PMO observations were performed from March to May 2020. During these observations, the system temperature was 400–900 K and 200–300 K for the TRAO and PMO telescopes, respectively.

The obtained CO(1–0) spectra were processed using the GILDAS CLASS software.[18] For the TRAO data, a beam efficiency of $\eta_{MB} = 0.40$ was adopted to convert the corrected antenna temperature into the main-beam brightness temperature ($T_{MB} = T_A^*/\eta_{MB}$). Meanwhile, no conversion was made for the PMO data because they were delivered in units of main-beam brightness temperature. The final spectra on 48″ scales were smoothed to a velocity resolution of 0.32 km s^{-1} and have a median root-mean-square (rms) noise level of 0.1 K. A comparison between the CO(1–0) emission and H I absorption spectra is presented in Appendix A.

To determine the presence of CO emission, we adopted a 3σ threshold and considered components whose peak-to-rms ratios are equal to or higher than three as detections. Once the presence of CO emission was confirmed, we fitted Gaussians to the spectra to derive the line parameters, such as the central velocity (ν_{CO}), full width at half maximum (FWHM; $\Delta\nu_{CO}$), and peak main-beam brightness temperature ($T_{peak,CO}$). The derived line parameters, as well as the CO integrated intensity ($I(CO)$; calculated by integrating CO(1–0) emission over a velocity range where the emission is clearly visible) and rms noise, are presented in Table 2.

Finally, we note that two of our target sources (3C092 and 3C108) were observed using both telescopes to check the calibration levels of the TRAO and PMO observations. The difference between the TRAO and PMO observations was 10%–20%, which is within the calibration uncertainty of $\sim$20% for the TRAO telescope at 115 GHz. This suggests that the obtained CO(1–0) spectra are well calibrated and can be used for further analyses.

3.3. Planck Data

We used Planck data to estimate the environmental conditions of the GNOMES LOSs such as the strength of UV radiation (I_{UV}), V-band dust extinction (A_V), and dust-to-gas ratio (DGR). Specifically, we employed the images of dust temperature (T_{dust}), spectral index (β), and dust opacity at 353 GHz (τ_{353}) from Planck Collaboration et al. (2016) and extracted the values of the 58 LOSs using the "dustmaps" Python package from Green (2018). These extracted values are on 5′ scales.

4. Environmental Conditions

4.1. Dust and Gas Properties

Before comparing the observed H I and CO properties, we probed the environmental conditions of the GNOMES LOSs based on the Planck data. We first calculated the dust abundances by converting the 353 GHz dust opacity into the V-band dust extinction:

$$\begin{aligned} A_V \text{ (mag)} &= R_V \, E(B-V) \\ &= 3.1 \times (1.5 \times 10^4 \, \tau_{353}). \end{aligned} \quad (5)$$

For this calculation, the total-to-selective extinction ratio $R_V = 3.1$ for the diffuse ISM is assumed (Mathis 1990). In addition, the conversion factor of 1.5×10^4 mag is adopted to translate the 353 GHz dust opacity τ_{353} into the reddening $E(B-V)$ based on Planck Collaboration et al. (2014). The derived A_V ranges from 0.2 mag to 4 mag with a median of 1 mag, suggesting that our LOSs probe diffuse to dense interstellar gas.

Next we estimated the DGR by deriving $A_V/N(H I)$ for diffuse LOSs where gas is dominated by atomic gas. To identify these LOSs, the following criteria were applied: (1) no CO or OH detection, and (2) $A_V < 0.5$ mag. The second threshold is motivated by observational and theoretical studies that found H_2 formation at $A_V \sim 0.5$ mag in the solar neighborhood (e.g., Lee et al. 2012; Sternberg et al. 2014). With the two criteria, we found 13 atomic-dominated LOSs and calculated $N(H I)$ by considering both CNM and WNM column

---

[17] For WNM components, lower and upper limits are provided on the spin temperature and peak optical depth, respectively.
[18] https://www.iram.fr/IRAMFR/GILDAS/

Table 2
Derived CO(1–0) Properties

Source	ν_{CO} (km s^{-1})	$\Delta\nu_{CO}$ (km s^{-1})	$T_{peak,CO}$ (K)	$I(CO)$ (K km s^{-1})	σ_{rms} (K)	Telescope
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4C+07.13	⋯	⋯	⋯	⋯	0.10	PMO
PKS0319+12	⋯	⋯	⋯	⋯	0.11	PMO
4C+11.15	6.96 ± 0.02	0.41 ± 0.06	1.14 ± 0.13	0.49 ± 0.04	0.07	PMO
4C+16.09	⋯	⋯	⋯	⋯	0.17	TRAO
3C090	⋯	⋯	⋯	⋯	0.10	PMO
NV0157+28	⋯	⋯	⋯	⋯	0.10	PMO
4C+27.07	⋯	⋯	⋯	⋯	0.10	PMO
4C+29.05	⋯	⋯	⋯	⋯	0.10	PMO
3C067	⋯	⋯	⋯	⋯	0.10	PMO
J035613+130535	⋯	⋯	⋯	⋯	0.09	PMO
4C+28.07	⋯	⋯	⋯	⋯	0.08	PMO
4C+28.06	⋯	⋯	⋯	⋯	0.10	PMO
3C096	⋯	⋯	⋯	⋯	0.08	PMO
5C06.237	⋯	⋯	⋯	⋯	0.10	PMO
4C+08.15	⋯	⋯	⋯	⋯	0.20	TRAO
3C068.2	⋯	⋯	⋯	⋯	0.07	PMO
4C+24.06	6.95 ± 0.09	1.02 ± 0.21	0.44 ± 0.08	0.50 ± 0.09	0.10	PMO
4C+18.11	⋯	⋯	⋯	⋯	0.10	PMO
4C+34.07	⋯	⋯	⋯	⋯	0.08	PMO
4C+17.23[a]	9.10 ± 0.01	0.64 ± 0.02	4.96 ± 0.11	3.43 ± 0.07	0.11	PMO
4C+17.23[a]	11.17 ± 0.02	0.92 ± 0.05	2.07 ± 0.09	2.04 ± 0.07	0.11	PMO
NV0232+34	⋯	⋯	⋯	⋯	0.08	PMO
B20218+35	⋯	⋯	⋯	⋯	0.07	PMO
4C+30.04[a]	0.69 ± 0.01	0.62 ± 0.03	2.64 ± 0.12	1.77 ± 0.06	0.10	PMO
4C+30.04[a]	0.89 ± 0.19	2.79 ± 0.53	0.37 ± 0.09	1.10 ± 0.13	0.10	PMO
B20326+27	⋯	⋯	⋯	⋯	0.10	PMO
3C114[a]	8.55 ± 0.11	0.70 ± 0.27	0.33 ± 0.10	0.25 ± 0.06	0.10	PMO
3C114[a]	9.48 ± 0.05	0.46 ± 0.15	0.66 ± 0.16	0.33 ± 0.05	0.10	PMO
4C+17.25	⋯	⋯	⋯	⋯	0.10	PMO
4C+26.12	⋯	⋯	⋯	⋯	0.11	TRAO
4C+17.26	⋯	⋯	⋯	⋯	0.11	PMO
PKS0446+11	⋯	⋯	⋯	⋯	0.12	PMO
4C+34.09	⋯	⋯	⋯	⋯	0.10	PMO
3C108[b]	6.14 ± 0.11	0.40 ± 0.21	0.69 ± 0.29	0.30 ± 0.07	0.13	TRAO
3C108[b]	9.42 ± 0.01	1.13 ± 0.02	11.10 ± 0.15	13.49 ± 0.07	0.13	TRAO
B20400+25	7.07 ± 0.06	1.01 ± 0.13	0.59 ± 0.07	0.69 ± 0.08	0.09	PMO
3C092	8.80 ± 0.01	1.66 ± 0.02	9.68 ± 0.10	17.22 ± 0.12	0.10	PMO
4C+21.17	10.17 ± 0.06	0.65 ± 0.13	0.66 ± 0.12	0.58 ± 0.09	0.12	PMO
4C+28.11	6.63 ± 0.01	1.10 ± 0.02	9.83 ± 0.12	11.78 ± 0.12	0.10	PMO
4C+25.14[a]	3.59 ± 0.09	1.33 ± 0.21	0.47 ± 0.06	0.67 ± 0.09	0.09	PMO
4C+25.14[a]	6.78 ± 0.01	0.74 ± 0.02	4.78 ± 0.10	3.80 ± 0.09	0.09	PMO
4C+25.14[a]	7.92 ± 0.02	1.14 ± 0.06	2.64 ± 0.07	3.24 ± 0.08	0.09	PMO
3C093.1	⋯	⋯	⋯	⋯	0.09	PMO
4C+07.16	⋯	⋯	⋯	⋯	0.09	PMO
PKS0509+152	⋯	⋯	⋯	⋯	0.12	PMO
J053239+073243	⋯	⋯	⋯	⋯	0.09	PMO
4C+14.14	2.17 ± 0.13	0.34 ± 0.42	0.60 ± 0.57	0.28 ± 0.08	0.15	PMO
3C132	⋯	⋯	⋯	⋯	0.10	PMO
4C+09.21	1.99 ± 0.26	2.59 ± 0.61	0.23 ± 0.05	0.64 ± 0.11	0.10	PMO
B20411+34	⋯	⋯	⋯	⋯	0.09	PMO
4C+35.07	⋯	⋯	⋯	⋯	0.11	PMO
3C138	⋯	⋯	⋯	⋯	0.15	PMO
PKS0528+134	9.63 ± 0.02	0.89 ± 0.06	2.64 ± 0.15	2.77 ± 0.15	0.17	PMO
3C115	⋯	⋯	⋯	⋯	0.11	TRAO
3C133	7.45 ± 0.02	0.84 ± 0.04	3.57 ± 0.15	3.18 ± 0.14	0.18	PMO
4C+01.17	⋯	⋯	⋯	⋯	0.11	TRAO
4C+27.14[a]	6.03 ± 0.04	1.00 ± 0.10	1.11 ± 0.08	1.19 ± 0.07	0.09	PMO
4C+27.14[a]	7.78 ± 0.01	1.27 ± 0.02	7.16 ± 0.07	9.77 ± 0.08	0.09	PMO
4C+23.14[b]	-3.70 ± 0.13	1.32 ± 0.30	0.36 ± 0.07	0.51 ± 0.09	0.10	TRAO
4C+23.14[b]	1.15 ± 0.13	0.95 ± 0.31	0.41 ± 0.08	0.42 ± 0.09	0.10	TRAO
4C+23.14[b]	2.37 ± 0.13	0.84 ± 0.30	0.39 ± 0.09	0.35 ± 0.07	0.10	TRAO
3C131[a]	4.79 ± 0.12	1.60 ± 0.31	0.51 ± 0.06	0.87 ± 0.10	0.10	PMO

<div align="center">**Table 2**</div>
<div align="center">(Continued)</div>

Source (1)	ν_{CO} (km s^{-1}) (2)	$\Delta\nu_{CO}$ (km s^{-1}) (3)	$T_{peak,CO}$ (K) (4)	$I(CO)$ (K km s^{-1}) (5)	σ_{rms} (K) (6)	Telescope (7)
3C131[a]	6.86 ± 0.02	1.31 ± 0.04	3.67 ± 0.07	5.16 ± 0.09	0.10	PMO
4C+17.33	⋯	⋯	⋯	⋯	0.09	TRAO
4C+33.10[b]	-2.29 ± 0.01	1.22 ± 0.02	5.53 ± 0.09	7.24 ± 0.09	0.10	PMO
4C+33.10[b]	5.91 ± 0.01	0.42 ± 0.06	3.09 ± 0.35	1.38 ± 0.09	0.10	PMO
4C+33.10[b]	6.61 ± 0.01	0.68 ± 0.02	6.24 ± 0.13	4.55 ± 0.07	0.10	PMO
PKS0531+19	⋯	⋯	⋯	⋯	0.09	TRAO
4C+17.34	⋯	⋯	⋯	⋯	0.11	TRAO
4C+29.16	⋯	⋯	⋯	⋯	0.18	TRAO

Notes. (1) Source name; (2) central velocity; (3) FWHM; (4) peak brightness temperature; (5) integrated intensity; (6) rms noise level; and (7) telescope that was used to obtained the spectrum.
[a] Multiple peaks are close enough in velocity to be considered as one component based on our threshold (Section 5.1).
[b] These sources are considered to have two distinct peaks.

densities:

$$N(\text{H I}) (\text{cm}^{-2}) = N_{CNM} + N_{WNM}$$
$$= 1.823 \times 10^{18} \int \left(\sum_0^{N-1} T_{s,n} \tau_{0,n} e^{-[(\nu-\nu_{0,n})/\delta\nu_n]^2} \right.$$
$$\left. + \sum_0^{K-1} T_{0,k} e^{-[(\nu-\nu_{0,k})/\delta\nu_k]^2} \right) d\nu \text{ (K km s}^{-1}). \quad (6)$$

Here the subscripts n and k refer to CNM and WNM components, τ_0 is the peak optical depth, ν_0 is the central velocity, T_0 is the peak brightness temperature, and $\delta\nu$ is the $1/e$ width of the component. The derived $A_V/N(\text{H I})$ in units of mag cm^2 has a range of $(0.3–0.5) \times 10^{-21}$ with a median of 0.4×10^{-21} and is in good agreement with typical Galactic DGR values (e.g., Bohlin et al. 1978; Liszt 2014; Lenz et al. 2017; Nguyen et al. 2018).

Finally, we estimated total gas column densities toward the GNOMES LOSs by dividing the Planck-based A_V by the representative DGR of 0.4×10^{-21} mag cm^2:

$$N(\text{H})_{A_V-\text{based}} (\text{cm}^{-2}) = N(\text{H I}) + 2N(\text{H}_2)$$
$$= \frac{A_V}{0.4 \times 10^{-21} \text{ mag cm}^2}. \quad (7)$$

4.2. UV Radiation

We estimated the strength of UV radiation in units of the Draine field using the expression derived for dust grains at high Galactic latitudes (e.g., Boulanger et al. 1996; Paradis et al. 2011):

$$I_{UV} = \left(\frac{T_{dust}}{17.5 \text{ K}} \right)^{\beta+4}. \quad (8)$$

Except for the four LOSs with relatively high dust temperatures (20–24 K), T_{dust} is mostly 18 K, resulting in the typical solar neighborhood condition of $I_{UV} \sim 1$.

4.3. Uncertainties

Our estimation of the strength of UV radiation and total gas column density likely suffers from several systematic uncertainties. First, the Planck dust data (T_{dust}, β, and τ_{353}) were derived based on the model of modified blackbody emission, and the main assumptions for this derivation, such as a single dust temperature along a LOS, could be invalid under some circumstances. Second, the conversion of τ_{353} into $N(\text{H})$ involves a few steps, which are likely to be reasonable for diffuse gas but could be less appropriate for H$_2$-dominated LOSs. For example, R_V could be higher than 3.1 in dense regions due to grain growth (e.g., Chapman & Mundy 2009; Steinacker et al. 2010). Grain growth could also cause an underestimation of the DGR (e.g., Roman-Duval et al. 2014). Finally, the different angular resolutions of the Planck and H I measurements could hinder a derivation of accurate gas and dust properties.

This discussion of the possible uncertainties demonstrates that various factors affect our derivation of the dust and gas properties. However, it is not straightforward to evaluate the impact of each factor based on the currently available data, and we hence proceeded while bearing in mind the uncertainty sources.

5. Results

5.1. Observed CO Properties

CO(1–0) is detected toward 19 sources (tan crosses in Figure 1), which suggests a detection rate of 33% at a rms level of 0.1 K. Among these sources, 10 show simple spectra with single Gaussians, while nine have multiple peaks. For these peaks, we examined the difference between their central velocities and considered them as one component if the velocity difference is smaller than the sum of their $2 \times$ FWHMs. Based on this threshold, only the peaks toward 3C108, 4C+23.14, and 4C+33.10 are regarded as sufficiently distinct components.

The derived central velocities are mostly between 7 km s^{-1} and 11 km s^{-1} with a few components at lower velocities (-4 km s^{-1} to 4 km s^{-1}), which suggests that the observed CO(1–0) emission is associated with Perseus, Taurus, California, and their surrounding regions (e.g., Ridge et al. 2006; Narayanan et al. 2008; Lada et al. 2009). The FWHM line widths are generally small, with a median of 1 km s^{-1}. Finally, the peak main-beam brightness temperature ranges from 0.2 K to 11.1 K, indicating that we are tracing diffuse ($\lesssim 1$ K) to dense ($\gtrsim 5$ K) molecular gas.
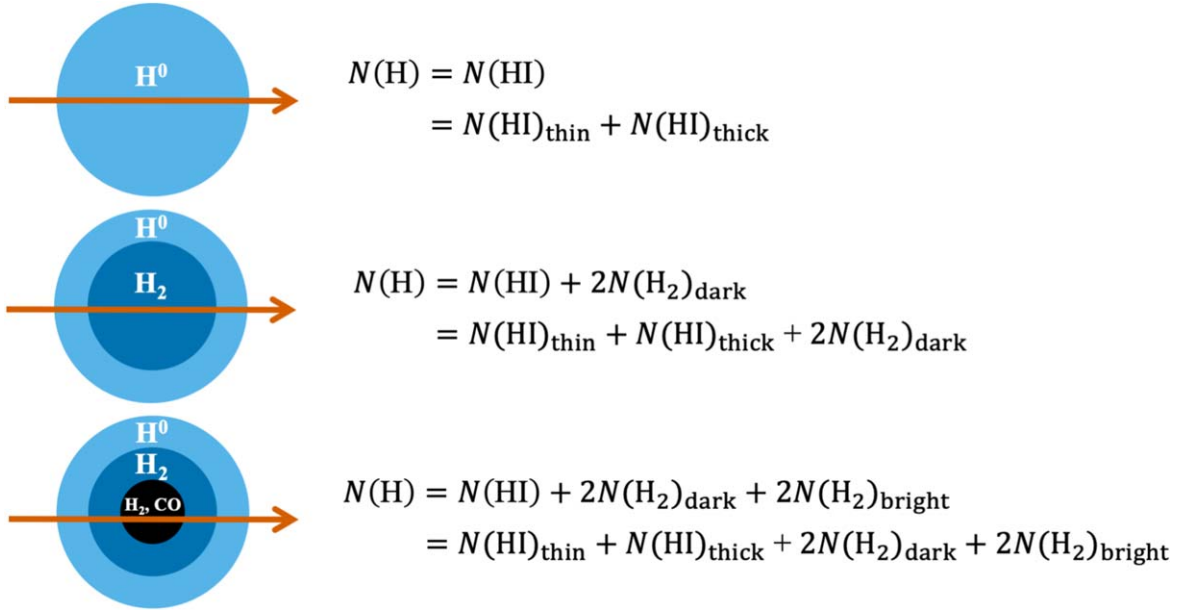
Figure 2. Illustration of LOSs with different gas phases. (Top) H I-only LOSs. (Middle) LOSs with H I and CO-dark H_2 gas. (Bottom) LOSs with H I, CO-dark H_2, and CO-bright H_2 gas.

5.2. Gas Phases

With the available multi-wavelength data, we can examine gas phases toward the GNOMES LOSs where

$$N(\mathrm{H}) = N(\mathrm{H\,I}) + 2N(\mathrm{H_2})$$
$$= N(\mathrm{H\,I})_{\mathrm{thin}} + N(\mathrm{H\,I})_{\mathrm{thick}} + 2N(\mathrm{H_2})_{\mathrm{dark}}$$
$$+ 2N(\mathrm{H_2})_{\mathrm{bright}}. \quad (9)$$

For our examination, we used A_V as a tracer of total gas column density and estimated $N(\mathrm{H})$ by dividing A_V by the representative DGR of 0.4×10^{-21} mag cm^2 (Section 4.1). In the case of atomic gas, its true column density is a sum of the following: (1) $N(\mathrm{H\,I})_{\mathrm{thin}}$, which is calculated by assuming optically thin emission; and (2) $N(\mathrm{H\,I})_{\mathrm{thick}}$, which is missing in the optically thin approximation due to a high opacity. Similarly, molecular gas consists of two types of H_2: (1) $N(\mathrm{H_2})_{\mathrm{dark}}$, which is invisible in CO(1–0) emission; and (2) $N(\mathrm{H_2})_{\mathrm{bright}}$, which is traced by CO(1–0) emission. H_2 with faint or no CO(1–0) emission (CO-dark H_2) is expected due to the different locations of H_2 and CO formation in interstellar clouds (A_V ∼0.5 mag and 2 mag) and exists along with C^+ and C^0 (e.g., Tielens & Hollenbach 1985; Grenier et al. 2005; Wolfire et al. 2010; Bolatto et al. 2013). Among the four components of gas in Equation (9), the optically thick H I and CO-dark H_2 together are called "dark gas" because they are not probed by traditional gas tracers such as H I and CO(1–0) emission. Our examination of the different gas phases is illustrated in Figure 2.

As the first step of our examination, we derived $N(\mathrm{H\,I}) = N(\mathrm{H\,I})_{\mathrm{thin}} + N(\mathrm{H\,I})_{\mathrm{thick}}$ based on Equation (6) and separated $N(\mathrm{H\,I})_{\mathrm{thin}}$ and $N(\mathrm{H\,I})_{\mathrm{thick}}$ for the observed 58 LOSs by:

$$N(\mathrm{H\,I})_{\mathrm{thin}} \ (\mathrm{cm}^{-2}) = 1.823 \times 10^{18} \int T_{\exp} \ (\mathrm{K \ km \ s^{-1}}) \quad (10)$$

and

$$N(\mathrm{H\,I})_{\mathrm{thick}} = N(\mathrm{H\,I}) - N(\mathrm{H\,I})_{\mathrm{thin}}. \quad (11)$$

The derived optically thin H I column densities make up 16%–99% (median of 62%) of the total $N(\mathrm{H})$, while the optically thick H I column densities constitute only 1%–38% (median of 12%). These results suggest that the observed LOSs are mostly H I-dominated and the contribution from the optically thick H I to the total $N(\mathrm{H})$ is small (Figure 3 and Table 3).

In addition, we calculated $2N(\mathrm{H_2})_{\mathrm{dark}}$ for the 39 CO non-detected LOSs, which include H I-only and H I + CO-dark H_2 LOSs, by

$$2N(\mathrm{H_2})_{\mathrm{dark}} = N(\mathrm{H})_{A_V-\mathrm{based}} - N(\mathrm{H\,I})$$
$$= N(\mathrm{H})_{A_V-\mathrm{based}} - (N(\mathrm{H\,I})_{\mathrm{thin}} + N(\mathrm{H\,I})_{\mathrm{thick}}) \quad (12)$$

and present its distribution in Figure 4.

Figure 4 shows that the CO-dark H_2 column density distribution is approximately Gaussian from -5×10^{20} cm^{-2} to 5×10^{20} cm^{-2} with a peak of ∼0 cm^{-2}, suggesting that the 39 LOSs are dominated by atomic-only LOSs and the dispersion of the Gaussian distribution likely results from a slight variation in the DGR. In other words, our adopted DGR of 0.4×10^{-21} mag cm^2 is indeed representative, and $2N(\mathrm{H_2})_{\mathrm{dark}}$ values larger than 5×10^{20} cm^{-2} are likely reliable. For 14 CO non-detected LOSs with $2N(\mathrm{H_2})_{\mathrm{dark}} > 5 \times 10^{20}$ cm^{-2}, we then found that the ratio of $2N(\mathrm{H_2})_{\mathrm{dark}}$ to $N(\mathrm{H})$ changes from 14% to 54% with a median of 31% (Figure 3 and Table 3). As compared to the CO-dark H_2, the contribution from the optically thick H I to the total $N(\mathrm{H})$ is minor (7%–34% with a median of 15%). This finding of the CO-dark H_2 as a major constituent of the dark gas in the solar neighborhood is in agreement with previous studies, such as Lee et al. (2015), Liszt et al. (2018), and Murray et al. (2018a). In addition, our median CO-dark H_2 fraction of 31% is
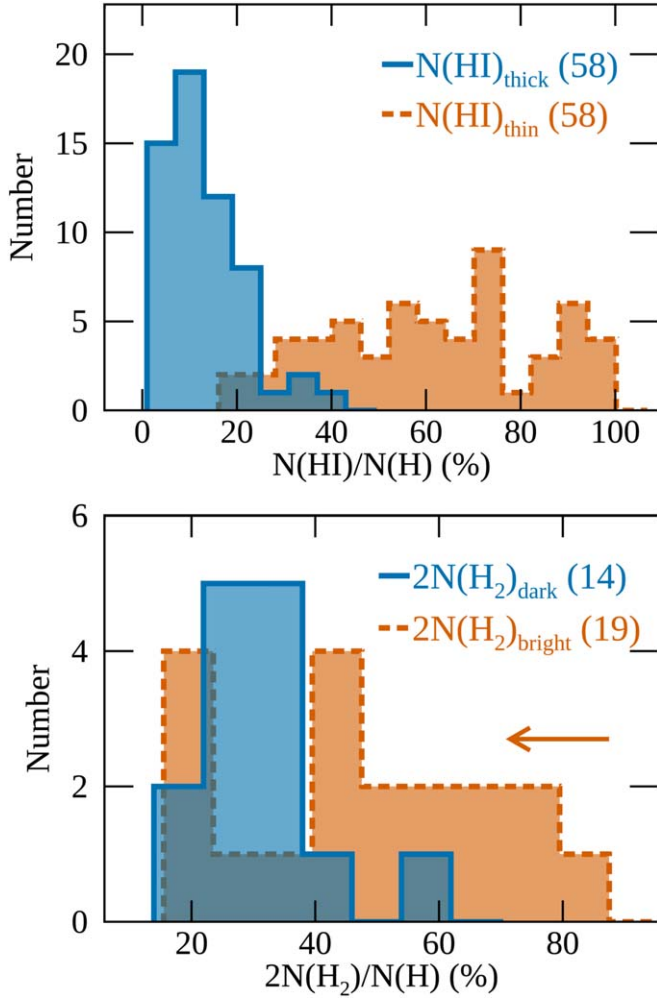
Figure 3. Fraction of each gas phase with respect to the total hydrogen. (Top) Optically thick and thin H I in solid blue and dashed tan. (Bottom) CO-dark and CO-bright H_2 in solid blue and dashed tan. The leftward arrow indicates that the derived fractions for CO-bright H_2 are upper limits.

consistent with the Galactic average value of ∼30% derived from the Herschel GOT C$^+$ survey (Langer et al. 2014).

Finally, we estimated upper limits on $2N(H_2)_{bright}$ for the 19 CO-detected LOSs by

$$2N(H_2)_{bright} = N(H)_{A_V-based} - N(H\,I) - 2N(H_2)_{dark}$$
$$< N(H)_{A_V-based} - N(H\,I) \qquad (13)$$

and summarize the results in Figure 3 and Table 3. As shown in Figure 2, the CO-detected LOSs probe CO-free H_2 shells, as well as CO-bright H_2 cores. However, separating these two components is not straightforward unless a CO-to-H_2 conversion factor X_{CO} is applied to the measured CO integrated intensity to calculate the CO-bright H_2 column density. Considering that X_{CO} could change by more than a factor of 100 over the measured $A_V \sim 0.5$–4 mag for the CO-detected LOSs (e.g., Lee et al. 2014), we do not take the X_{CO} approach and provide upper limits on $2N(H_2)_{bright}$ by assigning all the measured H_2 to CO-bright H_2 (inequality sign in Equation (13)). In this case, the ratio of the upper limit on the CO-bright H_2 column density to the total hydrogen column density ranges from 16%–81%, with a median of 44%.
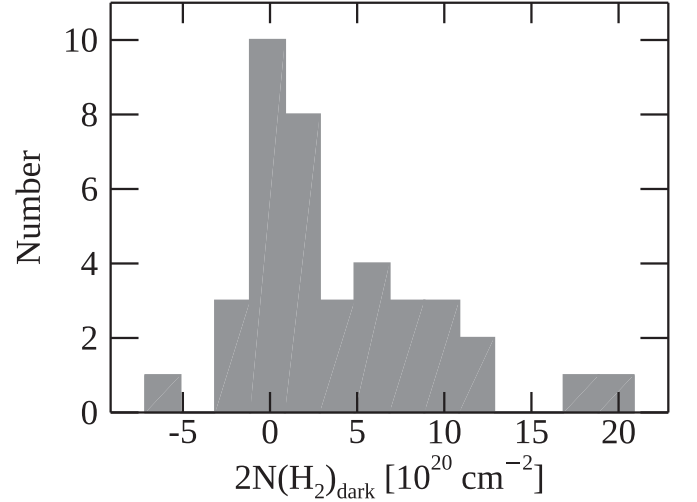
Figure 4. Histogram of the CO-dark H_2 column density for the 39 CO non-detected LOSs.

Table 3
Derived Properties of the Four Gas Phases

Gas Phase (1)	Minimum (cm^{-2}) (2)	Maximum (cm^{-2}) (3)	Median (cm^{-2}) (4)
$N(H\,I)_{thin}$ (58)	4.61×10^{20} 16%	2.97×10^{21} 99%	1.49×10^{21} 62%
$N(H\,I)_{thick}$ (58)	5.94×10^{18} 1%	1.67×10^{21} 38%	3.52×10^{20} 12%
$2N(H_2)_{dark}$ (14)	5.34×10^{20} 14%	2.03×10^{21} 54%	9.60×10^{20} 31%
$2N(H_2)_{bright}$[a] (19)	1.84×10^{20} 16%	6.35×10^{21} 81%	2.85×10^{21} 44%

Note. (1) Gas phase. The number of LOSs that were used for the derivation is indicated in parenthesis. (2) Minimum values of the derived properties. The column density and the fraction with respect to the total hydrogen are presented in the rows. (3) Same as the second column, but for maximum values. (4) Same as the second column, but for median values.
[a] The derived values are upper limits, as CO-dark H_2 was not considered for the derivation.

6. Conditions for the Formation of Molecular Gas: Observational Perspective

6.1. Kinematic Association between H I and CO

To investigate the conditions for the formation of molecular gas, we first compared H I and CO central velocities. For our analysis, we selected CNM and WNM components that are closest to the detected CO(1–0) emission in velocity and calculated absolute velocity differences between H I and CO. The cumulative distribution function (CDF) of these velocity differences is shown in Figure 5.

Figure 5 shows that the velocity difference between the CNM and CO is systematically smaller than that between the WNM and CO. Specifically, the CNM–CO velocity difference ranges from 0.01 km s^{-1} to 4.3 km s^{-1} (median of 0.4 km s^{-1}), while the WNM is offset from CO by 0.04–12.8 km s^{-1} (median of 1.7 km s^{-1}). This difference between the CNM and WNM becomes more significant when additional components are considered (e.g., including the first and second closest
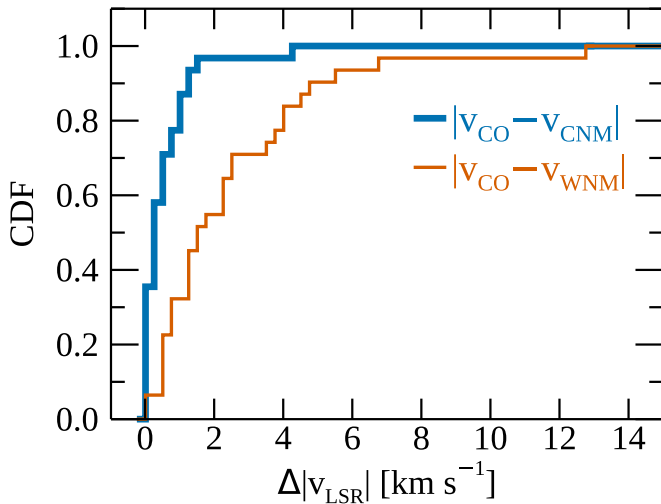
Figure 5. CDF of the absolute velocity difference between H I and CO (CNM and CO in thick blue; WNM and CO in thin tan). For these CDFs, H I components that are closest to the observed CO emission in velocity were considered.

Table 4
Number of H I Components

Classification	CNM	WNM
Individual Properties		
CO non-detection	146	128
CO detection	100	69
Case A	33	13
Case B	25	8
Integrated Properties		
CO non-detection	39	39
CO detection	19	19
Case A	22	22
Case B	22	22

components to CO results in median velocity differences of $1.3\,\mathrm{km\,s^{-1}}$ and $4.7\,\mathrm{km\,s^{-1}}$ for the CNM and WNM), demonstrating that the CNM is kinematically more closely associated with CO emission. If we take velocity as a proxy for position (e.g., CNM components at different velocities would be located in different places), then our result implies that CO-bright molecular gas likely forms in CNM environments.

6.2. Individual H I Properties

Next we examined the properties of individual H I components (T_s, τ_{CNM}, N_{CNM}, and N_{WNM}) in the presence of CO(1–0) emission. For our examination, we classified the observed Gaussian components into four groups: (1) CO non-detection (all components toward the 39 CO non-detected LOSs); (2) CO detection (all components toward the 19 CO-detected LOSs); (3) Case A (components whose central velocities fall between $\nu_{CO} - 2\Delta\nu_{CO}$ and $\nu_{CO} + 2\Delta\nu_{CO}$; subset of CO detection); and (4) Case B (similar to Case A, but H I central velocities are within $\pm\Delta\nu_{CO}$ from ν_{CO}; subset of CO detection and Case A). This classification is motivated to probe the individual H I properties required for the formation of CO-bright molecular gas. In particular, Cases A and B are designed to select H I components that are kinematically closely associated with CO emission with small velocity differences. The number of CNM and WNM components for each group is summarized in Table 4.

For each group, we examined the distributions of spin temperature, optical depth, CNM and WNM column density and present them in Figure 6 and Table 5. In general, we found that the CO non-detection and detection groups are almost indistinguishable in terms of their H I properties. Meanwhile, Cases A and B have several distinctive features compared to the CO non-detection and detection groups. For example, they do not have CNM components with $T_s > 200\,\mathrm{K}$ and show a factor of 2–5 smaller dispersion in T_s compared to the CO non-detection and detection groups. In addition, their minimum $\tau_{CNM} = 0.1$, $N_{CNM} = 2 \times 10^{19}\,\mathrm{cm^{-2}}$, and $N_{WNM} = 2 \times 10^{20}\,\mathrm{cm^{-2}}$ are an order of magnitude higher than those for the CO non-detection and detection groups. These distinctive features of Cases A and B are not pronounced in the comparison between the CO non-detection and detection

groups, mainly because Cases A and B are only a small fraction of the individual H I components (e.g., the Case B CNM and WNM are 25% and 12% of the CO detection CNM and WNM).

All in all, our result implies that CO-bright molecular gas forms in regions where individual CNM components evolve toward colder temperature and higher column density. However, only $\sim$20% of the CNM components with $T_s < 200\,\mathrm{K}$, $\tau_{CNM} > 0.1$, and $N_{CNM} > 2 \times 10^{20}\,\mathrm{cm^{-2}}$ are associated with CO emission (Cases A and B), suggesting that individual CNM components with low temperature and high column density are necessary but not sufficient for the formation of CO-bright molecular gas. This conclusion is consistent with what Rybarczyk et al. (2022) found from H I and HCO$^+$ observations of diffuse Galactic LOSs (see Appendix B for details).

6.3. Integrated H I Properties

Finally, we examined the integrated H I properties required for the formation of CO-bright molecular gas by comparing the four groups in terms of total CNM, WNM, CNM+WNM column densities, and CNM fraction (Figure 7 and Table 5). In contrast to the analysis in Section 6.2, these integrated H I properties were derived by considering all (CO detection and non-detection) or several (Cases A and B) Gaussian components. Specifically, we used Equation (6) and applied the relevant velocity limits for integration (whole LOS for the CO non-detection and detection groups and $\nu_{CO} \pm 2\Delta\nu_{CO}$ and $\nu_{CO} \pm \Delta\nu_{CO}$ for Cases A and B) to calculate the total CNM and WNM column densities. In addition, we defined the CNM fraction $f_{CNM} = N_{CNM}/N(\mathrm{H\,I})$, where $N(\mathrm{H\,I})$ is given by Equation (6).

Figure 7 shows that the CO detection group generally has slightly higher H I column densities than the CO non-detection group. For example, the median total CNM, WNM, CNM+WNM column densities of the CO detection group are a factor of 1.2–1.5 higher than those of the CO non-detection group. This difference in the integrated column densities is in contrast with the almost identical distributions of the individual H I properties for the two groups (Section 6.2) and implies that the total amount of gas along a LOS (and consequently associated dust extinction) could be one of the important factors for the formation of CO-bright molecular gas. An examination of the CO peak brightness temperature as a function of A_V (Figure 8) indeed reveals that CO emission is detected primarily toward LOSs with $A_V \gtrsim 0.5$–1 mag, which is
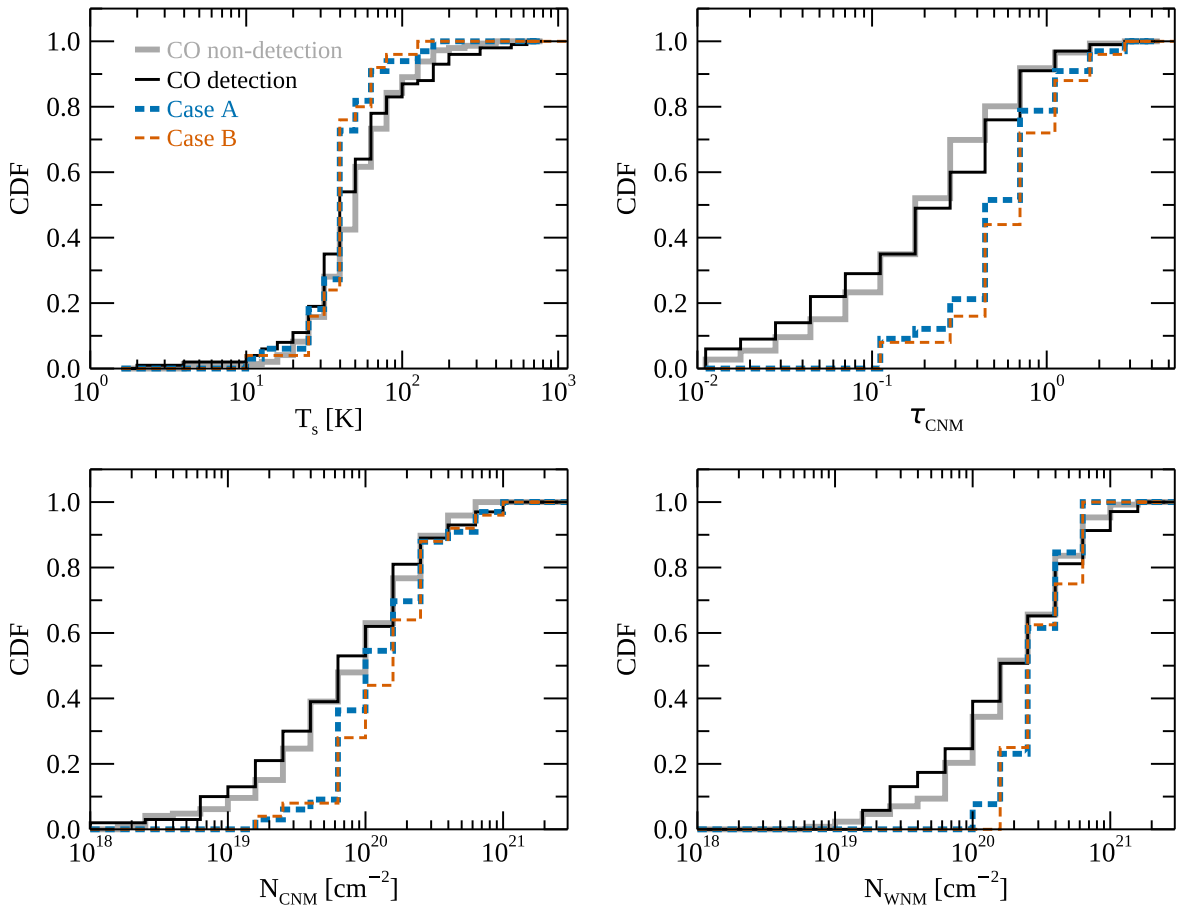
Figure 6. CDFs of the spin temperature, optical depth, CNM and WNM column density. In each panel, the four groups are shown in different colors and line styles: CO non-detection (thick solid gray), CO detection (thin solid black), Case A (thick dashed blue), and Case B (thin dashed tan).

comparable to the threshold dust extinction for CO formation in the solar neighborhood (e.g., Pineda et al. 2008; Lee et al. 2014, 2018).

Another interesting finding is that Cases A and B have systematically higher CNM fractions than the other groups (e.g., median CNM fraction of 0.4 for the CO non-detection and detection groups and 0.6 for Cases A and B). These higher CNM fractions could result from two cases: (1) an increase in the column density of individual CNM components; and (2) an increase in the relative number of CNM components. As for the first case, Figure 6 and Table 5 confirm that the column density of individual H I components increases toward CO more in the CNM than in the WNM. For example, from the CO detection to Case B, the median CNM and WNM column densities increase by a factor of 2.2 and 1.5, respectively. To evaluate the second case, we then estimated the CNM component density ($f_{\#CNM}$) by dividing the number of CNM components by the number of total H I components and compared its distribution between the four groups (Figure 7 and Table 5). Our analysis shows that the CNM component density is indeed systematically higher for Cases A and B than for the CO non-detection and detection groups, which is in line with our previous finding of the CNM being kinematically more closely associated with CO emission (Section 6.1). Based on these results, we conclude that an increase in both the individual CNM column density and the relative number of CNM components could contribute to the higher CNM fraction toward CO.

In summary, our comparison between H I and CO suggests that the formation of CO-bright molecular gas is favored in high column density environments that are able to provide significant shielding against dissociating UV radiation. In these environments, the CNM becomes colder (lower temperature) and more abundant (higher density), facilitating H_2 and consequently CO formation. We will discuss the conditions for the formation of molecular gas in Section 8.1.

7. Conditions for the Formation of Molecular Gas: Theoretical Perspective

In this section, we compare the observed CNM properties to the prediction from the S14 model with the aim of investigating the fundamental principles of the H I-to-H_2 transition. Specifically, our approach is to estimate the density expected for H_2 formation from S14 (n^{exp}) and to confront it with the CNM density inferred from our observations (n^{obs}_{CNM}). As for the theoretically expected density, we recall that the total H I column density of a plane-parallel slab of gas and dust in the S14 model is controlled by the dimensionless parameter αG (Equation (1)). Since αG is a function of I_{UV} and n (Equation (4); $\tilde{\sigma}_g \sim 1$ for our case of the solar neighborhood conditions), n^{exp} can be expressed as

$$n^{exp}\ (cm^{-3}) = \frac{18.4\ I_{UV}}{\exp\left(N_{CNM}/8.4 \times 10^{20}\ cm^{-2}\right) - 1}, \quad (14)$$

Table 5
Physical Properties of H I Components

Properties (1)	CO Non-detection (2)	CO Detection (3)	Case A (4)	Case B (5)
		Individual Properties		
T_s (K)	4.11−479.28	1.99−725.42	10.82−188.27	10.82−130.63
	53.83	48.13	46.21	46.21
τ_{CNM}	0.01−3.14	0.01−3.41	0.11−3.41	0.11−3.41
	0.27	0.29	0.68	0.78
N_{CNM} (10^{20} cm^{-2})	0.02−8.59	0.01−13.70	0.17−10.02	0.17−10.02
	1.07	0.87	1.50	1.91
N_{WNM} (10^{20} cm^{-2})	0.10−18.57	0.17−16.70	1.23−7.33	1.86−7.33
	2.22	2.49	3.52	3.79
		Integrated Properties		
N_{CNM} (10^{20} cm^{-2})	0.58−19.90	1.97−22.45	0.17−9.93	0.01−8.31
	5.48	7.99	2.82	1.69
N_{WNM} (10^{20} cm^{-2})	1.67−27.48	1.85−25.54	0.03−5.96	0.01−4.44
	9.58	11.90	2.41	1.29
N(H I) (10^{20} cm^{-2})	4.70−38.57	7.33−38.08	0.81−14.81	0.38−11.41
	16.99	24.21	5.51	2.96
f_{CNM}	0.12−0.69	0.19−0.77	0.05−0.99	0.01−0.99
	0.36	0.37	0.56	0.57
$f_{\#CNM}$	0.20−0.80	0.38−0.82	0.00−1.00[a]	0.00−1.00[a]
	0.56	0.57	0.75	1.00

Note. (1) Physical properties. Each row displays the range of values for each physical property, with the median value indicated below the range. (2) H I components toward the 39 CO non-detected LOSs. (3) H I components toward the 19 CO-detected LOSs. (4) H I components whose central velocities fall between $\nu_{CO} - 2\Delta\nu_{CO}$ and $\nu_{CO} + 2\Delta\nu_{CO}$. (5) H I components whose central velocities are in the range of $\nu_{CO} \pm \Delta\nu_{CO}$.
[a] There are two (Case A) and three (Case B) CO peaks where there is no associated CNM component. $f_{\#CNM}$ is set to zero accordingly, and these cases are still considered for the calculation of the median values.

where N(H I) is substituted with N_{CNM}. This substitution is motivated by the fact that our Planck-based I_{UV} estimates are mostly ~ 1 (Section 4.2). The nearly uniform I_{UV} values suggest isotropic UV radiation that is most likely attenuated by the widespread WNM. The impact of the WNM on the H I-to-H$_2$ transition is already taken into account in this manner, and we therefore proceeded by replacing N(H I) with N_{CNM}. As for the observationally inferred density, we took the thermal pressure $\log_{10}(P/k_B \text{ cm}^{-3} \text{K}) = 3.58 \pm 0.18$ from Jenkins & Tripp (2011) (estimated for the CNM based on Hubble Space Telescope observations of C I multiplets at UV wavelengths) and calculated n_{CNM}^{obs} by

$$n_{CNM}^{obs} \text{ (cm}^{-3}) = \left(\frac{10^{3.58 \pm 0.18}}{T_s} \right). \quad (15)$$

Since $n^{exp} = n_1 + 2n_2$ is the total number density, n^{exp} should be higher than n_{CNM}^{obs} for CO-detected LOSs.

7.1. Density versus CNM Column Density

We estimated n^{exp} and n_{CNM}^{obs} for the following three cases: (1) Entire LOS; (2) Case A; and (3) Case B. For each LOS, all CNM components are considered for the Entire LOS, while CNM components within $\nu_{CO} \pm 2\Delta\nu_{CO}$ and $\nu_{CO} \pm \Delta\nu_{CO}$ are examined for Cases A and B. We assessed these three cases mainly because of a lack of knowledge on the geometry of the CNM. For example, the Entire LOS would correspond to a case where all CNM components at different velocities belong to one large structure and absorb dissociating UV photons

(Figure 9(a)). Meanwhile, Cases A and B would be equivalent to a case where only CNM components near CO clumps provide shielding against UV photons (Figure 9(b)). While being simple pictures, these two scenarios cover small (Cases A and B) to large (Entire LOS) volume filling factors for the CNM. Finally, we used the Planck-based I_{UV} values for Equation (14) and the opacity-weighted mean spin temperature ($T_{s,\tau}$) for Equation (15):

$$T_{s,\tau} \text{ (K)} = \frac{\sum_{n=0}^{N-1} \tau_{0,n} T_{s,n}}{\sum_{n=0}^{N-1} \tau_{0,n}}. \quad (16)$$

The derived n^{exp} and n_{CNM}^{obs} values for the three cases are presented as a function of N_{CNM} in Figure 10.

We found that the three cases show similar trends. For example, the total densities expected from S14 are higher than the inferred CNM densities at low column densities ($N_{CNM} \lesssim 10^{20}$ cm^{-2}). In other words, the model is in agreement with the lower limits on the total densities constrained by our observations. On the contrary, the S14-based total densities are lower than the inferred CNM densities at high column densities ($N_{CNM} \gtrsim 10^{20}$ cm^{-2}), resulting in a discrepancy between the model and our observations. This discrepancy becomes more significant from Case B to Case A to the Entire LOS case and reaches up to one or two orders of magnitude at the highest column density of $\sim 10^{21}$ cm^{-2}.

7.2. Limitations and Implications

Taken at face value, the discrepancy at high CNM column densities implies that only a small fraction of the total CNM along a LOS participates in H$_2$ formation ($\lesssim 13\%$ on average;
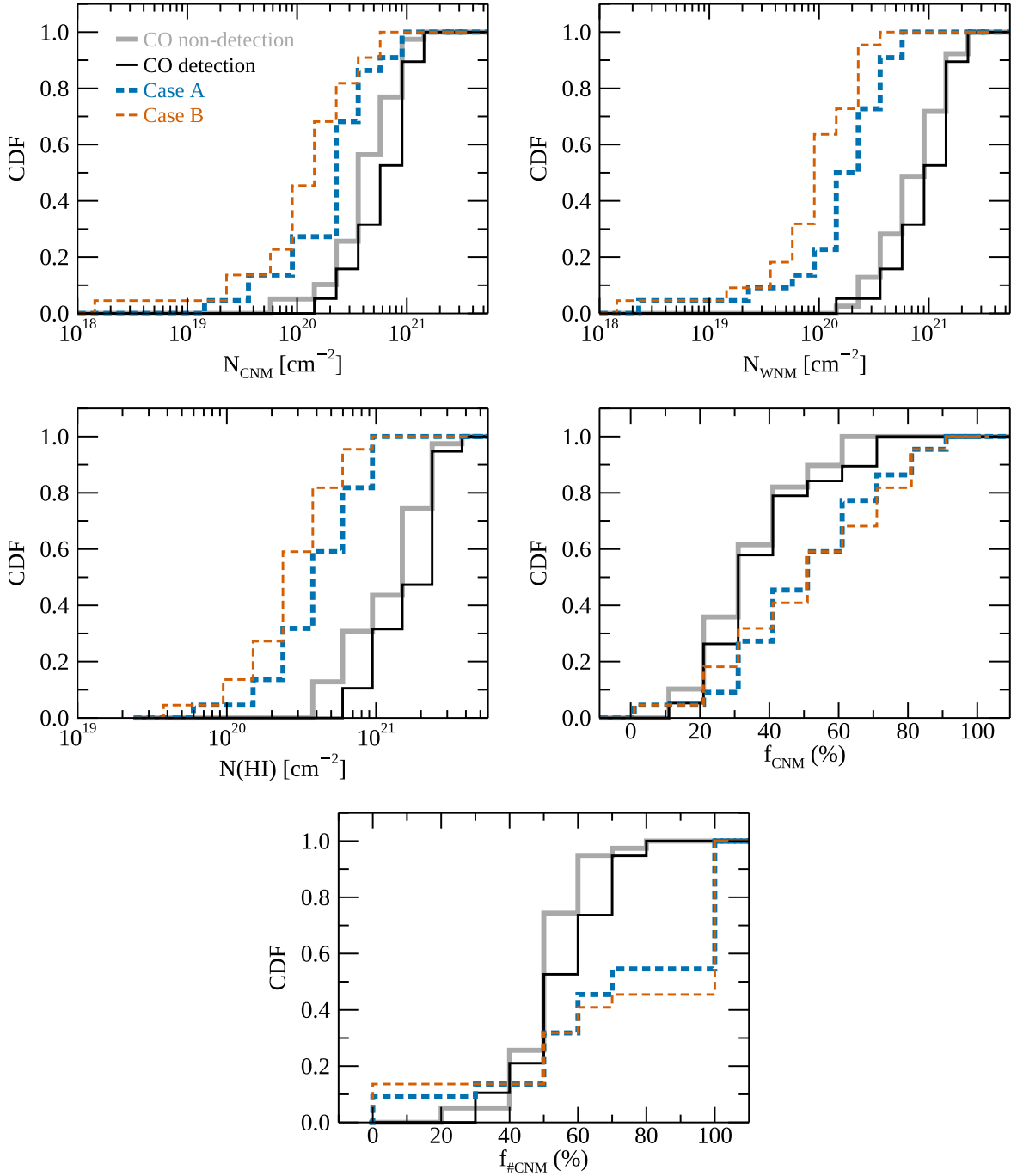
Figure 7. CDFs of the CNM, WNM, CNM+WNM column density, CNM fraction, and CNM component density. As in Figure 6, the four groups are shown in different colors and line styles: CO non-detection (thick solid gray), CO detection (thin solid black), Case A (thick dashed blue), and Case B (thin dashed tan). The column densities of Cases A and B are lower than those of the CO non-detection and detection by design (integrated over smaller velocity ranges) and are shown here only for the sake of completeness.

this fraction is estimated from the LOSs with large discrepancies at $N_{CNM} > 10^{20}\,cm^{-2}$ for the Entire LOS case). In other words, the CNM must be clumpy with a small volume filling factor. Although this is a reasonable interpretation, our analysis is not without limitations. In this section we discuss other sources of the discrepancy and their implications.

One possible source of the observed discrepancy is the assumed thermal pressure of 2500–5800 cm^{-3} K for the CNM. While we considered this factor of two variation in the thermal pressure, Goldsmith et al. (2018) recently found a larger

variation ($\sim 10^3$–$10^4\,cm^{-3}$ K) from SOFIA [C II] 158 μm observations of Galactic LOSs (3C131, one of our CO-detected LOSs, was found to have $P/k_B \sim$2500–3200 cm^{-3} K). If the thermal pressure of the CNM varies by an order of magnitude as the SOFIA observations suggest, then the discrepancy between the S14 prediction and our observations would certainly decrease, but it is likely that the discrepancy would still persist at the highest column density of $\sim 10^{21}\,cm^{-2}$.

Another possible source of the discrepancy is cosmic-rays. Cosmic-rays ionize atoms and molecules in collisions and can
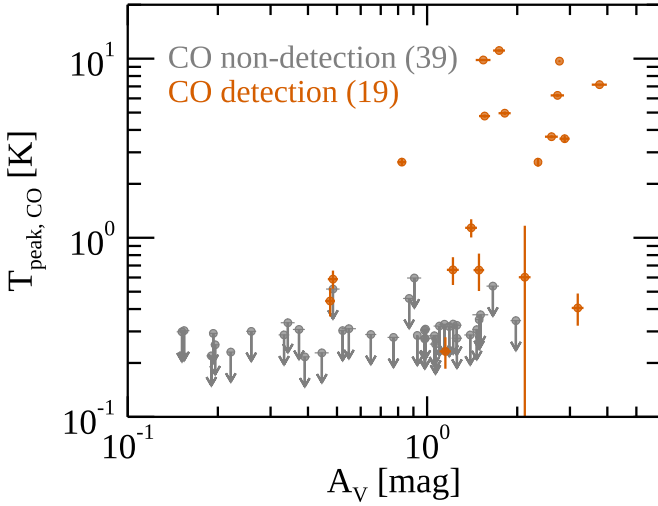
Figure 8. CO peak brightness temperature as a function of Planck-based A_V. The 19 CO-detected LOSs are shown in tan, while upper limits based on 3σ values are indicated as downward arrows for the 39 CO non-detected LOSs.
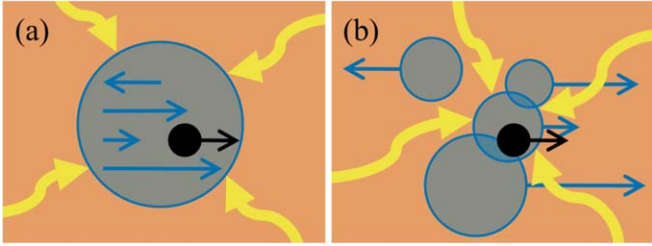

Figure 9. Possible distributions of the WNM, CNM, and CO (tan, blue, and black, respectively). Based on our finding of the kinematic association between H I and CO (Section 6.1), the WNM is represented as the diffuse background, while the CNM and CO are shown as the smaller embedded structures. The velocities of the CNM and CO are indicated as blue and black arrows with arbitrary sizes, and UV radiation is presented as yellow arrows. (Left) Scenario for the Entire LOS, where CNM components at different velocities belong to one large structure and contribute to the shielding of the CO core. (Right) Scenario for Cases A and B, where CNM components close to CO only provide shielding against UV radiation.

destruct H_2, as follows (e.g., Sternberg et al. 2021):

$$H_2 + CR \longrightarrow H_2^+ + e, \qquad (17)$$

$$H_2^+ + H_2 \longrightarrow H_3^+ + H. \qquad (18)$$
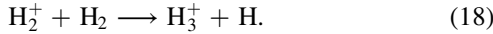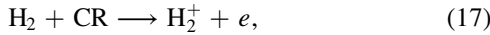
In addition, cosmic-rays can directly dissociate H_2:

$$H_2 + CR \longrightarrow H + H. \qquad (19)$$
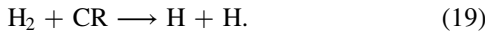
A preliminary study of the impact of cosmic-rays on the H I-to-H_2 transition suggests that a combination of UV photons and cosmic-rays could increase the total H I column density by up to a factor of 10 compared to the case with UV photons only (when examined over a reasonable parameter space with the density $n = 10^1$–$10^3 \, cm^{-3}$, total column density $N(H) = (1-4) \times 10^{21} \, cm^{-2}$, UV radiation field $I_{UV} = 1$, and cosmic-ray ionization rate $\zeta = (0.5-2) \times 10^{-16} \, s^{-1}$; A. Sternberg & S. Bialy 2023, in preparation). Interestingly, considering a realistic density increase by a factor of 10 from the envelope to the core of a cloud reduces the impact of cosmic-rays significantly, making it almost negligible at the envelope density of $\sim 10^2 \, cm^{-3}$ (comparable to the CNM densities inferred from our observations). These results suggest that
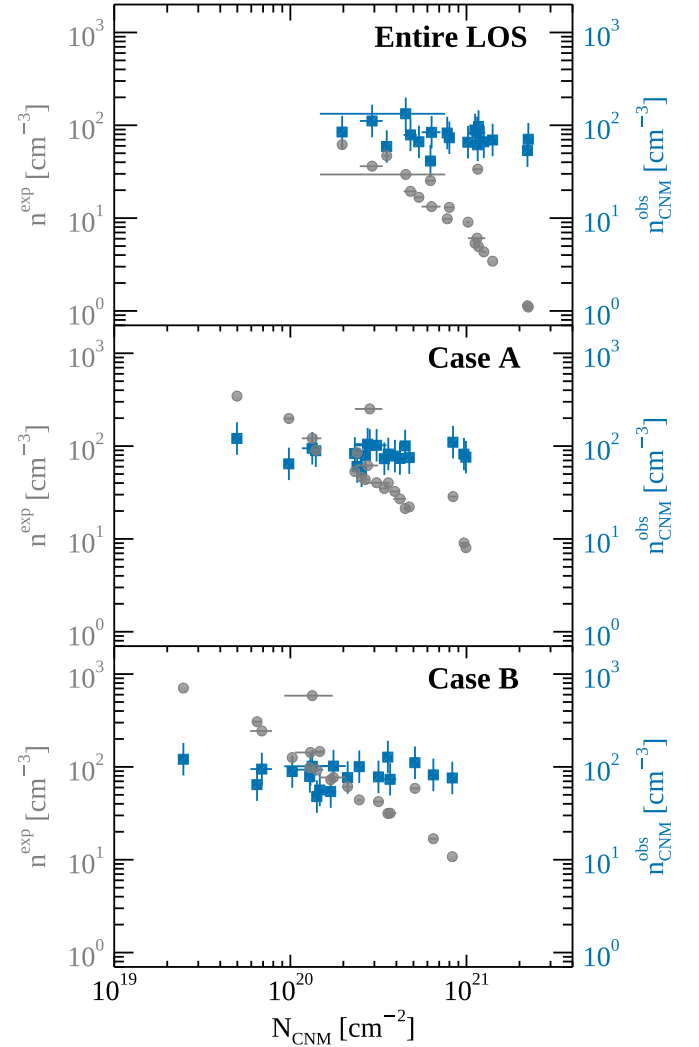
Figure 10. Comparison between the S14-based total number densities (n^{exp}; gray circles) and the observationally inferred CNM densities (n_{CNM}^{obs}; blue squares) as a function of the CNM column density for the 19 CO-detected LOSs. Since the predicted n^{exp} corresponds to the total gas number density ($n_1 + 2n_2$) at which H_2 formation is expected to occur, it should be higher than n_{CNM}^{obs}. Finally, the observed variation in the thermal pressure, $\log_{10}(P/k_B \, cm^{-3} \, K) = 3.58 \pm 0.18$, is indicated as the 1σ error bars for n_{CNM}^{obs}. (Top) Entire LOS. (Middle) Case A. (Bottom) Case B.

detailed studies are needed to properly evaluate the impact of cosmic-rays on H_2 formation.

Finally, the steady-state approximation in S14 could be invalid. A wide range of dynamical processes operate in the ISM, producing continuous flows of gas. For example, dense molecular clouds can undergo gravitational collapse on the freefall timescale $t_{ff} \sim 1 \, Myr \, (n/10^4 \, cm^{-3})^{-1/2}$. Similarly, interstellar turbulence dissipates its kinetic energy on the eddy turnover timescale $t_{turb} \sim 1 \, Myr \, (L/pc)^{1/2}$, where L is the eddy size (e.g., Chevance et al. 2023). For the CNM with $n \sim 10^2 \, cm^{-3}$ distributed within $\sim 100 \, pc$ scale molecular clouds, t_{ff} and t_{turb} are approximately 10 Myr, which are comparable to the H_2 formation timescale $t_{H_2} \sim (10^9/n) \, yr$ (e.g., Hollenbach et al. 1971). These rough estimates illustrate that the CNM could be heavily perturbed over time, making the steady-state approximation for H_2 formation inappropriate (e.g., Valdivia et al. 2016; Bialy et al. 2021).

In summary, we conclude that the CNM must be clumpy with a small volume filling factor if H_2 formation in the solar neighborhood is determined by a combination of UV radiation, gas density, and metallicity (αG), as the simple steady-state S14 model predicts. Otherwise, missing elements in the S14 model, such as cosmic-rays and dynamical processes, could play an important role and need to be considered more comprehensively for a better understanding of H_2 formation.

8. Discussion

8.1. Conditions for the Formation of Molecular Gas

In Section 6.1, we concluded that CO-bright molecular gas likely forms in CNM environments based on a close association between the CNM and CO in velocity. This conclusion is consistent with Savage et al. (1977), who measured a H_2 kinetic temperature of 45–128 K with a median of 77 K (comparable to our median spin temperature of ∼50 K) for the medium within 1 kpc of the Sun by analyzing Copernicus UV absorption observations. Similarly, Bellomi et al. (2020) compared UV measurements of the H I-to-H_2 transition to a suite of magnetohydrodynamic simulations and claimed that H_2 within 2 kpc of the Sun is built up in CNM structures with a size of ∼3–10 pc.

In Sections 6.2 and 6.3, we then went further and showed that the formation of CO-bright molecular gas is favored in high column density environments where the CNM becomes colder and more abundant (which is as expected). As for the conditions for more abundant CNM, Saury et al. (2014) examined a large set of hydrodynamic simulations and found that the fraction of the CNM increases with increasing initial density and decreasing turbulent velocity, implying that high densities ($\gtrsim 2\,\mathrm{cm}^{-3}$) along with a moderate level of gas compression are required for the formation of the CNM, and consequently molecular gas.

Finally, our finding of the minimum dust extinction $A_V \gtrsim 0.5$–1 mag for CO detection (Section 6.3) implies the importance of the total amount of gas available for the formation of CO-bright molecular gas. All things considered, we conclude that accumulating a large amount of atomic gas by dynamical processes (e.g., spiral arms, supernova explosions, and expanding shells that lead to gas compression) and building up cold and dense structures would be a key step in the formation of molecular gas. This conclusion is consistent with what previous observational and theoretical studies suggested (e.g., McKee & Ostriker 2007; Chevance et al. 2023).

8.2. H I Absorption as a Diagnostic Tool for Probing the Formation and Evolution of Molecular Clouds

While a range of dynamical processes (e.g., spiral arms on large scales and expanding shells and bubbles on small scales) certainly play a role in the formation and evolution of molecular clouds (e.g., McKee & Ostriker 2007), it remains unclear exactly how they operate and which process dominates. As an accessible tracer of atomic gas, the raw ingredient of molecular clouds, H I emission has been frequently employed to address this issue. For example, Fukui et al. (2009) found that the H I mass of molecular clouds in the Large Magellanic Cloud (LMC) increases with evolutionary stages of star formation and estimated an H I accretion rate of 0.05 $M_\odot\,\mathrm{yr}^{-1}$ based on H I line widths. In addition, Tahani et al. (2022)

examined the difference in velocity between H I and CO emission for the Perseus molecular cloud and interpreted a systematic positive offset of $\nu_{CO} - \nu_{H I}$ as an indication of the formation of molecular gas behind compressed H I bubbles. In comparison to H I emission that traces all three phases of neutral atomic gas and often exhibits broad and featureless spectra, H I absorption mostly arises from the CNM (which is more closely associated with molecular gas) and shows relatively narrow and structured spectra, making it an excellent probe for the formation and evolution of molecular clouds. As an example, we showed that there is an absolute velocity difference of 0.01–4.3 km s^{-1} with a median of 0.4 km s^{-1} between the CNM and CO (Section 6.1). Considering that the CNM also has a comparable velocity difference of 0.06–2.64 km s^{-1} with a median of 0.5 km s^{-1} with OH absorption (estimated from 10 of our 58 GNOMES LOSs where OH absorption is clearly detected; Petzler et al. 2023), this velocity difference between the CNM and CO is most likely real (not due to different beam sizes) and could suggest that the CNM and CO-bright molecular gas are in slightly different regions and/or dynamically decoupled (e.g., Soler et al. 2019; Beuther et al. 2020; Wang et al. 2020). Unfortunately, our GNOMES LOSs are scattered over a relatively large area of sky and cannot provide insights into how the CNM is distributed and moves about in individual molecular clouds.

The power of H I absorption as a diagnostic tool for probing the formation and evolution of molecular clouds could be harnessed by getting a large number of H I absorption spectra over a fine grid of continuum sources located behind molecular clouds. These spectra could then be analyzed with synthetic H I data from numerical simulations of multiphase gas (e.g., Kim & Ostriker 2017; Seifried et al. 2022), enabling us to examine the signature of the formation and evolution process imprinted on the properties of the CNM (e.g., kinematics and distribution). Such observations as we propose will be routinely carried out by next generation radio telescopes with a wide field of view, including the Square Kilometre Array (SKA), as Dickey et al. (2022) recently demonstrated with the Australian Square Kilometre Array Pathfinder (ASKAP).

9. Summary

This paper presents a detailed study on the formation of molecular gas in the solar neighborhood. To probe the conditions for the H I-to-H_2 transition, H I emission and absorption spectra toward 58 LOSs at $b < -5°$ (Arecibo) were analyzed along with CO(1–0) and dust data (TRAO, PMO, and Planck). These multi-wavelength data were also compared to the one-dimensional steady-state H_2 formation model of Sternberg et al. (2014) to provide insights into the fundamental principles of the H I-to-H_2 transition. Our key results are as follows.

1. Among the observed 58 sources, 19 sources show clear CO(1–0) emission, suggesting a detection rate of 33% at a rms level of 0.1 K (angular and spectral resolutions of 48″ and 0.32 km s^{-1}, respectively).

2. The decomposition of gas into atomic and molecular phases shows that the observed LOSs are mostly H I-dominated. In addition, the CO-dark H_2, not the optically thick H I, is found as a major constituent of the dark gas in the solar neighborhood.

3. The CNM shows a systematically smaller velocity difference from CO emission than the WNM. When CO-closest components are considered, a median value of the absolute velocity difference between the CNM and CO is 0.4 km s^{-1}, as opposed to 1.7 km s^{-1} for the WNM and CO. This implies that the CNM is kinematically (and spatially if we take velocity as a proxy for position) more closely associated with CO.

4. When CO-associated components (ones within CO velocity ranges) are considered, the CNM and WNM exhibit distinctive properties. Namely, the CO-associated components have the spin temperature $T_s < 200$ K, optical depth $\tau_{CNM} > 0.1$, and column densities $N_{CNM} > 2 \times 10^{19}$ cm^{-2} and $N_{WNM} > 2 \times 10^{20}$ cm^{-2}. This suggests that CO-bright molecular gas forms in environments where individual CNM components evolve toward colder temperature and higher column density.

5. The CO-associated components have higher total column densities (V-band dust extinction $A_V \gtrsim 0.5$ mag) and CNM fractions (median of 0.6) than those outside CO emission, indicating that high column density environments where the CNM becomes more abundant facilitate the formation of CO-bright molecular gas.

6. A comparison with the prediction from Sternberg et al. (2014) infers that the CNM must be clumpy with a small volume filling factor. An alternative possibility would be that missing ingredients in the model, such as cosmic-rays and dynamical processes, play an important role in the H I-to-H$_2$ transition in the solar neighborhood.

Appendix A
Comparison between the GNOMES H I and CO Spectra

In Figure A1, we compare the H I absorption and CO emission spectra toward the 19 CO-detected LOSs. The measured optical depth (τ_{CNM}) and main-beam brightness temperature ($T_{MB,CO}$) are aligned at peaks for ease of comparison and are presented in blue and gray, respectively.
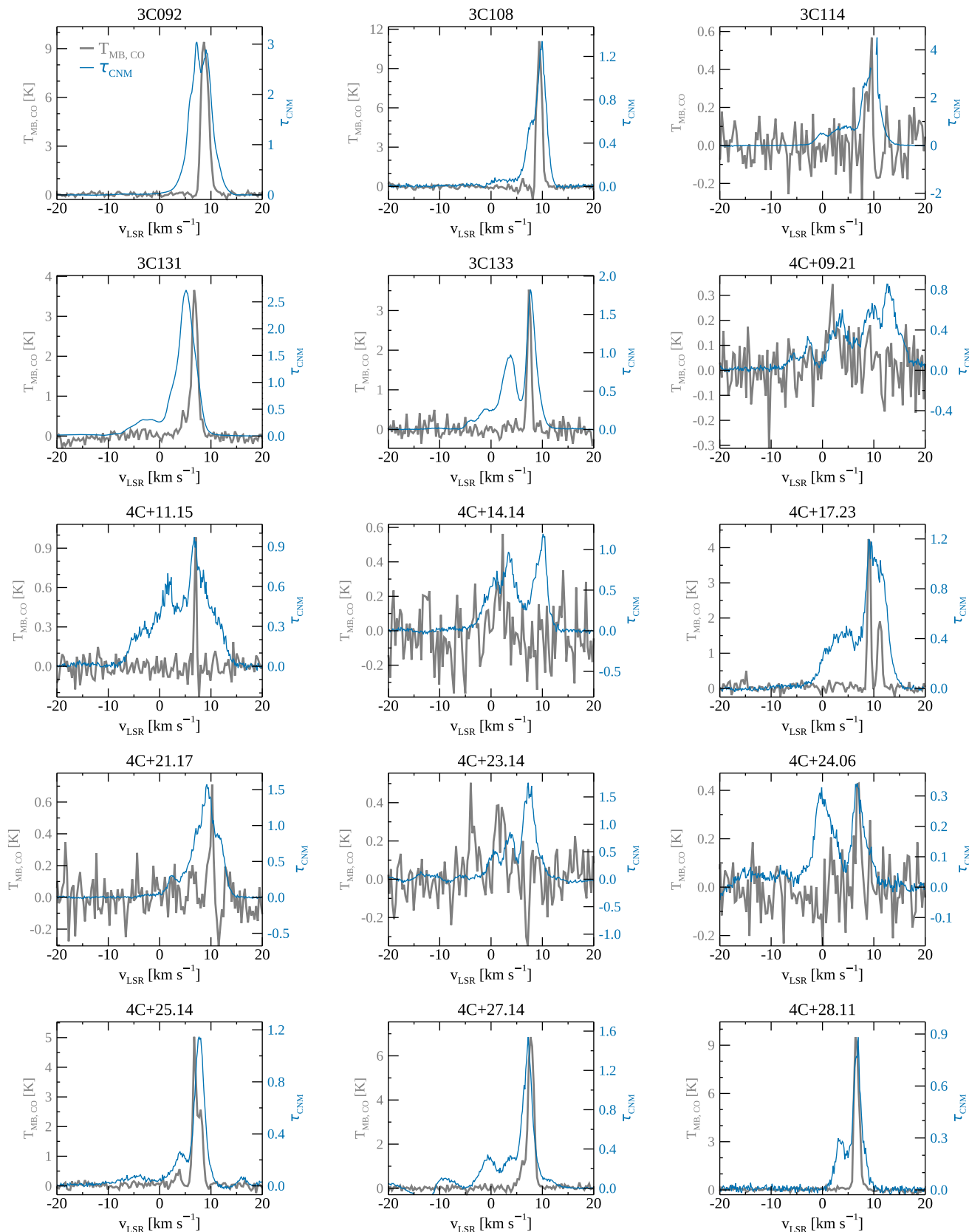
Figure A1. Comparison between the H I absorption (thin blue) and CO emission (thick gray) spectra toward the 19 CO-detected LOSs. The spectra are aligned at peaks for ease of comparison.
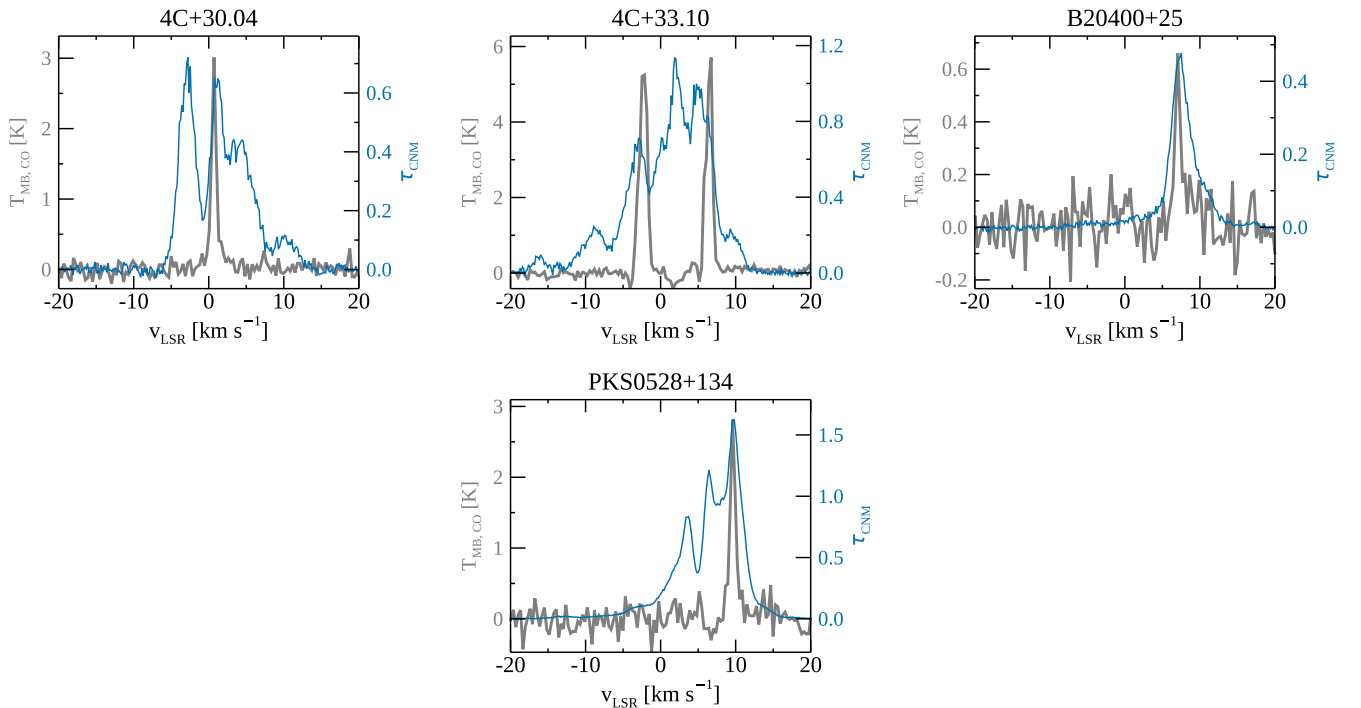
Figure A1. (Continued.)

Appendix B
Comparison to the ALMA-NOEMA/21-SPONGE Survey

Recently, Rybarczyk et al. (2022) observed 20 LOSs from the 21-SPONGE survey (21 cm Spectral Line Observations of Neutral Gas with the Karl G. Jansky Very Large Array; Murray et al. 2015, 2018a) using the Atacama Large Millimeter/submillimeter Array (ALMA) and the Northern Extended Millimeter Array (NOEMA) and obtained HCO^+, HCN, HNC, and C_2H absorption spectra. By comparing the observed molecular species with the existing H I and dust properties, the authors found that molecular absorption is clearly detected toward LOSs with $A_V \gtrsim 0.25$ mag. In addition, they revealed that molecular gas is preferentially associated with cold ($T_s < 80$ K) and optically thick ($\tau_{CNM} > 0.1$) CNM structures.

While both the GNOMES and ALMA-NOEMA/21-SPONGE surveys attempt to probe the conditions for the formation of molecular gas, our approach is uniquely different. First, unlike the ALMA-NOEMA/21-SPONGE survey targeting random LOSs throughout the Milky Way, we focus on individual molecular clouds and their surrounding environments. Second, we employ the most commonly used tracer of molecular gas, CO(1–0) emission, for our analyses. To examine the difference between the two surveys, we selected 11 LOSs at $b < -5°$ from the ALMA-NOEMA/21-SPONGE survey, and extracted 64 CNM and 67 WNM components. In addition, we identified H I components that are kinematically closest (minimum absolute velocity difference) to the observed CO emission (GNOMES; 19 LOSs) and HCO^+ absorption (ALMA-NOEMA/21-SPONGE; eight LOSs), and compare their properties in Figure B1 and Table B1.

Figure B1 shows that the CNM and WNM at $b < -5°$ from the two surveys have in general systematically different properties, i.e., the GNOMES components have lower spin temperatures, higher optical depths, and higher column densities (black and gray CDFs). This difference persists for T_s and N_{WNM}, when we focus on H I components that are closely associated with molecular gas (tan and pink CDFs). One of the likely reasons for this difference is that the GNOMES survey probes higher column density environments by concentrating on molecular clouds and their surroundings. For example, the CO-detected GNOMES LOSs have a median A_V of 2 mag, which is a factor of two higher than that for the HCO^+-detected 21-SPONGE LOSs. Another possibility is that CO and HCO^+ formation requires different conditions. In summary, bearing in mind the small number of LOSs in the ALMA-NOEMA/21-SPONGE survey, we conclude that the two surveys sample slightly different populations of the CNM and WNM, while showing consistent results regarding the evolution of H I properties toward molecular gas. The two surveys are thus highly complementary to each other.
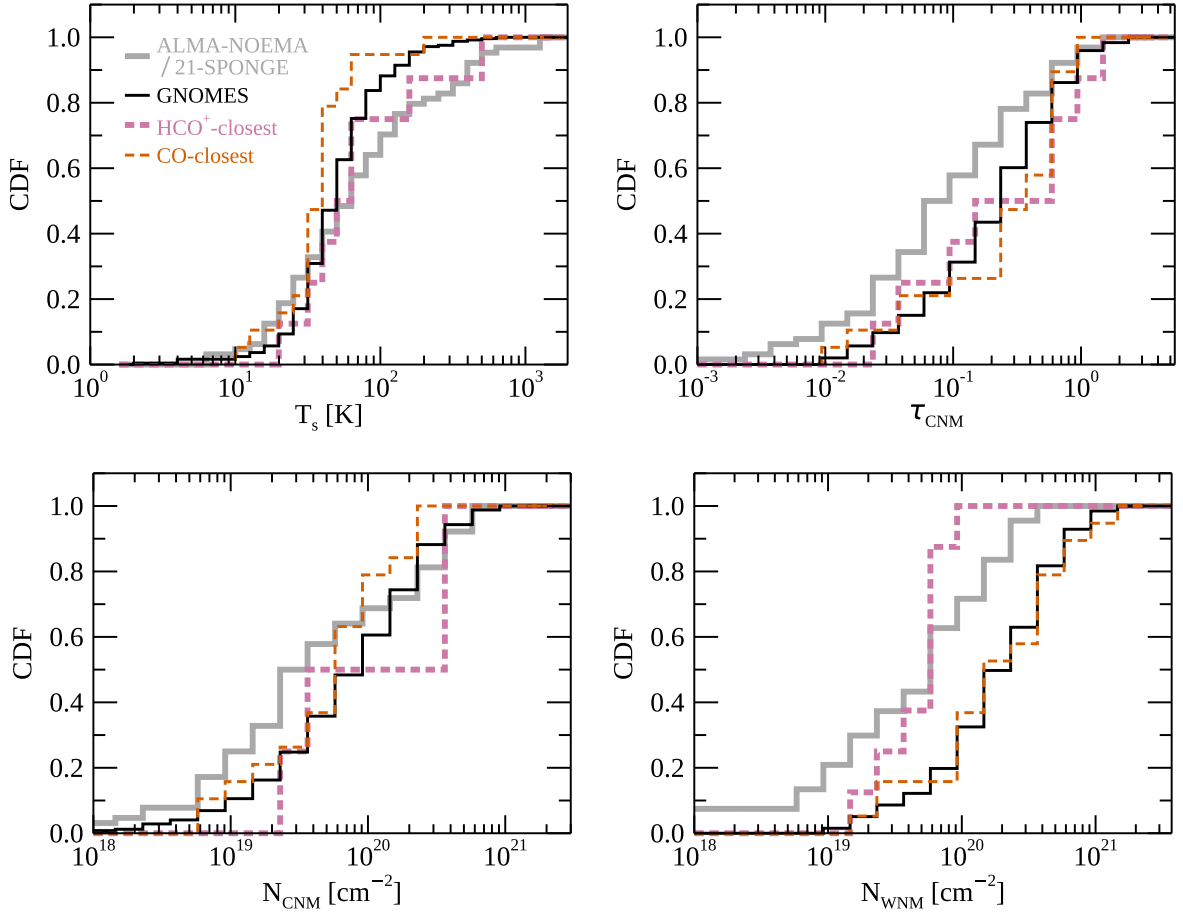
Figure B1. CDFs of the spin temperature, optical depth, CNM and WNM column density. All H I components toward LOSs at $b < -5°$ from the GNOMES and ALMA-NOEMA/21-SPONGE surveys are shown in thin black and thick gray (solid lines). Among these components, those in close proximity to CO and HCO$^+$ in velocity are selected and presented in thin tan and thick pink (dashed lines).

Table B1
Physical Properties of Individual H I Components

(1)	Properties	T_s (K)	τ_{CNM}	N_{CNM} (10^{20} cm^{-2})	N_{WNM} (10^{20} cm^{-2})
			GNOMES		
(2)	All	1.99−725.42 52.28	0.01−3.41 0.27	0.01−13.70 1.00	0.10−18.57 2.32
(3)	CO-closest	10.82−228.91 40.11	0.01−1.38 0.45	0.07−2.94 0.81	0.20−16.11 2.13
			ALMA-NOEMA/21-SPONGE		
(4)	All	7.19−1551.67 65.47	9.37E-4−1.68 0.10	3.63E-3−8.86 0.37	2.32E-18−5.37 0.75
(5)	HCO$^+$-closest	20.46−619.19 65.07	0.02−1.65 0.62	0.28−5.36 3.88	0.16−0.95 0.70

Note. (1) Physical properties. The ranges are given, and the median values are provided below the ranges. (2) H I components toward the 58 GNOMES LOSs. (3) H I components that are closest to the measured 19 $T_{peak,CO}$. (4) H I components toward the 11 ALMA-NOEMA/21-SPONGE LOSs. (5) H I components that are closest to the observed 8 τ_{peak,HCO^+}.

ORCID iDs

Gyueun Park ⓘ https://orcid.org/0000-0002-7374-7864
Min-Young Lee ⓘ https://orcid.org/0000-0002-9888-0784
Shmuel Bialy ⓘ https://orcid.org/0000-0002-0404-003X
Blakesley Burkhart ⓘ https://orcid.org/0000-0001-5817-5944
J. R. Dawson ⓘ https://orcid.org/0000-0003-0235-3347
Carl Heiles ⓘ https://orcid.org/0000-0002-7456-8067
Di Li ⓘ https://orcid.org/0000-0003-3010-7661
Claire Murray ⓘ https://orcid.org/0000-0002-7743-8129
Hiep Nguyen ⓘ https://orcid.org/0000-0002-2712-4156
Anita Hafner ⓘ https://orcid.org/0000-0001-6179-0606
Daniel R. Rybarczyk ⓘ https://orcid.org/0000-0003-3351-6831
Snežana Stanimirović ⓘ https://orcid.org/0000-0002-3418-7817

References

Barriault, L., Joncas, G., Lockman, F. J., & Martin, P. G. 2010, MNRAS, 407, 2645
Bellomi, E., Godard, B., Hennebelle, P., et al. 2020, A&A, 643, A36
Beuther, H., Wang, Y., Soler, J., et al. 2020, A&A, 638, A44
Bialy, S. 2020, ApJ, 903, 62
Bialy, S., & Sternberg, A. 2016, ApJ, 822, 83
Bialy, S., & Sternberg, A. 2019, ApJ, 881, 160
Bialy, S., Zucker, C., Goodman, A., et al. 2021, ApJL, 919, L5
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Boulanger, F., Abergel, A., Bernard, J. P., et al. 1996, A&A, 312, 256
Browning, M. K., Tumlinson, J., & Shull, J. M. 2003, ApJ, 582, 810
Chapman, N. L., & Mundy, L. G. 2009, ApJ, 699, 1866
Chevance, M., Krumholz, M. R., McLeod, A. F., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Initsuka et al. (San Francisco, CA: ASP)
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Dickey, J. M., Dempsey, J. M., Pingel, N. M., et al. 2022, ApJ, 926, 186
Douglas, K. A., & Taylor, A. R. 2007, ApJ, 659, 426
Draine, B. T. 1978, ApJS, 36, 595
Draine, B. T., & Bertoldi, F. 1996, ApJ, 468, 269
Fukui, Y., Kawamura, A., Wong, T., et al. 2009, ApJ, 705, 144
Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273
Goldsmith, P. F., Pineda, J. L., Neufeld, D. A., et al. 2018, ApJ, 856, 96
Green, G. 2018, JOSS, 3, 695
Grenier, I. A., Casandjian, J.-M., & Terrier, R. 2005, Sci, 307, 1292
Heiles, C., & Troland, T. H. 2003a, ApJS, 145, 329
Heiles, C., & Troland, T. H. 2003b, ApJ, 586, 1067
Hollenbach, D. J., & Tielens, A. G. G. M. 1997, ARA&A, 35, 179
Hollenbach, D. J., Werner, M. W., & Salpeter, E. E. 1971, ApJ, 163, 165
Jenkins, E. B., & Tripp, T. M. 2011, ApJ, 734, 65
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kim, C.-G., & Ostriker, E. C. 2017, ApJ, 846, 133
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216

Lada, C. J., Lombardi, M., & Alves, J. F. 2009, ApJ, 703, 52
Langer, W. D., Velusamy, T., Pineda, J. L., Willacy, K., & Goldsmith, P. F. 2014, A&A, 561, A122
Lee, C., Leroy, A. K., Bolatto, A. D., et al. 2018, MNRAS, 474, 4672
Lee, M.-Y., Stanimirović, S., Douglas, K. A., et al. 2012, ApJ, 748, 75
Lee, M.-Y., Stanimirović, S., Murray, C. E., Heiles, C., & Miller, J. 2015, ApJ, 809, 56
Lee, M.-Y., Stanimirović, S., Wolfire, M. G., et al. 2014, ApJ, 784, 80
Lenz, D., Hensley, B. S., & Doré, O. 2017, ApJ, 846, 38
Liszt, H. 2014, ApJ, 780, 10
Liszt, H., Gerin, M., & Grenier, I. 2018, A&A, 617, A54
Liszt, H. S. 2007, A&A, 461, 205
Mathis, J. S. 1990, ARA&A, 28, 37
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
Murray, C. E., Peek, J. E. G., Lee, M.-Y., & Stanimirović, S. 2018a, ApJ, 862, 131
Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2015, ApJ, 804, 89
Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2018b, ApJS, 238, 14
Narayanan, G., Heyer, M. H., Brunt, C., et al. 2008, ApJS, 177, 341
Nguyen, H., Dawson, J. R., Lee, M.-Y., et al. 2019, ApJ, 880, 141
Nguyen, H., Dawson, J. R., Miville-Deschênes, M. A., et al. 2018, ApJ, 862, 49
Paradis, D., Reach, W. T., Bernard, J.-P., et al. 2011, AJ, 141, 43
Petzler, A., Dawson, J. R., Nguyen, H., et al. 2023, PASA, 40, e015
Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, ApJ, 679, 481
Planck Collaboration, Abergel, A., Ade, P. A. R., et al. 2014, A&A, 571, A11
Planck Collaboration, Aghanim, N., Ashdown, M., et al. 2016, A&A, 596, A109
Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002, ApJ, 577, 221
Radhakrishnan, V., Murray, J. D., Lockhart, P., & Whittle, R. P. J. 1972, ApJS, 24, 15
Reach, W. T., Koo, B.-C., & Heiles, C. 1994, ApJ, 429, 672
Ridge, N. A., Di Francesco, J., Kirk, H., et al. 2006, AJ, 131, 2921
Roman-Duval, J., Gordon, K. D., Meixner, M., et al. 2014, ApJ, 797, 86
Rybarczyk, D. R., Stanimirović, S., Gong, M., et al. 2022, ApJ, 928, 79
Saury, E., Miville-Deschênes, M. A., Hennebelle, P., Audit, E., & Schmidt, W. 2014, A&A, 567, A16
Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, ApJ, 216, 291
Seifried, D., Beuther, H., Walch, S., et al. 2022, MNRAS, 512, 4765
Shull, J. M., Danforth, C. W., & Anderson, K. L. 2021, ApJ, 911, 55
Soler, J. D., Beuther, H., Rugel, M., et al. 2019, A&A, 622, A166
Stanimirović, S., Murray, C. E., Lee, M.-Y., Heiles, C., & Miller, J. 2014, ApJ, 793, 132
Steinacker, J., Pagani, L., Bacmann, A., & Guieu, S. 2010, A&A, 511, A9
Sternberg, A., & Dalgarno, A. 1995, ApJS, 99, 565
Sternberg, A., Gurman, A., & Bialy, S. 2021, ApJ, 920, 83
Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10
Tahani, M., Lupypciw, W., Glover, J., et al. 2022, A&A, 660, A97
Tielens, A. G. G. M., & Hollenbach, D. 1985, ApJ, 291, 722
Valdivia, V., Hennebelle, P., Gérin, M., & Lesaffre, P. 2016, A&A, 587, A76
van Dishoeck, E. F., & Black, J. H. 1986, ApJS, 62, 109
Wang, Y., Bihr, S., Beuther, H., et al. 2020, A&A, 634, A139
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152
Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278