The column densities of molecular gas across cosmic time: bridging observations and simulations

Roland Szakacs, Céline Péroux, Martin A. Zwaan, Dylan Nelson, Eva Schinnerer, Natalia Lahén, Simon Weng, Alejandra Y. Fresco

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **512,** 4736–4751 (2022) https://doi.org/10.1093/mnras/stac510
Advance Access publication 2022 February 24

The column densities of molecular gas across cosmic time: bridging observations and simulations

Roland Szakacs,[1]★ Céline Péroux,[1,2] Martin A. Zwaan,[1] Dylan Nelson,[3] Eva Schinnerer,[4] Natalia Lahén,[5] Simon Weng[1,6,7] and Alejandra Y. Fresco[8]

[1]*European Southern Observatory (ESO), Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany*
[2]*Aix Marseille Université, CNRS, CNES, LAM, F-13007 Marseille, France*
[3]*Universitat Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany*
[4]*Max Planck Institute for Astronomy, K'onigstuhl 17, D-69117 Heidelberg, Germany*
[5]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85740 Garching, Germany*
[6]*Sydney Institute for Astronomy, School of Physics, University of Sydney, Sydney, NSW 2006, Australia*
[7]*ATNF, CSIRO Astronomy and Space Science, PO Box 76, Epping, NSW 1710, Australia*
[8]*Max-Planck-Institut für Extraterrestrische Physik (MPE), Giessenbachstr. 1, D-85748 Garching bei München, Germany*

Accepted 2022 February 16. Received 2022 February 16; in original form 2021 November 26

ABSTRACT

Observations of the cosmic evolution of different gas phases across time indicate a marked increase in the molecular gas mass density towards $z \sim 2$–3. Such a transformation implies an accompanied change in the global distribution of molecular hydrogen column densities ($N_{\rm H_2}$). Using observations by PHANGS-ALMA/SDSS and simulations by GRIFFIN/IllustrisTNG we explore the evolution of this H_2 column density distribution function [$f(N_{\rm H_2})$]. The H_2 (and H I) column density maps for TNG50 and TNG100 are derived in post-processing and are made available through the IllustrisTNG online API. The shape and normalization of $f(N_{\rm H_2})$ of individual main-sequence star-forming galaxies are correlated with the star formation rate (SFR), stellar mass (M_*), and H_2 mass ($M_{\rm H_2}$) in both observations and simulations. TNG100, combined with H_2 post-processing models, broadly reproduces observations, albeit with differences in slope and normalization. Also, an analytically modelled $f(N)$, based on exponential gas discs, matches well with the simulations. The GRIFFIN simulation gives first indications that the slope of $f(N_{\rm H_2})$ might not majorly differ when including non-equilibrium chemistry in simulations. The $f(N_{\rm H_2})$ by TNG100 implies that higher molecular gas column densities are reached at $z = 3$ than at $z = 0$. Further, denser regions contribute more to the molecular mass density at $z = 3$. Finally, H_2 starts dominating compared to H I only at column densities above $\log(N_{\rm H_2}/{\rm cm}^{-2}) \sim 21.8$–22 at both redshifts. These results imply that neutral atomic gas is an important contributor to the overall cold gas mass found in the ISM of galaxies including at densities typical for molecular clouds at $z = 0$ and 3.

Key words: ISM: atoms – ISM: evolution – ISM: molecules – galaxies: evolution – quasars: absorption lines.

1 INTRODUCTION

While the total amount of baryons in the Universe ($\Omega_{\rm baryons} = \rho_{\rm baryons}/\rho_{\rm crit,\,0} \sim 4$ per cent, where $\rho_{\rm crit,\,0}$ is the critical density of the Universe) is well established from measurements of anisotropies in the Cosmic Microwave Background (Planck Collaboration VI 2016) and from primordial nucleosynthesis (Cooke, Pettini & Steidel 2018), the contribution and evolution of different gas phases remain to be probed. Especially constraints on the evolution of the phases most closely linked to star formation, namely the neutral atomic and molecular gas phases, are limited. None the less, recent observations have shown first indications of how these gas phases are evolving.

The neutral atomic gas phase shows little evolution with redshift, with its comoving baryonic mass density only slightly declining as the redshift decreases [$\rho_{\rm neutral\text{-}gas} \sim (1+z)^{0.57\pm0.04}$] (e.g. Wolfe, Gawiser & Prochaska 2005; Noterdaeme et al. 2009; Crighton et al. 2015; Jones et al. 2018; Péroux & Howk 2020; Tacconi, Genzel & Sternberg 2020; Walter et al. 2020). This can be traced by the column density distribution function [$f(N_{\rm H\,I})$] across cosmic time, which describes the number of H I systems per unit column density per unit distance interval. $f(N_{\rm H\,I})$ quantifies the distribution of H I column densities on the sky and by integrating $f(N_{\rm H\,I})$ one can compute the comoving H I mass density. While $f(N_{\rm H\,I})$ of various shapes can result in the same $\rho_{\rm neutral\text{-}gas}$, H I-absorption in quasar spectra and emission-line measurements have revealed that the $f(N_{\rm H\,I})$ shows little to no evolution, either in shape or in normalization (e.g. Zwaan et al. (e.g. Péroux et al. 2005; Zwaan et al. 2005; Zafar et al. 2013; Ho, Bird & Garnett 2021).

Observations calculating the comoving molecular mass density, on the other hand, have indicated a more radical evolution of the gas phase crucially needed for star formation. The comoving mass density of H_2 rises until cosmic noon ($z \sim 2$–3) where it peaks and drops towards $z = 0$ (e.g. (e.g. Liu et al. 2019; Popping et al. 2019; Riechers et al. 2019; Decarli et al. 2020; Péroux & Howk 2020; Tacconi et al. 2020; Walter et al. 2020). Given this evolution

★ E-mail: Roland.Szakacs@eso.org

of the H_2 comoving mass density over cosmic time, changes in the normalization or shape of $f(N_{H_2})$ can be expected.

Globally, the neutral atomic gas mass density is higher than that of the molecular phase (Péroux & Howk 2020; Tacconi et al. 2020; Walter et al. 2020), but $f(N)$ helps reveal in which type of objects the neutral and molecular gas lies. H I-absorbers can be split into different categories from the Ly α forest for column densities $N_{HI} \leq 1.6 \times 10^{17} cm^{-2}$, to Lyman-limit systems (LLSs; $1.6 \times 10^{16} \leq N_{HI} \leq 10^{19} cm^{-2}$), to sub-damped Ly α absorbers (sub-DLAs, $10^{19} \leq N_{HI} \leq 2 \times 10^{20} cm^{-2}$), up to Damped Ly α absorbers (DLAs; $N_{HI} \geq 2 \times 10^{20} cm^{-2}$). The association between these systems and their origin is still challenging, but various works have kinematically associated LLSs, sub-DLAs, and DLAs to environments like parts of the extended rotating discs, inflows and outflows of galaxies (e.g. Rahmani et al. 2018a,b; Schroetter et al. 2019; Zabl et al. 2020; Szakacs et al. 2021). H I emission-line studies on the other hand (e.g. Zwaan et al. 2005; Braun 2012; French et al. 2021) can easily associate column densities with regions of galaxies like the interstellar medium (ISM) as the galaxies are completely imaged down to a given sensitivity instead of individual pencil beams. While the gas mass densities and $f(N)$ are global properties including multiple objects, comparing $f(N_{H1})$ and $f(N_{H_2})$ gives an indication in which regions of galaxies (e.g. the ISM, CGM, Molecular Clouds) neutral atomic or molecular gas dominates on average. The typical cold gas column densities for these regions are the following: Molecular Clouds: $N \geq \times 10^{20.8} cm^{-2}$ (e.g. Spilker, Kainulainen & Orkisz 2021), ISM: $N \geq \times 10^{19} cm^{-2}$, CGM:$N \geq \times 10^{14}$–$10^{19} cm^{-2}$ (e.g. van de Voort et al. 2019). Therefore, this helps us understand if neutral atomic gas is an important mass contributor in the ISM compared to molecular gas or if it is only substantial in the haloes surrounding galaxies.

Today's state-of-the-art cosmological simulations enable the study of physical processes of galaxy formation for both the dark matter and baryonic component of the Universe. The results of these simulations are compared to observables to learn how well the model fits. A limitation of these simulations is that due to their large volume, the scales at which these physical processes and observables can be resolved is limited so that subgrid models are used. The advantage that these simulations offer is the large statistical sample, as thousands of galaxies are simulated. Recently, there have been considerable efforts in modelling the cold gas phase by post-processing these simulations (e.g. Lagos et al. 2015; Diemer et al. 2018; Popping et al. 2019). While properties of cold gas in these simulations show various levels of (dis)agreement with observations [e.g. a higher cosmic mass density of H I and H_2 in IllustrisTNG compared to observations at $z = 0$ (Diemer et al. 2019), tensions concerning the the cosmic metal density evolution in neutral gas in EAGLE, IllustrisTNG, and L-GALAXIES 2020 (Yates, Péroux & Nelson 2021), the lower molecular mass as a function of stellar mass and number of H_2 rich galaxies in IllustrisTNG compared to the ASPECS survey (Popping et al. 2019)], other observables, like the H I column density distribution function have been accurately reproduced (Rahmati et al. 2013). Therefore, further studies and comparisons of these and similar observables, like the $f(N_{H_2})$, are needed to improve the models and to design future observations.

The goal of this study is to probe the evolution of $f(N_{H_2})$ across cosmic time. For this we compare data from observations on one hand and isolated and cosmological (magneto-)hydrodynamical simulations on the other hand. In the past, $f(N_{H_2})$ has been studied using CO emission lines at low-z (Zwaan & Prochaska 2006) and more recently by studying composite H_2 QSO absorption spectra at $z \sim 3$ (Balashev & Noterdaeme 2018). High-resolution CO emission-line observations of local galaxies by the PHANGS-ALMA survey (Leroy et al. 2021) now enable us to derive $f(N_{H_2})$ using emission lines from galaxies on scales of Giant Molecular Clouds (GMCs). Further, state-of-the-art hydrodynamical simulations including non-equilibrium chemical networks tracking H_2 on-the-fly in high-resolution dwarf galaxy simulations (Hu et al. 2014, 2016, 2017; Lahén et al. 2019, 2020a,b; Hislop et al. 2021) and post processing the TNG100 cosmological magnetohydrodynamical simulation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018) enable bridging observations and simulations. Finally, we aim to compare $f(N_{HI})$ and $f(N_{H_2})$ to provide indications of the regions of galaxies (e.g the ISM, CGM, Molecular Clouds) in which the molecular or neutral atomic gas phases dominate.

This paper is organized as follows: Section 2 describes the column density distribution function $f(N)$. Section 3 presents the observational setup as well as the simulations used for the analysis of $f(N_{H_2})$. Section 4 describes the resolution dependence of $f(N_{H_2})$. Section 5 presents the $f(N_{H_2})$ of individual galaxies in the PHANGS-ALMA survey and their correlations with integrated physical properties of the galaxy. Section 6 presents the results of the key goal of this manuscript. We describe the redshift evolution of the $f(N_{H_2})$ derived from both observations and simulations and study their differences and similarities across cosmic time. Further, we compare $f(N_{H_2})$ with $f(N_{HI})$ in order to explore at which densities neutral atomic gas dominates over molecular gas in and surrounding galaxies. In Section 7, we discuss our results from the previous sections. Finally, in Section 8, we give a summary of the findings. Throughout this paper, we adopt an $H_0 = 67.74 km\ s^{-1} Mpc^{-1}$, $\Omega_M = 0.3089$, and $\Omega_\Lambda = 0.6911$ cosmology.

2 QUANTIFYING THE DISTRIBUTION OF COLUMN DENSITIES OBSERVED ON THE SKY

Column densities of different chemical species or different phases are not distributed uniformly on the sky as low-density gas is more frequent within our Universe. One way to quantify the distribution of column densities is the so-called column density distribution function $f(N_s)$. It is defined such that $f(N_s)dN_s dX$ is the number of systems with column densities between N_s and $N_s + dN_s$ over a distance interval dX, where s is the species one is studying (e.g. H I or H_2). While in the past $f(N_s)$ have been mostly studied using absorption systems, high-resolution data of emission lines in galaxies enable an alternative way of studying the column density distribution function. Using emission-line observations one can calculate the $f(N_s)$ as follows (e.g. Zwaan et al. 2005; Zwaan & Prochaska 2006):

$$f(N_s) = \frac{c}{H_0} \frac{\sum_i \Phi(x_i) w(x_i) A_i(\log(N_s))}{N_s \ln(10) \Delta \log(N_s)}. \quad (1)$$

We bin the galaxies of our samples by their stellar mass, with a bin size of $\Delta \log(M_{*,i}/M_\odot) = 0.2$. $\Phi(M_{*,i})$ is the stellar mass function with $M_{*,i}$ being the central stellar mass value of the bin i the corresponding galaxy is in. $w(M_{*,i}) = 1/(N_{gal,i})$ is a weighting function taking into account the varying number of galaxies across the range $\log(M_{*,i}/M_\odot) - \Delta \log(M_{*,i}/M_\odot)/2$ to $\log(M_{*,i}) + \Delta \log(M_{*,i}/M_\odot)/2$ by calculating the reciprocal of the number of galaxies within the stellar mass bin i. $A_i(\log(N_s))$ is the area function describing the area corresponding to a column density in the range $\log(N_s)$ to $\log(N_s) + \Delta \log(N_s)$ for stellar mass bin i in Mpc^2. We use $\Delta \log(N_{H_2}) = 0.1$ in our calculations of $f(N_{H_2})$. Finally, we convert the number of systems per Mpc into that per unit redshift using c/H_0. The column

densities in this work are in units of H_2 molecules cm^{-2} for molecular gas and H atoms cm^{-2} for neutral atomic gas. The space densities and areas are proper.

3 BRIDGING OBSERVATIONS AND SIMULATIONS

In order to study the H_2 column density distribution [$f(N_{H_2})$] at $z = 0$ and 3, we use an approach that bridges observations and simulations. We study how the state-of-the-art simulations compare to recent observations and explore if a similar evolution of the $f(N_{H_2})$ can be seen in both of the approaches. On the observational side we use data from the PHANGS-ALMA survey[1] (see Section 3.1, Leroy et al. 2021) at $z = 0$ and data from Balashev & Noterdaeme (2018) (SDSS; see Section 3.2) at $z = 3$. On the simulation side we use TNG100 of the IllustrisTNG project[2] (see Section 3.3; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018) at both redshifts $z = 0$ and 3 and a high-resolution isolated dwarf galaxy simulation including a non-equilibrium chemical network from the GRIFFIN Project[3] (see Section 3.4, Lahén et al. 2019, 2020a,b) meant to represent a low-redshift dwarf galaxy.

3.1 Resolved Molecular Gas in Local Galaxies

State-of-the-art mm- and radio-telescopes like the Atacama Large Millimeter/submillimeter Array (ALMA) have enabled the astronomical community to study the coldest gas in the Universe with unprecedented spatial and spectral resolution. One of the surveys making use of these technological advances is the PHANGS-ALMA survey (Leroy et al. 2021). This survey is the first cloud-scale (∼100 pc) survey aimed at studying the physics of molecular gas within the local galaxy population and targets galaxies that lie on or near the $z = 0$ main sequence of star-forming galaxies with a stellar mass range of $10^9 < M_* < 10^{11}$ $M_\odot$. PHANGS-ALMA quantifies the physics of star formation and feedback at giant molecular cloud scales and further connects them to galaxy-scale properties and processes (Leroy et al. 2021). Further, additional state-of-the-art multiwavelength data are provided by the PHANGS-MUSE (Emsellem et al. 2021) and PHANGS-*HST* surveys (Lee et al. 2021), which will study the ionized gas, stellar populations, and characterize stellar clusters of the objects observed by the PHANGS-ALMA survey.

We make use of the highly resolved CO(2–1) data of the PHANGS-ALMA Survey (Leroy et al. 2021) in order to constrain the global and local $f(N_{H_2})$ in the range $\log(N_{H_2}/cm^{-2}) \sim 19.5$–24 at $z = 0$. We use a pixel-by-pixel analysed sample consisting of 70 galaxies from Sun et al. (2020). The stellar mass distribution of the sample can be seen in Fig. 1. In summary, the CO(2–1) data were analysed by Sun et al. (2020) as follows: The cubes were convolved to a common spatial resolution of 150 pc and 1 kpc. Then the data cubes were masked to only include voxels that contain emission detected with high confidence. Those cubes were finally integrated to create integrated intensity maps. The integrated maps were then used to

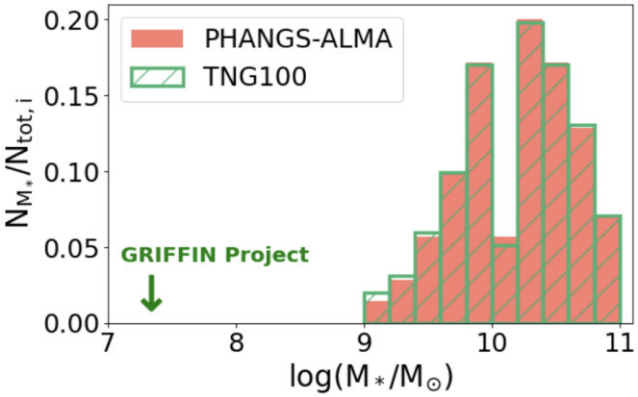

Figure 1. The stellar mass distribution of the PHANGS-ALMA survey (red) and the matching TNG100 sample (green-hatched). For each stellar mass bin we additionally only select galaxies in TNG100 with similar SFRs as found in the corresponding stellar mass bin in PHANGS-ALMA.

derive the molecular gas surface density for each pixel as follows:[4]

$$\Sigma_{mol} = \alpha_{CO}\ R_{21}^{-1}\ I_{CO}, \tag{2}$$

here $R_{21} = 0.65$ is the CO(2–1)-to-CO(1–0) line ratio (Leroy et al. 2013; den Brok et al. 2021) and α_{CO} is the metallicity-dependent CO-to-H_2 conversion factor taken as

$$\alpha_{CO} = 4.35\ Z'^{-1.6}\ M_\odot\ pc^{-2}\ (K\ km\ s^{-1})^{-1}, \tag{3}$$

where Z' is the local ISM metallicity in units of the solar value. The local Z' is estimated using the global stellar mass, effective radius and the stellar mass metallicity relation by Sánchez et al. (2019) combined with a metallicity gradient (Sánchez et al. 2014). For more details, see Sun et al. (2020). For the error calculation, we additionally compute the surface density using the constant $\alpha_{CO} = 4.3\ M_\odot\ pc^{-2}\ (K\ km\ s^{-1})^{-1}$ of the Milky Way (Bolatto, Wolfire & Leroy 2013). Measurement uncertainties are omitted as they are negligible compared to the uncertainties of the different α_{CO} conversion factors used.

With this sample we are able to constrain $f(N_{H_2})$ at $z = 0$. We convert the derived surface densities to column densities using

$$N_{H_2} = \frac{\Sigma_{H_2}}{M_{H_2\text{-molecule}}}, \tag{4}$$

with Σ_{H_2} in units of $kg\ cm^{-2}$.

We then follow equation (1) to calculate $f(N_{H_2})$ and use two stellar mass functions by Weigel, Schawinski & Bruderer (2016) as our space density function. The first stellar mass function is that of the entire sample, and the second is one for late-type galaxies only as the PHANGS sample mostly consists of late-type galaxies on the star-forming main sequence (see table 5 in Weigel et al. 2016 for the Schechter parameters).

Sun et al. (2018) estimates the 100 per cent completeness surface density limit for a subsample of galaxies in the PHANGS-ALMA sample to be $\log(\Sigma_{H_2}/M_\odot pc^{-2}) = 10$–100 at 120-pc resolution. This translates to a column density completeness limit of $\log(N_{H_2}/cm^{-2}) = 20.8$–21.8. At 150-pc resolution, the completeness limit is expected to be lower. We therefore use a conservative estimate of

[1] sites.google.com/view/phangs/home.
[2] tng-project.org.
[3] wwwmpa.mpa-garching.mpg.de/~naab/griffin-project/.

[4] Surface density table for 150 pc can be found at canfar.net/storage/list/phangs/RELEASES/Sun_etal_2020b, datafileB1. 1 kpc table provided by authors of Pessa et al. (2021).

$\log(N_{H_2}/cm^{-2}) = 21.6$ for 100 per cent completeness of the full PHANGS-ALMA sample.

3.2 Absorption lines as a probe for the H_2 column density distribution at high redshifts

At high redshifts, it is currently challenging to observe H_2 directly or resolve CO emission lines in galaxies at spatial scales similar to the PHANGS-ALMA survey. Therefore one has to resort to another approach to study the H_2 column density distribution. H_2 imprints resonant electronic absorption bands in the UV and so studying absorption systems is a promising way of studying $f(N_{H_2})$ at high redshifts. H_2 absorption lines are usually found within Ly α absorption systems, so called Damped-Lyman Alpha absorbers (DLAs). It is time consuming to detect these H_2 absorbers, due to the low detection rate of $\leq$10 per cent. For these reasons, Balashev & Noterdaeme (2018) use composite spectra of DLAs by Mas-Ribas et al. (2017), which are based on ~27 000 DLAs from SDSS (Noterdaeme et al. 2012) in order to detect the weak mean signature of H_2 at $z \sim 3$. Balashev & Noterdaeme (2018) revert to these composite H_2 spectra in order to fit an $f(N_{H_2})$ in the range $\log(N_{H_2}/cm^{-2}) = 18$–$22$ on which, in turn, they fit the observed composite line profiles.

3.3 Cosmological simulations providing large statistical samples

Cosmological simulations provide large statistical samples for studies of galaxy evolution. One of these simulations is TNG100 of the IllustrisTNG project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018). TNG100 is a state-of-the-art gravomagnetohydrodynamics (MHD) cosmological simulation including a comprehensive model for galaxy formation physics (Weinberger et al. 2017; Pillepich et al. 2018) within a 75 000 ckpc h^{-1} sized box using the AREPO code (Springel 2010). IllustrisTNG aims to study the physical processes that drive galaxy formation and to study how galaxies evolve within large-scale structures.

We aim to exploit the large sample size of the TNG100 simulation in order to compare the observed $f(N_{H_2})$ at $z = 0$ and 3. While TNG50 offers higher resolution, we choose TNG100 as our fiducial model due to two reasons: (1) The SMF, which is an important parameter in our calculations, is closer to observations for TNG100 than for TNG50; and (2) to enable future comparisons with the EAGLE cosmological simulation (Schaye et al. 2015), as TNG100 is the closest in terms of resolution to the EAGLE 100 Mpc box simulation.

The molecular gas phase in current large-scale cosmological simulations is challenging to assess. Using chemical networks to track H_2 on-the-fly is computationally time consuming due to the complex physics involved and the high resolution needed in order for the H_2 mass fraction to converge within the forming molecular clouds (~0.12 pc, Seifried et al. 2017). In order to capture the unresolved physics, one has to revert to subgrid models, which split the cold hydrogen component in the simulations into a neutral atomic and molecular component. We use the H_2 post-processing catalogues of Popping et al. (2019) for TNG100, for which three different models are available. The used models are by Blitz & Rosolowsky (2006), Gnedin & Kravtsov (2011), Krumholz (2013). The model by Blitz & Rosolowsky (2006) is a pressure-based empirical fit based on a sample of 14 local spiral and dwarf galaxies that have measured atomic, molecular and stellar surface densities. Using this sample, they find a nearly linear relation between the hydrostatic pressure and the ratio of molecular to atomic gas. Gnedin & Kravtsov (2011) designed a phenomenological model for the formation of molecular hydrogen, which is dependent on the gas density, dust-to-gas ratio, and the far-UV radiation flux. This model was tested on cosmological simulations by Gnedin, Tassis & Kravtsov (2009). Finally, the model by Krumholz (2013) is a column density, metallicity, and radiation field dependent relation for splitting the cold hydrogen component in simulations.

In our study, we select central galaxies at $z = 0$ and 3. We add the ability to generate post-processed H_2 column density (N_{H_2}) maps based on Popping et al. (2019) for TNG50 and TNG100 using the subgrid models mentioned above, which split the cold hydrogen within those galaxies into atomic and molecular components, to the IllustrisTNG online API.[5] We use this functionality to create H_2 column density (N_{H_2}) maps at 150-pc and 1-kpc resolution. These maps are generated by projecting gas cells as adaptively sized SPH kernels. The kernel size parameter is set to $h_{kernel} = 2.5 r_{cell}$. With r_{cell} being the cell size determined by using the Voronoi cell volume: $r_{cell} = (3V_{cell}/4\pi)^{1/3}$. We use the same projection direction for every galaxy (z-axis of the simulation) and only consider gas cells gravitationally bound to the selected subhalos within a fixed 200 × 200 kpc box. This method reproduces $f(N_{H_2})$ derived from the full box of TNG100 when using the same resolution for the derivation as in Klitsch et al. (2019). Therefore, we do not expect this choice to affect $f(N_{H_2})$ as compared to using a box size dependent on halo properties. At $z = 0$, we selected a PHANGS-ALMA survey-like sample within TNG100. For this, we select ~570 galaxies within a stellar mass of 10^9 and 10^{11} $M_\odot$. We match the PHANGS-ALMA sample stellar mass distribution (see Fig. 1). Further, for each 0.2-dex stellar mass bin, we select galaxies with similar star formation rates (SFRs) as in the PHANGS-ALMA sample. We note that when selecting a sample of ~700 central galaxies with stellar masses between 10^9 and $10^{12.6}$ $M_\odot$ and no star formation selection criterion, we derive very similar results in TNG100 as for the PHANGS-ALMA-like sample, with the only differences being the slightly higher column densities reached in the larger sample (~0.2 dex higher) and a slightly higher normalization at column densities above $\log(N_{H_2}/cm^{-2})$ ~21.5, which is likely due to the larger size of the additional galaxies. At $z = 3$, we select ~550 galaxies with stellar masses between 10^9 and $10^{11.8}$ $M_\odot$. Therefore, this includes galaxies between the resolution limit of TNG100 up to the highest stellar mass limit of TNG100. We do not set any constraints on the SFR of the galaxies, as the observational $f(N_{H_2})$ is based on H_2 absorption line studies, where we do not have any SFR information.

Following equation (1), we use the derived H_2 column density maps to calculate $f(N_{H_2})$. For the space density function, we use the stellar mass function (SMF) of the simulation box itself. At $z = 0$, the Schechter parameters are $\log(M^*/M_\odot) = 11.27$, $\log(\Phi_1^*/h^3\ Mpc^{-3}) = -3.31$, $\log(\Phi_2^*/h^3\ Mpc^{-3}) = -3.28$, $\alpha_1 = -1.36$, $\alpha_2 = -1.36$. At $z = 3$, the double Schechter parameters are $\log(M^*/M_\odot) = 10.83$, $\log(\Phi_1^*/h^3\ Mpc^{-3}) = -3.84$, $\log(\Phi_2^*/h^3\ Mpc^{-3}) = -3.58$, $\alpha_1 = -0.29$, $\alpha_2 = -1.64$.

3.4 Molecular gas in highly resolved simulations of individual galaxies

An alternative approach to studying molecular gas in simulations is to use highly resolved simulations of individual isolated galaxies,

[5] tng-project.org/api/.

which include non-equilibrium chemical networks that track H_2 on-the-fly throughout the simulation. Although this is currently mostly limited to dwarf galaxies, the advantage of these simulations is a more accurate representation of H_2 due to a non-equilibrium chemical network.

One of these simulations is the high-resolution isolated dwarf simulation from the GRIFFIN Project (Lahén et al. 2019, 2020a,b) with a stellar mass of $\log(M_*/M_\odot) \sim 7.3$. The simulation is based on the smoothed particle hydrodynamics tree code GADGET-3 (Springel 2005) with the gas dynamics modelled using the SPH implementation SPHGal (Hu et al. 2014, 2016, 2017). The simulation resolves individual massive stars at sub-pc resolutions and includes a non-equilibrium chemical network based on Nelson & Langer (1997), Glover & Mac Low (2007a), and Glover & Clark (2012). The chemical network follows the abundances of six chemical species for cooling processes at low temperatures ($<3 \times 10^3$ K, most importantly H_2). Further, the simulation includes star formation, an interstellar radiation field, and stellar feedback prescriptions. A detailed discussion of the isolated dwarf simulation is given in Hu et al. (2016), Hu et al. (2017).

For the calculation of $f(N_{H_2})$, we time- and inclination-average the isolated dwarf galaxy simulation. Therefore, we produce H_2 column density maps with all possible lines of sight and slightly varying total H_2 masses using the analysis tool PYGAD (Röttgers et al. 2020). First, we create H_2 column density maps by using snapshots over a time range of ∼300 Myr. For each of these snapshots, we create H_2 column density maps at resolution of 150 pc with inclinations between 0° and 90° in $\Delta\cos(i) = 0.05$ steps. We then follow equation (1) to calculate $f(N_{H_2})$ by using these H_2 column density maps and use the Weigel et al. (2016) SMF of the entire sample for the normalization of $f(N_{H_2})$ following the prescription described in Section 2 (for the Schechter parameters, see table 5 in Weigel et al. 2016). The $f(N_{H_2})$ is therefore calculated using a single stellar mass bin (as the stellar mass of the simulated dwarf galaxy does not evolve much over time). However, galaxies of this stellar mass are not represented in the PHANGS-ALMA and TNG100 sample, so we can not directly compare the $f(N_{H_2})$ of similar galaxies.

4 A RESOLUTION-DEPENDENT H_2 COLUMN DENSITY DISTRIBUTION FUNCTION

We test how $f(N_{H_2})$ depends on the resolution of the data used for its calculation. First, we study how $f(N_{H_2})$ depends on the resolution of the simulation by comparing $f(N_{H_2})$ in TNG50 and TNG100 from the Illustris project. Then we compare how the resolution of the N_{H_2}-maps from both observations and simulations affects $f(N_{H_2})$.

4.1 $f(N_{H_2})$ – dependence on the resolution of simulations

Here we compare the $f(N_{H_2})$ derived from TNG100 with TNG50 (Pillepich et al. 2019; Nelson et al. 2019b) using the Gnedin & Kravtsov (2011) H_2 model. TNG50 has a box length of 51.7 Mpc and 2×2160^3 resolution elements, while TNG100 has a box length of 110.7 Mpc and 2×1820^3 resolution elements. Therefore, TNG50 gives us an indication how a higher resolution simulation affects $f(N_{H_2})$.

In Fig. 2 (left-hand panel), the $f(N_{H_2})$ derived from TNG100 and TNG50 at $z = 0$ using a 150-pc resolution of the post-processed column density map are displayed. TNG50 extends to higher column densities compared to TNG100. The finer resolution reaches higher gas densities and in turn higher column densities. Further, at column densities above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 22$, the $f(N_{H_2})$ in TNG50 initially displays a steep drop with a subsequent flattening of the $f(N_{H_2})$. These differences indicate that the H_2 column densities are not converged in this region. Given these difference,s we would expect higher resolution simulations to reach even higher column densities, and possibly also affect the shape in the region above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 22$. We note that IllustrisTNG uses the subgrid model of Springel & Hernquist (2003) for the star-forming ISM. Independent of resolution, the subgrid model begins star formation at ISM densities of 0.1 cm^{-3} preventing the simulation from resolving the cold gas phase and subsequently the formation of molecular clouds. Due to this, the model itself is limited by the subgrid ISM model and a higher resolution is only sensitive up to the limitations of the model. The resolution tests, however, indicate that the subgrid model is not the limiting factor in terms of densities reached at the resolution of TNG100 since $f(N_{H_2})$ is not converged at high column densities. However, modifying the model to treat the multiphase ISM more realistically will likely affect the results.

4.2 $f(N_{H_2})$ – dependence on the resolution of N_{H_2} maps

We compare how $f(N_{H_2})$ depends on the resolution of observed and simulated H_2 column density (N_{H_2}) maps. We calculate $f(N_{H_2})$ using 150 pc and 1 kpc map resolution CO(2–1) data from the PHANGS-ALMA survey and a sample of galaxies from TNG100.

The $f(N_{H_2})$ for these data sets and resolutions are displayed in Fig. 2 (right-hand panel). Differences in both shape and the column density range are found between the different map resolutions in both TNG100 and the PHANGS-ALMA survey. The $f(N_{H_2})$ derived from PHANGS-ALMA shows a more substantial map resolution dependence compared to TNG100. There are two factors that together cause this higher map resolution dependence of the PHANGS-ALMA data. (1) Creating N_{H_2} maps using larger pixel sizes averages the column densities over larger regions. This leads to lower mean observed column densities, as very high column densities are usually detected at GMC scales, which are smaller than 1 Kpc (Leroy et al. 2021). This effect is especially apparent at the high-column-density end, as in the 1 kpc map resolution data column densities above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 22.8$ are diluted by this effect. (2) Observational data are limited by their sensitivity and completeness. Coarser resolution data have a reduced noise and are thus more complete (see the comparison of native resolution versus 150-pc map resolution data in PHANGS-ALMA; Leroy et al. 2021). Therefore, the coarser map resolution observations are sensitive to lower column densities compared to finer resolution observations. This effect is especially significant below $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 22.5$ in Fig. 2 (right-hand panel). TNG100 does not suffer from these sensitivity and incompleteness effects at lower column densities and therefore the map resolution dependence is less drastic. TNG100 is only affected by beam smearing. The core of the distribution $[\log(N_{H_2}/\mathrm{cm}^{-2}) = 21\text{–}22]$ is robust to resolution effects in both simulations and observations.

5 DOES THE H_2 COLUMN DENSITY DISTRIBUTION OF INDIVIDUAL GALAXIES DEPEND ON THEIR PHYSICAL PROPERTIES?

We study the $f(N_{H_2})$ of individual objects in the PHANGS-ALMA survey and TNG100 ($z = 0$) in order to explore how it depends on integrated physical properties of the galaxies. We calculate the individual $f(N_{H_2})$ using equation (1), but set the normalization parameters [$\Phi(x_i)$ and $w(x_i)$] equal to 1. The individual column density distributions, colour-coded according to the integrated SFR,

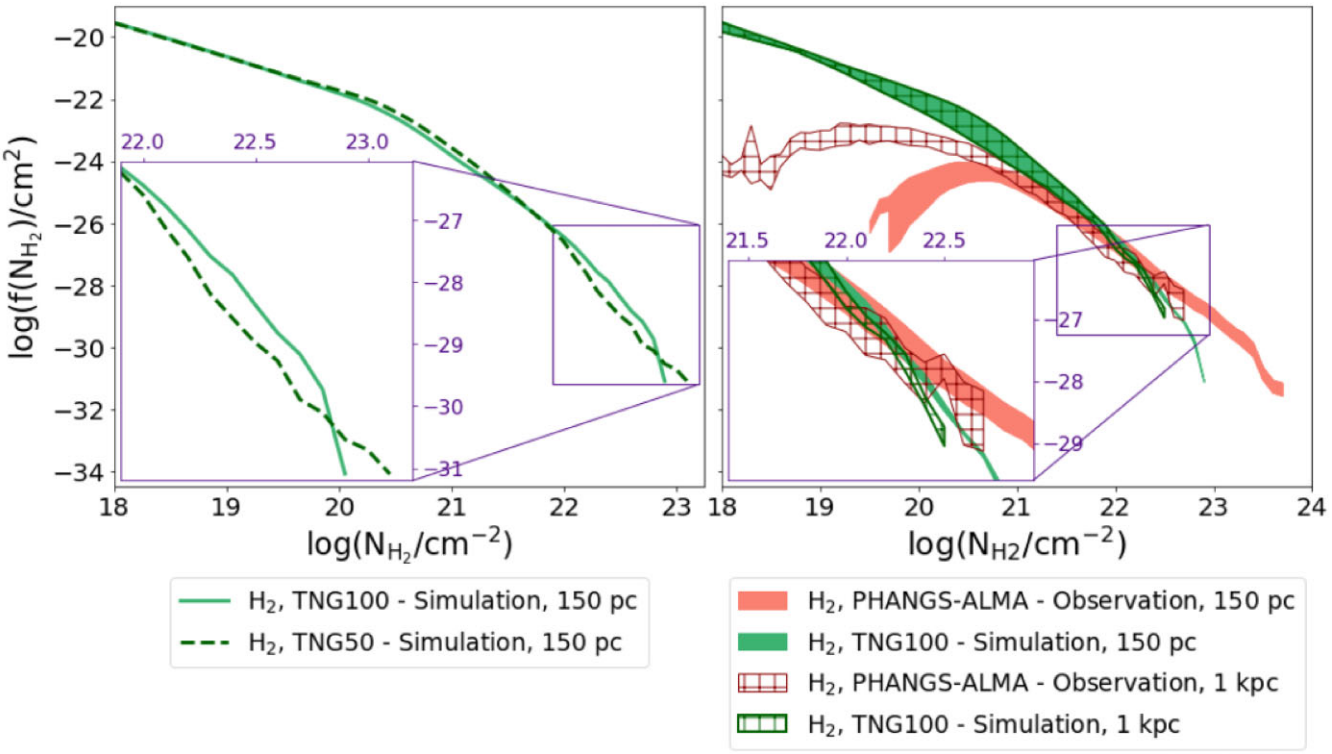

Figure 2. Left-hand panel: $f(N_{H_2})$ derived from TNG100 and TNG50. Higher column densities are reached in TNG50, as a higher resolution enables the simulation to reach higher gas densities. Especially at column densities above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 22$, the $f(N_{H_2})$ differs. This indicates that H_2 in TNG100 is not converged for those column densities. Right-hand panel: $f(N_{H_2})$ derived from PHANGS-ALMA (red bands) and TNG100 (green bands) data at map resolutions of 150 pc (filled) and 1 kpc (hatched). The $f(N_{H_2})$ both in observations and simulations show a map resolution dependence. In TNG100 this effect arises due to the averaging of column densities over a larger area. Very high column densities are usually detected on small scales (much smaller than 1 kpc), leading to a dilution of high column densities. In PHANGS-ALMA, this effect is additionally combined with sensitivity and incompleteness specifics of the observations. The core of the distribution [$\log(N_{H_2}/\mathrm{cm}^{-2}) = 21$–$22$] is robust to resolution effects in both simulations and observations. In general, high-resolution observations and simulations are needed to resolve column densities typically found in very dense environments like molecular clouds.

stellar mass (M_*), and H_2 mass (M_{H_2})[6] of the corresponding galaxy are displayed in Fig. 3. For these calculations, we use column density maps with a resolution of 150 pc. The colour coding of the plots reveals a connection between $f(N_{H_2})$ and the physical parameters mentioned. We note that the integrated molecular masses of galaxies in TNG100 are generally higher compared to PHANGS-ALMA. This is to be expected, as TNG100 probes the full disc and is not limited by observational sensitivity and incompleteness limits when compared to the PHANGS-ALMA sample. Further, Leroy et al. (2021) estimate that, on average, ∼30 per cent of molecular gas is missed by PHANGS-ALMA due to the limited field of view when compared to WISE3 luminosities. Additionally, TNG100 is possibly overestimating H_2 within the simulation at $z = 0$ (Diemer et al. 2019).

The $f(N_{H_2})$ of individual galaxies have very similar shapes in both observation and simulation. This is possibly related to the galaxies in the sample, which are main-sequence star-forming galaxies. These types of galaxies mostly have rotating discs and are hypothesized to have radially exponential gas profiles (Leroy et al. 2008; Stevens et al. 2019). While the disciness and exponential gas profiles of the galaxies under consideration still need to be established, the similar $f(N_{H_2})$ could indeed stem from similar gas profiles within these galaxies. In Section 6, we further explore this possibility by comparing an analytical $f(N_{H_2})$ model assuming radially exponential gas discs with simulated results.

While the shapes of the $f(N_{H_2})$ are similar, the $f(N_{H_2})$ also show a correlation with integrated physical parameters of the galaxies. The colour coding in Fig. 3 indicates that the $f(N_{H_2})$ are correlated with the integrated SFR, M_*, and M_{H_2} of the galaxies. The higher the SFR, M_*, and M_{H_2}, the more massive these galaxies are, leading to a higher normalization of $f(N_{H_2})$. Further, higher column densities are detected in more massive galaxies, implying that more dense gas is formed in larger galaxies. The higher abundance of denser gas in more massive galaxies could also lead to higher SFRs, as more gas is found at densities suitable for star formation (e.g. above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 21$, Clark & Glover 2014). This is possibly related to the correlation between the SFR surface density and molecular gas surface densities in galaxies (e.g. Bigiel et al. 2008; Feldmann 2020). We however note that the correlation we find is related to the integrated SFR of the galaxy and not the SFR surface density. In Appendix A, we explore these correlations using the PHANGS-ALMA sample and provide a way to approximate $f(N_{H_2})$ given physical parameters. It, however, remains unclear which galaxy property is the governing parameter for the shape of $f(N_{H_2})$, as SFR, M_*, and M_{H_2} all have similar correlation strengths with parameters of the gamma distribution used to fit $f(N_{H_2})$ in the Appendix.

[6]SFR and M_* are taken from sites.google.com/view/phangs/sample. M_{H_2} is calculated by summing up the surface density of individual pixels multiplied by pixel area (table found in datafileB1 at canfar.net/storage/list/phangs/RELEASES/Sun_etal_2020b).

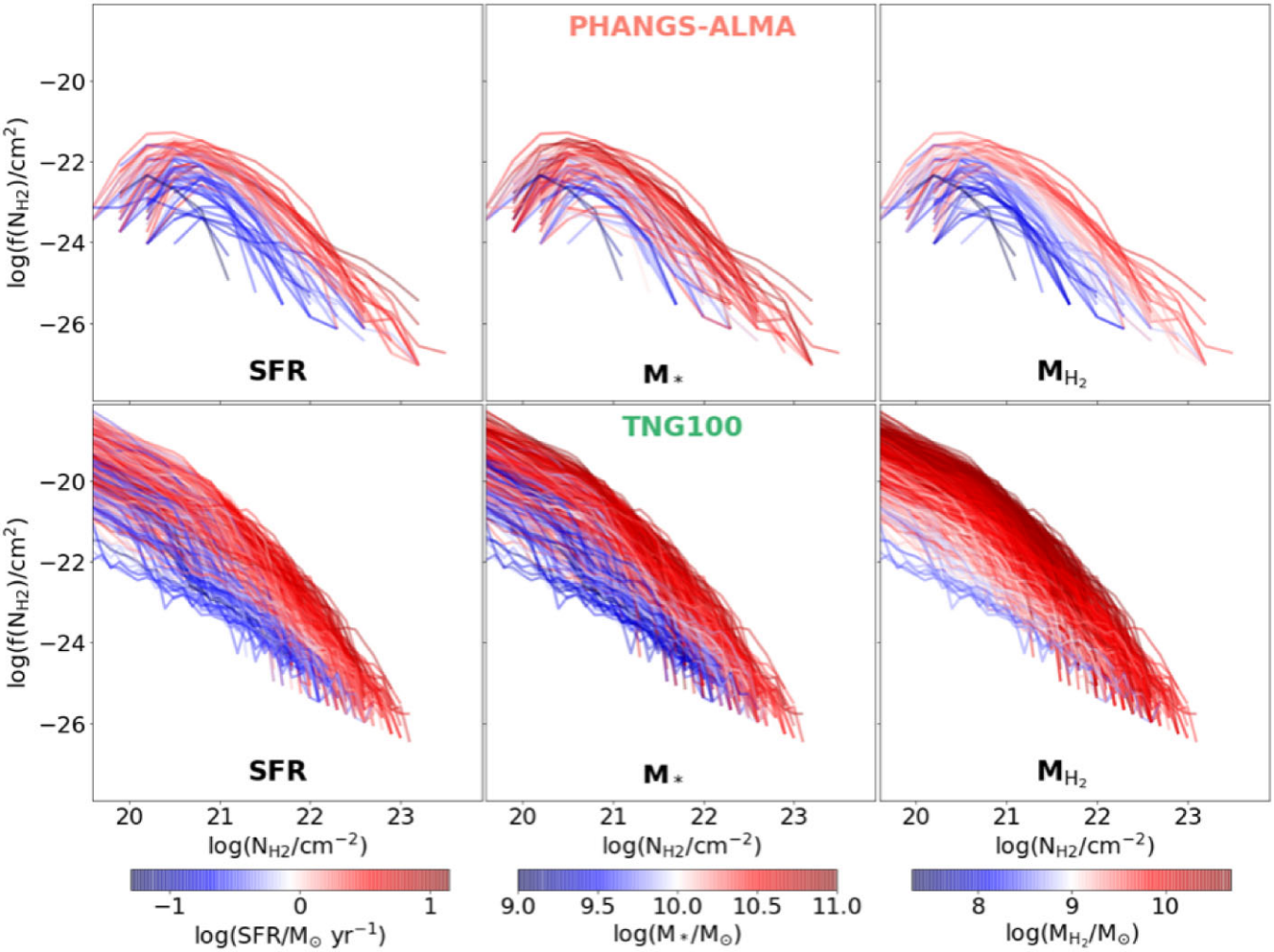

Figure 3. H_2 column density distributions ($f(N_{H_2})$) of individual PHANGS-ALMA and TNG100 ($z = 0$) galaxies. The plots display the correlation of the individual $f(N_{H_2})$ on the integrated SFR, M_*, and M_{H_2} of the galaxies. The flattening and steep drop of the PHANGS-ALMA $f(N_{H_2})$ at lower column densities is due to the incompleteness and sensitivity limit of the observations. The individual $f(N_{H_2})$ in both PHANGS-ALMA and TNG100 have similar shapes. This is likely due to the sample consisting of main-sequence star-forming galaxies. These galaxies mostly have rotating discs and are hypothesized to have.
radially exponential gas profiles. The $f(N_{H_2})$ correlate with integrated physical parameters (SFR, M_*, and M_{H_2}) of the galaxies. The higher these parameters, the larger are the galaxies, leading to a higher normalization of $f(N_{H_2})$. Further, higher column densities are detected in more massive galaxies, implying that more dense gas is found in larger galaxies.

6 THE REDSHIFT EVOLUTION OF THE H_2 COLUMN DENSITY DISTRIBUTION IN SIMULATIONS AND OBSERVATIONS

We study $f(N_{H_2})$ at $z = 0$ and 3 using both observations and simulations. First, we study how recent observations compare to the state-of-the-art simulation TNG100 at both $z = 0$ and 3 and how the isolated dwarf galaxy simulation from the GRIFFIN Project fits into the column density distribution at $z = 0$. Then we discuss the evolution of $f(N_{H_2})$ from $z = 0$ to 3. Finally, we examine how $f(N_{H_2})$ compares to $f(N_{H1})$ at $z = 0$ and 3 to explore in which regions of galaxies the neutral atomic gas is dominating over the molecular gas.

6.1 $f(N_{H_2})$ at $z = 0$

In Fig. 4 (left-hand panel), the $f(N_{H_2})$ from both observations, simulations, and an analytical model at $z = 0$ are displayed. For TNG100, we plot a band (green band) encompassing the three post-processing methods described in Section 3.3. Note that below $\log(N_{H_2}/\mathrm{cm}^{-2}) < 18$, the post-processing results for H_2 become unreliable as post-processing the simulations with different SPH kernel smoothing lengths leads to highly different results in that region. This region is represented by bands filled with the ⋆ symbol. The dwarf galaxy simulation $f(N_{H_2})$ (green line) is based on the results from the on-the-fly chemical network included in the simulation. The red band encompasses the $f(N_{H_2})$ from the PHANGS-ALMA survey using varying assumptions. It includes calculations using a stellar mass function based on the full sample and late-type galaxy only sample in Weigel et al. (2016). Further, we calculate the $f(N_{H_2})$ with both a metallicity-dependent α_{CO} (see Section 3.1) and a constant $\alpha_{CO} = 4.3\ M_\odot\ (\mathrm{K\ km\ s^{-1}\ pc^2})^{-1}$ (Bolatto et al. 2013). We note that the drop of the PHANGS $f(N_{H_2})$ at column densities below $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 21$ is not of physical origin, but due to the sensitivity and incompleteness limit of the observations, leading to the observations not probing the full disc. We also include an analytical model (black line) used to estimate $f(N)$

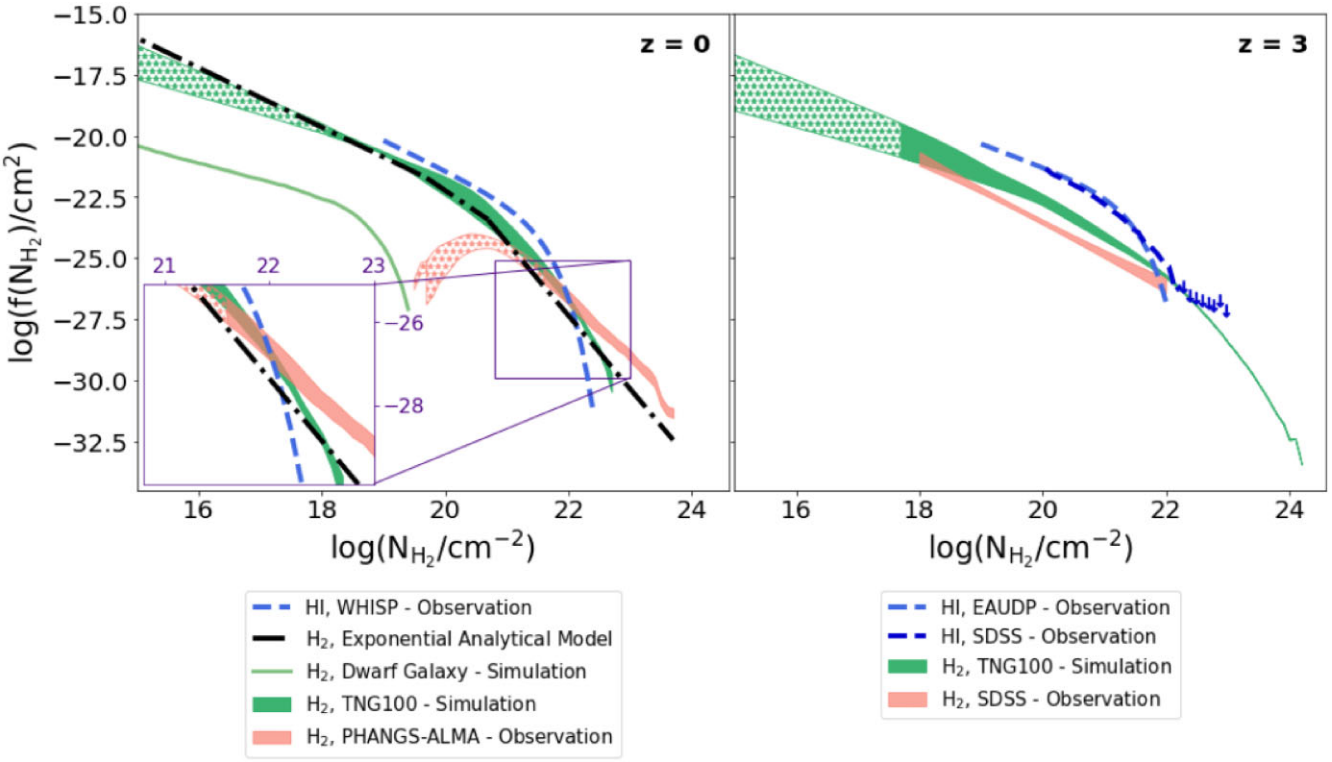

Figure 4. $f(N_{\rm H\,I})$ and $f(N_{\rm H_2})$ derived from both simulations and observations at $z = 0$ and 3. The column densities at which the $f(N_{\rm H_2})$ become unreliable due to incompleteness or simulation specifics are indicated by regions filled with the $\star$ symbol. Left-hand panel: $f(N_{\rm s})$ of TNG100, the PHANGS-ALMA survey (red band), the WHISP sample (blue; Zwaan et al. 2005), an analytical model (black line; Zwaan 2000), and the simulated dwarf galaxy from the GRIFFIN project (green line) at $z = 0$. The $f(N_{\rm H_2})$ by TNG100 broadly reproduces the observations by the PHANGS-ALMA survey. The analytical $f(N_{\rm H_2})$ based on the assumption of an radially exponential gas profile in galaxies is a good approximation for $f(N_{\rm H_2})$ for both observations and simulations. The $f(N_{\rm H_2})$ derived from a simulated dwarf galaxy including a non-equilibrium chemical network displays similar slopes compared to TNG100. Right-hand panel: $f(N_{\rm s})$ of TNG100 (green band), SDSS [(red band; Balashev & Noterdaeme 2018), (dark blue line; Ho et al. 2021)], and the EAUDP sample (dark blue line; Zafar et al. 2013) at $z = 3$. The $f(N_{\rm H_2})$ from TNG100 and the observational results based on the SDSS sample have matching slopes. The normalization between the two shows an $\sim$1-dex difference, possibly arising due to differences in selection and environments probed.

assuming radially exponential gas discs averaged over all possible inclinations (Zwaan 2000).[7] The analytical model is approximated by three linear functions.

6.1.1 *TNG100 broadly reproduces observations*

We study how TNG100 (green band, Fig. 4, left-hand panel) reproduces the $f(N_{\rm H_2})$ observed by the PHANGS-ALMA survey (red band, Fig. 4, left-hand panel). While there are differences in the $f(N_{\rm H_2})$, the observations are broadly reproduced by TNG100 in the column density ranges where simulation and observation specifics do not hinder a fair comparison.

One difference is that TNG100 does not reach as high a column density as observed by the PHANGS-ALMA survey. This is due to limitations of the simulation. Given the resolution of TNG100 gas densities that can be reached at given redshifts are limited (see Section 4.1). Further, for the regions below $\log(N_{\rm H_2}/{\rm cm^{-2}}) \sim 21.6$, TNG100 $f(N_{\rm H_2})$ shows a higher normalization than observed with PHANGS-ALMA. This can be explained by the sensitivity and incompleteness limit of the observations below those column densities. In the region between $\log(N_{\rm H_2}/{\rm cm^{-2}}) \sim 21.6$–22.2 both $f(N_{\rm H_2})$ are overlapping, albeit with TNG having a steeper slope when approximated as a linear function in log space ($\beta_{\rm PHANGS} \sim 2.3$–2.35, $\beta_{\rm TNG} \sim 3.15$–3.6).

Given that the two $f(N_{\rm H_2})$ are based on vastly different methods of calculating the column densities (one being post-processed H_2 from a cosmological magnetohydrodynamical simulation, and one observations of CO(2–1), which are converted into H_2), the similarity between the two $f(N_{\rm H_2})$ is remarkable. Nonetheless further tests and studies are needed to explore the inconsistencies between observations and simulations, especially at the high-column-density end of $f(N_{\rm H_2})$. Higher resolution simulations would likely extend the $f(N_{\rm H_2})$ to higher column densities. Additionally, alternatives to the Springel & Hernquist (2003) subgrid star formation prescription in order to resolve the cold gas phase in the simulations might be needed to reach the high column densities detected in observations. Including non-equilibrium chemistry in these simulations could also give more accurate representations of H_2 in the simulations (e.g. Maio, Péroux & Ciardi 2022). Finally, deeper observations would enable fair comparisons of $f(N_{\rm H_2})$ below $\log(N_{\rm H_2}/{\rm cm^{-2}}) \sim 21.6$.

6.1.2 *An analytical model closely matching TNG100*

We compare the analytical model by Zwaan (2000) (black line, Fig. 4, left-hand panel) with the results from TNG100 (green band, Fig. 4, left-hand panel) at $z = 0$ to study how well the simulated

[7] With the N_0 parameter of the model, which determines the knee of the curve, set to $10^{20.7}$ cm^{-2}. For H_2, this is an ad-hoc choice.

predictions match the analytical model. We approximate the analytical model using three ad-hoc linear functions in the following H_2 column density ranges: $\log(N_{H_2}/cm^{-2}) \leq 20$, $\log(N_{H_2}/cm^{-2}) = 20$–$21$, $\log(N_{H_2}/cm^{-2}) \geq 21$ and therefore use these regions for a comparison. Generally, the analytical model assuming radially exponential gas discs in galaxies produces comparable results to calculating $f(N_{H_2})$ using the post-processed H_2 column densities of galaxies within TNG100.

Below $\log(N_{H_2}/cm^{-2}) \leq 20$ the normalizations are comparable, but the TNG100 $f(N_{H_2})$, depending on the post-processing prescription used, has a slightly lower slope compared to the analytical model ($\beta_{TNG} \sim -0.75$ to -1.15, $\beta_{ana} \sim -1.22$). In the range of $\log(N_{H_2}/cm^{-2}) = 20$–$21$, we find similar results, with the normalization matching, but a slightly higher slope in TNG100 ($\beta_{TNG} \sim -1.94$ to -2.05, $\beta_{ana} \sim -1.71$). Further, approximating the slope of TNG100 at $\log(N_{H_2}/cm^{-2}) \geq 21$ using a linear function in log space leads to similar results, with TNG100 producing higher slopes for $f(N_{H_2})$ in this region ($\beta_{TNG} \sim -3.4$ to -3.8, $\beta_{ana} \sim -3.0$). Finally, the analytical model predicts slightly less systems in this column density range compared to TNG100. We note that Zwaan (2000) also proposes an analytical model based on Gaussian gas profiles. This model results in slopes of $\beta_{ana,\,gauss} \sim -1$ at column densities below $\log(N_{H_2}/cm^{-2}) = 20.7$ and $\beta_{ana,\,gauss} \sim -3$ at column densities above this threshold.

We conclude that while TNG100 produces $f(N_{H_2})$ with slightly higher slopes and in some parts different normalizations, the $f(N_{H_2})$ derived from the analytical model is still comparable and a good approximation. Since radially exponential gas discs are also a good approximation for disc galaxies in TNG100 (e.g. H I discs described in Stevens et al. 2019), it appears natural that an analytical model making the assumption of radially exponential gas discs yields similar results. While the results of an analytical model using exponential gas discs matches predictions by TNG100 well, a Gaussian distribution within gas discs of galaxies yields similar results. Therefore, further studies of the distribution in gas discs and their relevance to $f(N_{H_2})$ are required to fully understand how the gas disc distribution and $f(N_{H_2})$ relate.

6.1.3 *A dwarf galaxy simulation producing similar slopes compared to TNG100*

We compare the $f(N_{H_2})$ of the simulated dwarf galaxy from the GRIFFIN project, which includes a non-equilibrium chemical network tracking H_2 on the fly (green line, Fig. 4, left-hand panel) with the results of TNG100 (green band, Fig. 4, left-hand panel) at $z = 0$. This helps us understand the impact for $f(N_{H_2})$ when running simulations at sub-pc resolution including a non-equilibrium chemical network in an isolated environment.

The $f(N_{H_2})$ only probes one galaxy with a stellar mass of $\log(M_*/M_\odot) \sim 7.3$.[8] This leads to a number of differences when compared to a sample of galaxies. Due to the limited mass and size, the dwarf galaxy in the simulation only reaches column densities up to $\log(N_{H_2}/cm^{-2}) \sim 19.5$. The slope of both $f(N_{H_2})$ is consistent. For the dwarf simulation, the logarithmic slope is $\beta_{dwarf} \sim -0.7$ before the drop off at $\log(N_{H_2}/cm^{-2}) \sim 18$. The slope found in TNG100 at those column densities is $\beta_{TNG} \sim -0.7$ to -1.1. It is surprising that the slope of $f(N_{H_2})$ of a single galaxy is so similar the slope of a large sample of galaxies with varying sizes, especially given the different methods for deriving molecular gas in these simulations. While it is difficult to disentangle the effects that the different galaxy properties and derivation methods of molecular gas have on the slope of $f(N_{H_2})$, this is possibly a first indication that the slope $f(N_{H_2})$ is not affected by non-equilibrium chemistry. Especially since the slope of $f(N_{H_2})$ for individual main-sequence star-forming galaxies in TNG100 is similar below $\log(N_{H_2}/cm^{-2}) \lesssim 20$ and not majorly affected by galaxy properties. In order to further our understanding of how non-equilibrium chemistry might affect $f(N_{H_2})$ a larger sample size of highly resolved simulated galaxies spanning a wider range of stellar masses would be needed. Alternatively, running and comparing the dwarf galaxy simulation by GRIFFIN without non-equilibrium chemistry with the current GRIFFIN model would also help disentangling the effects that non-equilibrium chemistry and galaxy properties have on the slope of $f(N_{H_2})$. This is an interesting avenue to explore in the future.

6.1.4 *Which column densities contribute most to the H_2 mass density at $z = 0$?*

As a final analysis of $f(N_{H_2})$ at $z = 0$, we study which column densities contribute the most to the overall mass density (ρ_{mol}) in both TNG100 and the PHANGS-ALMA survey. Disentangling which column densities contribute the most to the mass density helps us understand in which regions of galaxies (e.g. the ISM, CGM, Molecular Clouds) most of the molecular gas is detected. Further, we can interpret if most of the gas is in regions suitable for star formation or not.

In Fig. 5 (left-hand panel), we plot the mass densities as a function of H_2 column density. The red band corresponds to the PHANGS-ALMA results and the green band to the TNG100 results. For TNG100, the highest mass density contribution stems from column densities in the range $\log(N_{H_2}/cm^{-2}) \sim 20.5$–$20.7$. Therefore the majority of molecular gas in TNG100 is found at column densities typical for the ISM of galaxies, but below densities of molecular clouds (e.g. Spilker et al. 2021) as opposed to less dense and diffuse regions surrounding galaxies, like the CGM.

Using numerical models, Clark & Glover (2014) predict that star formation is possible in regions where the mean area averaged column density exceeds $\log(N_{H_2}/cm^{-2}) \sim 21$. TNG100 predicts the mass density peak slightly below the star formation threshold advocated by Clark & Glover (2014) and therefore in a region not suitable for star formation. This fraction of the gas could be either in regions where the molecular gas has been depleted due to star formation, or in regions that are possibly in the process of collapsing into denser regions.

In PHANGS-ALMA, we find an overall flatter distribution of the H_2 mass densities in the regions where the observations are complete and when compared to TNG100. The highest contribution to the overall mass density is in the range of $\log(N_{H_2}/cm^{-2}) \sim 21.2$–$21.5$. This is at densities detected in the ISM and typical for molecular clouds. We note that for the 1-kpc resolution PHANGS-ALMA data, the highest contribution shifts to $\log(N_{H_2}/cm^{-2}) \sim 21$. However, it is not trivial to quantify how much this is an effect of higher completeness at lower resolutions compared to averaging over a larger area.

The mass density peak in PHANGS-ALMA is detected at densities above the star formation threshold advocated by Clark & Glover (2014). While this is inconsistent with results by TNG100, we note that the observations of PHANGS-ALMA are incomplete in this region. It is therefore conceivable that deeper observations of

[8]The stellar mass does not evolve much over the course of the simulation

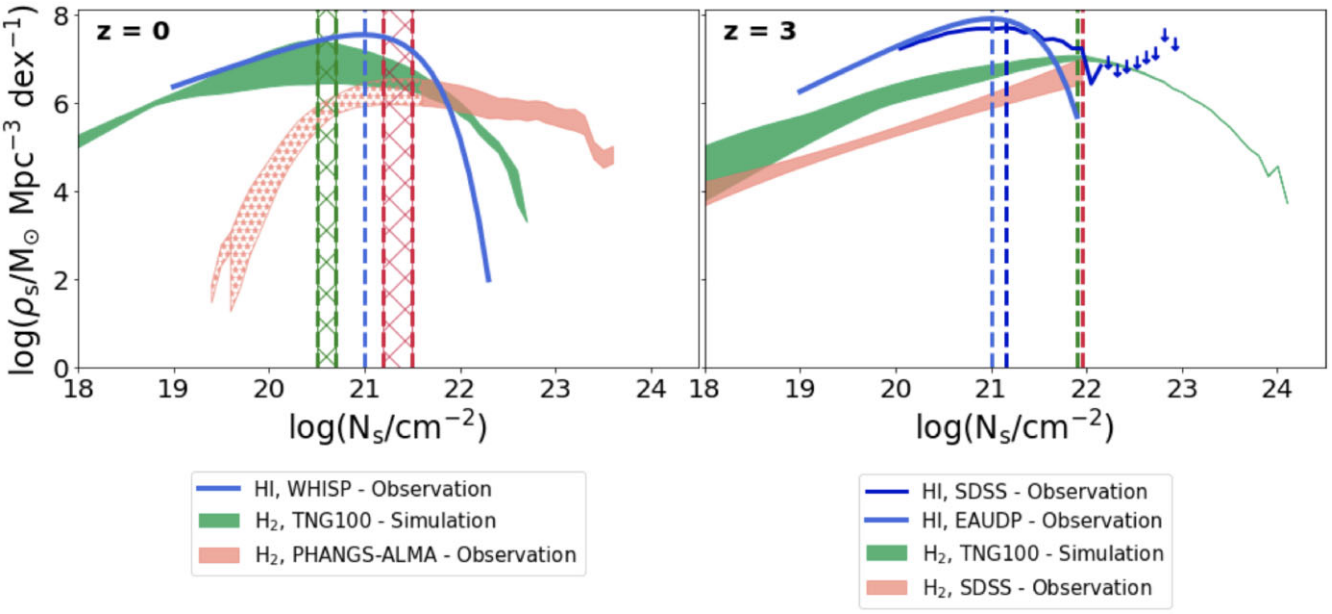

Figure 5. The mass density contribution per dex column density (ρ_s) of H I and H_2 derived from observations and simulations. Overall, at $z = 3$, the highest mass contribution of H_2 comes from denser gas compared to $z = 0$ in both simulations and observations. Further, the mass density distributions suggest that H I dominates over H_2 at most column densities, making it an important contributor to the cold gas mass density of galaxies. H_2 starts dominating compared to H I at column densities above $\log(N_{H_2}/cm^{-2}) \sim 21.8$–22 at both redshifts. Left-hand panel: mass densities of H_2 and H I against column densities at $z = 0$ for TNG100 (green band), the PHANGS-ALMA survey (red band), and the WHISP sample (blue line; Zwaan et al. 2005). The H_2 highest mass density contribution can be constrained between $\log(N_{H_2}/cm^{-2}) \sim 20.5$–21.5 (column density regions typical for the ISM and, in part, molecular clouds) Right-hand panel: mass densities of H_2 and H I against column densities at $z = 3$ for TNG100 (green band), SDSS [(red band; Balashev & Noterdaeme 2018), (dark blue line; Ho et al. 2021)], and the EAUDP sample (dark blue line; Zafar et al. 2013). The highest H_2 mass density contribution can be constrained at $\log(N_{H_2}/cm^{-2}) \sim 22$ (column densities typical for molecular clouds).

molecular gas in these galaxies may shift the observed column density contributions to lower column densities.

In conclusion, when combining results by observations and simulations, the highest H_2 mass density contribution is found at column densities detected within the ISM of galaxies and partly in regions observed in local molecular clouds.

6.2 $f(N_{H_2})$ at $z = 3$

6.2.1 TNG100 broadly reproducing observations

Here we compare the $f(N_{H_2})$ based on composite SDSS H_2 absorption spectra (Balashev & Noterdaeme 2018) to the results of TNG100 at $z = 3$. In Fig. 4 (right-hand panel), we display these $f(N_{H_2})$. The data from Balashev & Noterdaeme (2018) (red band) only include column densities of $\log(N_{H_2}/cm^{-2}) = 18$–22 and we therefore can only compare the slopes before the steeper drop of $f(N_{H_2})$ at higher column densities. Both $f(N_{H_2})$ have similar slopes in this region ($\beta_{balashev} \sim 1.13$–1.45, $\beta_{TNG} \sim 0.88$–1.47). The normalization of the observed $f(N_{H_2})$ is ~ 1 dex lower than that predicted by TNG100 in most column density regions.

This may be caused by the different methods of deriving $f(N_{H_2})$. The observed $f(N_{H_2})$ is based on absorption line studies of DLAs, which are typically observed at high impact parameters surrounding galaxies (e.g. Péroux et al. 2011; Christensen et al. 2014; Krogager et al. 2017), while TNG100 relies on post-processed H_2 column density maps, which include all regions of galaxies. Further, these observations might be biased towards galaxies in group environments (Hamanowicz et al. 2020). Given this, while there are still inconsistencies between simulation and observation, the two $f(N_{H_2})$ are remarkably close in slope. Further studies, including high-spatial-resolution molecular gas observations or post-processing TNG100 using ray-casting codes at typical impact parameters of absorption line systems, might help alleviate some of these inconsistencies and are an interesting avenue for future studies.

We stress that another model by Krogager & Noterdaeme (2020) using the fraction of cold gas absorption in strong H I selected absorbers derived by Balashev & Noterdaeme (2018) predicts an $f(N_{H_2})$ with a knee at $\log(N_{H_2}/cm^{-2})) \sim 21$ and a highest column density of $\log(N_{H_2}/cm^{-2}) \sim 23$. While the model is also using the Blitz & Rosolowsky (2006) method for splitting the cold gas into a neutral and molecular fraction as for TNG100, the results are inconsistent with the predictions made by TNG100, which estimates the knee of $f(N_{H_2})$ to occur at $\log(N_{H_2}/cm^{-2}) \sim 22$ and includes column densities beyond $\log(N_{H_2}/cm^{-2}) = 24$.

In conclusion, both observations and simulations have $f(N_{H_2})$ with well matching slopes in the overlapping regions. However, they differ in normalization by ~ 1 dex. Thus, the results by observations and simulations are in tension for the overlapping H_2 column density regions at $z = 3$.

6.2.2 Which column densities contribute most to the H_2 mass density at $z = 3$?

As a final analysis of $f(N_{H_2})$ at $z = 3$, we study which column densities contribute the most to the overall mass density (ρ_{mol}) in both TNG100 and the $f(N_{H_2})$ by Balashev & Noterdaeme (2018) derived from composite H_2 spectra.

In Fig. 5 (right-hand panel), we plot the mass densities for each H_2 column density. The peak of the H_2 mass density is not reached by the

Balashev & Noterdaeme (2018) data (red band), meaning that we can only set a limit of $\log(N_{H_2}/cm^{-2}) \gtrsim 22$. This is in regions typically observed within molecular clouds. Further, it is well above the density threshold for star formation. The TNG100 results show that the highest mass density contribution is at densities of $\log(N_{H_2}/cm^{-2}) \sim 21.9$, so slightly below the limit that one can set with observations.

6.3 Denser molecular gas found at high redshifts

In this section, we study how $f(N_{H_2})$ evolves from redshift $z = 0$ to 3. The $f(N_{H_2})$ for both redshifts is shown in Fig. 4.

In TNG100 (green bands), the slopes below $\log(N_{H_2}/cm^{-2}) = 20$ are similar ($\beta_{TNG,\,z=0} \sim 0.75$–$1.14$, $\beta_{TNG,\,z=3} \sim 0.73$–$1.10$) and show little to no evolution. At column densities above tha, differences start to arise. At $z = 0$, TNG100 predicts molecular gas up to column densities of $\log(N_{H_2}/cm^{-2}) \sim 23$. At $z = 3$, column densities beyond $\log(N_{H_2}/cm^{-2}) \gtrsim 24$ are reached in TNG100. This indicates that denser H_2 gas exists in the earlier Universe. It is the case that physical densities are intrinsically higher in the high-redshift versus low-redshift Universe. At the same time, this prediction from TNG100 could be affected by its finite numerical resolution. Further, there is a steeper drop off at high column densities at $z = 0$ compared to $z = 3$ in TNG100, where the $f(N_{H_2})$ is flatter at high column densities. Due to limitations in the observations, we cannot make similar statements at the high column densities using observations. We, however, find that in the overlapping region $f(N_{H_2})$ of both the SDSS sample (Balashev & Noterdaeme 2018) and the PHANGS-ALMA survey are similar. The $f(N_{H_2})$ of Balashev & Noterdaeme (2018) is a good continuation of the $f(N_{H_2})$ found in the PHANGS-ALMA survey. We therefore expect larger differences in the $f(N_{H_2})$ to arise at higher column densities. This would mean that the largest differences of $f(N_{H_2})$ arise at the densest molecular regions in the Universe. Observations at $z = 3$ with higher column densities are needed in order to test if the predictions by TNG100 are correct.

Fig. 5 shows that the column densities contributing the most to the molecular gas mass densities are shifting towards higher column densities at $z = 3$. When combining the results from observations and simulations we find the following: While at $z = 0$ the highest contribution is found at column densities of $\log(N_{H_2}/cm^{-2}) \sim 20.5$–$21.5$, at $z = 3$ it is found at column densities of $\log(N_{H_2}/cm^{-2}) \sim 21.9$–$22$. When assuming that the column density relates to the density of the gas, denser gas found at higher redshifts is in line with observations of the SFR across cosmic time, which is higher at $z = 3$ compared to $z = 0$ (Madau & Dickinson 2014; Tacconi et al. 2020). The shape of the cosmic molecular mass density as a function of redshift is similar to the shape of the SFR density, making a coupling of these two quantities likely. Therefore, one would expect that more molecular gas found in denser regions leads to a higher global SFR in galaxies (Péroux & Howk 2020). When assuming that the column density relates to the density of the gas, this is exactly what we observe when studying the column density distributions at $z = 0$ and 3.

6.4 Is H_2 dominating the higher column densities?

We compare $f(N_{H_2})$ and $f(N_{HI})$ at $z = 0$ and 3 to study the column densities at which H_2 overtakes H I. In the following sections we compare these derived densities with the combined results of $f(N_{H_2})$ derived from both observation and simulation. At $z = 0$, we compare the $f(N_{H_2})$ with $f(N_{HI})$ derived by Zwaan et al. (2005). At $z = 3$, we compare the $f(N_{H_2})$ with the $f(N_{HI})$ derived by Zafar et al. (2013) and Ho et al. (2021).

6.4.1 *H I and H_2 Column Density Distributions at $z = 0$*

The $f(N_{HI})$ at $z = 0$ from Zwaan et al. (2005) is based on 21-cm maps of 355 galaxies of the WHISP sample (van der Hulst, van Albada & Sancisi 2001). The WHISP sample covers galaxies of all Hubble types from S0 to Im and a considerable luminosity range and were selected using the Uppsala General Catalogue (UGC) of galaxies (Nilson 1973). The median spatial resolution reached by these observation is ∼1.4 kpc.

The $f(N_s)$ (left-hand panel in Fig. 4) and ρ_s (left-hand panel in Fig. 5) at $z = 0$ show that H_2 starts to dominate the mass density at column densities above $\log(N_{H_2}/cm^{-2}) \sim 21.8$–$22$.[9] This is consistent with results from Schaye (2001), who predicted that H I clouds with $N_{HI} \gtrsim 10^{22} cm^{-2}$ transform to molecular clouds before reaching higher column densities. Similar predictions have also been made more recently by Altay et al. (2011) and Bird et al. (2014) using (magneto-)hydrodynamical simulations.

These results imply that while molecular gas dominates the high column densities above $\log(N_s/cm^{-2}) \gtrsim 22$, H I dominates the majority of the column density regions found within the interstellar medium (including column density regimes typical for molecular clouds), making neutral gas an important contributor to the cold gas mass found within galaxies at $z = 0$.

6.4.2 *H I and H_2 column density distributions at $z = 3$*

The two $f(N_{HI})$ at $z \sim 3$ are based on H I-absorption systems (sub-DLAs and DLAs). The calculation therefore relies on pencil beam observations of H I-column densities as studying 21-cm H I in emission is not feasible at this redshift. The $f(N_{HI})$ by Ho et al. (2021) is based on the Sloan Digital Sky Survey Data Release 16 which were analysed using Gaussian processes, where DLAs are detected using Bayesian model selection. While SDSS-DR16 includes redshifts between $z = 2$ and 5, we only use the results of the $z = 2.5$–3 integration for our comparison. The $f(N_{HI})$ by Zafar et al. (2013) is based on the ESO UVES advanced data products (EUADP) sample and includes measurements in the $z \sim 1.5$–3.1 range. The $f(N_{HI})$ of both samples show comparable results up to $\log(N_{HI}/cm^{-2}) \sim 22$. Above this density, SDSS results display a possible flattening of the $f(N_{HI})$. This flattening would be inconsistent with the predictions of the maximum N_{HI} by Schaye (2001), but the Gaussian process analysis shows that the $f(N_{HI})$ in that region is also consistent with 0 and therefore not well constrained. We further note that while the SDSS-DR16 sample is larger than the EAUDP sample, the resolution is lower. The lower resolution could lead to blending at higher column densities, which would lead to measurements of column densities above $\log(N_{HI}/cm^{-2}) \sim 22$.

The $f(N_s)$ (right-hand panel in Fig. 4) and ρ_s (right-hand panel in Fig. 5) at $z = 3$ show that H_2 starts to dominate the mass density at column densities between $\log(N_{H_2}/cm^{-2}) \sim 21.5$–$22$. As for $z = 0$, neutral gas is an important contributor to the global mass in a wide range of regions found in the ISM including higher density regions typical of molecular clouds.

[9]We note that the TNG100 ρ_{H_2} band implies that at $z = 0$, the H_2 mass density is roughly equal in the $\log(N_{H_2}/cm^{-2}) \sim 19$–$20.5$ column density region. We attribute this to a possible over-prediction of H_2 (and H I) in the simulation compared to observations at $z = 0$ (Diemer et al. 2019). Deeper observations are needed to quantify how high the contribution of molecular gas is at these densities.

6.4.3 H I – an important contributor to the cold gas mass of galaxies

In conclusion, Fig. 5 indicates that H I dominates over H_2 at most column densities. The H I column density contributing most to the overall mass density (blue vertical lines) has a higher mass contribution than H_2 at both redshifts. H I could therefore be an important contributor to the cold gas mass of galaxies at $z = 0$ and 3.

The column density contributing the most to the overall H I gas mass density is at $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \sim 21$ for both redshifts. In contrary the highest contributing column density of H_2 evolves with redshift. It is $\log(N_{\rm H_2}/{\rm cm}^{-2}) \sim 22$ at $z = 3$ and less than $\log(N_{\rm H_2}/{\rm cm}^{-2}) \sim 21.5$ at $z = 0$. We note that the molecular phase of the gas cycle is likely to be shorter than the neutral atomic phase as indicated by cold gas depletion time-scales (Péroux & Howk 2020). Therefore, the molecular gas phase is more dynamic and variations in the gas densities are to be expected across cosmic time.

The H I column density contributing the most to the H I mass density is $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \sim 21$. These high column densities are not found in diffuse gas (e.g. the CGM), but are typical of column densities found the ISM.

At both $z = 0$ and 3, H_2 starts to dominate the mass density at column densities in the $\log(N_{\rm H_2}/{\rm cm}^{-2}) \sim 22$ range therefore showing little to no evolution of this observable. This is consistent with the predictions made by Schaye (2001), suggesting that little to no gas is found in the neutral phase at column densities above $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \gtrsim 22$ due to the clouds turning molecular at those column densities.

7 DISCUSSION

Given the evolution of the H_2 comoving mass density over cosmic time (e.g. Riechers et al. 2019; Decarli et al. 2020; Péroux & Howk 2020), changes in the normalization or shape of $f(N_{\rm H_2})$ are expected. The $f(N_{\rm H_2})$ derived from both observations and simulations corroborate this hypothesis with various changes of the $f(N_{\rm H_2})$ across cosmic time. In general, the combined results of observations and simulations imply that molecular gas is more often found in systems of higher column densities at $z = 3$ when compared to $z = 0$. These changes in the $f(N_{\rm H_2})$ are in line with the higher comoving molecular mass densities detected at $z = 3$. Combined with the higher SFR density detected around cosmic noon (e.g. Madau & Dickinson 2014), the results imply that the overall denser molecular gas at higher redshifts lead to a higher global SFR. While we study global properties in this work, these results are similar to findings of local observations of nearby star-forming galaxies where a correlation between the SFR surface density and H_2 surface density is well established (e.g. the molecular Schmidt law in Bigiel et al. 2008).

Rahmati et al. (2013) have demonstrated that observed $f(N_{\rm H\,I})$ can be accurately reproduced using the cosmological hydrodynamical simulation EAGLE (Schaye et al. 2015). Similarly, at $z = 3$, the cosmological simulation Illustris (Genel et al. 2014; Vogelsberger et al. 2014a,b; Sijacki et al. 2015) reproduces $f(N_{\rm H\,I})$ of observations (Noterdaeme et al. 2009; Prochaska, O'Meara & Worseck 2010; Zafar et al. 2013) accurately. However, there are still tensions between Illustris and observations below $z = 3$ (Bird et al. 2014). However, Villaescusa-Navarro et al. (2018) demonstrate that these tensions are not apparent in the successor of Illustris. Comparing results by TNG100 of the IllustrisTNG project with observations Villaescusa-Navarro et al. (2018) find that $f(N_{\rm H\,I})$ is accurately reproduced at $z \lesssim 5$.

While it has been demonstrated that $f(N_{\rm H\,I})$ is consistent with observations in different (magneto-)hydrodynamical cosmological simulations, there are still a number of inconsistencies for $f(N_{\rm H_2})$, despite the broad similarities of simulated and observed $f(N_{\rm H_2})$. At $z = 0$, Klitsch et al. (2019) demonstrate that TNG100 predict more low-column-density molecular gas compared to constraints by the ALMACAL survey (e.g. Oteo et al. 2016; Bonato et al. 2018; Klitsch et al. 2018) and, similarly to this work, does not reach the high column densities detected in observations (Zwaan & Prochaska 2006). These shortcomings are, in part, due to TNG100 not resolving the cold gas phase of the ISM. These simulation specifics stem from limitations in resolution and subgrid star formation models. Further, at $z = 0$, TNG100 might over predict H_2 compared to observational findings (Diemer et al. 2019), especially when not taking observational apertures into account (Popping et al. 2019).

At $z = 3$, we find an ∼1-dex difference in normalization for $f(N_{\rm H_2})$, which could arise due to the difference in selection and environments probed. The observational $f(N_{\rm H_2})$ at $z = 3$ is based on DLA studies. DLAs mostly trace the outskirts of galaxies (e.g. Péroux et al. 2011; Christensen et al. 2014; Krogager et al. 2017) and are often associated with group environments (Hamanowicz et al. 2020), while in TNG100, the full disc with no constraints on the environment of the galaxies is probed. Therefore, further efforts, on both the observational and simulation side are needed. On the simulation side more accurate representations of the cold gas phase is needed, including different subgrid models of star formation, higher resolution and the inclusion of non-equilibrium chemistry. On the observational side, we need better constraints of $f(N_{\rm H_2})$, especially at $z = 3$. Preferably, this could be achieved by a combination of high-spatial-resolution galaxy observations and a larger sample of H_2 absorption line systems at $z = 3$.

Non-equilibrium chemistry networks (e.g. Glover & Mac Low 2007b; Glover & Clark 2012; Gong, Ostriker & Wolfire 2017) have recently been used to model the cold gas phase in simulations on the fly. Such models have been implemented in simulations of individual regions of galactic discs (e.g. Walch et al. 2015; Hu, Sternberg & van Dishoeck 2021; Rathjen et al. 2021), isolated galaxies e.g. Hu et al. 2016; Richings & Schaye 2016; Lahén et al. 2019), and more recently in cosmological simulations (Maio et al. 2022). These studies have shown that non-equilibrium chemistry, e.g. heavily influences the H_2 mass fraction at low metallicities (Hu et al. 2021), affect the chemical make-up of outflows (Richings & Schaye 2016) and more accurately reproduce cosmological H_2 mass densities of observations (Maio et al. 2022). As a first attempt to study how and if non-equilibrium chemistry affects $f(N_{\rm H_2})$, we compare the time- and inclination-averaged $f(N_{\rm H_2})$ derived from a dwarf galaxy simulation by the GRIFFIN Project with the $f(N_{\rm H_2})$ derived by TNG100. The normalization of the dwarf galaxy $f(N_{\rm H_2})$ in the overlapping column density region is lower than for TNG100 $f(N_{\rm H_2})$ due to the highly different stellar masses that are probed. Interestingly, the slope of the $f(N_{\rm H_2})$ is similar, even though the samples and cold gas models are vastly different. We cannot disentangle the effects that non-equilibrium chemistry and the different samples have on $f(N_{\rm H_2})$ with our current study. None the less, this could be a first indication that non-equilibrium chemistry might not affect the slope of $f(N_{\rm H_2})$, especially since the slope of $f(N_{\rm H_2})$ for individual main-sequence star-forming galaxies in TNG100 is similar below $\log(N_{\rm H_2}/{\rm cm}^{-2}) \lesssim 20$ and not majorly affected by galaxy properties. However, comparisons between simulation runs of the same galaxy with and without non-equilibrium chemistry could help understand if and how $f(N_{\rm H_2})$ is affected by non-equilibrium chemistry. Further,

studies with larger samples, similar to Maio et al. (2022), are needed to further investigate how non-equilibrium chemistry might affect $f(N_{H_2})$.

The global $f(N_{H_2})$ and that of individual main-sequence star-forming galaxies give first indications that its shape could be related to the gas distribution within gas discs. Exponential gas distributions have not only been observed in disc galaxies (e.g. Leroy et al. 2008), but also reproduced in simulated ones (e.g. in TNG100, Stevens et al. 2019). An analytical model, based on exponential gas distribution in discs (Zwaan 2000) broadly reproduces $f(N_{H_2})$ of simulations and observations and is giving a first indication that these two distributions are related. Nonetheless, analytical models with e.g. Gaussian gas distributions in gas discs yield similar results. Therefore, it currently remains unclear how closely coupled the shape of $f(N_{H_2})$ and the gas distribution in gas discs are. Further studies are needed for a complete understanding to confirm the hypothesis of this connection between these two observables.

At $z = 0$, observations have shown that neutral atomic hydrogen dominates the total mass of the neutral ISM, with $M_{H\,I} \sim 2$–$10 M_{mol}$ (e.g. Saintonge et al. 2011; Saintonge & Catinella 2022). In studies at higher redshifts, it is often assumed that the neutral atomic component can be omitted and H_2 is assumed to be the dominant gas component in galaxies (e.g. between $z = 0.4$ and 4; Tacconi et al. 2018). In part, this is due to technical limitations, as the H I 21-cm emission line is not observable at higher redshifts with current instruments. Further, the molecular mass density peaks within this redshift range, while the neutral atomic mass density remains fairly constant across cosmic time, possibly making molecular gas an important contributor to the overall gas mass of galaxies within this redshift range (especially around cosmic noon). However, it still remains unclear what the contribution of the neutral atomic gas phase is to galaxies at higher redshifts. Heintz et al. (2021) have given first indications of the contribution of H I at higher redshifts, by exploiting [C II] as a tracer for neutral atomic gas. The results indicate that at $z = 4$–6, the contribution of H I is substantial, with the H I mass being equal to the dynamical mass of galaxies. At $z \sim 2$, the contribution of H I is found to be less substantial, with the H I mass being between 0.2 and 1 dex lower than the dynamical mass of galaxies. Therefore, at $z \sim 2$, the contribution by molecular gas or the stellar component is possibly higher. Comparing $f(N_{H_2})$ and $f(N_{H\,I})$ we, however, find that H I is an important contributor to the overall cold gas mass found in the ISM of galaxies (see Section 5) at both redshifts $z = 0$ and 3. We therefore caution from omitting the neutral atomic gas component in studies at these redshifts.

8 CONCLUSIONS

In this work we study the H_2 column density distribution [$f(N_{H_2})$] at redshift $z = 0$ and 3 using observations and simulations. On the observational side we use data from the PHANGS-ALMA survey (Leroy et al. 2021) at $z = 0$ and from an H_2 absorption line study by Balashev & Noterdaeme (2018) at $z = 3$ based on SDSS data. On the simulation side, we use data from TNG100 of the IllustrisTNG project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018) at both redshifts $z = 0$ and 3 and a high-resolution isolated dwarf galaxy simulation including a non-equilibrium chemical network by the GRIFFIN project (Lahén et al. 2019, 2020a,b) meant to represent a low-redshift dwarf galaxy.

In summary our analysis includes the following studies:

(i) We study how the integrated properties of galaxies in the PHANGS-ALMA sample shape the $f(N_{H_2})$ of individual objects.

(ii) We contrast the $f(N_{H_2})$ from observations and simulations to test how predictions made by TNG100 match observations.

(iii) We study how well analytical models match results by TNG100.

(iv) We compare results from a simulation including non-equilibrium chemistry (GRIFFIN Project) with results from the post-processed simulation TNG100.

(v) We study the evolution of $f(N_{H_2})$ from $z = 3$ to 0.

(vi) We explore which column densities contribute most to the overall H_2 and H I mass density at $z = 0$ and 3.

(vii) We investigate how the $f(N_{H_2})$ compare to $f(N_{H\,I})$ based on the WHISP sample (Zwaan et al. 2005), EAUDP sample (Zafar et al. 2013), and SDSS data (Ho et al. 2021) to examine in which regions of galaxies molecular gas dominates over neutral atomic gas.

In conclusion our findings are the following:

(i) The shapes of the $f(N_{H_2})$ of individual galaxies in the PHANGS-ALMA and the TNG100 sample at $z = 0$ are similar. This is possibly related to the galaxies in the sample. The sample consists of main-sequence star-forming galaxies, which typically have rotating discs and are hypothesized to have radially exponential gas profiles. (Leroy et al. 2008; Stevens et al. 2019). The radially exponential gas profiles could potentially be the cause of the similar $f(N_{H_2})$ observed for individual galaxies. Further, the normalization of $f(N_{H_2})$ and highest observed H_2 column densities depend on the integrated SFR, stellar mass (M_*), and H_2 mass (M_{H_2}) of the galaxy. More massive galaxies lead to a higher normalization of the $f(N_{H_2})$ of individual galaxies. The $f(N_{H_2})$ indicates that more massive galaxies produce more dense gas.

(ii) TNG100 broadly reproduces the $f(N_{H_2})$ we observe at both $z = 0$ and 3, albeit with some key differences. At $z = 0$ TNG100 produces steeper slopes for the $f(N_{H_2})$ compared to PHANGS-ALMA. Further, observations detect column densities up to $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 24$ at $z = 0$. Such high column densities are not present in TNG100 at that redshift. This is potentially due to resolution effects and the star formation subgrid interstellar medium model, both of which could inhibit the formation of high column densities of cold gas phases. At $z = 3$, the normalization of the $f(N_{H_2})$ is higher in the simulations compared to observations for the majority of the regions. This is likely due to the different environments probed by SDSS observations. None the less, the slopes $f(N_{H_2})$ in TNG100 and from observations are in good agreement at $z = 3$.

(iii) The dwarf galaxy simulation from the GRIFFIN project produces similar slopes as TNG100 for $f(N_{H_2})$ in the overlapping column density region. It is surprising that the slope of $f(N_{H_2})$ of a single simulated galaxy including a non-equilibrium chemistry network is so similar to the slope of a large sample of galaxies where H_2 was derived using post-processing prescriptions. This could be a first indication that non-equilibrium chemistry might not majorly affect the slope of $f(N_{H_2})$. However, further studies are needed to understand how and if non-equilibrium chemistry affects $f(N_{H_2})$.

(iv) The slopes of $f(N_{H_2})$ below $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 20$ show little to no evolution from $z = 3$ to 0. As indicated by the $f(N_{H_2})$ derived from TNG100, we expect an evolution of the $f(N_{H_2})$ to arise at higher column densities.

(v) The mass density distributions of the neutral atomic and molecular gas phase indicate that H I dominates over H_2 at most column densities and shows that H I could be an important contributor to the cold gas mass of galaxies at $z = 0$ and 3.

(vi) The H_2 column density contributing most to the overall molecular gas density evolves with redshift. When combining data from observations and simulations, we find that the shift is from

$\log(N_{H_2}/cm^{-2}) \sim 20.5$–$21.5$ at $z = 0$ to $\log(N_{H_2}/cm^{-2}) \sim 21.9$–$22$ at $z = 3$. We therefore find that more gas in denser regions is found at $z = 3$ compared to $z = 0$. These results are in line with observations of the SFR across cosmic time, which is higher at $z = 3$ compared to $z = 0$. The shape of the cosmic molecular mass density as a function of redshift is similar to the shape of the SFR density, making a coupling of these two quantities likely. Therefore, one would expect that more molecular gas found in denser regions leads to a higher global SFR of in galaxies (Péroux & Howk 2020). When assuming that the column density relates to the density of the gas, this is exactly what we observe when studying the column density distributions at both redshifts.

(vii) Contrary to H_2, the column density contributing most to the H I gas mass density [$\log(N_{HI}/cm^{-2}) \sim 21$] does not evolve with redshift. Given that the molecular phase of the gas cycle is likely to be shorter than the neutral atomic gas phase, more variations in the molecular gas densities are expected. The highest column density contribution of H I is therefore found in regions of the ISM, and not in more diffuse regions like the CGM.

(viii) H_2 starts dominating compared to H I at column densities above $\log(N_{H_2}/cm^{-2}) \sim 21.8$–$22$ at both redshifts. This is consistent with results by Schaye (2001), who predicted that H I clouds with $\log(N_{HI}/cm^{-2}) \gtrsim 22$ do not occur due to the clouds turning molecular before reaching higher column densities. Further, this implies that neutral gas is an important contributor to the overall gas mass found in the ISM of galaxies, including column density regions typical for molecular clouds.

(ix) In order to further constrain the evolution of $f(N_{H_2})$, additional observations and simulations are needed: At $z = 0$, deeper observations are needed to constrain the low-density end of $f(N_{H_2})$. At $z = 3$, high-spatial-resolution molecular gas observations of galaxies would enable the study of the high-column-density end of $f(N_{H_2})$ and also probe more central regions of galaxies compared to absorption line studies. On the simulational side, efforts on resolving the cold gas phase within simulations are needed to constrain the high-column-density end of $f(N_{H_2})$ at $z = 0$. This may necessitate higher resolution simulations together with physical models for interstellar medium gas that aim to resolve the coldest phases. Further, the use of non-equilibrium chemical networks could provide a more accurate representation of the cold gas phase (e.g. Maio et al. 2022).

ACKNOWLEDGEMENTS

The Authors thank Jiayi Sun for the help with data from the PHANGS-ALMA survey and his comments and suggestions. We also would like to thank Sergei Balashev, Chia-Yu Hu, Gergö Popping, and Thorsten Naab for the discussions, suggestions and comments that greatly helped to improve this paper. RS thanks ESO and the IMPRS program for the support of his PhD. DN acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1). The computations for the GRIFFIN project were carried out at CSC – IT Center for Science Ltd. in Finland and the MPA cluster FREYA hosted by The Max Planck Computing and Data Facility (MPCDF) in Garching, Germany.

DATA AVAILABILITY

Data directly related to this publication and its figures will be made available on request from the corresponding author. The IllustrisTNG simulations are publicly available and accessible in their entirety at www.tng-project.org/data (Nelson et al. 2019a).

REFERENCES

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Balashev S. A., Noterdaeme P., 2018, MNRAS, 478, L7
Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Bonato M. et al., 2018, MNRAS, 478, 1512
Braun R., 2012, ApJ, 749, 87
Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, MNRAS, 445, 225
Clark P. C., Glover S. C. O., 2014, MNRAS, 444, 2396
Cooke R. J., Pettini M., Steidel C. C., 2018, ApJ, 855, 102
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Decarli R. et al., 2020, ApJ, 902, 110
Diemer B. et al., 2018, ApJS, 238, 33
Diemer B. et al., 2019, MNRAS, 487, 1529
Emsellem E. et al., 2021, preprint (arXiv:2110.03708)
Feldmann R., 2020, Commun. Phys., 3, 226
French D. M. et al., 2021, ApJ, 923, 24
Genel S. et al., 2014, MNRAS, 445, 175
Glover S. C. O., Clark P. C., 2012, MNRAS, 421, 116
Glover S. C. O., Mac Low M.-M., 2007a, ApJS, 169, 239
Glover S. C. O., Mac Low M.-M., 2007b, ApJ, 659, 1317
Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88
Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55
Gong M., Ostriker E. C., Wolfire M. G., 2017, ApJ, 843, 38
Hamanowicz A. et al., 2020, MNRAS, 492, 2347
Heintz K. E., Watson D., Oesch P. A., Narayanan D., Madden S. C., 2021, ApJ, 922, 147
Hislop J. M., Naab T., Steinwandel U. P., Lahén N., Irodotou D., Johansson P. H., Walch S., 2021, MNRAS, 509, 5938
Ho M.-F., Bird S., Garnett R., 2021, MNRAS, 507, 704
Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, MNRAS, 443, 1173
Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528
Hu C.-Y., Naab T., Glover S. C. O., Walch S., Clark P. C., 2017, MNRAS, 471, 2151
Hu C.-Y., Sternberg A., van Dishoeck E. F., 2021, ApJ, 920, 26
Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2
Klitsch A., Péroux C., Zwaan M. A., Smail I., Oteo I., Biggs A. D., Popping G., Swinbank A. M., 2018, MNRAS, 475, 492
Klitsch A. et al., 2019, MNRAS, 490, 1220
Krogager J.-K., Noterdaeme P., 2020, A&A, 644, L6
Krogager J. K., Møller P., Fynbo J. P. U., Noterdaeme P., 2017, MNRAS, 469, 2959
Krumholz M. R., 2013, MNRAS, 436, 2747
Lagos C. d. P. et al., 2015, MNRAS, 452, 3815
Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2019, ApJ, 879, L18
Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., Steinwandel U. P., Moster B. P., 2020a, ApJ, 891, 2
Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2020b, ApJ, 904, 71
Lee J. C. et al., 2021, ApJS, 58, 10
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Leroy A. K. et al., 2013, AJ, 146, 19
Leroy A. K. et al., 2021, ApJS, 257, 61
Liu D. et al., 2019, ApJ, 887, 235

Madau P., Dickinson M., 2014, ARA&A, 52, 415
Maio U., Péroux C., Ciardi B., 2022, A&A, 657, A47
Marinacci F. et al., 2018, MNRAS, 480, 5113
Mas-Ribas L. et al., 2017, ApJ, 846, 4
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson R. P., Langer W. D., 1997, ApJ, 482, 796
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019a, Comput. Astrophys. Cosmol., 6, 2
Nelson D. et al., 2019b, MNRAS, 490, 3234
Nilson P., 1973, Uppsala General Catalogue of Galaxies (Uppsala Astron. Obs. Ann.), Vol. 6. Uppsala University Publications, Uppsala, Sweden
Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087
Noterdaeme P. et al., 2012, A&A, 547, L1
Oteo I., Zwaan M. A., Ivison R. J., Smail I., Biggs A. D., 2016, ApJ, 822, 36
Péroux C., Howk J. C., 2020, ARA&A, 58, 363
Péroux C., Dessauges-Zavadsky M., D'Odorico S., Sun Kim T., McMahon R. G., 2005, MNRAS, 363, 479
Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011, MNRAS, 410, 2237
Pessa I. et al., 2021, A&A, 650, A134
Pillepich A. et al., 2018, MNRAS, 473, 4077
Pillepich A. et al., 2019, MNRAS, 490, 3196
Planck Collaboration VI, 2016, A&A, 594, A13
Popping G. et al., 2019, ApJ, 882, 137
Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392
Rahmani H. et al., 2018a, MNRAS, 474, 254
Rahmani H. et al., 2018b, MNRAS, 480, 5046
Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427
Rathjen T.-E. et al., 2021, MNRAS, 504, 1039
Richings A. J., Schaye J., 2016, MNRAS, 458, 270
Riechers D. A. et al., 2019, ApJ, 872, 7
Röttgers B., Naab T., Cernetic M., Davé R., Kauffmann G., Borthakur S., Foidl H., 2020, MNRAS, 496, 152
Saintonge A., Catinella B., 2022, preprint (arXiv:2202.00690)
Saintonge A. et al., 2011, MNRAS, 415, 32
Sánchez S. F. et al., 2014, A&A, 563, A49
Sánchez S. F. et al., 2019, MNRAS, 484, 3042
Schaye J., 2001, ApJ, 562, L95
Schaye J. et al., 2015, MNRAS, 446, 521
Schroetter I. et al., 2019, MNRAS, 490, 4368
Seifried D. et al., 2017, MNRAS, 472, 4797
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Spilker A., Kainulainen J., Orkisz J., 2021, A&A, 653, 36
Springel V., 2005, MNRAS, 364, 1105
Springel V., 2010, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V. et al., 2018, MNRAS, 475, 676
Stevens A. R. H., Diemer B., Lagos C. D. P., Nelson D., Obreschkow D., Wang J., Marinacci F., 2019, MNRAS, 490, 96
Sun J. et al., 2018, ApJ, 860, 172
Sun J. et al., 2020, ApJ, 901, L8
Szakacs R. et al., 2021, MNRAS, 505, 4746
Tacconi L. J. et al., 2018, ApJ, 853, 179
Tacconi L. J., Genzel R., Sternberg A., 2020, ARA&A, 58, 157
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Walch S. et al., 2015, MNRAS, 454, 238
Walter F. et al., 2020, ApJ, 902, 111
Weigel A. K., Schawinski K., Bruderer C., 2016, MNRAS, 459, 2150
Weinberger R. et al., 2017, MNRAS, 465, 3291
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Yates R. M., Péroux C., Nelson D., 2021, MNRAS, 508, 3535
Zabl J. et al., 2020, MNRAS, 492, 4576
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141
Zwaan M. A., 2000, PhD thesis
Zwaan M. A., Prochaska J. X., 2006, ApJ, 643, 675
Zwaan M. A., van der Hulst J. M., Briggs F. H., Verheijen M. A. W., Ryan-Weber E. V., 2005, MNRAS, 364, 1467
den Brok J. S. et al., 2021, MNRAS, 504, 3221
van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, L85
van der Hulst J. M., van Albada T. S., Sancisi R., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, ASP Conf. Ser. Vol. 240, Gas and Galaxy Evolution. Astron. Soc. Pac., San Francisco, p. 451

APPENDIX A: $f(N_{H_2})$ DEPENDENCE ON PHYSICAL PROPERTIES

We explore how the integrated properties of galaxies in the PHANGS-ALMA sample shape the $f(N_{H_2})$ of individual objects. Namely we study the dependence of the $f(N_{H_2})$ on the SFR, stellar mass (M_*), and H_2 mass (M_{H_2}).

The colour coding in Fig. 3 already displays the dependence of the $f(N_{H_2})$ on the three parameters mentioned above. In order to quantify this relationship, we fit a gamma distribution of the form:

$$f(N_{H_2}) = \frac{f^*}{N^*} \left(\frac{N_{H_2}}{N^*}\right)^{-\beta} e^{-N_{H_2}/N^*}, \quad \text{(A1)}$$

to the computed individual $f(N_{H_2})$ of the 150-pc resolution PHANGS-ALMA sample. Note that there is no physical motivation for fitting a gamma distribution to the individual $f(N_{H_2})$, it simply provides good fits of the $f(N_{H_2})$ for a minimal number of parameters.

The individual $f(N_{H_2})$ is largely determined by the parameter N^* as the second free parameter f^* correlates with N^* (slope: -0.74 ± 0.09, intercept: 14.3 ± 2.0, Pearson-r: 0.7, p-value [calculated using a Kolmogorov–Smirnov test] <0.05) and β in turn correlates with f^* (slope: -0.55 ± 0.06, intercept 0.11 ± 0.13, Pearson-r:0.73, p-value <0.05). Fig. A1 displays the dependence of f^* on N^* and β on f^*.

As already indicated in Fig. 3, $f(N_{H_2})$ depends on physical parameters of the galaxies within the PHANGS-ALMA sample. This is quantified in Fig. A2, where we show the relationship of the free parameter N^* of the gamma distribution with SFR, M_*, and Σ_{H_2}. As N^* largely determines the $f(N_{H_2})$ of a galaxy, it is implied that these physical properties of a galaxy affect the $f(N_{H_2})$ of a galaxy. The three studied properties of the galaxies show the following correlation and fit parameters (in log space): SFR–N^*: slope: 0.90 ± 0.15 intercept: 21.90 ± 0.07, Pearson-r $= 0.59$, p-value < 0.05; M_*–N^*: slope: 1.21 ± 0.05 intercept: -5.5 ± 1.1 Pearson-r $= 0.62$, p-value < 0.05 and M_{H_2}–N^*: slope: 0.79 ± 0.11, intercept: 14.9 ± 1.0 Pearson-r: 0.64, p-value < 0.05.

Using the SFR, M_*, or M_{H_2} of a galaxy one could approximate its $f(N_{H_2})$ using these correlations. We note that our tests have shown that while these fits approximate the global $f(N_{H_2})$ well when using the PHANGS-ALMA sample, they often fail for individual galaxies because they are degenerate. We therefore caution from using these fits to predict individual $f(N_{H_2})$.

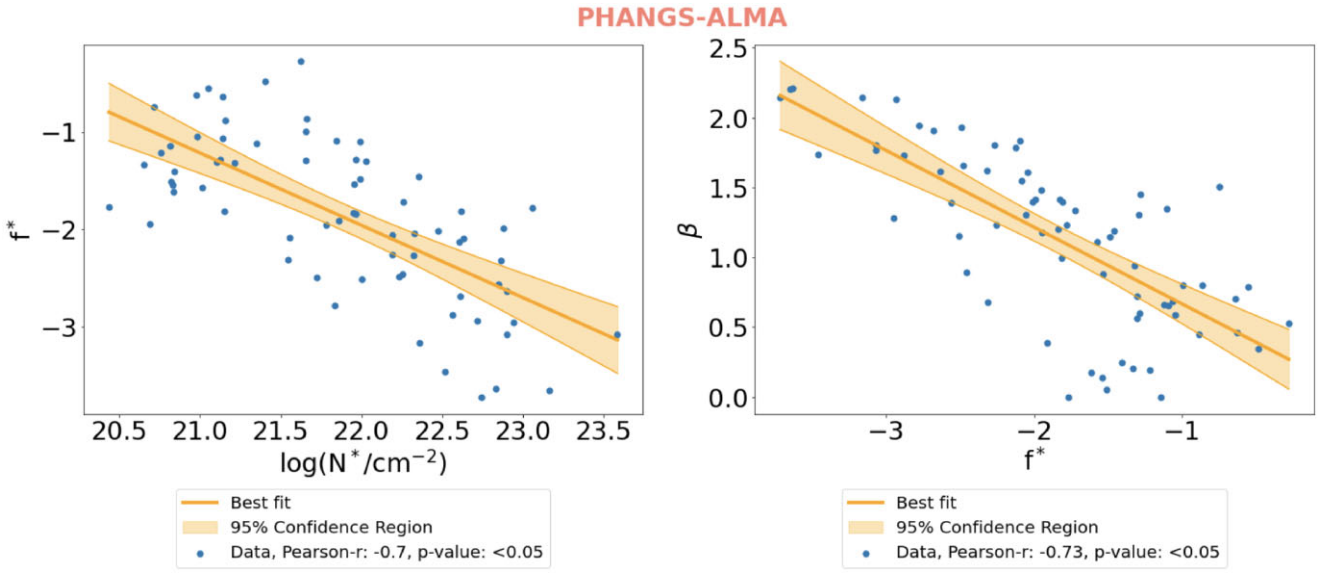

Figure A1. Correlations between the free parameter f^* and the free parameter N^* and between f^* and the slope β of the gamma distribution fits on the $f(N_{H_2})$ in the PHANGS-ALMA sample. The blue dots indicate the fit values of individual galaxies, the orange line the best fit, and the orange band the 95 per cent confidence region of the fit. Both samples show strong correlations with Pearson-rs of ∼0.7 and p-values < 0.05.

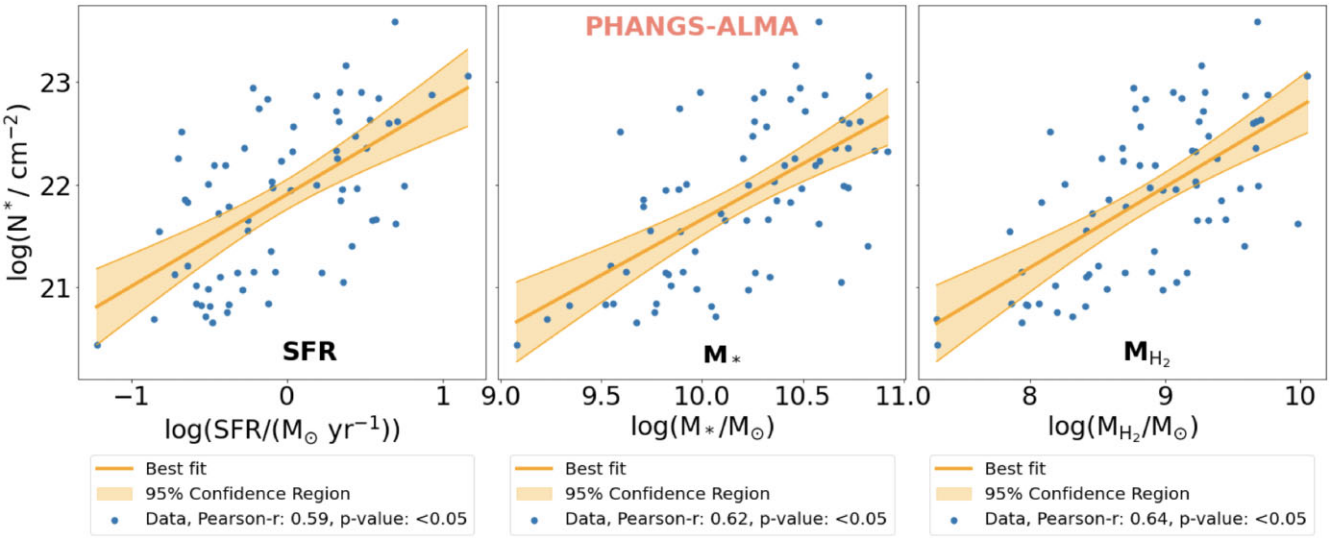

Figure A2. Dependence on different physical properties of the free parameter N^* of the gamma distribution fits on the $f(N_{H_2})$ within the PHANGS-ALMA sample. The blue dots indicate the fit values and properties of individual objects, the orange line the best fit, and the orange band the 95 per cent confidence region of the fit. The physical properties (SFR, M_*, and M_{H_2}) of the galaxies correlate with the free parameter N^*, albeit with a significant scatter.

This paper has been typeset from a TEX/LATEX file prepared by the author.