Ionized carbon as a tracer of the assembly of interstellar clouds

Schneider, N.; Bonne, L.; Bontemps, S.; Kabanovic, S.; Simon, R.; Ossenkopf-Okada, V.; ... ;
Tielens, A.G.G.M.

Article |

Ionized carbon as a tracer of the assembly of interstellar clouds

Nicola Schneider [1] ✉, Lars Bonne [2] ✉, Sylvain Bontemps [3], Slawa Kabanovic [1], Robert Simon [1], Volker Ossenkopf-Okada [1], Christof Buchbender [1], Jürgen Stutzki [1], Marc Mertens[1], Oliver Ricken [4], Timea Csengeri[3] & Alexander G.G.M. Tielens[5,6]

Molecular hydrogen clouds are a key component of the interstellar medium because they are the birthplaces for stars. They are embedded in atomic gas that pervades the interstellar space. However, the details of how molecular clouds assemble from and interact with the atomic gas are still largely unknown. As a result of new observations of the 158 μm line of ionized carbon [CII] in the Cygnus region within the FEEDBACK program on SOFIA (Stratospheric Observatory for Infrared Astronomy), we present compelling evidence that [CII] unveils dynamic interactions between cloud ensembles. This process is neither a head-on collision of fully molecular clouds nor a gentle merging of only atomic clouds. Moreover, we demonstrate that the dense molecular clouds associated with the DR21 and W75N star-forming regions and a cloud at higher velocity are embedded in atomic gas, and all components interact over a large range of velocities (roughly 20 km s^{-1}). The atomic gas has a density of around 100 cm^{-3} and a temperature of roughly 100 K. We conclude that the [CII] 158 μm line is an excellent tracer to witness the processes involved in cloud interactions and anticipate further detections of this phenomenon in other regions.

Molecular clouds are a crucial component of the interstellar medium (ISM) of galaxies as they are the birth sites of stars and planetary systems. However, the processes by which these clouds are assembled from the large atomic hydrogen (HI) reservoir in galaxies is still not well understood. Some models are based on a subtle equilibrium between gravity, turbulence and magnetic fields, for example, ref.[1]. An external increase of pressure or turbulence due to stellar feedback or spiral arm density waves then randomly triggers a quasi-static, slow build-up of density, leading to the formation of pockets of gas of molecular hydrogen (H$_2$). Other models, for example, ref.[2], propose that cloud formation is more dynamic and driven by large-scale motions in the galaxy, but still closely linked to the local transition from warm ($T \cong 5{,}000$ K), tenuous, mostly atomic gas to dense, cooler ($T \lesssim 100$ K), partly molecular gas. In

this simple two-phase model of the ISM, only the warm and cold neutral medium (WNM and CNM, respectively) are thermally stable. Gas at intermediate temperatures is not in equilibrium and, depending on its density, will either cool down and become denser and fully molecular or heat up to join the WNM. In addition, stellar feedback effects such as radiation, winds and supernova explosions generate turbulence and complicate the picture. It is thus challenging to find the right observational tracers for both the dynamic interaction between gas flows and the thermal and chemical transitions between WNM and CNM.

In simulations, dynamic cloud formation scenarios are idealized by low velocity ($\lesssim$10 km s^{-1}) converging flows, for example, refs.[3–6], which convert diffuse HI gas into dense H$_2$ gas. A recent study[7] showed that only flows with hydrogen densities approximately equal to 100 cm^{-3}

[1]Institute of Physics, University of Cologne, Cologne, Germany. [2]SOFIA Science Center, NASA Ames Research Center, Moffet Field, CA, USA. [3]Laboratory of Astrophysics at Bordeaux, University of Bordeaux, Pessac, France. [4]Max Planck Institute for Radioastronomy, Bonn, Germany. [5]Department of Astronomy, University of Maryland, College Park, MD, USA. [6]Leiden Observatory, Leiden University, Leiden, The Netherlands. ✉e-mail: nschneid@ph1.uni-koeln.de; lbonne@usra.edu
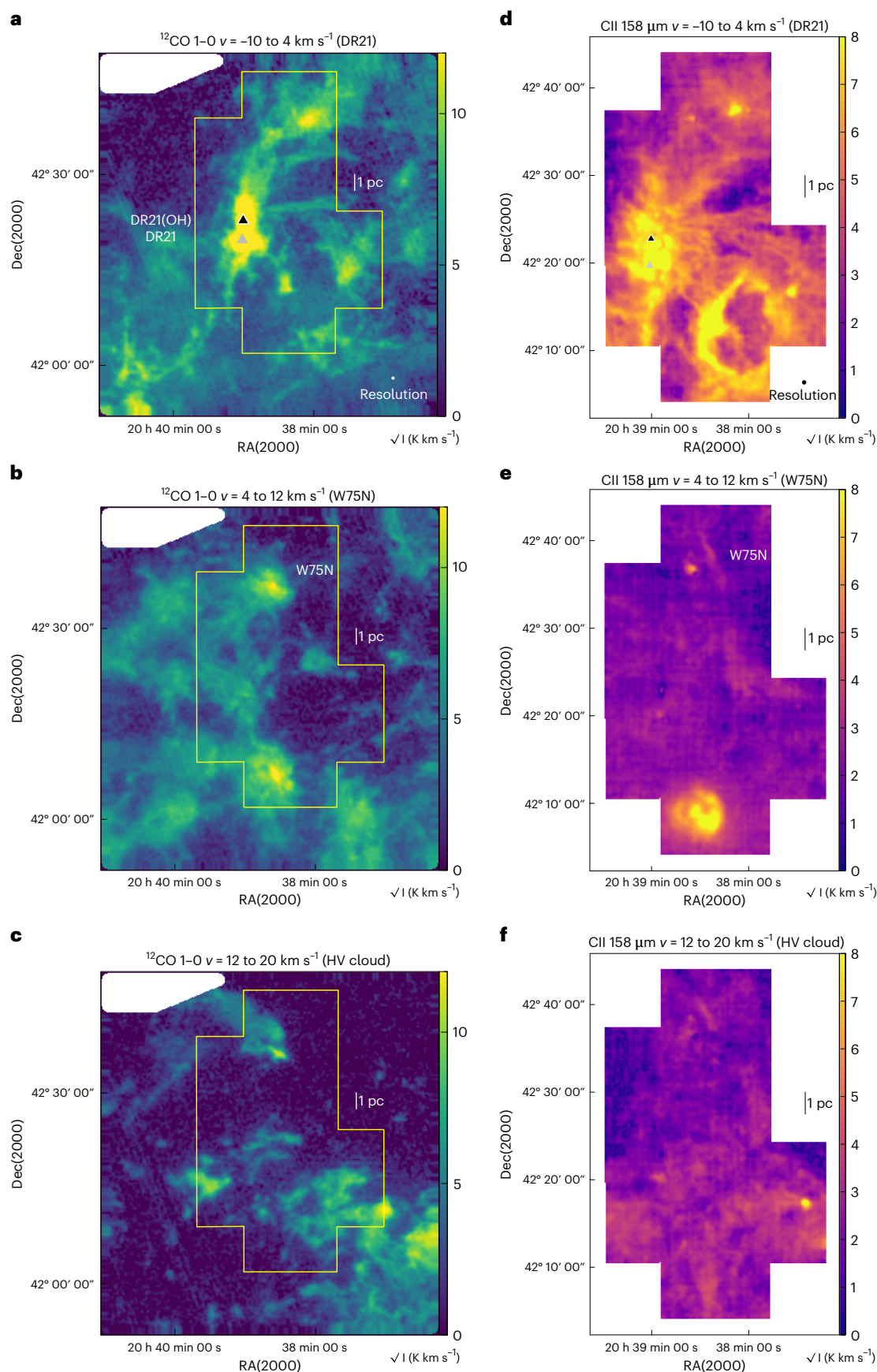
Fig. 1 | Velocity integrated CO and [CII] emission in the DR21, W75N and HV range. a–c, ^{12}CO 1 → 0 maps obtained with the Nobeyama telescope[26] in the three main velocity ranges of the Cygnus X region: −10–4 (**a**), 4–12 (**b**) and 12–20 km s^{-1} (**c**). The embedded massive star-forming sites DR21, DR21(OH) and W75N are indicated. **d–f,** The yellow polygon outlines the area mapped in [CII]: −10–4 (**d**), 4–12 (**e**) and 12–20 km s^{-1} (**f**). The colour wedges give the CO and [CII] intensities in square-root values. The resolution of the maps (30″) is indicated in **a** and **d**. RA, right acension; Dec, Declination.
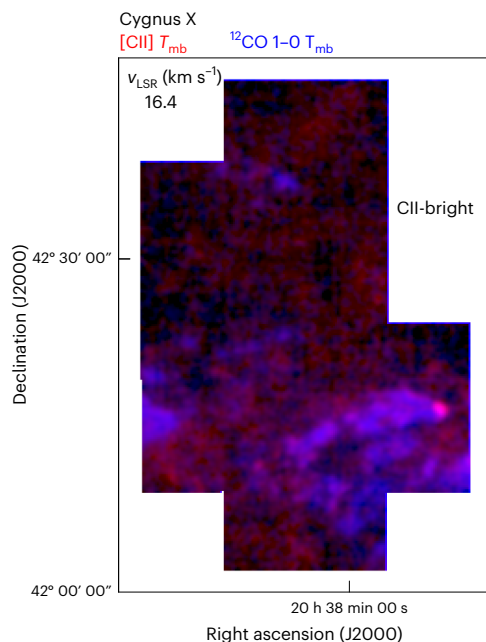
Fig. 2 | Video snapshot of [CII] and CO emission. The image displays a single velocity channel (16.4 km s^{-1}) of [CII] emission in red (0 to 16 K km s^{-1}) and CO emission in blue (0 to 25 K km s^{-1}), respectively, plotted as square-root values. Areas that are dark in CO (no or little emission) are bright in [CII] and reveal the large spatial extent of [CII] emission. The full video with all velocity channels can be found at https://astro.uni-koeln.de/stutzki/research/feedback/animations.

that collide with velocities $\simeq$20 km s^{-1} manage to build up massive structures in which stellar proto-clusters can form. In models with even higher density, the gas flows are already molecular before they collide and are then referred to as cloud–cloud collisions[8–10]. Observations with velocities ≳20 km^{-1} are reported in refs. [11,12]. However, these different scenarios result in contrasting observational predictions. Colliding HI flow models[6] anticipate many velocity components in the lines of ionized carbon ([CII]) and much less in the rotational transitions of carbon monoxide (CO). [CII] emission has its origin in the atomic gas and from non-thermal contributions of multiple molecular clump surfaces at different velocities along the line of sight, while CO only arises from the molecular component. Cloud–cloud collision simulations[8] produce two main molecular velocity components visible in CO, with a bridge of emission in velocity space in between the two components. [CII] emission stems mostly from the envelope of the molecular cloud and the surrounding ambient ISM gas that does not participate in the collision[9].

How can these differing views be confronted with observations? The 21 cm line of HI can be observed in emission and absorption, but it mostly fills the interstellar space so that velocity information is highly blurred. CO lines are used as a proxy for H$_2$ in dense, fully molecular clouds. However, because H$_2$ self-shields more effectively from ultraviolet (UV) photodissociation than CO, there is a gas component that is mostly dark in CO, but bright in H$_2$ (ref. [13]). Fortunately, the [CII] fine structure line at 158 μm is perfectly suited to determine the physical conditions in atomic and CO-dark molecular gas[14–16] and is hence an excellent observational tracer for molecular cloud formation frameworks.

We here present observations in the [CII] 158 μm and CO 1 → 0 spectral lines of Cygnus X, a region with a coherent network of molecular clouds[17], extending over roughly 100–200 pc. It includes the massive star-forming regions DR21 and W75N at a distance of 1.5 ± 0.1 and 1.3 ± 0.1 kpc, respectively, determined with maser parallax measurements[18] and the rich Cyg OB2 association with 169 OB stars[19]. Massive star formation occurs in parts of the clouds, for example, the young

stellar outflow source DR21, the cluster-forming site DR21(OH) and the cluster of early type B stars in W75N (ref. [20]). Cygnus X is not exceptional in terms of mass[21], but for its stellar content. Using ^{12}CO 1 → 0 observations, it was suggested[22,23] that the DR21 molecular cloud, which has a systemic velocity $v = -3$ km s^{-1} with respect to the local standard of rest, collides with the W75N molecular cloud ($v = 9$ km s^{-1}). The [CII] observations reported here indicate a different scenario with an interaction between atomic and molecular gas over a large range of velocities (roughly 20 km s^{-1}).

As part of the SOFIA (Stratospheric Observatory for Infrared Astronomy) FEEDBACK legacy program, the Cygnus X region was observed with the heterodyne array receiver upGREAT (Methods)[24] in the 158 μm line of [CII] at an angular resolution of 14.1″, corresponding to 0.1 pc at a distance of 1.4 kpc. This map of roughly 950 arcmin^2 (158 pc^2) size is, together with the SOFIA map of Orion A[25], a very large, high angular resolution [CII] map showing a very massive star-forming region that extends well into the outskirts of the molecular clouds. It provides data at sub-km s^{-1} spectral resolution so that the dynamic assemblage of molecular clouds can be traced in detail.

We smoothed the [CII] map to an angular resolution of 30″ and regridded to 10″. The data were resampled to a velocity resolution of 0.5 km s^{-1} and have a mean noise temperature of 0.3 K per channel (Methods and Extended Data Fig. 1). We also use ^{12}CO 1 → 0 data from the Nobeyama Cygnus survey[26], smoothed to 30″ resolution on a 10″ grid and a mean noise per 0.5 km s^{-1} wide channel of 0.6 K, and HI data from the Canadian Galactic Plane survey[27]. The HI data have an angular resolution of 1′ and a root mean square (r.m.s.) noise of 3 K in a 0.82 km s^{-1} channel.

Results

Distribution of CO and [CII] emission in Cygnus X

Figure 1 shows the ^{12}CO 1 → 0 and [CII] line integrated emission distribution in the three main velocity ranges in Cygnus X[17]. These are the DR21 range (−10–4 km s^{-1}), the W75N range (4–12 km s^{-1}) and emission between 12 and 20 km s^{-1} that we call the high-velocity (HV) component. The CO and [CII] emission is mostly concentrated in bright photodissociation regions (PDRs) among which the DR21 ridge and the W75N cloud are well-known star-forming sites[28,29]. The bright [CII] emission features will be discussed elsewhere, here we focus on the low surface brightness [CII] emission in which the molecular clouds are embedded, in particular on areas that appear devoid of CO emission in the W75N and HV velocity ranges. Figure 1e,f clearly shows that in those regions where CO emission is lower than its 3σ noise level of 3.6 K km s^{-1}, [CII] intensities are typically ≳5 K km s^{-1} (3σ = 1.8 K km s^{-1}) (see Methods and Extended Data Fig. 2 for details). We thus consider these [CII]-bright areas to be CO-dark but recognize that there may be faint CO emission below the detection limit. Substantial [CII] emission at locations of CO-dark gas is also seen in individual velocity channels, portrayed by a video that scans through all velocities. A one channel snapshot at +16.4 km s^{-1} is shown in Fig. 2. The rather homogeneous [CII] distribution argues against an origin from photoevaporation flows from the surfaces of the molecular clouds that would be more structured and intense towards the clouds.

From scatter plots (Fig. 3a,b) of [CII] and CO emission, we calculate a mean [CII] intensity of roughly 5 K km s^{-1} in the CO-dark and [CII]-bright regime and derive average [CII] and CO spectra (Fig. 3c,d) from these pixels (which are obviously not the same for each velocity range). In total, 29% (6%) of the area in the HV (W75N) range is CO-dark and [CII]-bright, compared to 63% (88%) for [CII] and CO-bright gas. These values, however, strongly depend on the total area that was mapped and are subject to a selection effect because the Cygnus [CII] mapping focused on the bright PDR regions and less on the cloud outskirts. The [CII] spectra show a velocity bridge of emission between the clouds, that is, DR21 at −3 km s^{-1}, W75N at 9 km s^{-1} and the HV cloud at +15 km s^{-1}. CO is also present, but clearly weaker, in particular in the HV
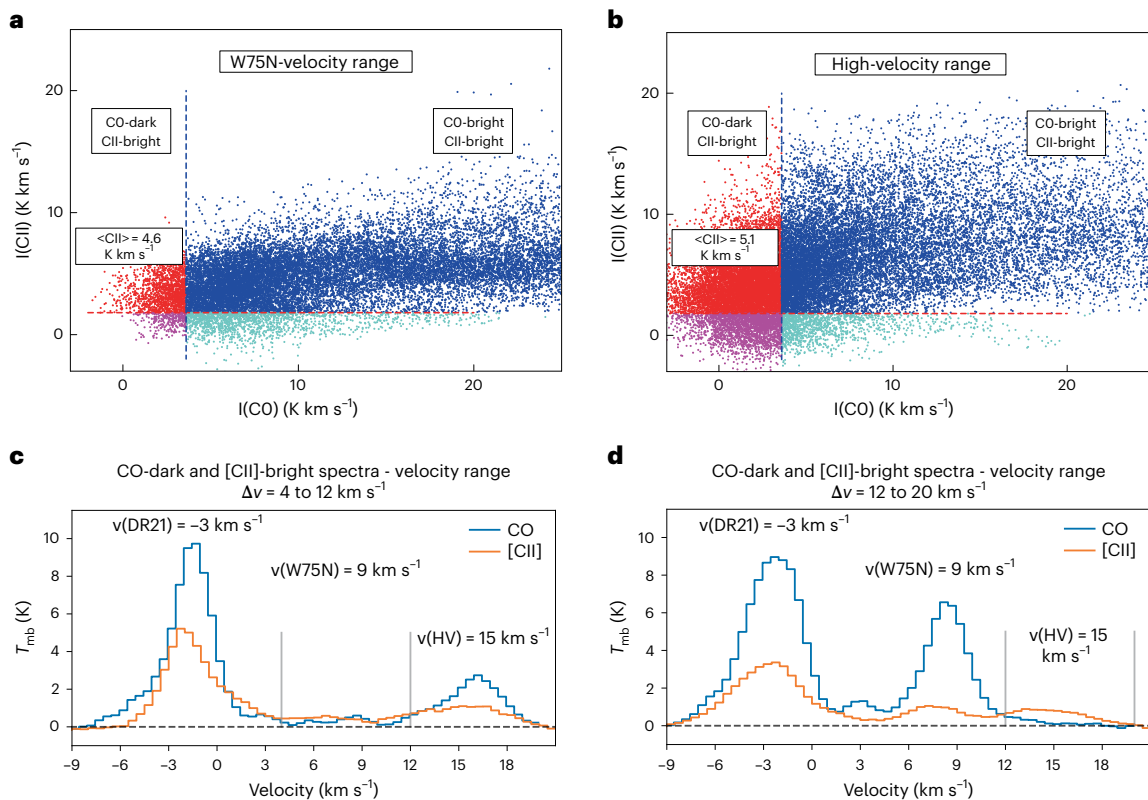
Fig. 3 | Scatter plots and spectra of [CII] and CO 1 → 0 emission. a,b, The pixel-by-pixel correlations in the W75N (**a**) and HV (**b**) velocity ranges. The 3σ noise levels for [CII] emission (1.8 K km s⁻¹) and CO emission (3.6 K km s⁻¹) are indicated by red and blue dashed lines, respectively. The upper left area with red pixels indicates the intensity space that is below the CO noise level but bright in [CII], with an average value of 4.6 K km s⁻¹ (standard deviation 1.5 K km s⁻¹) and 5.1 K km s⁻¹ (standard deviation 2.2 K km s⁻¹) for the W75N and HV velocity ranges,

respectively. **c,d,** The average spectra over all pixels in the map identified as CO-dark and [CII]-bright in the scatter plot in the W75N (**c**) and HV (**d**) velocity ranges, indicated by grey vertical lines. The r.m.s. noise of the spectra for both velocity ranges is 0.075 K for [CII] and 0.15 K for CO, respectively. The [CII] and CO line integrated emissions are 4.7 and 2.2 K km s⁻¹ for the W75N range and 5.4 and 1.3 K km s⁻¹ for the HV range, respectively.
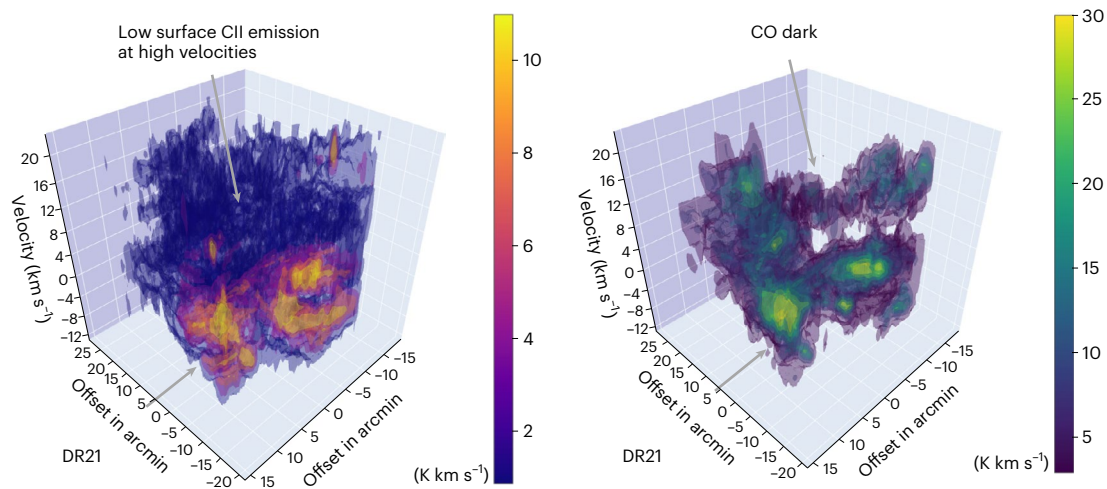

Fig. 4 | 3D iso-surface rendering of position–velocity cuts for [CII] and CO 1 → 0 emission over the entire area observed in [CII]. The *x* and *y* axes are offsets in arcmin from the central map position; the *z* axis is velocity in km s⁻¹. The emission starts at the 5σ level for both tracers. The bright star-forming cloud

DR21 and other dense molecular clouds are embedded in a large-scale cloud structure only visible in [CII] (dark blue). An interactive version of these plots is found at https://astro.uni-koeln.de/stutzki/research/feedback/animations.

range. The kinematic connection in [CII] becomes particularly evident in three-dimensional (3D) position–velocity plots displayed for [CII] and CO in Fig. 4. The strongest emission in both tracers is concentrated

in the −3 km s⁻¹ component from the DR21 ridge including the DR21 outflow and in the 9 km s⁻¹ component from the W75N cloud. These bright clouds and PDR regions (yellow in the [CII] image) are embedded
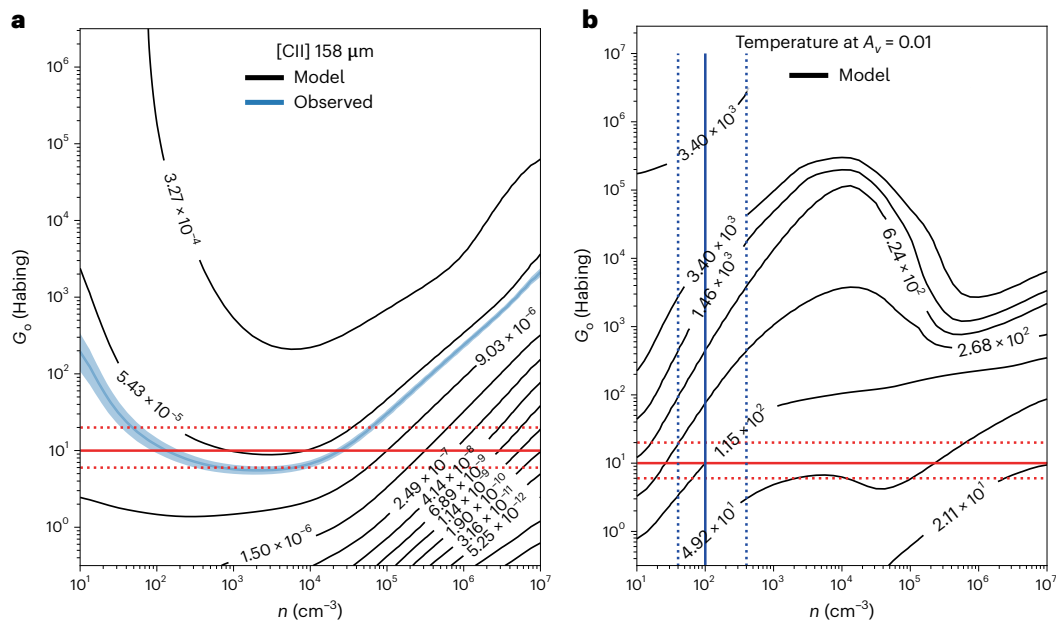
Fig. 5 | PDR model results. The panels show the parameter space of hydrogen density n and FUV field calculated from the PDR toolbox. **a**, The dark blue isocontour of the observed [CII] integrated intensity of 5 K km s^{-1} and the noise r.m.s. of 0.6 K km s^{-1} in dashed light blue. The estimated FUV field of 10 G$_\circ$ is indicated by a red line. The dotted red lines indicate approximately a FUV field of double and half the value. **b**, The isocontours of surface temperature from the PDR model. For an average density of roughly 100 cm^{-3} (blue line) and a FUV field of roughly 10 G$_*$ (red line), we obtain a temperature of 115 K. The dashed lines for density and temperature indicate how these values change if the FUV field is higher or lower.

Table 1 | CO-dark, [CII]-bright gas physical properties

I_{CII}[a] (K km s^{-1})	FUV field (G$_o$)	Density (cm^{-3})	Temperature (K)	N(CII)[b] 10^{18} (cm^{-2})	N(H)[c] 10^{21} (cm^{-2})
roughly 5	10	100	100	0.606	3.78

[a] Average of the C$^+$ line integrated intensity in CO-dark regions. We call [CII]-bright all regions with [CII] emission above the 3σ noise level. [b] The C$^+$ column density N(CII) is calculated with equation (1) using the C$^+$ line integrated intensities. [c] The total hydrogen column density N(H)=N(HI)+2N(H$_2$) is estimated from N(CII), applying the abundance C/H=1.6×10^{-4} (ref. [51]).

in a pervasive medium emitting in [CII] (in dark blue), which is not or only a little visible in CO.

Summarizing, we conclude that instead of a head-on collision between a −3 and +9 km s^{-1} molecular cloud[22,23], we witness here an interaction of several partly atomic flows (seen in [CII]) and partly molecular flows (seen in CO). The next section quantifies this scenario by calculating the physical properties of the interacting gas.

Physical properties of the interacting gas

We estimate the density and temperature of the gas detected with [CII] at velocities v > 4 km s^{-1} using predictions[30] from the PDR toolbox (Methods) for a [CII] line integrated intensity of 5 K km s^{-1}. From a census of the 169 OB stars of Cyg OB2, we derive a Habing field of roughly 10 G$_o$ (Extended Data Fig. 3), where G$_o$ is the mean interstellar radiation field. The PDR model (Fig. 5a) indicates hydrogen densities of roughly 100 cm^{-3}, which is typical for diffuse gas at the transition from atomic to molecular. We exclude here the high-density solution (>10^4 cm^{-3}) because, then, significant CO emission should have been detected, which is not the case. We note that all numbers have an uncertainty mostly because of the adopted value of the far ultraviolet (FUV) field. With the derived densities, we obtain a surface temperature (Fig. 5b) of 115 K for the PDR gas layer. This is an upper limit for the kinetic temperature T_{kin} of the gas, since the temperature drops entering deeper PDR layers. To narrow down T_{kin}, we performed a study

of HI self-absorption (HISA) towards DR21 (Methods and Extended Data Figs. 4 and 5) and obtained a gas temperature of roughly 100 K. We use this value to calculate C$^+$ and hydrogen column densities, N(CII) and N(H), respectively (Methods and Extended Data Fig. 6), and give all input values and results in Table 1. N(H) consists of an atomic and molecular part, and the relative fractions are variable because the formation of H$_2$ depends on the local radiation field and density, and on turbulent mixing motions[31] that cause large- and small-scale density fluctuations. We estimate (Methods) that roughly 23% of the gas in the W75N range and roughly 14% in the HV range is molecular. This is qualitatively in good agreement with results from colliding HI flow simulations[6], predicting that about 20% of hydrogen is in the form of H$_2$ at densities around 100 cm^{-3} for the initial phases of cloud formation. Our values also conform with the results of ref. [16] who find that ≲20% of [CII] comes from the molecular phase. Their simulation set-up represents a section of the Milky Way disc in which turbulence is injected by supernova explosions but the dynamic effect of gas accretion on to the clouds from the larger scale, galactic environment is retained. However, they investigate only the earliest phases of cloud formation with an UV field of 1.7 G$_o$ and lower temperatures of roughly 50 K. The masses (Methods) contained in the atomic gas are 7,800 M_{sun} for the W75N range and 9,900 M_{sun} for the HV range, respectively. This is an important mass reservoir for building up more molecular clouds, comparable to the fully molecular cloud DR21 (roughly 15,000 M_{sun}, ref. [29]). The time scale for cloud assembly is given by the relative velocity of the components and their size. The column densities of the W75N and the HV cloud translate into a size of 12 pc for a density of 100 cm^{-3}, leading to an assembly time of 1.3 Myr on the basis of their separation in velocity space by about 10 km s^{-1}. In a quasi-static scenario, molecular cloud formation would take much longer, about 10 Myr at a density of 100 cm^{-3}, on the basis of the formation rate of molecular H$_2$ of 3 × 10^{-17} cm^3 s^{-1} (refs. [32,33]). Faster cloud formation with significant fractions of H$_2$ can be explained, however, from colliding flow simulations that temporarily create pockets of gas with higher density[34].

Discussion

The observed levels of [CII] intensity in the W75N and HV velocity ranges (excluding bright local sources such as the W75N protocluster) are consistent with the low FUV illumination of roughly 10 G_0 we estimated from the census of the stars from Cyg OB2 at a distance of roughly 1.6 kpc (ref. [35]). The HV gas can thus not stem from the Cygnus Rift, a dark extinction feature at 600 pc with no notable excitation sources[17]. Recently, ref. [36] used the GAIA2 data release in combination with extinction to build 3D maps of the dust in the local arm and surrounding regions and confirmed that there is no active star formation taking place in the Rift.

Maser parallax measurements[18] indicate that W75N (1.3 ± 0.1 kpc) is slightly but clearly in front of DR21 (1.5 ± 0.1 kpc) showing that the densest parts of these two molecular clouds could not have collided head-on. Observed absorption features in ^{12}CO, HCO^+, CH^+, SH^+ towards DR21 (refs. [22,28,37]) support this 3D view (see Methods for details on the Cygnus X complex). In addition, our HISA study towards DR21 identifies broad absorption in the velocity range of −5 to 20 km s^{-1} (Methods). Accordingly, the atomic clouds at red-shifted velocities with $v > 4$ km s^{-1} (W75N, HV) must be located in front of DR21 and the dynamics we traced in [CII] indicates that all three of them are clearly on collisional trajectories. Our scenario of molecular cloud + HI envelopes interaction, visible through [CII], indicates that the DR21, W75N and HV components are not too far separated but should be located within a similar volume with a radius of presumably 20–50 pc. More precise distance estimates would help to test our view.

The composition of the gas seen in low surface brightness C^+ emission is about 20% molecular and 80% atomic. These values can be compared with the findings from the GOTC+ survey[14]. They observed galactic sightlines with many bright PDRs along the line-of-sight and derived that roughly 47% of [CII] emission arises from PDRs, roughly 28% from CO-dark gas, roughly 21% from cold atomic gas and roughly 4% from ionized gas. Our observations reveal a large reservoir of CO-dark gas of several thousand solar masses, comparable in mass to the active star-forming regions DR21 and W75N, that was previously unrecognized. We show that the DR21, W75N and HV molecular clouds cross each other, interacting mostly through their low-density enveloping atomic gas layers. The collision of small-scale HI flows forms dense, molecular gas in the compressed layer of oblique shocks such as those proposed in refs. [38,39] but the residual HI is still there and forms a reservoir from which more mass is accreted. We expect that many of these compressed layers show a flattened sheet-like structure, as increasingly seen in observations[40,41], and work on a study to test this scenario. We note that it depends strongly on the density if the interacting gas is also seen in CO (at the same velocity as the [CII]). For the W75N range, we indeed observe a bridge of emission for CO and [CII] at the same velocities, most probably because the molecular fraction is higher. Other studies[10], only using CO, already showed this molecular interaction, also at HVs[11]. To a certain extent, it is also possible that the reason we detect higher [CII] velocities is that some of the CO is already shocked due to the compression and thus at lower velocities. The preshock HV gas, however, is at low densities and only visible in [CII].

Our scenario leads to a continuous assembly of more molecular material on very short time scales over roughly 1 million years. It is not likely, but cannot fully be excluded, that the molecular clouds formed in the collision zones of interacting expanding HI shells[42] because there are no clear observational signatures for such HI shells and the relative velocities we observe in Cygnus are too large to be only driven by an expanding bubble.

We note that the magnetic field, and in particular its orientation, can also play an important role in this sort of cloud assembly.[5,43] derived that a general alignment of the collision axis and mean magnetic field direction is necessary for direct cloud formation. Recently, ref. [44] showed that under an inclination angle of 45°, it is possible to accumulate enough mass allowing the formation of massive clusters with O-type stars. Alternatively, the observed interaction between the HI/CO complexes have not formed directly the seeds of the CO clouds. Instead, we propose that the crossing of the HI/CO clouds we observe have triggered an acceleration of the concentration of these seeds into the dense and massive star-forming clouds through the scenario proposed for Musca by ref. [39]. For the quiescent Musca cloud, ref. [39] found that molecular gas formed preferentially at the convergence point of the magnetic field, bent behind the shock front. For Cygnus X, magnetic field measurements would thus be crucial to address this important point.

The implications of such a generalized scenario to explain cloud interactions, and consequently star formation, both in Musca and in Cygnus, are important because they suggest a certain degree of universality of cloud formation as a result of interactions between at least partly diffuse HI clouds. This is in line with the scenario of the interplay of feedback bubbles and gravity leading to the formation of dense structures in the multi-phase ISM[41]. The main differences between Musca and Cygnus X are the larger initial densities in Cygnus X and the velocity of the collision with roughly 20 km s^{-1} for Cygnus X and less than 10 km s^{-1} in Musca. Streams of diffuse gas with relative velocities of less than 10 km s^{-1} are easily justified by turbulent HI gas in the galaxy since the sound speed in the WNM is of this order of magnitude. The origin of gas streams of more than 20 km s^{-1} is more difficult to explain. They can be driven by the complex interplay between gravity and stellar feedback effects, and the thermodynamic response of the multi-phase ISM. In galaxy-wide simulations[45] spiral density waves lead to high-velocity collisions of flows that can form massive OB clusters such as the ones seen in Cygnus X. In any case, our finding will have major ramifications for our understanding of molecular cloud assemblage in the Milky Way and other galaxies. Further observations of the extended [CII] emission in the FEEDBACK sample will reveal whether this sort of interaction is common to other giant molecular cloud regions. In the future, the GUSTO and ASTRHOS balloon projects will measure the [CII] emission in the Milky Way and in the Large and Small Magellanic Clouds. A promising other tracer for detecting flows of partly atomic, partly molecular gas is atomic carbon ([CI]). It is predicted to also trace the CO-dark gas component[6] and the [CI] 1-0 line is observable from the ground. The GEco project on the upcoming CCAT-prime/FYST telescope will perform extended surveys in this line[46].

Methods

Observations and data reduction

The Cygnus X region was observed during several flights in November 2019, February 2021, November 2021 and April 2022 for the legacy program FEEDBACK, using the 2.7 m telescope onboard the SOFIA. The dual-frequency heterodyne array receiver upGREAT[24] was tuned to the [CII] 157.7 μm line in the low-frequency array (LFA) 2 × 7 pixel array and the [OI] 63 μm line (not shown) was observed in the high-frequency 7 pixel array. The LFA has an array size of 72.6″, and a pixel spacing of 31.8″. The half-power beam width at 158 μm is 14.1″, determined by the instrument and telescope optics and confirmed by observations of planets. For each flight series, the main-beam efficiencies η_{mb} for each pixel for the LFA and high-frequency array channels were determined, the average value for the LFA is 0.65 and we use this value here. The entire map region was split into multiple square 'tiles' with 435.6″ on one side and each square was covered four times. The on-the-fly (OTF) scan speed was selected to attain Nyquist sampling of the LFA beam (dump every 5.2″). The total time for one OTF line was 25.2 s, that is, together with the OFF observation, within the measured Allan variance stability time of the system. The first two coverages were done once horizontally and vertically with the array rotated 19° against the scan direction, so that scans by seven pixels are equally spaced. The second two coverages were then shifted by 36″ in both directions to achieve the best possible coverage for the [OI] line in the LFA array mapping mode. Overall, each tile took roughly 50 min to complete.

The horizontal and vertical scanning can cause some striping in the maps. To reduce these effects, we applied a principal component analysis (PCA) on the data. In short, our method based on PCA uses the information of systematic variations in the baseline from a large set of observations from the emission-free OFF position that are regularly taken during each scan. These variations are caused by time-dependent instabilities in the backends, receiver, telescope optics and atmosphere. We produce an 'OFF−OFF' spectra by subtracting subsequent OFF positions from each other and calibrate the data in the same way as the ON−OFF spectra that contain the astronomical source emission. We then identify systematic 'eigenspectra' in the OFF−OFF spectra that account for all or at least most of the structure in the baseline. Using a linear combination of the strongest components, we reconstruct the ON−OFF spectra with the best-fit coefficients for each component. We scale each component by the coefficients that were found and subtract those from the ON−OFF spectra. This procedure removes the systematic variations found in the OFF−OFF spectra but does not make changes to the astronomical line in terms of intensity, width, position and so on in the ON−OFF spectra. We further improve the data quality by using a sophisticated algorithm to determine a set of spectra that are free of emission from the PCA-corrected spectra for each frontend and SOFIA flight. With this set of emission-free spectra we use a second PCA-correction equal to the one described above, but using now these emission-free spectra to determine the baseline 'components'. With these components, we correct systematic variations that happen on time scales of the individual OTF-dumps, which are much shorter than the 10 s integration time for OFF spectra. The final PCA map was then compared to one obtained by removing a polynomial baseline of order 3, performing difference maps, ratio maps, scatter plots and so on, and we found no systematic effects.

We then spatially smoothed the resulting [CII] map to an angular resolution of 30″ in a Nyquist sampled 10″ grid to emphasize the large-scale emission distribution and not focus on small-scale variations. Some striping effects are still visible, but they do not produce systematic effects in the data analysis (scatter plots, for example). The map centre position is at right acension (2000) = 20 h 38 min 39.3 s, Declination (2000) = 42° 20′ 39.3″ and an emission-free OFF position at right acension (2000) = 20 h 39 min 48.34 s, Declination (2000) = 42° 57′ 39.11″ was used. A fast Fourier transform spectrometer with 4 GHz instantaneous bandwidth[47] with a velocity resolution of 0.04 km s^{-1} (from the hardware selected frequency resolution of 0.244 MHz) served as backend. Here, we use data resampled to 0.5 km s^{-1} resolution. For further technical details, see ref. [48]. All spectra are presented on a main-beam brightness temperature scale T_{mb}, that is, corrected for the main-beam efficiency. The r.m.s. noise of this final data set (and the one of CO emission) was then determined by fitting a 0th order baseline to each spectrum, excluding the windows with emission. The noise is homogeneous across the maps and follows a roughly Gaussian distribution (Extended Data Fig. 1) with a peak at 0.34 K for [CII] and 0.64 K for CO, respectively. As a second method, we determined the noise by spatially averaging the pixels in emission-free channels and obtained values of 0.31 K for [CII] and 0.62 for CO, respectively. For the line integrated intensities, we calculated the σ noise level for the 8 km s^{-1} wide velocity integration range by $\sigma = \sqrt{(16)} \times 0.6 \times 0.5 = 1.2$ K km s^{-1} in which 0.6 K is the noise, 0.5 km s^{-1} is the channel width and 8 km s^{-1} correspond to 16 channels. The 3σ CO level is thus 3.6 K km s^{-1}. The equivalent calculation for [CII] delivers a 3σ level of 1.8 K km s^{-1}. In addition, we show in Extended Data Fig. 2 a PDF of the [CII] bright, CO-dark intensities together with a PDF of the r.m.s. noise to demonstrate that the observed W75N and HV intensities are clearly offset from the noise. We also note that the smoothing is the same for all data and cannot change any structural behaviour.

A potential problem could be that a part of the emission that we find in the W75N velocity range might stem from a [^{13}CII] hyperfine transition at DR21 source velocities. The [^{13}CII] transition splits into three hyperfine components with a relative strength of $s_{2\rightarrow1} = 0.625$, $s_{1\rightarrow0} = 0.25$ and $s_{1\rightarrow1} = 0.125$ caused by the unpaired spin from the additional neutron. The three satellites are velocity shifted $\Delta v_{2\rightarrow1} = 11.2$ km s^{-1}, $\Delta v_{1\rightarrow0} = -65.2$ km s^{-1} and $\Delta v_{1\rightarrow1} = 63.2$ km s^{-1} with respect to the [CII] fine structure line[49]. The $F_{2\rightarrow1}$ component of [^{13}CII] emission from the bright PDRs at DR21 velocities (−3 km s^{-1}) may thus appear at roughly 8 km s^{-1}, close to the systemic velocity of W75N (roughly 9 km s^{-1}). We estimated this contribution by comparing the [CII] emission in the DR21 velocity range with the emission in the W75N velocity range for the [CII]-bright, CO-dark gas and found no correlation between the two quantities at a DR21 [CII] line strength of roughly 15 K km s^{-1} (note that the average [CII] intensity is 5 K km s^{-1}). For optically thin [CII], this translates into an expected [^{13}CII] $F_{2\rightarrow1}$ line strength of 0.16 K km s^{-1}. Even if optical depth effects may increase this value, it is so small compared to our noise limit that it certainly does not change our quantitative estimates and simultaneously explains that there is no correlation detected.

PDR modelling

We use the observed [CII] intensities using the plane-parallel models provided by the PDR toolbox found at https://dustem.astro.umd.edu (ref. [30]). In short, these models solve the radiative transfer equation with chemical balance and thermal equilibrium for a plane-parallel PDR layer exposed to a UV radiation field, cosmic rays and soft X-rays incident on one side. A given set of gas phase elemental abundances and grain properties is calculated, and the emergent [CII] line integrated intensities as a function of density n and radiation field G_o in Habing units are given. The beam filling is assumed to be unity, which we consider a good approach for the W75N and HV emission because the [CII] arises from extended, mostly diffuse gas. We here use the model WK2020 that features some updates of the photorates and dependence with PDR depth, ^{13}C chemistry and line emission, and O collision rates. However, there is nearly no difference in the [CII] model prediction compared to the 2006 model. The specific parameters for the WK2020 model are listed in Table 1 in ref. [30]. Some important values are the cosmic ray ionization rate per H nucleus of 2×10^{-16} s^{-1} and the formation rate of H_2 on dust of 6×10^{-17} s^{-1}.

We focus here on modelling only the [CII] emission, although the plane-parallel PDR model also predicts a CO (1 → 0) brightness. However, this value is very sensitive to the assumed total depth of the cloud. The [CII] intensity and the surface temperature trace surface properties while the CO emission only arises from the layers deeper in the cloud where CO can form, this means those that are not CO-dark. As the toolbox models assume a gas column of visual extinction $A_v = 7$, indicating a large fraction of CO-bright gas, while the observed gas rather has a column of 3.8×10^{21} cm^{-2} we expect, and see, a significant over-prediction of the CO 1-0 intensity by the model.

Considering the [CII] intensity, we derive a density of $n = 100$ cm^{-3} for a FUV field of 10 G_o. Taking into account the uncertainties in the FUV field, we assume a field that is double (roughly 20 G_o) and half (roughly 5 G_o) of the average value and then derive a density range of roughly 40 to 400 cm^{-3}. The corresponding surface temperatures are then roughly 200 K for $n = 40$ cm^{-3} and roughly 90 K for $n = 400$ cm^{-3}, respectively. From our HISA study, however, we obtain a temperature of 90–120 K with an average of 108 K, which points towards a density of roughly 100 cm^{-3} and thus to a FUV field of 10 G_o. A significantly lower UV field would push the densities towards higher values of more than 10^3 cm^{-3} but we note that the observed [CII] isocontour is a very flat curve for high densities. In any case, we would have detected CO emission in densities above 10^3 cm^{-3}. An even higher UV field has a smaller impact on the density, but would still move values below roughly 40 cm^{-3}. For simplicity, we will use a common temperature of 100 K and a density of 100 cm^{-3}, but work out the values for column density fractions using these extreme limits of density and temperature.

Calculation of physical properties

We assume that the main collision partner for C^+ is atomic hydrogen, and that the [CII] line is optically thin, so that the line emission must be subthermally excited, in view of the low densities. The [CII] column density N(CII) is then calculated[50] from

$$N(\text{CII}) = \frac{I_{\text{CII}}10^{16}}{3.43} \times \left[1 + 0.5 \times \exp\left(\frac{91.25}{T_{\text{kin}}}\right)\left(1 + \frac{2.4 \times 10^{-6}}{C_{\text{ul}}}\right)(\text{cm}^{-2})\right] \quad (1)$$

with the line integrated [CII] emission I_{CII} in K km s^{-1}, the kinetic temperature T_{kin} (K) and the de-excitation rate C_{ul} (s^{-1})

$$C_{\text{ul}} = n \times R_{\text{ul}} \quad (2)$$

with hydrogen density n (cm^{-3}) and de-excitation rate coefficient R_{ul}, derived with

$$R_{\text{ul}} = 7.6 \times 10^{-10}\,\text{cm}^3\text{s}^{-1}(T_{\text{kin}}/100)^{0.14} \quad (3)$$

The total hydrogen column density $N(\text{H}) = N(\text{HI}) + 2N(\text{H}_2)$ is estimated from N(CII), assuming all carbon is in the form of C^+ and applying the abundance C/H $= 1.6 \times 10^{-4}$ (ref. [51]). For the nominal values of $T = 100$ K and $n = 100$ cm^{-3}, we obtain $N(\text{CII}) = 0.61 \times 10^{18}$ cm^{-2} and $N(\text{H}) = 3.78\,10^{21}$ cm^{-2}. This total hydrogen column density corresponds very well to the one estimated from our HISA study, assuming that most of the gas seen in [CII] is atomic. We thus consider a density of roughly 100 cm^{-3} as the most likely value for the atomic gas. Nevertheless, with a lower density of $n = 40$ cm^{-3} and $T = 200$ K, we calculate $N(\text{CII}) = 0.85 \times 10^{18}$ cm^{-2} and $N(\text{H}) = 5.31 \times 10^{21}$ cm^{-2}. With $n = 400$ cm^{-3} and $T = 90$ K, we derive $N(\text{CII}) = 0.20 \times 10^{18}$ cm^{-2} and $N(\text{H}) = 1.22 \times 10^{21}$ cm^{-2}.

The mass of the atomic gas is then estimated by

$$M = \frac{N_{\text{CII}}\,A\,m_{\text{H}}}{\text{C/H}}\,[M_{\text{sun}}] \quad (4)$$

with the area A in cm^2, the mass of hydrogen m_{H} in kg and the C/H abundance 1.6×10^{-4} (ref. [51]).

Determination of the molecular fraction

The diffuse gas at HVs (>4 km s^{-1}) is partly molecular and partly atomic. We here give a rough estimate of the molecular fraction, which is defined[50] by

$$f(\text{H}_2) = \frac{2N(\text{H}_2)}{N(\text{H})} \quad (5)$$

The molecular hydrogen column density $N(\text{H}_2)$ is derived using the average CO line intensities (I_{CO}) from the spectra in the W75N velocity range (2.2 K km s^{-1}) and the HV range (1.3 K km s^{-1}). Both values have an error of 1.2 K km s^{-1}. With a CO-to-H$_2$ conversion factor $X_{\text{CO}} = N(\text{H}_2)/I_{\text{CO}}$ of 2×10^{20} cm^{-2} (K km s^{-1})$^{-1}$, commonly used in the literature, we obtain H$_2$ column densities of 0.44×10^{21} and 0.26×10^{21} cm^{-2} for the W75N and HV velocity range, respectively. The total hydrogen column density $N(\text{H})$ is derived from the [CII] column densities, given in Table 1, with a value of 3.78×10^{21} cm^{-2} in each velocity range. With these values, the molecular fraction is then calculated to be $f(\text{H}_2) = 0.23$ for the W75N velocity range and $f(\text{H}_2) = 0.14$ for the HV range, respectively, with an average value of $f(\text{H}_2) = 0.19$. The portion of molecular gas is thus nearly twice as much in the W75N velocity range compared to the HV range. We note that the H$_2$ column densities and molecular fractions are lower limits since we use the canonical value of the X_{CO} factor that is mostly valid for evolved molecular clouds. In case the atomic gas has a lower or higher density because of a different incident FUV field, the molecular fractions change accordingly. We estimate that for $n = 40$ cm^{-3}, $f(\text{H}_2) = 0.17$ for

the W75N velocity range and $f(\text{H}_2) = 0.10$ for the HV range, respectively. In the high-density case with $n = 400$ cm^{-3}, $f(\text{H}_2) = 0.72$ for the W75N velocity range and $f(\text{H}_2) = 0.43$ for the HV range, respectively. These values are more extreme and less likely than the fractions we obtained with the nominal values but we note that all observational and model values have their uncertainties.

Evaluation of the FUV field

The FUV field in Cygnus X was derived from a census of the stars of Cyg OB2, using the compilation of ref. [19]. They listed 169 OB stars including 52 O-type and three Wolf–Rayet stars. To determine the UV luminosity of the cluster, we assume that the spectral radiance of each star can be represented by a black-body:

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \quad (6)$$

with the wavelength λ, the speed of light c, the Boltzmann constant k_B, the Planck constant h and the temperature of each star T. To extract the UV portion of the luminosity L, we integrate the Planck function over the UV range between 910 and 2,066 Å, which corresponds to a photon energy range of 6 to 13.6 eV. The ratio of the integrated spectral radiance over the UV range to the entire black-body spectrum gives us the UV luminosity:

$$L_{\text{UV}} = \frac{\pi \int_{\lambda_{910}}^{\lambda_{2066}} B(\lambda, T)\,\text{d}\lambda}{\sigma T^4}L \quad (7)$$

with the Stefan–Boltzmann constant σ. The superposition of the stellar UV flux of all stars considered gives the UV field at every point (RA, Declination) of the grid:

$$F_{\text{UV}} = \sum_i \frac{L_{\text{UV},i}}{4\pi R_i^2} \quad (8)$$

where R_i is the radial distance to each star. We assumed the most recent distance estimate, based on GAIA, of 1.6 kpc (ref. [35]) for every star in the cluster to the observer, although there can be line-of-sight distance differences between the individual stars. Extinction of the UV field by gas of the ISM is not taken into account, thus the determined UV field is an upper limit. The resulting FUV field is shown in Extended Data Fig. 3 where it becomes obvious that the FUV field varies between roughly 5 and 20 G$_o$ across the [CII] map. We use a value of 10 G$_o$ for our PDR modelling. We note that a closer distance of Cyg OB2 of roughly 1.45 kpc (ref. [52]) would increase the UV field to values between roughly 10 and 30 G$_o$ with little influence on the PDR modelling.

HISA

Studying self-absorption of atomic hydrogen towards a strong continuum source is a way to determine the atomic column density in front of the source. We follow the approach of ref. [40] to derive the amount of cold absorbing HI in front of DR21, which emits strongly in the 1.4 GHz continuum. Both, the HI 21 cm line data and the continuum, stem from the Canadian Galactic Plane survey[27] and have an angular resolution of 1 . First, the method requires to estimate the HI emission by finding a mostly absorption free position (off) as a reference. For that, 100 HI emission spectra around DR21 within a $20' \times 20'$ square were plotted (Extended Data Fig. 4a) and compared to the average HI spectrum towards DR21, showing absorption across a large velocity range. We are mostly interested in the range $v = 4$ to 20 km s^{-1}. All off spectra show a similar line profile with a flat-top shape peaking at a temperature of roughly 100 K at the absorption dip of the on spectrum. However, the various small absorption dips in the off spectra indicate that the absorbing hydrogen cloud is also extended. As a best candidate we chose a

spectrum that shows the least absorption features. The location of this off spectrum is plotted in Extended Data Fig. 5, indicated by a blue circle of radius 1′, having a central position right acension (J2000) = 20 h 39 min 34.07 s, Declination (J2000) = 42° 29′ 38.00″. The off spectrum is shown in Extended Data Fig. 4b in blue, together with a Gaussian fit optimized to fit the wings of the off spectrum. It could represent the HI emission if some emission is still absorbed at our off position. We calculated the velocity resolved optical depth for both choices, the off spectrum and the fitted one. This helps us to assess the uncertainties in the derivation of the column density. The optical depth as a function of velocity v is given by

$$\tau_{\text{HISA}}(v) = -\ln\left(1 - \frac{T_{\text{on-off}}(v)}{T_{\text{HISA}} - T_{\text{off}}(v) - T_{\text{cont}}}\right) \quad (9)$$

with the continuum temperature T_{cont} (437 K as an average) for a HISA temperature T_{HISA}. The corresponding column number density of the absorbing HI layer, N_{HISA}, can be determined by integration over $\tau_{\text{HISA}}(v)$:

$$\frac{N_{\text{HISA}}}{\text{cm}^{-2}} = 1.8224 \times 10^{18} \frac{T_{\text{HISA}}}{\text{K}} \int \tau_{\text{HISA}}(v) \, d\left(\frac{v}{\text{km s}^{-1}}\right) \quad (10)$$

We integrate over the velocity range of 4 to 20 km s^{-1}. The temperature dependent HISA column number density is shown in Extended Data Fig. 6 for both, the off position and the Gaussian fit. It becomes obvious that the differences are small. To further constrain the possible range of HISA column density, we determine the maximum HISA temperature by the temperature at the absorption dip minimum at each spectrum of the HI map:

$$T_{\text{HISA,max}} = T_{\text{on,min}} + T_{\text{cont}} \quad (11)$$

The resulting temperature map is shown in Extended Data Fig. 5 and confirms what we already concluded from Extended Data Fig. 4, that is, that the HISA temperature cannot be higher than roughly 100 K, which results in a HISA column density N_{HISA} of 3.5×10^{21} cm^{-2}. For a density of 100 cm^{-3}, this translates into a HI layer of 11 pc that compares well to other dominant molecular structures in Cygnus X, such as, for example, the DR21 ridge with a size of roughly 7 pc.

The Cygnus X star-forming complex and the distance problem
The Cygnus X star-forming region has for a long time been noticed as an outstanding area because it lies around $l = 90°$ where the local galactic arm, the Perseus arm and the outer galaxy are found along the same line of sight, covering distances between 1 and 8 kpc. Since radial velocities around the tangent point in Cygnus X are close to zero, they do not provide reliable distances. This is why the Cygnus X region was long proposed to be an accumulation of clouds at various distances along the different spiral arms[53]. However, on the basis of CO data and arguments of UV illumination, ref. [17] proposed that most of the observed velocity components seen in CO were associated and part of a single complex despite the very large differences in velocities (around −5 to 18 km s^{-1}). The main molecular cloud regions are DR21 (around −3 km s^{-1}) and W75N (around 9 km s^{-1}). At that time, it was not clear how such large relative velocities could co-exist spatially and inside a single event of star formation. The scenario of a single complex for Cygnus X[17] has been confirmed by maser parallax measurements[18], obtaining roughly the same distance for DR21 and W75N despite their velocity difference of roughly 12 km s^{-1}. Our study now shows that the W75N and the HV atomic/molecular clouds are located in front of DR21 ([CII] emission and HI absorption) and interact with each other.

Data availability
The calibrated and polynominal baseline-corrected [CII] data cubes at a velocity resolution of 0.2 km s^{-1} are provided by the IRSA/IPAC archive and are found at https://irsa.ipac.caltech.edu/applications/sofia under the project no. 07_0077 (FEEDBACK, PIs are A.G.G.M. Tielens and N. Schneider). The PCA reduced data set will also be made available on the IRSA/IPAC archive together with other data sets from the FEEDBACK program as level 4 data products.

References
1. Krumholz, M. R. & McKee, C. F. A general theory of turbulence-regulated star formation, from spirals to ultraluminous infrared galaxies. *Astrophys. J.* **630**, 250 (2005).
2. Hartmann, L., Ballesteros-Paredes, J. & Bergin, E. A. Rapid formation of molecular clouds and stars in the solar neighborhood. *Astrophys. J.* **562**, 852–868 (2001).
3. Koyama, H. & Inutsuka, S.-I. Molecular cloud formation in shock-compressed layers. *Astrophys. J.* **532**, 980–993 (2000).
4. Vazquez-Semadeni, E., Ryu, D., Passot, T., Gonzalez, R. F. & Gazol, A. Molecular cloud evolution. I. Molecular cloud and thin cold neutral medium sheet formation. *Astrophys. J.* **643**, 245–259 (2006).
5. Inoue, T. & Inutsuka, S. Two-fluid MHD simulations of converging HI flows in the ISM II. Are molecular clouds generated directly from a warm neutral medium? *Astrophys. J.* **704**, 161 (2009).
6. Clark, P. C., Glover, S. C. O., Ragan, S. E. & Duarte-Cabral, A. Tracing the formation of molecular clouds via [CII], [CI], and CO emission. *Mon. Not. R. Astron. Soc.* **486**, 4622–4637 (2019).
7. Dobbs, C. L., Liow, K. Y. & Rieder, S. The formation of young massive clusters by colliding flows. *Mon. Not. R. Astron. Soc. Lett.* **496**, L1–L5 (2020).
8. Haworth, T. J. et al. Isolating signatures of major cloud-cloud collisions using position-velocity diagrams. *Mon. Not. R. Astron. Soc.* **450**, 10–20 (2015).
9. Bisbas, T. G., Tanaka, K. E. I., Tan, J. C., Wu, B. & Nakamura, F. GMC collisions as triggers of star formation. V. Observational signatures. *Astrophys. J.* **850**, 23 (2017).
10. Fukui, Y., Habe, A., Inoue, T., Enokiya, R. & Tachihara, K. Cloud-cloud collisions and triggered star formation. *Publ. Astron. Soc. Jpn* **73**, S1–S34 (2021).
11. Fukui, Y. et al. Molecular clouds toward the super star cluster NGC3603; possible evidence for a cloud-cloud collision in triggering the cluster formation. *Astrophys. J.* **780**, 36 (2014).
12. Torii, K. et al. Cloud-cloud collision as a trigger of the high mass star-formation: a molecular line study in RCW120. *Astrophys. J.* **806**, 7 (2015).
13. Wolfire, M. G., Hollenbach, D. & McKee, C. F. The dark molecular gas. *Astrophys. J.* **716**, 1191–1207 (2010).
14. Pineda, J. L., Langer, W. D., Velusamy, T. & Goldsmith, P. F. A Herschel [CII] Galactic plane survey I. The global distribution of ISM gas components. *Astron. Astrophys.* **554**, A103 (2013).
15. Beuther, H. et al. Carbon in different phases ([CII], [CI], and CO) in infrared dark clouds: cloud formation signatures and carbon gas fractions. *Astron. Astrophys.* **571**, A53 (2014).
16. Franeck, A. et al. Synthetic [CII] emission maps of a simulated molecular cloud in formation. *Mon. Not. R. Astron. Soc.* **481**, 4277–4299 (2018).
17. Schneider, N. et al. A new view of the Cygnus X region. KOSMA ^{13}CO 2-1, 3-2, and ^{12}CO 3-2 imaging. *Astron. Astrophys.* **458**, 855–871 (2006).
18. Rygl, K. L. J. et al. Parallaxes and proper motions of interstellar masers toward the Cygnus X star-forming complex. I. Membership of the Cygnus X region. *Astron. Astrophys.* **539**, A79 (2012).
19. Wright, N. J., Drew, J. E. & Mohr-Smith, M. The massive star population of Cygnus OB2. *Mon. Not. R. Astron. Soc.* **449**, 741–760 (2015).

20. Reipurth, B. & Schneider, N. in *Handbook of Star Forming Regions: The Northern Sky* Vol. I (ed. Reipurth, B.) 36 (ASP Monograph Publications, 2008).

21. Nguyen-Luong, Q. et al. The scaling relations and star formation laws of mini-starburst complexes. *Astrophys. J.* **833**, 23 (2016).

22. Dickel, J. R., Dickel, H. R. & Wilson, W. J. The detailed structure of CO in molecular cloud complexes. II. The W75-DR21 region. *Astrophys. J.* **223**, 840–853 (1978).

23. Dobashi, K., Shimoikura, T., Katakura, S., Nakamura, F. & Shimajiri, Y. Cloud-cloud collision in the DR 21 cloud as a trigger of massive star formation. *Publ. Astron. Soc. Jpn* **71**, 12 (2019).

24. Risacher, C. et al. The upGREAT dual frequency heterodyne arrays for SOFIA. *J. Astron. Instrum.* **7**, 1840014 (2018).

25. Pabst, C. et al. Disruption of the Orion molecular core 1 by wind from the massive star zeta Orionis C. *Nature* **565**, 618 (2019).

26. Yamagishi, M. et al. Nobeyama 45 m Cygnus-X CO survey. I. Photodissociation of molecules revealed by the unbiased large-scale CN and C18O maps. *Astrophys. J. Supp.* **235**, 9 (2018).

27. Taylor, A. R. et al. The Canadian Galactic Plane survey. *Astrophys. J.* **125**, 3145 (2003).

28. Schneider, N. et al. Dynamic star formation in the massive DR21 filament. *Astron. Astrophys.* **520**, A49 (2010).

29. Hennemann, M. et al. The spine of the swan: a Herschel study of the DR21 ridge and filaments in Cygnus X. *Astron. Astrophys.* **543**, L3 (2012).

30. Pound, M. W. & Wolfire, M. G. The PhotoDissociation Region toolbox: software and models for astrophysical analysis. *Astronomical. J.* **165**, 25 (2023).

31. Bialy, S., Burkhart, B. & Sternberg, A. The HI-to-H2 transition in a turbulent medium. *Astrophys. J.* **843**, 92 (2017).

32. Jura, M. Formation and destruction rates of interstellar H2. *Astrophys. J.* **191**, 375 (1974).

33. Glover, S. C. O. & Mac Low, M.-M. Simulating the formation of molecular clouds. II. Rapid formation from turbulent initial conditions. *Astrophys. J.* **659**, 1317 (2007).

34. Clark, P. C., Glover, S. C. O., Klessen, R. S. & Bonnell, I. A. How long does it take to form a molecular cloud? *Mon. Not. R. Astron. Soc.* **424**, 2599 (2012).

35. Apellaniz, J. M. et al. The Villafranca catalog of Galactic OB groups II. From GAIA DR2 to EDR3 and ten new systems with O stars. *Astron. Astrophys.* **657**, A131 (2022).

36. Lallement, R. et al. GAIA-2MASS 3D maps of Galactic interstellar dust within 3 kpc. *Astron. Astrophys.* **625**, A135 (2022).

37. Godard, B. et al. Comparative study of CH+ and SH+ absorption lines observed towards distant star-forming regions. *Astron. Astrophys.* **540**, A87 (2012).

38. Inoue, T. et al. The formation of massive molecular filaments and massive stars triggered by a magnetohydrodynamic shock wave. *Publ. Astron. Soc. Jpn* **70**, 53 (2018).

39. Bonne, L. et al. Formation of the Musca filament: evidence for asymmetries in the accretion flow due to a cloud-cloud collision. *Astron. Astrophys.* **644**, A27 (2020).

40. Kabanovic, S. et al. Self-absorption in [C II], ^{12}CO, and H I in RCW120. Building up a geometrical and physical model of the region. *Astron. Astrophys.* **659**, A36 (2022).

41. Pineda, J. E. et al. From bubbles and filaments to cores and disks: gas gathering and growth of structure leading to the formation of stellar systems. Preprint at https://ui.adsabs.harvard.edu/abs/2022arXiv220503935P (2022).

42. Inutsuka, S. et al. The formation and destruction of molecular clouds and galactic star formation. An origin for the cloud mass function and star formation efficiency. *Astron. Astrophys.* **580**, A49 (2015).

43. Inoue, T. & Inutsuka, S. Formation of turbulent and magnetized molecular clouds via accretion flows of HI clouds. *Astrophys. J.* **759**, 35 (2012).

44. Abe, D. et al. The effect of shock wave duration on star formation and the initial condition of massive cluster formation. *Astrophys. J.* **940**, 106 (2022).

45. Dobbs, C. L., Bending, T. J. R. & Pettitt, A. R. The formation of clusters and OB associations in different density spiral arm environments. *Mon. Not. R. Astron. Soc.* **517**, 675 (2022).

46. Simon, R. et al. The cycling of matter from the interstellar medium to stars and back. *Bull. Am. Astron. Soc.* **51**, https://baas.aas.org/pub/2020n3i367 (2019).

47. Klein, B. et al. High resolution wide-band fast Fourier transform spectrometers. *Astron. Astrophys.* **542**, L3 (2012).

48. Schneider, N. et al. FEEDBACK: a SOFIA legacy program to study stellar feedback in regions of massive star formation. *Publ. Astron. Soc. Pac.* **132**, 104301 (2020).

49. Ossenkopf, V. et al. Herschel/HIFI observations of [CII] and [13CII] in photon-dominated regions. *Astron. Astrophys.* **550**, A57 (2013).

50. Goldsmith, P. F., Langer, W. D., Pineda, J. L. & Velusamy, T. Collisional excitation of the [CII] fine structure transition in interstellar clouds. *Astrophys. J. Supp.* **203**, 13 (2012).

51. Sofia, U. J., Lauroesch, J. T., Meyer, D. M. & Cartledge, S. I. B. Interstellar carbon in translucent sight lines. *Astrophys. J.* **605**, 272–277 (2004).

52. Hanson, M. M. A study of Cygnus OB2: pointing the way toward finding our galaxy's super-star clusters. *Astrophys. J.* **597**, 957 (2003).

53. Piepenbrink, A. & Wendker, H. J. The Cygnus X region. XVII. H 110 alpha and H2 CO line surveys with the 100 m-RT. *Astron. Astrophys.* **191**, 313–322 (1988).

Acknowledgements

Author contributions

N.S. and A.G.G.M.T. are the principal investigators (PIs) of the FEEDBACK project and prepared the SOFIA proposal. J.S. is the PI of the GREAT instrument. M.M. and O.R. are GREAT instrument scientists. R.S., C.B. and N.S. (plus other members of the GREAT team) performed the observations and reduced the [CII] data. C.B. developed the (self)-PCA method. N.S., L.B. and S.B. led the data interpretation and write-up. V.O.-O., S.K., A.G.G.M.T. and T.C. contributed in discussions. S.K. performed the FUV field calculations and the HISA study and provided plots for the error calculation. N.S. and V.O.-O. applied the data to the PDR toolbox and interpreted the results.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41550-023-01901-5.

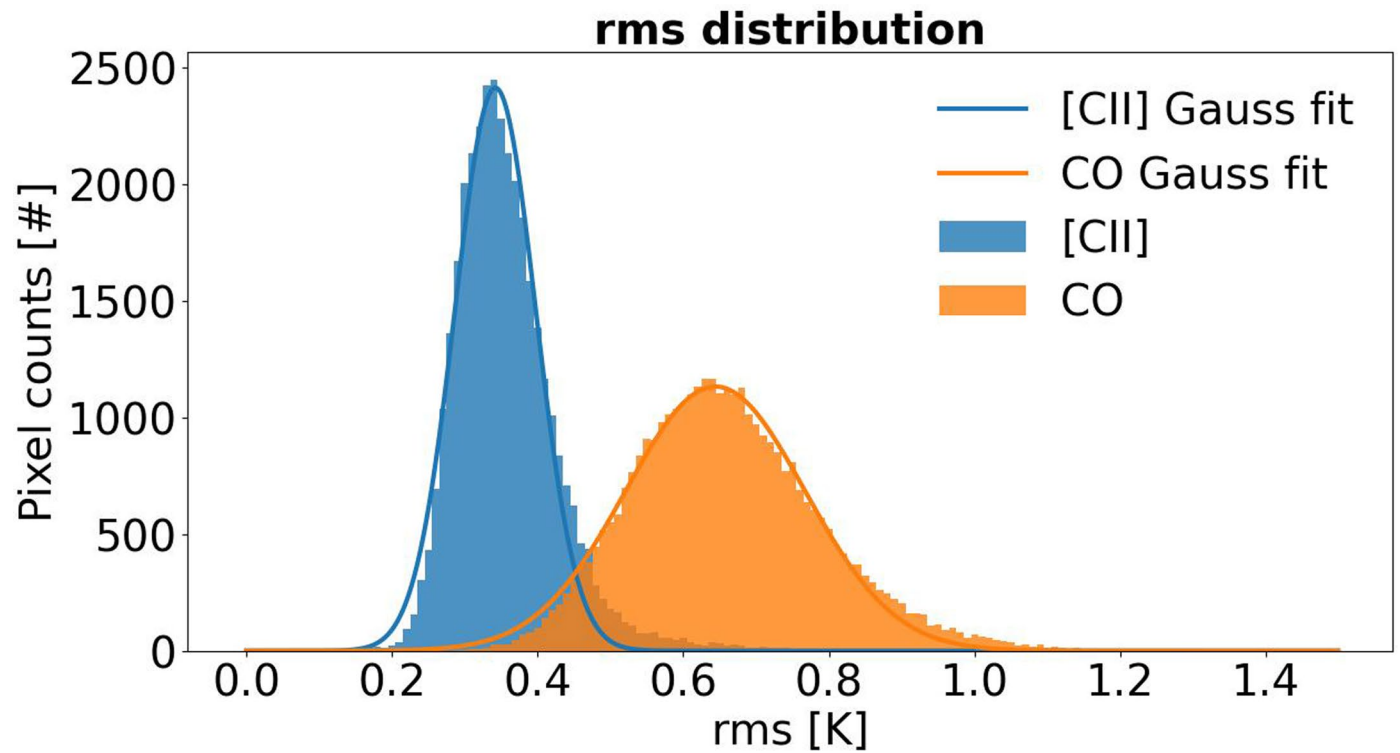**Correspondence and requests for materials** should be addressed to Nicola Schneider or Lars Bonne.

Peer review information *Nature Astronomy* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended Data Fig. 1 | Rms noise distribution. The noise is binned in 0.01 K bins for all map pixels in CII (blue) and CO (orange). The peaks of the Gaussian distributions are at 0.34 K and 0.64 K, respectively.
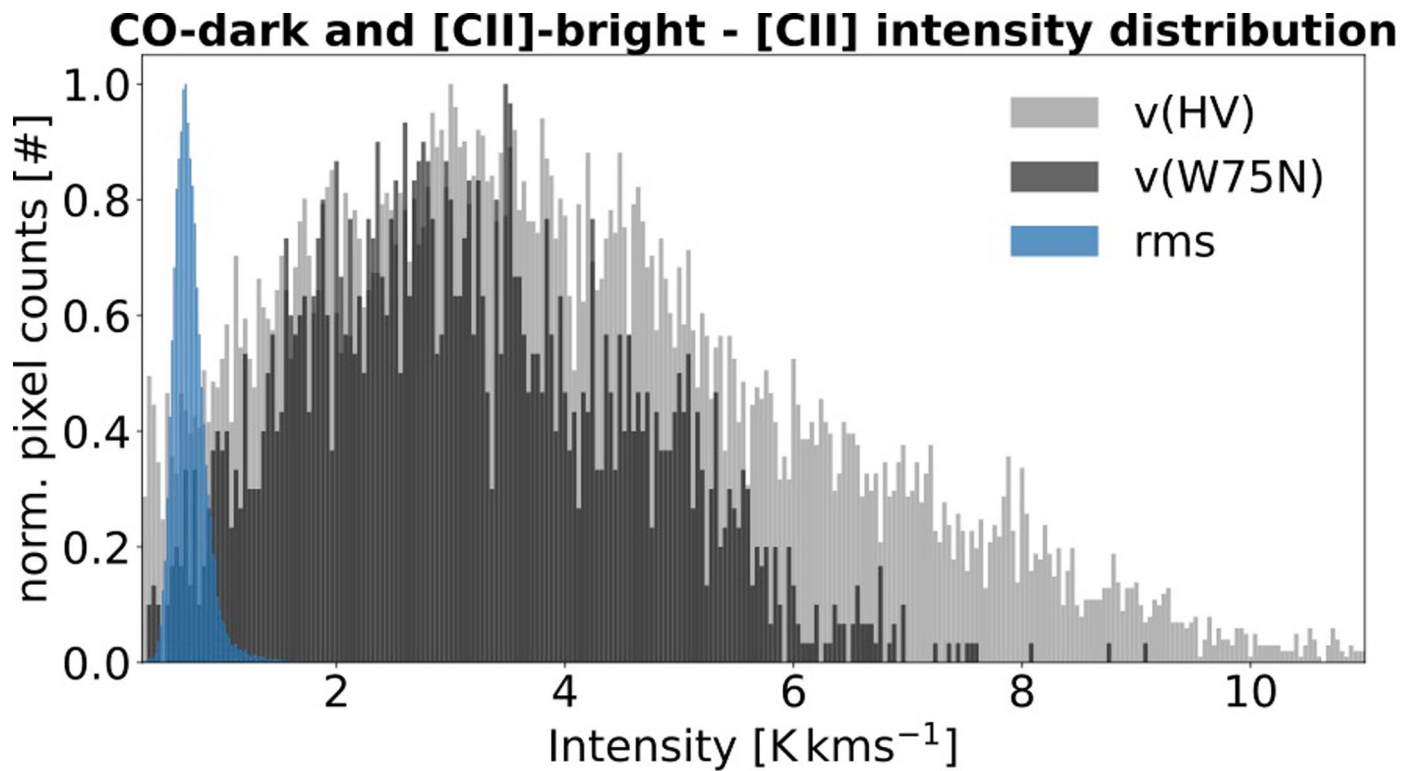
Extended Data Fig. 2 | Binned intensity normalized PDF. The binning for the probability distribution function is 0.02 K km/s for the rms noise and 0.04 K km/s for the CII intensity, respectively. The W75N pixels are indicated in dark grey, the HV range in light grey and the rms in blue.
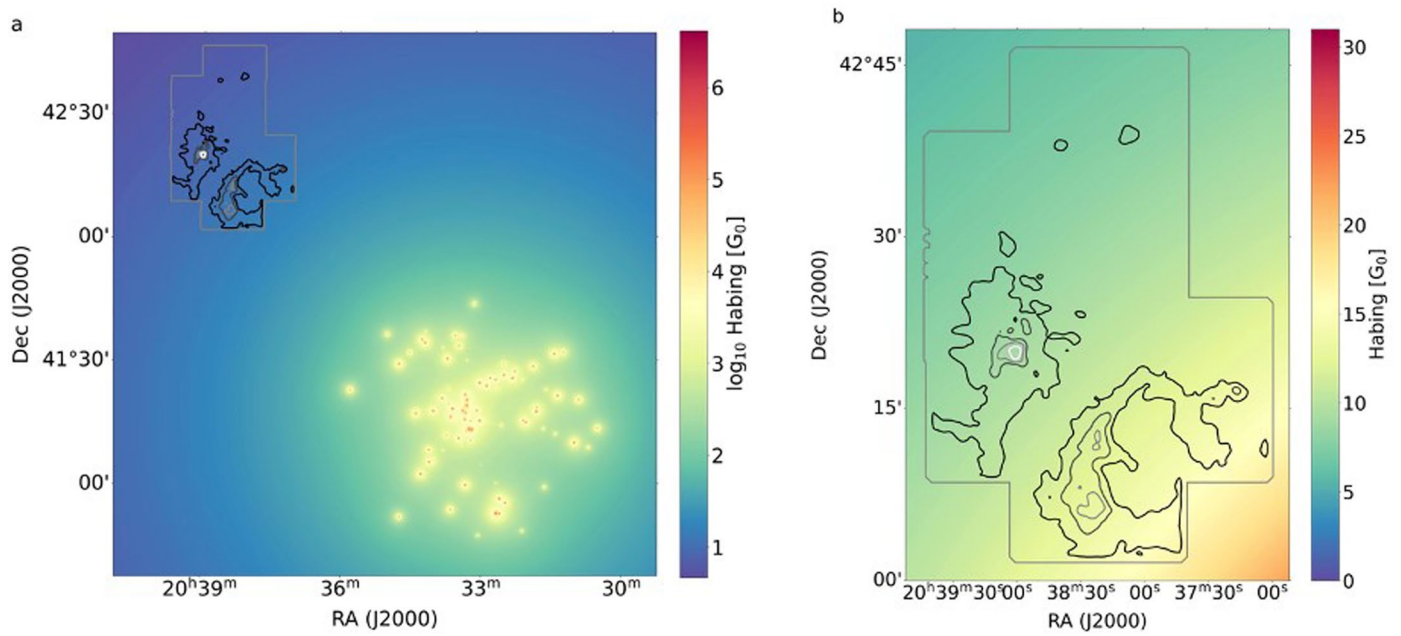
Extended Data Fig. 3 | Far-UV field in the Cygnus region. a, The large-scale FUV field on a log-scale in Habing units Go, determined from a census of the O and B-stars from the Cyg OB2 cluster that are indicated in the panel. The grey box outlines the area shown in panel **b**. The observed CII emission is overlaid with black contours, corresponding to 50 K km/s to 210 K km/s by 40 K km/s. **b**, The FUV field in the DR21 and W75N region on a linear scale with the CII contours identical to the ones in panel **a**.
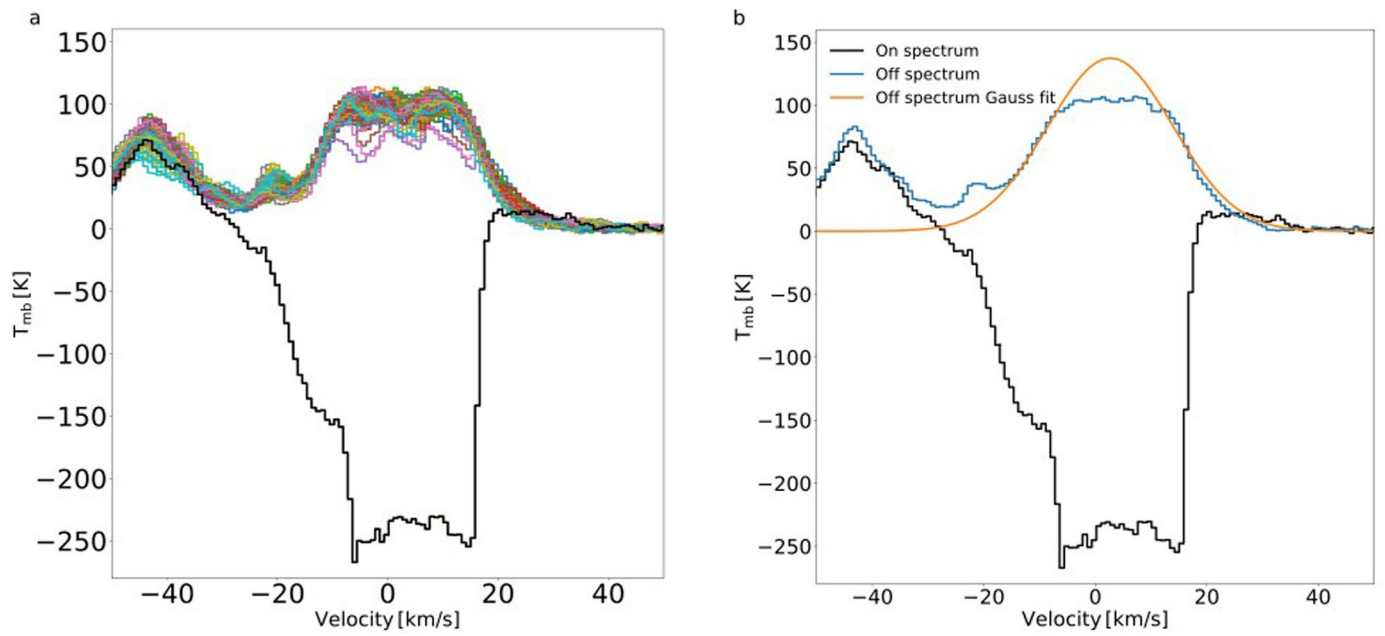
Extended Data Fig. 4 | HI spectra toward DR21. a, The HI absorption spectrum is shown in black, the 100 colored spectra represent the HI emission within a square of 20′ × 20′ around DR21 in a grid of 1′. **b**, The black curve is again the absorption HI spectrum toward DR21, the blue curve is the HI emission spectrum (off), and the orange curve shows the Gaussian fit to the blue off spectrum.
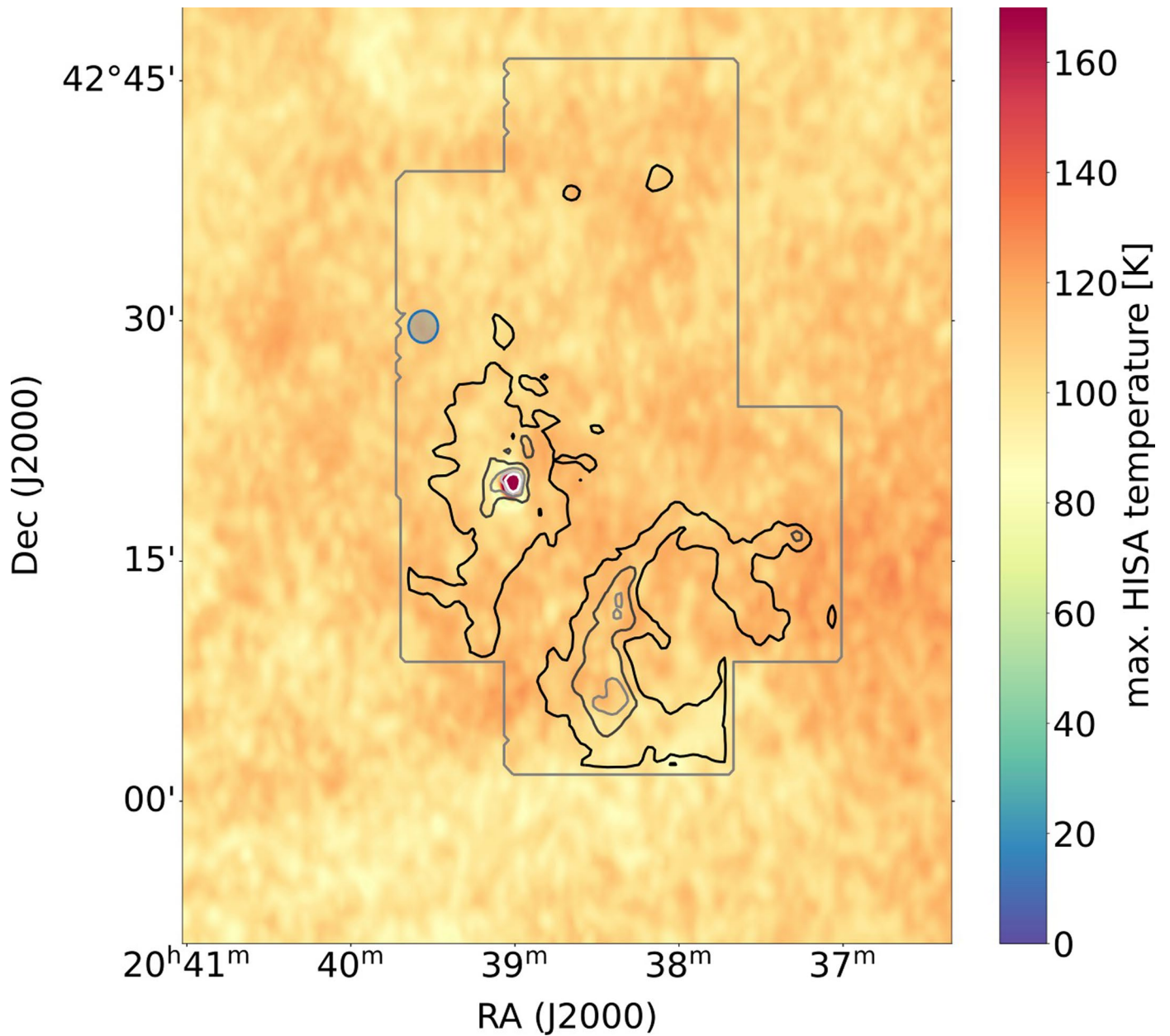
Extended Data Fig. 5 | Maximum possible HISA temperature around DR21. The strong continuum source DR21 stands out as a red spot with high temperatures of 175 K. Overall, the HISA temperature ranges between 90 and 120 K with an average of 100 K. The overlaid CII contours go from black to white with six contour levels corresponding to 50 K km/s to 210 K km/s by 40 K km/s. The blue filled circle indicates the location of the off position.
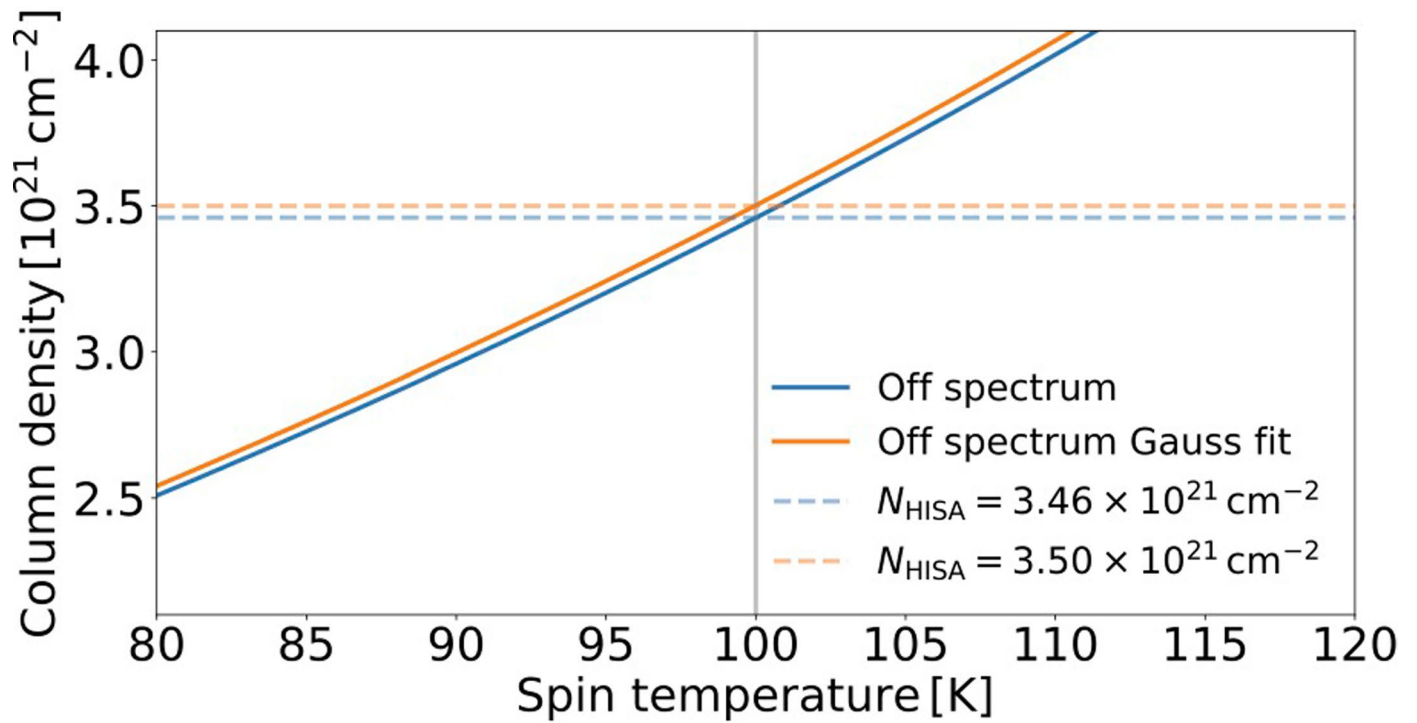
Extended Data Fig. 6 | HISA column density as a function of temperature. The solid blue curve shows the results for the off position and the solid orange curve for the Gaussian fit to the off position. The blue and orange dashed lines indicate the column density at 100 K for the off position and the Gaussian fit, respectively.