MNRAS **452,** 3815–3837 (2015) doi:10.1093/mnras/stv1488

Molecular hydrogen abundances of galaxies in the EAGLE simulations

Claudia del P. Lagos,[1,2]★ Robert A. Crain,[3,4] Joop Schaye,[4] Michelle Furlong,[5] Carlos S. Frenk,[5] Richard G. Bower,[5] Matthieu Schaller,[5] Tom Theuns,[5] James W. Trayford,[5] Yannick M. Bahé[6] and Claudio Dalla Vecchia[7,8]

[1]*International Centre for Radio Astronomy Research (ICRAR), M468, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia*
[2]*European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany*
[3]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[4]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[5]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[6]*Max-Planck-Institut fur Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[7]*Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain*
[8]*Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísico Franciso Sánchez s/n, E-38206 La Laguna, Tenerife, Spain*

Accepted 2015 July 2. Received 2015 June 1; in original form 2015 March 16

ABSTRACT

We investigate the abundance of galactic molecular hydrogen (H_2) in the 'Evolution and Assembly of GaLaxies and their Environments' (EAGLE) cosmological hydrodynamic simulations. We assign H_2 masses to gas particles in the simulations in post-processing using two different prescriptions that depend on the local dust-to-gas ratio and the interstellar radiation field. Both result in H_2 galaxy mass functions that agree well with observations in the local and high-redshift Universe. The simulations reproduce the observed scaling relations between the mass of H_2 and the stellar mass, star formation rate and stellar surface density. Towards high redshifts, galaxies in the simulations display larger H_2 mass fractions and lower H_2 depletion time-scales, also in good agreement with observations. The comoving mass density of H_2 in units of the critical density, Ω_{H_2}, peaks at $z \approx 1.2$–1.5, later than the predicted peak of the cosmic star formation rate activity, at $z \approx 2$. This difference stems from the decrease in gas metallicity and increase in interstellar radiation field with redshift, both of which hamper H_2 formation. We find that the cosmic H_2 budget is dominated by galaxies with $M_{H_2} > 10^9\,M_\odot$, star formation rates $> 10\,M_\odot\,yr^{-1}$ and stellar masses $M_{stellar} > 10^{10}\,M_\odot$, which are readily observable in the optical and near-IR. The match between the H_2 properties of galaxies that emerge in the simulations and observations is remarkable, particularly since H_2 observations were not used to adjust parameters in EAGLE.

Key words: ISM: abundances – galaxies: evolution – galaxies: formation – galaxies: ISM.

1 INTRODUCTION

The past few years have brought impressive developments in surveys of molecular gas in resolved and unresolved galaxies, locally and at high redshift (e.g. Leroy et al. 2008; Daddi et al. 2010; Genzel et al. 2010; Davis et al. 2011; Saintonge et al. 2011; Young et al. 2011; Genzel et al. 2013; Boselli et al. 2014). Molecular hydrogen (H_2) is an important component of the interstellar medium (ISM) and, in galaxies like our own, it amounts to a few per cent of the stellar mass (Putman, Peek & Joung 2012). The exquisite resolution and sensitivity of studies of local galaxies have enabled important conclusions to be reached. For example, on kiloparsec scales, observations show a nearly linear correlation between the star formation rate (SFR) surface density and the surface density of H_2 (Bigiel et al. 2008, 2010; Leroy et al. 2008; Schruba et al. 2011). At a global level (i.e. integrated over the galaxy), the H_2 mass correlates very well with the total SFR (Saintonge et al. 2011; Boselli et al. 2014). These observations place new constraints on galaxy formation models.

Molecular hydrogen is very difficult to observe in the ISM of galaxies because it lacks a dipole moment, making its emission extremely weak at the typical temperature of the cold ISM. A widely used tracer of H_2 is the carbon monoxide (hereafter 'CO') molecule, which is the second most abundant molecule after H_2, and is easily excited (see Carilli & Walter 2013 for a recent review on how CO has been used to trace H_2 at $z \lesssim 6$). Direct CO detections are available for relatively large samples in the local Universe ($z < 0.1$),

★ E-mail: claudia.lagos@icrar.org

from which it has been possible to derive the $z \approx 0$ CO(1−0) luminosity function (Keres, Yun & Young 2003), where (1−0) is the lowest energy rotational transition. From this luminosity function, and adopting a Milky Way (MW) like CO(1−0)-H_2 conversion factor (Bolatto, Wolfire & Leroy 2013), it has been possible to derive the H_2 mass function and the cosmic density of H_2, $\rho_{H_2} = 4.3 \pm 1.1 \times 10^7\, h\, M_\odot\, Mpc^{-3}$, at $z \approx 0.05$ (Keres et al. 2003; Obreschkow & Rawlings 2009).

Campaigns to obtain constraints on ρ_{H_2} at redshifts higher than $z \approx 0$ have used blind CO surveys tuned to find CO emission at $z > 1$. For example, Aravena et al. (2012) performed a Jansky Very Large Array[1] blind survey towards a candidate cluster at $z \sim 1.5$, from which they obtained two detections and placed constraints on the number density of bright CO(1–0) galaxies. Walter et al. (2014) carried out a blind CO survey with the Plateau de Bure Interferometer and placed constraints on both the CO(1–0) luminosity function and ρ_{H_2} in the redshift range $z \approx$ 1–3. Both surveys indicate that ρ_{H_2} increases by an order of magnitude from $z = 0$ to $z \approx 2$, but systematic uncertainties still dominate the measurements.

There is also a wealth of literature using dust mass as a proxy for H_2 mass through the dust-to-gas mass ratio dependence on metallicity (e.g. Boselli, Lequeux & Gavazzi 2002; Santini et al. 2014; Swinbank et al. 2014; Béthermin et al. 2015). Using this technique, and with the availability of IR photometry for large samples of galaxies from *Herschel*, one can obtain more statistically meaningful samples than is possible with direct CO imaging. So far results from both direct CO detections and dust-derived H_2 masses are in good agreement. For both techniques, scaling relations between the H_2 mass and other galaxy properties have been explored locally and at high redshift. It has been established that the H_2-to-stellar mass ratio anticorrelates with stellar mass and that such a relation is present from $z = 0$ to at least $z \approx 2.5$ (e.g. Saintonge et al. 2011, 2013; Tacconi et al. 2013; Bothwell et al. 2014; Dessauges-Zavadsky et al. 2015; Santini et al. 2014). These studies have also conclusively shown that there is an overall tendency for an increasing H_2-to-stellar mass ratio with increasing redshift at fixed stellar mass (e.g. Geach et al. 2011; Saintonge et al. 2013). This evidence suggests that gas in galaxies at high redshift is denser.

The availability of high-fidelity observations of molecular gas in the local and distant Universe has opened up new means of testing galaxy formation models and simulations. This opportunity has been exploited using semi-analytic models which, when coupled with a star formation law that relates the SFR directly to the content of H_2, reproduce many of the above correlations (e.g. Fu et al. 2010; Lagos et al. 2011b,a, 2014b Popping et al. 2014). However, all these models assume a constant H_2 to SFR conversion efficiency, even though there is observational evidence contradicting this (e.g. Saintonge et al. 2013; Dessauges-Zavadsky et al. 2015; Huang & Kauffmann 2014).

Hydrodynamic simulations are well suited to investigate the relation between SFR, H_2 mass, stellar mass and other galaxy properties because the gas dynamics is followed self-consistently and no relations between the global (galaxy-wide) and local properties of the gas are necessarily imposed (which is the case in semi-analytic models, where galaxies are unresolved). Thus, one can use observations of H_2 to test whether the simulations reproduce the scaling relations of the gas component of galaxies and isolate the physical drivers of the observed relations.

[1] https://science.nrao.edu/facilities/vla

Many of the scaling relations discussed above, for example those between the surface density of SFR, H_2, and other properties of the ISM of galaxies (such as hydrostatic pressure), have been explored in simulations of small portions of the ISM and in individual galaxy simulations in which H_2 formation has been implemented (e.g. Pelupessy, Papadopoulos & van der Werf 2006; Robertson & Kravtsov 2008; Gnedin, Tassis & Kravtsov 2009; Pelupessy & Papadopoulos 2009; Christensen et al. 2012; Glover & Clark 2012; Walch et al. 2015). For example, the physical drivers of the relation between the surface density of SFR and H_2 have been explored in Feldmann, Gnedin & Kravtsov (2011), Gnedin & Kravtsov (2011, hereafter GK11), and Glover & Clark (2012). Feldmann et al. (2011) and Gnedin & Kravtsov (2011) show that the latter relation is not fundamental in the sense that large variations are expected for varying physical conditions (e.g. gas metallicity and interstellar radiation field – ISRF). In addition, Glover & Clark (2012) argued against the importance of molecular cooling (e.g. from H_2 and CO) as the low temperatures and high-densities needed to form stars are reached even in situations where the gas is forced to remain atomic (see also Schaye 2004). We can implement some of the relations derived from the high-resolution simulations above in lower resolution, cosmological hydrodynamic simulations to explore the scaling relation between the gas properties and other galaxy properties.

Recently, significant advances have been achieved in cosmological hydrodynamical simulations, which can now produce a galaxy population with properties that are broadly consistent with many observations (see Vogelsberger et al. 2014 for the Illustris simulation and Schaye et al. 2015 for the Evolution and Assembly of GaLaxies and their Environments, EAGLE, simulations). For example, EAGLE reproduces the stellar mass function of galaxies at $z \approx 0.1$, the SFR history of the Universe, and the properties of star-forming galaxies, in large cosmological volumes, ($\sim (100\,Mpc)^3$), with enough particles to resolve the Jeans scale in the warm ISM.

Genel et al. (2014) presented molecular gas scaling relations for the Illustris simulation, and showed that it reproduces the observed correlation between H_2 and stellar mass reasonably well if the star-forming gas (which has high enough density to allow star formation to take place) is used as a proxy for the H_2 mass. However, estimating the H_2 fraction from the neutral gas introduces a large systematic uncertainty because the range of temperatures, densities and metallicities of the star-forming gas indicates that, contrary to the assumption of Genel et al. (2014), it is not always H_2-dominated.

The OWLS simulation suite described in Schaye et al. (2010), a forerunner of EAGLE, explored lower resolution and smaller cosmological volumes than EAGLE. Duffy et al. (2012a) presented a comparison between OWLS and the H_2 mass function and scaling relations for galaxies with $M_{H_2} > 10^{10}\,M_\odot$, which was the mass range probed by their simulation. The authors found that the H_2 content of galaxies decreases with decreasing redshift, but their galaxies tended to be more H_2-rich than observed and the mass range probed was very limited. We extend this work using more sophisticated, theoretically motivated ways of calculating the H_2 fraction, and apply them to EAGLE, which has a much larger dynamical range than OWLS and, unlike OWLS, reproduces key properties of the observed galaxy population.

Here, we present a comprehensive comparison between EAGLE and observations of H_2 on galaxy scales. Compared with, for example, the work of Duffy et al. (2012b) and Genel et al. (2014), we extract a more realistic H_2 fraction for individual gas particles, based on their gas metallicity, density, pressure, SFR and Jeans length. The methods to compute the H_2 fraction of gas particles that we use here employ simple, analytic prescriptions that are based on

Table 1. EAGLE simulations used in this paper. The columns list: (1) the name of the simulation, (2) comoving box size, (3) number of particles, (4) initial particle masses of gas and (5) dark matter, (6) comoving gravitational softening length, and (7) maximum proper comoving Plummer-equivalent gravitational softening length. Units are indicated below the name of column. The top two simulations are of intermediate resolution, while the bottom two are high-resolution simulations. In EAGLE, we adopt (6) as the softening length at $z \geq 2.8$, and (7) at $z < 2.8$. At $z = 2.8$, 2.66 ckpc $\equiv$ 0.7 pkpc and 1.33 ckpc $\equiv$ 0.35 pkpc.

(1)	(2),	(3)	(4)	(5)	(6)	(7)
Name	L	# particles	Gas particle mass	DM particle mass	Softening length	Max. gravitational softening
units	(cMpc)		($M_\odot$)	($M_\odot$)	(ckpc)	(pkpc)
Ref-L025N0376	25	2×376^3	1.81×10^6	9.7×10^6	2.66	0.7
Ref-L050N0752	50	2×752^3	1.81×10^6	9.7×10^6	2.66	0.7
Ref-L100N1504	100	2×1504^3	1.81×10^6	9.7×10^6	2.66	0.7
Ref-L025N0752	25	2×752^3	2.26×10^5	1.21×10^6	1.33	0.35
Recal-L025N0752	25	2×752^3	2.26×10^5	1.21×10^6	1.33	0.35

more detailed studies. This kind of parametrization avoids the use of expensive radiative transfer and high-resolution simulations that are required to account for direct H_2 formation within a multiphase ISM. However, we note that in parallel to our simple approach, the development of detailed chemistry networks and their application to individual galaxies is under way (e.g. Glover & Jappsen 2007; Richings, Schaye & Oppenheimer 2014a,b). In future, the application of such networks to entire galaxies will enable us to test how well the phenomenological approach that we adopt here describes the ISM in galaxies, and to identify regimes within which our approach is inaccurate. An analysis of the atomic hydrogen content of galaxies in EAGLE will be presented in Crain et al. (in preparation), Bahé et al. (submitted) and Rahmati et al. (2015).

This paper is organized as follows. In Section 2, we give a brief overview of the simulation and the subgrid physics included in EAGLE. In Section 3, we describe how we partition the gas into ionized, atomic and molecular components, the methods we use to do this and their main differences, with detailed equations provided in Appendix A. Comparisons with local Universe and high-redshift surveys are presented in Sections 4 and 5, respectively. We discuss the results and present our conclusions in Section 6. In Appendix B, we present 'weak' and 'strong' convergence tests (terms introduced by Schaye et al. 2015), focusing on the H_2 properties of galaxies.

2 THE EAGLE SIMULATION

The EAGLE simulation suite[2] (described in detail by Schaye et al. 2015, hereafter S15, and Crain et al. 2015) consists of a large number of cosmological simulations with different resolution, box sizes and physical models, adopting the cosmological parameters of the Planck Collaboration XVI (2014). S15 introduced a reference model, within which the parameters of the subgrid models governing energy feedback were calibrated to ensure a good match to the $z = 0.1$ galaxy stellar mass function, whilst also reproducing the observed sizes of present-day disc galaxies. Furlong et al. (2015) presented the evolution of the galaxy stellar mass function for the EAGLE simulations and found that the agreement extends to $z \approx 7$.

The EAGLE suite includes simulations adopting the reference model, run within cosmological volumes of $(25\,\mathrm{cMpc})^3$, $(50\,\mathrm{cMpc})^3$ and $(100\,\mathrm{cMpc})^3$, where cMpc denotes comoving megaparsecs. In Table 1, we summarize technical details of the simulations used in this work, including the number of particles, volume, particle masses, and spatial resolution. The label of each simulation indicates box size and particle number. For example, L0100N1504 is a simulation of length 100 cMpc on a side, using 1504^3 particles of dark matter and an equal number of baryonic particles. Note that the spatial resolution of the intermediate-resolution simulations (such as L0100N1504) was chosen to (marginally) resolve the Jeans length in the warm ISM (with a temperature of $\approx 10^4$ K; see S15 for details). In Table 1, pkpc denotes proper kiloparsecs.

The EAGLE simulations were performed using an extensively modified version of the parallel *N*-body smoothed particle hydrodynamics (SPH) code GADGET-3 (Springel et al. 2008; see also Springel 2005 for the publicly available code GADGET-2). GADGET-3 is a Lagrangian code, where a fluid is represented by a discrete set of particles, from which the gravitational and hydrodynamic forces are calculated. SPH properties, such as the density and pressure gradients, are computed by interpolating across neighbouring particles. The main modifications to the standard GADGET-3 code for the EAGLE project include updates to the hydrodynamics, as described in Dalla Vecchia (in preparation; see also Appendix APPENDIX A: in S15), which are collectively referred to as 'Anarchy'. The impact of these changes on cosmological simulations are discussed in Schaller et al. (in preparation). Among the main features of Anarchy are (i) the pressure-entropy formulation of SPH described in Hopkins (2013), (ii) the artificial viscosity switch of Cullen & Dehnen (2010), (iii) an artificial conduction switch described in Price (2008), (iv) a C2 Wendland (1995) kernel with 58 neighbours to interpolate SPH properties across neighbouring particles, and (v) the timestep limiter from Durier & Dalla Vecchia (2012), which is required to model feedback events accurately.

Another major aspect of the EAGLE project is the use of state-of-the-art subgrid models that capture the unresolved physics. We briefly discuss the subgrid physics modules adopted by EAGLE in Section 2.1. In order to distinguish models with different parameter sets, a prefix is used. For example, Ref-L100N1504 corresponds to the reference model adopted in a simulation with the same box size and particle number of L100N1504. A complete description of the model can be found in S15 and an analysis of the impact of variations in the parameters of the subgrid physics on galaxy properties is given in Crain et al. (2015). Crain et al. (2015) also presented the motivation for the parameters adopted in each physics module in EAGLE.

S15 introduced the concept of 'strong' and 'weak' convergence tests. Strong convergence refers to the case where a simulation is re-run with higher resolution (i.e. better mass and spatial resolution) adopting exactly the same subgrid physics and parameters. Weak convergence refers to the case when a simulation is re-run with higher resolution but the subgrid parameters are recalibrated

[2] See http://eagle.strw.leidenuniv.nl and http://icc.dur.ac.uk/Eagle/ for images, movies and data products.

to recover, as far as possible, similar agreement with the adopted calibration diagnostic (in the case of EAGLE, the $z = 0.1$ galaxy stellar mass function and disc sizes of galaxies). S15 argue that simulations that do not resolve the cold ISM, nor the detailed effects of feedback mechanisms, require calibration of the subgrid model for feedback, and that such simulations should be recalibrated if the resolution is changed, because in practice changing the resolution leads to changes in the physics of the simulation as better spatial resolution allows higher gas densities to be reached. In this spirit, S15 introduced two higher resolution versions of EAGLE, both in a box of $(25\,\mathrm{cMpc})^3$ and with 2×752^3 particles, Ref-L025N0752 and Recal-L025N0752. These simulations have better spatial and mass resolution than the intermediate resolution simulations by factors of 2 and 8, respectively.

Comparisons between the simulations Ref-L025N0376 and Ref-L025N0752 represent strong convergence tests, whilst comparisons between the simulations Ref-L025N0376 and Recal-L025N0752 test weak convergence. In the simulation Recal-L025N0752, the values of four parameters have been slightly modified with respect to the reference simulation. These are related to the efficiency of feedback associated with star formation and active galactic nuclei (AGN), which are adjusted to reach a similar level of agreement with the galaxy stellar mass function at $z = 0.1$ as obtained for the Ref-L025N0376 simulation (see S15 for details). In Appendix B, we present strong and weak convergence tests explicitly involving the H_2 abundance of galaxies.

2.1 Subgrid physics included in EAGLE

(i) *Radiative cooling and photoheating.* Radiative cooling and heating rates are computed on an element-by-element basis for gas in ionization equilibrium exposed to the cosmic microwave background radiation (CMB), and the model for the UV and X-ray background radiation from Haardt & Madau (2001). Cooling tables are produced with CLOUDY (version 07.02; Ferland et al. 1998) for the 11 elements that dominate the cooling rate (i.e. H, He, C, N, O, Ne, Mg, S, Fe, Ca, Si), which are followed individually. For more details see Wiersma, Schaye & Smith (2009a) and S15.

(ii) *Star formation.* Gas particles that have cooled to reach densities greater than $n_{\rm H}^*$ are eligible for conversion to star particles, where $n_{\rm H}^*$ is defined as

$$n_{\rm H}^* = 10^{-1}\,\mathrm{cm}^{-3}\left(\frac{Z}{0.002}\right)^{-0.64}, \quad (1)$$

which has a dependence on gas metallicity, Z, as described in Schaye (2004) and S15. Gas particles with densities greater than $n_{\rm H}^*$ are assigned an SFR, $\dot{m}_\star$ (Schaye & Dalla Vecchia 2008):

$$\dot{m}_\star = m_{\rm g}\,A\,(1\,\mathrm{M}_\odot\,\mathrm{pc}^{-2})^{-n}\left(\frac{\gamma}{G}\,f_{\rm g}\,P\right)^{(n-1)/2}, \quad (2)$$

where $m_{\rm g}$ is the mass of the gas particle, $\gamma = 5/3$ is the ratio of specific heats, G is the gravitational constant, $f_{\rm g}$ is the mass fraction in gas (which is unity for gas particles), P is the total pressure. $A = 1.515 \times 10^{-4}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$ and $n = 1.4$ are chosen to reproduce the observed Kennicutt–Schmidt relation (Kennicutt 1998), scaled to a Chabrier initial mass function (IMF). A global temperature floor, $T_{\rm eos}(\rho)$, is imposed, corresponding to a polytropic equation of state,

$$P \propto \rho_{\rm g}^{\gamma_{\rm eos}}, \quad (3)$$

where $\gamma_{\rm eos} = 4/3$. Equation (3) is normalized to give a temperature $T_{\rm eos} = 8 \times 10^3$ K at $n_{\rm H} = 10^{-1}\,\mathrm{cm}^{-3}$, which is typical of the warm ISM (e.g. Richings et al. 2014b).

(iii) *Stellar evolution and enrichment.* The contribution from mass and metal loss from asymptotic giant branch (AGB) stars, massive stars and supernovae (both core collapse and type Ia) are tracked using the yield tables of Portinari, Chiosi & Bressan (1998), Marigo (2001), and Thielemann et al. (2003). The mass and metals that are lost from stars are added to the gas particles that are within the SPH kernel of the given star particle (see Wiersma et al. 2009b and S15 for details). The adopted stellar IMF is that of Chabrier (2003), with minimum and maximum masses of 0.1 and 100 $\mathrm{M}_\odot$.

(iv) *Stellar feedback.* The method used in EAGLE to represent energetic feedback associated with star formation (which we refer to as 'stellar feedback') was introduced by Dalla Vecchia & Schaye (2012), and consists of a stochastic selection of neighbouring gas particles that are heated by a temperature of $10^{7.5}$ K. A fraction of the energy from core-collapse supernovae is injected into the ISM 30 Myr after the star particle forms. This fraction, $f_{\rm th}$, depends on the local gas metallicity and density as

$$f_{\rm th} = f_{\rm th,min} + \frac{f_{\rm th,max} - f_{\rm th,min}}{1 + \left(\frac{Z}{0.1\,Z_\odot}\right)^{n_Z}\left(\frac{n_{\rm H,birth}}{n_{\rm H,0}}\right)^{-n_{\rm n}}}, \quad (4)$$

where Z is the star particle metallicity, $n_{\rm H,\,birth}$ is the gas density of the star particle's parent gas particle, when the star particle was formed. The parameters $f_{\rm th,\,min}$ and $f_{\rm th,\,max}$ are fixed at 0.3 and 3, respectively, for the simulations analysed here. In the reference model, $n_Z = n_{\rm n} = 2/\ln(10)$ and $n_{\rm H,0} = 0.67\,\mathrm{cm}^{-3}$, while in the recalibrated model $n_{\rm n}$ is reduced to $1/\ln(10)$ and $n_{\rm H,\,0}$ to $0.25\,\mathrm{cm}^{-3}$. In principle, $f_{\rm th,\,max} > 1$ may seem unphysical. However, the mechanism of stellar feedback in EAGLE includes several forms of feedback in the ISM: supernovae, stellar winds, radiation pressure, etc. Moreover, a subgrid efficiency greater than unity can be justified if there are numerical radiative losses. S15 quoted the mean and median values of $f_{\rm th}$ at $z = 0.1$ which are 1.06 and 0.7 for the Ref-L100N1504 simulations, and 1.07 and 0.93 for the Recal-L025N0752, respectively.

The metallicity dependence in equation (4) aims to take into account the faster cooling that takes place in regions of high metallicity, while the density dependence compensates for numerical radiative losses that occur at high density as a result of finite resolution (see Crain et al. 2015 and S15 for the motivation of this model). The need for this phenomenological model comes from the fact that stellar feedback is not simulated from first principles.

(v) *Black hole growth and AGN feedback.* When haloes become more massive than $10^{10}\,\mathrm{M}_\odot\,h^{-1}$, they are seeded with black holes of mass $10^5\,\mathrm{M}_\odot\,h^{-1}$. Subsequent gas accretion episodes make black holes grow at a rate that is computed following the modified Bondi–Hoyle accretion rate of Rosas-Guevara et al. (2013). In this modification, the angular momentum of the gas can reduce the accretion rate compared to the standard Bondi–Hoyle rate if the tangential velocity of the gas is similar to, or larger than, the local sound speed. A viscosity parameter, $C_{\rm visc}$, controls the degree of modulation of the Bondi rate in high circulation flows, and it is a free parameter of the model. The value of $C_{\rm visc}$ in the reference simulations is $C_{\rm visc} = 2\pi$, and in the recalibrated, higher resolution simulation it is $C_{\rm visc} = 2\pi \times 10^3$ (see Crain et al. 2015 for analysis of the effect of $C_{\rm visc}$ on galaxy properties). The Eddington limit is imposed as an upper limit to the accretion rate on to black holes. In addition, black holes can grow by merging with other black holes.

For AGN feedback, a similar model to the stochastic model of Dalla Vecchia & Schaye (2012) is applied. Particles surrounding

the black hole are chosen randomly and heated by a temperature $\Delta T_{\rm AGN} = 10^{8.5}$ K in the reference simulation and $\Delta T_{\rm AGN} = 10^{9}$ K in the recalibrated simulation (see Table 1 for the prefixes).

Throughout the paper, we make extensive comparisons between stellar mass, SFR, H_2 mass and metallicity. Following S15, all these properties are measured in three-dimensional apertures of 30 pkpc. The effect of the aperture is minimal, even if we double its size to 60 pkpc differences are within 15 per cent and much smaller for most galaxies.

3 CALCULATING H I AND H_2 GAS FRACTIONS

We describe how we calculate the neutral gas fraction of each gas particle in Section 3.1 and how much of that neutral gas is in molecular form in Section 3.2. We discuss two prescriptions for calculating the H_2 fraction and some basic predictions of each of them. For clarity, we introduce the number density of hydrogen, $n_{\rm H} \equiv n_{\rm H\,II} + n_{\rm H\,I} + 2\,n_{\rm H_2}$, and of neutral hydrogen, $n_{\rm n} \equiv n_{\rm H\,I} + 2\,n_{\rm H_2}$. We also refer to the fraction $f_{\rm H_2} \equiv \Sigma_{\rm H_2}/\Sigma_{\rm n}$, where $\Sigma_{\rm H_2}$ and $\Sigma_{\rm n}$ are the mass surface densities of H_2 and neutral hydrogen, respectively.

3.1 Calculating the neutral fraction

Before we calculate the mass of H_2 associated with each gas particle, we need to determine the transition from H II to neutral hydrogen. Rahmati et al. (2013a) studied the neutral gas fraction in cosmological simulations by coupling them to a full radiative transfer calculation with the aim of predicting the neutral gas column densities. Rahmati et al. (2013a) presented fitting functions to their results, which enable the calculation of the neutral fraction on a particle-by-particle basis from the gas temperature and density, and the total ionization rate (photoionization plus collisional ionization). This is presented in appendix A of Rahmati et al. (2013a). Here, we do not repeat their calculations but instead refer the reader to Rahmati et al. (2013a). Using these fitting functions we calculate $\eta = n_{\rm n}/n_{\rm H}$.

3.2 Calculating the H_2 gas fraction

In this section, we briefly describe the prescriptions of Gnedin & Kravtsov (2011) and Krumholz (2013) to calculate the H_2 gas fraction for individual gas particles, which we apply to EAGLE. Both prescriptions give H_2 fractions that depend on the gas metallicity and the ISRF.

We use the SPH kernel-smoothed metallicities rather than the particle metallicity with the aim of alleviating the consequences of the lack of metal mixing that arises from the fact that metals are fixed to particles (Wiersma et al. 2009b). We assume the ISRF to be proportional to the SFR surface density, $\Sigma_{\rm SFR}$ (see Appendix A for details).

3.2.1 The prescription of Gnedin & Kravtsov (2011) applied to EAGLE

Gnedin & Kravtsov (2011) developed a phenomenological model for H_2 formation and applied it to a set of zoom-in cosmological simulations at high resolution that also included gravity, hydrodynamics, non-equilibrium chemistry combined with equilibrium cooling rates for metals, and a three-dimensional, on the fly, treatment of radiative transfer, using an Adaptive Mesh Refinement (AMR) simulation. The highest spatial resolution achieved in these simulations is 260 comoving parsecs (cpc), and the gas mass of individual resolution elements ranges from 10^3 to 10^6 $M_\odot$. From the outcome of the simulations, the authors parametrized in analytic formulae the fraction of H_2-to-total neutral gas, $f_{\rm H_2}$, as dependent on the dust-to-gas ratio and the ISRF. We assume the dust-to-gas mass ratio scales with the local metallicity, and the radiation field scales with the local surface density of star formation, which we estimate from the properties of gas particles assuming local hydrostatic equilibrium (Schaye 2001; Schaye & Dalla Vecchia 2008). The equations used to implement this prescription into EAGLE are given in Appendix A. We refer to this prescription as 'GK11'. Gnedin & Draine (2014, hereafter GD14) presented an updated version of the GK11 model taking into account the line blending in the Lyman–Werner bands. We find that this model gives results that are different from the GK11 prescription by ≈50 per cent, in a way that the resulting H_2 masses after applying the GD14 prescription are lower than when applying the GK11 prescription. This difference decreases with redshift, and even reverses at $z \gtrsim 3.5$, with the H_2 masses in the GD14 model being higher than in the GK11 model. The differences between the GD14 and GK11 prescriptions are of a similar magnitude as the differences between the GK11 and Krumholz (2013) prescriptions (Section 3.2.4). However, applying the GD14 or GK11 prescriptions leads to the same conclusions we present in this work, and therefore we will hereafter only use the GK11 prescription. In Appendix A1, we show comparisons between the GK11 and GD14 prescriptions.

In order to visualize how the H_2 column density is distributed with respect to the stars, we show in Fig. 1 five randomly chosen galaxies embedded in haloes of different masses. The optical images were created using three monochromatic radiative transfer simulations with the code SKIRT (Baes et al. 2011) at the effective wavelengths of the Sloan Digital Sky Survey (SDSS) u, g and r filters. Dust extinction was implemented using the metal distribution directly from the simulations, and assuming that 30 per cent of the metal mass is locked up in dust grains. Only material within a spherical aperture of radius of 30 pkpc was included in the radiative transfer calculation. The method will be presented in detail in Trayford et al. (in preparation).

The variety of morphologies in the EAGLE simulations is well captured in Fig. 1 (see also Fig. 2 in S15), where we see early-type galaxies (top row), S0-like galaxies (second row) and galaxy mergers (fourth row). In the case of the early-type galaxy example in the top row of Fig. 1, one can see that the H_2 is relatively concentrated compared to the stars, and is also connected to a dust lane. This is in qualitative agreement with spatially resolved observations of early-type galaxies (Young et al. 2011), in which H_2 is typically concentrated in the centre and in a relatively thin disc.

3.2.2 The prescription of Krumholz (2013) applied to EAGLE

Krumholz (2013) developed a theoretical model for the transition from H I to H_2, that depends on the total column density of neutral hydrogen, the gas metallicity and the ISRF. The fraction $f_{\rm H_2}$ is determined by the balance between photodissociation of H_2 molecules by the ISRF and the formation of molecules on the surfaces of dust grains.

A key property in the Krumholz (2013) model is the density of the cold neutral medium (CNM). At densities $n_{\rm H} \gtrsim 0.5\,{\rm cm}^{-3}$, the transition from H I to H_2 is mainly determined by the minimum density that the CNM must have to ensure pressure balance with the warm neutral medium (WNM, which is H I dominated). The

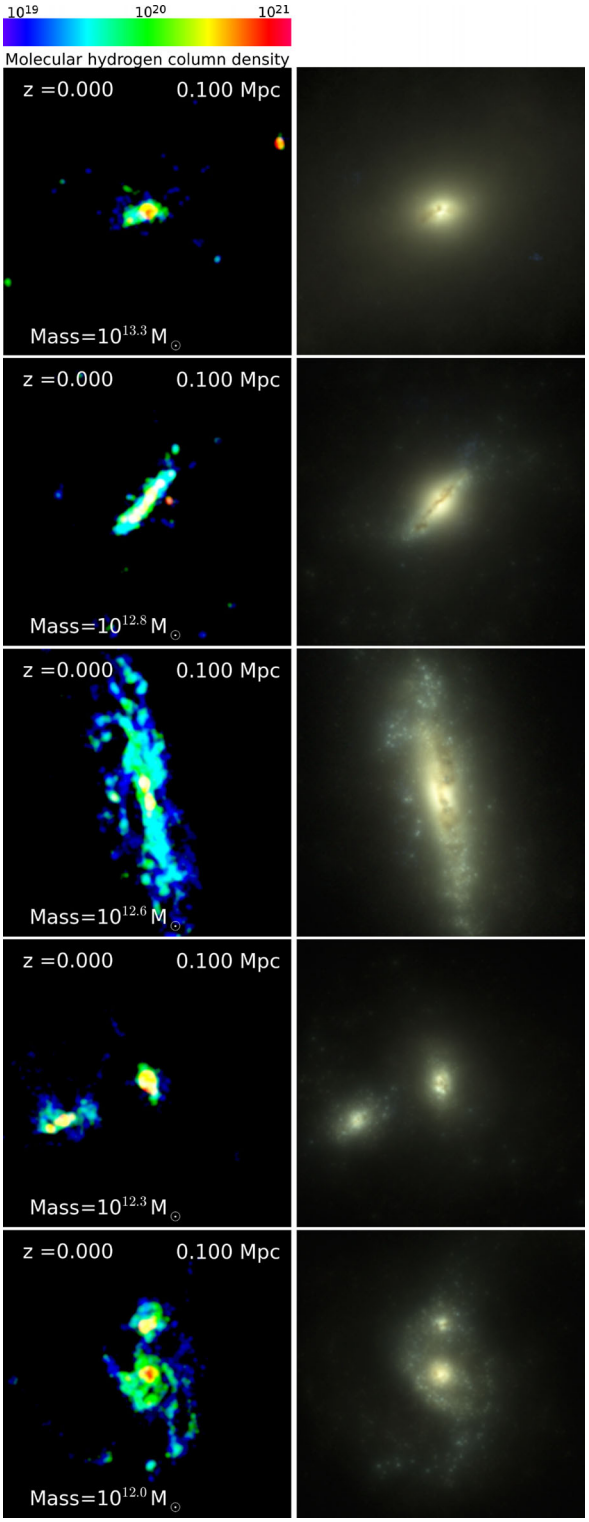

Figure 1. Visualization of five galaxies at $z = 0$ in Ref-L025N0376, which have been randomly chosen (and oriented) from different ranges of subhalo masses (indicated at the bottom of the left-hand panels). Maps are coloured by H_2 column density (calculated using the GK11 prescription; left-hand panels), and the right-hand panels show the stellar light based on monochromatic *u*-, *g*- and *r*-band SDSS filter means and accounting for dust extinction (see the text for details). Colours in the H_2 column density maps are as in the colour bars at the top of the top, where column densities are in units of cm^{-2}. Particles are smoothed by 1ckpc in the N_{H_2} maps. Map sizes are 0.1×0.1 pMpc^2.

assumption is that the CNM is supported by turbulence, while the WNM is thermally supported (see also Wolfire et al. 2003). At $n_H \lesssim 0.5\,\mathrm{cm}^{-3}$, the transition from H I to H_2 is mainly determined by the hydrostatic pressure, which has three components: the self-gravity of the WNM ($\propto \Sigma_{H\,I}^2$), the gravity between the CNM and WNM ($\propto \Sigma_{H\,I}\Sigma_{H_2}$), and the gravity between the WNM and the stellar plus dark matter component ($\propto \Sigma_{H\,I}\Sigma_{sd}$, where Σ_{sd} is the surface density of stars plus dark matter). Note that the exact value of n_H at which the transition between these two regimes takes place is a strong function of gas metallicity. The equations used to implement this prescription into EAGLE are given in Appendix A. We will refer to the Krumholz (2013, hereafter K13) prescription as 'K13'.

3.2.3 *Resolution limit on the H_2 galaxy masses*

We define the resolution limit of the intermediate-resolution simulations in terms of the subhalo masses rather than H_2 masses and present the resolution analysis in Appendix B. We find that haloes with $M_{Tot} > 10^{10}\,M_\odot$, where M_{Tot} is the dark plus baryonic subhalo mass, are converged in their H_2 abundances in the intermediate-resolution simulations.

3.2.4 *Differences between the GK11 and K13 prescriptions*

The lower two panels of Fig. 2 show the predicted ratios of the total H_2 mass to total neutral hydrogen gas mass (atomic plus molecular) as a function of stellar mass when the GK11 (top panel) and K13 (bottom panel) prescriptions are applied. There is a weak positive correlation between the H_2 to total neutral hydrogen mass ratio and stellar mass in both the Ref-L100N1504 and Recal-L025N752 simulations when the K13 prescription applied, but such a trend is not seen in the case of the GK11 prescription. The width of the distributions (as shown by the 16th and 84th percentiles) seen in both the Ref-L100N1504 and Recal-L025N752 simulations when the GK11 prescription is applied is much larger at $M_{stellar} > 10^{10}\,M_\odot$ than when the K13 prescription is applied. Most of the dispersion seen in Fig. 2 is due to gas metallicity. For example, for the simulation Ref-L100N1504 with the GK11 prescription applied and at $M_{stellar} \approx 10^8\,M_\odot$, the metallicity increases from $0.1\,Z_\odot$ at $M_{H_2}/M_{neutral} \approx 0.05$ to $1\,Z_\odot$ at $M_{H_2}/M_{neutral} \approx 0.8$ (see also Fig. A4 for the effect of metallicity on f_{H_2}). Observations favour a positive correlation between $M_{H_2}/M_{neutral}$ and stellar mass (Leroy et al. 2008), similar to the predictions of the Recal-L025N752 simulation. The very weak dependency predicted by the Ref-L100N1504 simulation contradicts observational results. We find that this is due to the gas metallicities in the Ref-L100N1504 simulation being higher than in the Recal-L025N752 simulation and observations (Schaye et al. 2015). We discuss this in Section 4.1.1.

The top panel of Fig. 2 shows the ratio between the H_2 that is locked up in star-forming particles to the total H_2, which also includes the contribution from non-star-forming particles. The GK11 prescription applied to EAGLE results in a relative contribution to the total H_2 mass from the H_2 locked up in non-star-forming particles that is on average 20–40 per cent and does not show a strong dependence on stellar mass. However, when the K13 prescription is applied, this contribution decreases to 5–20 per cent (see the discussion in Appendix A for the contribution to the total H_2 mass of gas particles that are on the equation of state). This difference is significant, given that the errors on the median assessed via bootstrapping are very small ($\lesssim$0.05 dex). The source of the difference between the GK11 and K13 prescriptions, is that the K13 prescription leads

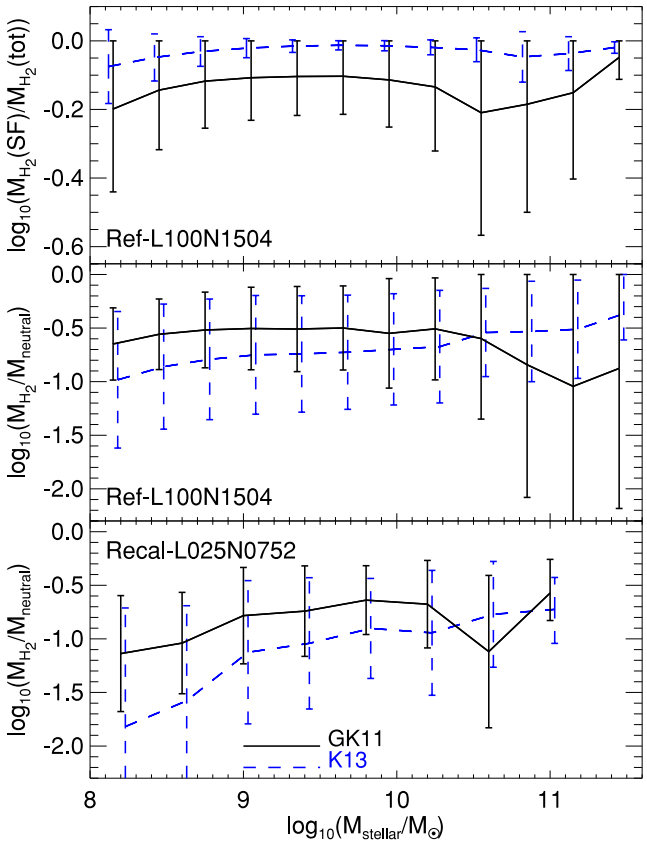

Figure 2. Top panel: the ratio of the H_2 mass that is locked in star-forming particles to total H_2 mass, as a function of stellar mass at $z = 0$ in the simulation Ref-L100N1504 after applying the GK11 (solid line) and K13 (dashed line) prescriptions. Lines and error bars indicate the median and 16th and 84th percentiles. Middle panel: the ratio of the H_2 mass to total neutral hydrogen mass, as a function of stellar mass at $z = 0$ in the simulation Ref-L100N1504 after applying the GK11 (solid line) and K13 (dashed line) prescriptions. Lines and error bars are as in the top panel. Bottom panel: as in the middle panel but for the simulation Recal-L025N0752.

to $f_{H_2} \equiv 0$ at gas surface densities $\Sigma_H \lesssim 1\,M_\odot\,pc^{-2}$ in ISM conditions that are MW-like, while leading to $f_{H_2} \equiv 0$ at higher Σ_H when the ISRF increases. On the contrary, the GK11 prescription leads to small, but non-zero H2 fractions at these low gas surface densities.

It has been observed in nearby galaxies that there is a component of the molecular gas that is not associated with star formation and that is characterized by a high-velocity dispersion consistent with the H I disc (Caldú-Primo et al. 2013; Pety et al. 2013). This component contributes ≈30–50 per cent of the total H_2 mass inferred for those galaxies. This observed diffuse component could be associated with the H_2 locked up in non-star-forming particles in EAGLE, as their contribution to the total H_2 in galaxies (at least when the GK11 prescription is applied) is of a similar magnitude as the observed contribution from the diffuse H_2 to the total inferred H_2 mass. This suggests that EAGLE can shed some light on to the nature of such a component of the ISM of nearby galaxies. We investigate this in more detail in a separate paper (Lagos et al. in preparation). Additional differences between the GK11 and K13 prescriptions, as well as comparisons with previous work can be found in Appendix A.

Fig. 2 shows that the simulations Ref-L100N1504 and Recal-L025N0752 display significant differences both for the GK11 and K13 prescriptions, particularly in galaxies with stellar masses $M_{stellar} < 10^{10}\,M_\odot$. This is mainly driven by the gas metallicity in the Recal-L025N0752 simulation being lower on average than in the Ref-L100N1504 simulation (see fig. 13 in S15 for the mass–metallicity relations (MZR) predicted by the simulations compared here). For example, at a stellar mass of $10^9\,M_\odot$, Recal-L025N0752 simulation predicts a gas metallicity 0.4 dex lower than the metallicity of galaxies of the same mass in the Ref-L100N1504 simulation, on average. However, for $M_{stellar} > 10^{10}\,M_\odot$ differences are small. In Section 4.1.1, we discuss the effect the gas metallicities have on the H_2 masses of galaxies in EAGLE.

4 ABUNDANCE AND SCALING RELATIONS OF H_2 IN THE LOCAL UNIVERSE

In this section, we compare the results of EAGLE with observations of H_2 in the local Universe. We focus on the H_2 mass function in Section 4.1 and on H_2 scaling relations in Section 4.2.

Note that in previous papers using EAGLE, simulation results and observations are compared at $z = 0.1$ in order to best correspond to the median redshift of optical surveys. However, here we compare the EAGLE predictions with observations at $z = 0$ because observations of CO, which here we convert to H_2, in the local Universe correspond to galaxies at $z < 0.1$.

4.1 The H_2 mass function

The curves in Fig. 3 show the $z = 0$ H_2 mass function calculated using either the GK11 or K13 prescriptions, and the mass function of the hydrogen component of the star-forming gas, for the simulation Ref-L100N1504. These are compared to observations at $z \approx 0$, which are shown as symbols with 1σ error bars. We can see that the star-forming gas mass is a poor proxy for the H_2 mass. The median ratio between H_2 mass and the hydrogen component of star-forming gas is ≈0.7, but fluctuations around this value are large, with the 16th and 84th percentiles being 0.33 and 1.25, respectively. We qualitatively discussed the large systematics introduced by the assumption that the star-forming gas is a good proxy for H_2 mass in Section 1 in the context of previous work (e.g. Genel et al. 2014), and here we have quantified them.

As we discussed in Section 1, H_2 is not directly observed, but constraints on the H_2 mass function at $z = 0$ have been provided by surveys of CO(1–0) in local galaxies. The CO(1–0)-to-H_2 conversion factor is the largest uncertainty in the observations. The conversion factor, X, is defined as

$$\frac{N_{H_2}}{cm^{-2}} = X \times 10^{-20} \left(\frac{I_{CO(1-0)}}{K\,km\,s^{-1}} \right), \quad (5)$$

where N_{H_2} is the column density of H_2 and I_{CO} is the integrated CO(1–0) line intensity per unit surface area. Systematic variations of X are both theoretically predicted and observationally inferred. For example, variations are expected as a function of gas metallicity and the ISRF, such that X increases in low metallicity and/or intense ISRF environments (e.g. Bell, Viti & Williams 2007; Meijerink, Spaans & Israel 2007; Bayet et al. 2009; Shetty et al. 2011; Narayanan et al. 2012; Feldmann et al., 2012; Bisbas, Papadopoulos & Viti 2015). In addition to this, different values of X have been measured using different methods. For example, in the MW, variations in the method for measuring X lead to values that differ by a factor of ≈3 (see Bolatto et al. 2013, for a recent review). This systematic error clearly dominates the uncertainty in the H_2 mass over the uncertainties in the measurement of the CO emission line. We keep these

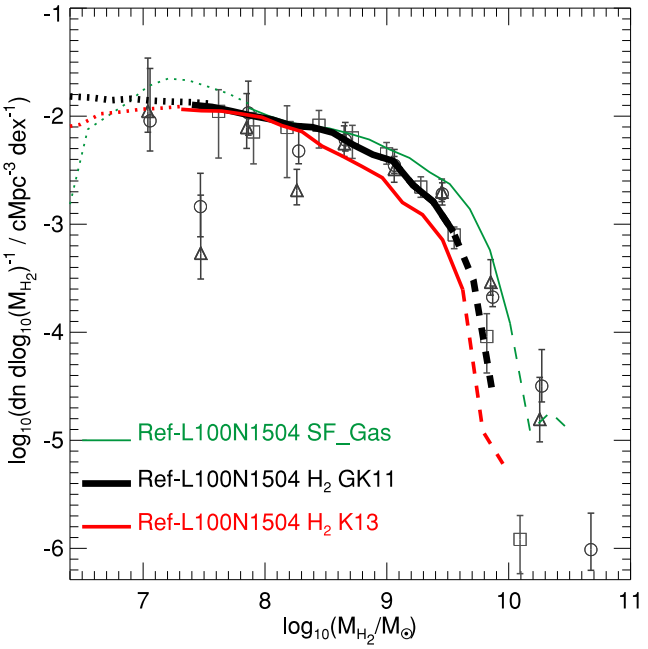

Figure 3. The mass functions of H_2 (thick lines) and of the hydrogen component of the star-forming gas (thin line) in the simulation Ref-L100N1504 at $z = 0$. For H_2, we show the mass functions after applying the GK11 (black line) and K13 (red line) prescriptions. The part of the mass functions that is affected by low number statistics (less than 10 objects per dex bin) is shown as dashed lines, while the part that is affected by resolution is shown as dotted lines (see Appendix B for details). Observational estimates correspond to the Keres et al. (2003) *B*-band (triangles) and 60 μm (circles) selected samples of galaxies, adopting the MW like H_2-to-CO(1–0) conversion factor X = 2 as universal (see equation 5 for the definition of *X*). We also show the observational estimates of Obreschkow & Rawlings (2009) in square symbols, in which a luminosity-dependent *X* was applied to the observations of Keres et al. (2003). The star-forming gas mass function differs significantly from the H_2 mass functions obtained using either the GK11 or the K13 prescriptions at $M_{H_2} > 5 \times 10^8\,M_\odot$, and can therefore not be used as a reliable tracer of the H_2 mass function. The H_2 mass function predicted by EAGLE is in good agreement with the observations.

systematics in mind when comparing the EAGLE predictions with the observations.

For the entire range of observed H_2 masses, $10^7\,M_\odot \lesssim M_{H_2} \lesssim 10^{10}\,M_\odot$, the H_2 mass function of EAGLE, calculated either using the GK11 or the K13 prescriptions, is in good agreement with the observations. We see the agreement is the best with the H_2 mass function derived by Obreschkow & Rawlings (2009) from the CO(1–0) surveys of Keres et al. (2003) using a luminosity-dependent H_2-to-CO(1–0) conversion factor. Compared to the H_2 mass functions derived using a fixed $X = 2$, we find some tension at $M_{H_2} \gtrsim 5 \times 10^9\,M_\odot$. However, the differences are well within the systematic uncertainties of the observational measurements, which are well illustrated by the difference between the observational inferences of Obreschkow & Rawlings (2009) and Keres et al. (2003).

4.1.1 Variations due to gas metallicity

S15 compared the MZR predicted by EAGLE with observations and showed that the intermediate-resolution simulation Ref-L100N1504 exhibits ISM metallicities that are higher than the observations by up to 0.3 dex at a stellar mass of $\lesssim 10^9\,M_\odot$. The higher resolution simulation, Recal-L025N0752, on the other hand, displays ISM metallicities that are lower than the lower resolution simulations at fixed stellar mass, and are in good agreement with the observations.

The gas metallicity is important for two reasons. First, the GK11 and K13 prescriptions have an explicit dependence on gas metallicity, as dust is the catalyst for H_2 formation, and the abundance of dust is assumed to be proportional to the gas metallicity. Secondly, the formation of CO, and therefore the reliability of CO as a tracer of H_2, depends on the gas metallicity (e.g. Bolatto et al. 2013). Observations suggest a strong dependence of *X* on gas metallicity. For example, Boselli et al. (2002) reported the following relation between *X* and *Z*,

$$\log_{10}(X) = 0.5^{+0.2}_{-0.2} - 1.02^{+0.05}_{-0.05}\log_{10}(Z/Z_\odot) \quad (6)$$

In order to quantify the effect that the predicted higher metallicities have on the H_2 content of galaxies in EAGLE, we first repeat the calculation presented in Section 3.2 but forcing galaxies to follow the observed MZR. In practice, we compare the neutral gas mass-weighted metallicity of individual EAGLE galaxies with the observed median gas metallicity of Zahid et al. (2014) at the same stellar mass and estimate how much higher the predicted metallicity is compared to the median in the observations. If the difference is greater than 50 per cent of the observed median, we rescale the metallicities of all gas particles by the factor needed to bring the galaxy metallicity into agreement with the observations, and recalculate the H_2 fraction and the 30 pkpc aperture H_2 mass. The 50 per cent factor is included to allow for a dispersion around the median of the same magnitude as the 1σ distributions reported in the observations. We convert the metallicities of Zahid et al. (2014) to our adopted metal abundances: $12 + \log_{10}(\mathrm{O/H})_\odot = 8.69$ and $Z_\odot = 0.0127$. For stellar masses $M_{\rm stellar} < 10^9\,M_\odot$, we extrapolate the MZR of Zahid et al. (2014) to get an 'observed' metallicity at that stellar mass. The H_2 mass functions using the GK11 prescription before and after the rescaling of gas metallicities are shown for the simulations Ref-L100N1504 and Recal-L025N0752 in the top panel of Fig. 4.

The differences in the H_2 mass functions before and after the metallicity rescaling are mild. The metallicity rescaling has no visible effect on the H_2 mass function of Recal-L025N0752, while for Ref-L100N1504, the number density of galaxies with $M_{H_2} \lesssim 5 \times 10^8\,M_\odot$ decreases slightly after rescaling. When integrating over all H_2 masses, the average density of H_2 in Ref-L100N1504 decreases by 5 per cent after the gas metallicity rescaling. This difference is well within the error bars of the observations (see Section 5.4). However, larger differences are obtained for the relation between the $M_{H_2}/M_{\rm neutral}$ ratio and stellar mass (Fig. 2), in which the gas metallicity rescaling leads to $M_{H_2}/M_{\rm neutral}$ ratios lower by up to ≈0.5 dex at $10^8\,M_\odot \lesssim M_{\rm stellar} \lesssim 5 \times 10^8\,M_\odot$, for both prescriptions. A steeper relation is then obtained, which removes the tension with the observations described in Section 3.2.4.

As a second experiment, we use equation (6) to convert the predicted H_2 masses in EAGLE to CO(1–0) luminosities and construct the CO(1–0) luminosity function, in order to show the effect of a metallicity-dependent *X*. This is shown in the bottom panel of Fig. 4, along with the observations of CO(1–0) of Keres et al. (2003). The level of agreement between the Ref-L100N1504 and Recal-L025N0752 simulations and the observations is similarly good to that for the H_2 mass functions, and the effect of rescaling the gas metallicity is also similar. This shows that the main effect of gas metallicities is in the H_2 formation rather than its later conversion to CO(1–0). Note that the CO(1–0) luminosity functions predicted by

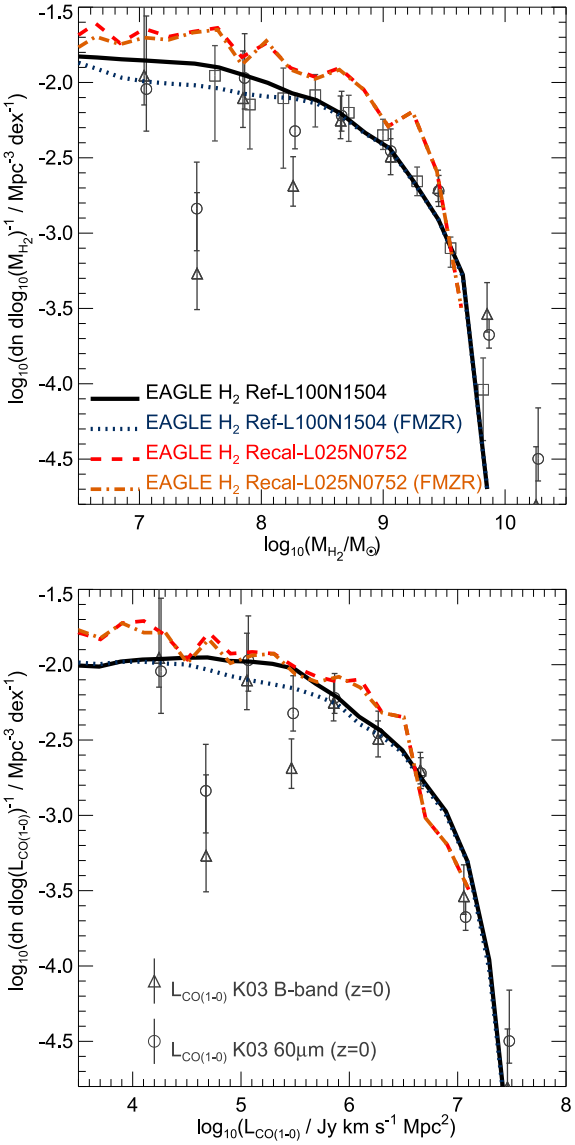

Figure 4. Top panel: the H_2 mass function at $z = 0$ for the simulations Ref-L100N1504 and Recal-L025N0752 using the GK11 prescription, before and after the gas metallicities of galaxies in EAGLE are rescaled to satisfy the observed MZR of Zahid et al. (2014). We refer to the simulations after rescaling as 'FMZR'. As shown in Fig. 3, the effects of resolution in the Ref-L100N1504 start to appear at H_2 masses $\lesssim 5 \times 10^7\,M_\odot$. The effect of sampling, on the other hand, appears at number densities of -2.5 in the Recal-L025N0752 and -4.3 in the Ref-L100N1504 (in the units of the y-axis). Observations are as in Fig. B1. The rescaling of gas metallicities has a significant effect at $M_{H_2} < 10^9\,M_{H_2}$. Bottom panel: CO(1–0) luminosity function for the same models of the top panel, and applying the conversion factor of equation (6) to convert the predicted H_2 masses into CO(1–0) luminosities. Observations correspond to the B band and 60 microns selected samples of Keres et al. (2003), as labelled.

the simulations Ref-L100N1504 and Recal-L025N0752 are closer to each other than the predicted H_2 mass functions. This is because the simulation Recal-L025N0752 has, on average, lower gas metallicities than the Ref-L100N1504 simulation, which in practice leads to higher X values.

The effects of the higher gas metallicities on the H_2 masses of galaxies is mild and therefore we continue to use the predicted gas metallicities (rather than the rescaled ones) for the rest of the paper.

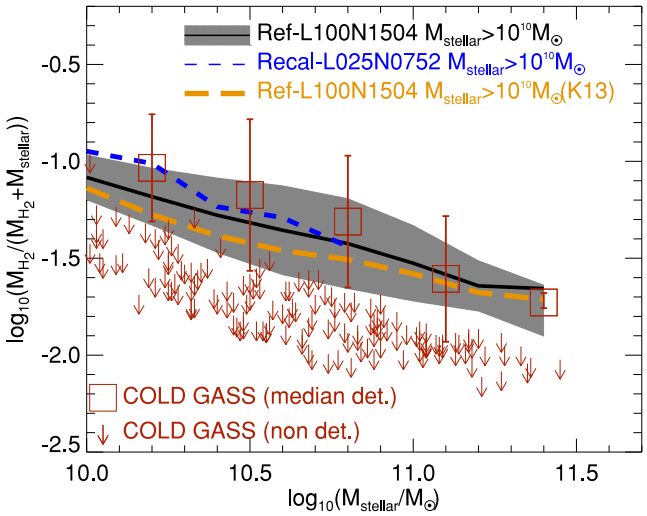

Figure 5. Molecular hydrogen fraction, $M_{H_2}/(M_{H_2} + M_{stellar})$, as a function of stellar mass for the simulations Ref-L100N1504 (solid line and filled region) and Recal-L025N0752 (dashed line with error bars) for galaxies with $M_{stellar} > 10^{10}\,M_\odot$ and with H_2 masses above the sensitivity limit of COLD GASS at $z = 0$. Here, we use the GK11 prescription to calculate H_2 masses. We also show the simulation Ref-L100N1504 when the K13 prescription is applied (long-dashed line) for the same redshift and stellar mass selection. Lines show the medians of the simulated galaxies for which the H_2 mass exceeds the COLD GASS detection limit, while the filled region shows the 16th and 84th percentiles for the Ref-L100N1504 simulation when the GK11 prescription is applied. The dispersions in the Ref-L100N1504 simulation with the K13 prescription and in the Recal-L025N0752 simulation are of similar size. Squares with error bars show the median and 16th and 84th percentiles of galaxies in the COLD GASS survey that have the CO(1–0) emission line detected, while arrows show upper limits for the non-detections (Saintonge et al. 2011). The predictions from EAGLE are in very good agreement with the data.

4.2 H_2 Scaling relations

Measurements of the H_2 and H I mass content, as well as other galaxy properties, are available for relatively large samples of local galaxies (running into the hundreds), enabling the characterization of scaling relations between the cold gas and the stellar mass content (e.g. Bothwell, Kennicutt & Lee 2009; Catinella et al. 2010; Saintonge et al. 2011; Boselli et al. 2014). Recently, H_2 and H I have been studied in homogeneous samples of relatively massive galaxies by Saintonge et al. (2011) and Catinella et al. (2010), respectively, with the aim of measuring the fundamental relations between the stellar content of galaxies and their cold gas mass. The survey of Saintonge et al. (2011), the CO Legacy Database for the *GALEX* Arecibo SDSS Survey (COLD GASS) is the first stellar mass-limited CO survey. The strategy used was to select all galaxies with $M_{stellar} > 10^{10}\,M_\odot$ at $z < 0.05$ from the SDSS and follow-up a subsample of those at millimetre (mm) wavelengths to detect CO(1–0). Saintonge et al. (2011) integrated sufficiently long to enable H_2 gas fractions, $M_{H_2}/M_{stellar} > 0.015$, at stellar masses $M_{stellar} > 10^{10.6}\,M_\odot$, to be detected, or H_2 masses $M_{H_2} > 10^{8.8}\,M_\odot$ in galaxies with stellar masses, $10^{10}\,M_\odot < M_{stellar} < 10^{10.6}\,M_\odot$. The simple and well-defined selection function of COLD GASS makes this survey an instructive testbed for EAGLE.

Fig. 5 shows the H_2 fraction, $M_{H_2}/(M_{H_2} + M_{stellar})$, as a function of stellar mass for galaxies with $M_{stellar} > 10^{10}\,M_\odot$ and H_2 gas fractions above the sensitivity limit defined by Saintonge et al. (2011) for COLD GASS at $z = 0$ in EAGLE. Note that both

observations and simulation results account for molecular hydrogen only (without helium correction). We show the median and 16th and 84th percentiles for galaxies in COLD GASS, for which the CO(1–0) emission line is detected as squares with error bars, while we show the upper limits for the non-detections as arrows. Note that uncertainties in the measurement of the integrated flux in the CO(1–0) emission line are small (approximately 1 per cent). Here, the systematic errors are dominant, and they are mainly related to how the CO flux is converted into the H_2 mass, as we discussed in Section 4.1.

From Fig. 5, we see that EAGLE predicts that the ratio $M_{H_2}/(M_{H_2} + M_{stellar})$ decreases with $M_{stellar}$. The amplitude of this relation is slightly higher, by $\lesssim$0.1 dex, when the GK11 prescription is applied than when the K13 prescription is applied. The high-resolution simulation Recal-L025N0752 predicts a trend that is similar to that in Ref-L100N1504, although with systematically higher $M_{H_2}/(M_{H_2} + M_{stellar})$ (except for the largest galaxies, for which we have only a small number in the Recal-L025N0752 simulation). This is due to the slightly higher SFRs and H_2 masses predicted in the higher resolution simulation at a fixed halo mass. S15 showed that the Ref-L100N1504 simulation predicts SFRs that are lower than the observations by ≈0.1–0.15 dex for galaxies with $M_{stellar} > 10^{10}\,M_{\odot}$, while the Recal-L025N0752 simulation predicts higher SFRs, in better agreement with the observations. Such differences also appear in Fig. 5, where the median in the Ref-L100N1504 simulation is ≈0.1 dex lower than the median H2 fraction in COLD GASS. This offset is smaller than the scatter in the observations. Thus, the agreement seen here is fully consistent with the analysis of the specific SFRs in S15. The differences between the Ref-L100N1504 and Recal-L025N0752 simulations are also similar to those discussed for the H_2-subhalo mass relation in Appendix B.

The predictions from the EAGLE simulations shown in Fig. 5 agree very well with the observations. Galaxies in COLD GASS for which the CO(1–0) emission line is detected lie on the main sequence of star-forming galaxies (in the SFR–stellar mass plane), which is also the case for galaxies in EAGLE that have H_2 masses above the detection limit of COLD GASS. This is a significant outcome, since none of these observations were used to constrain the calibrated parameters governing the efficiency of star formation and feedback, nor those governing the partitioning of gas into ionized, atomic and molecular components in EAGLE. This is the first time that a hydrodynamic simulation performs so well when confronted by observational inferences of H_2.

In Fig. 6, we show the SFR as a function of the H_2 mass for EAGLE galaxies with $M_{stellar} > 10^{10}\,M_{\odot}$ and H_2 gas fractions above the sensitivity limit of COLD GASS at $z = 0$ and for the COLD GASS observations. In EAGLE there is a tight correlation between these two quantities, because star formation and most of the H_2 coexist in the same gas phase (see Fig. 2). This is also the reason why the relation between SFR and M_{H_2} is the tightest of all the scaling relations discussed in this section. Simulation Recal-L025N0752 displays a lower normalization than Ref-L100N1504 (with the same prescription applied) at H_2 masses $M_{H_2} \gtrsim 10^{9.5}\,M_{\odot}$ by a factor of ≈1.5, consistent with the typically greater H_2 mass associated with galaxies at a fixed stellar mass in the Recal-L025N0752 simulation, as discussed in Appendix B. Note that the tight relation between the SFR and the H_2 mass is not surprising, because the star formation law adopted in EAGLE relates the SFR of particles with their gas mass, above a given density threshold. However, the component of the H_2 mass that is non-star-forming gas, in addition to the dependence on gas metallicity, result in differences between the different simulations with the K13 or GK11 prescriptions to look different in the SFR–M_{H_2} relation (with differences being of ≈0.2 dex), even though the star formation law adopted is the same in every case.

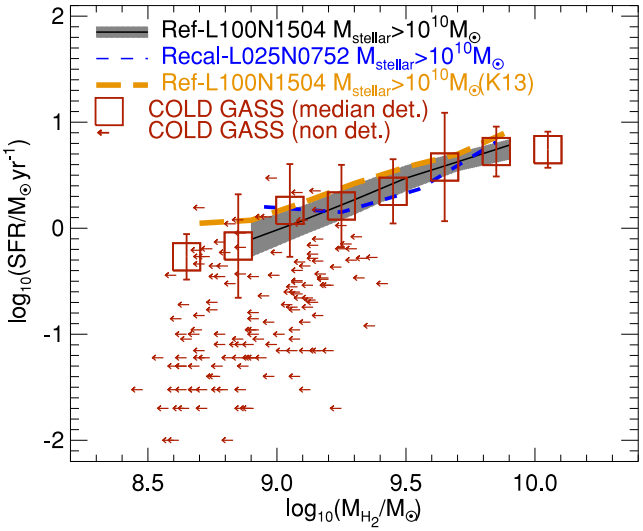

Figure 6. The SFR as a function of the molecular hydrogen mass at $z = 0$ for galaxies with $M_{stellar} > 10^{10}\,M_{\odot}$ and with H_2 gas fractions above the sensitivity limit of COLD GASS in the simulations Ref-L100N1504 (solid line and filled region) and Recal-L025N0752 (dashed line) with the GK11 prescription, and the simulation Ref-L100N1504 with the K13 prescription (long dashed line), where lines show the medians. Observations are as in Fig. 5. The simulations and observations are in good agreement.

The relation between SFR and the H_2 mass predicted by EAGLE agrees with the reported detections in COLD GASS. Galaxies for which the CO(1–0) emission line is not detected have significantly lower SFRs. EAGLE predicts that the H_2 masses of these galaxies should be an order of magnitude (or more) below the detection limit of COLD GASS. This is because EAGLE predicts that the tight relation between SFR and H_2 extends down to very low values in both quantities. This can be seen in Fig. 7, where we lower the threshold in $M_{H_2}/M_{stellar}$ to $M_{H_2}/M_{stellar} > 0$ (while maintaining the threshold in stellar mass).

In Fig. 7, we also show the observations of the ATLAS^{3D} survey (Cappellari et al. 2011), which covers a similar range in stellar mass as COLD GASS but only includes early-type galaxies (ellipticals and lenticulars). The complementary aspect of ATLAS^{3D} relative to COLD GASS is that it detects much fainter CO(1–0) fluxes (see Young et al. 2011, 2014). SFRs in ATLAS^{3D} come from a combination of UV and mid-IR photometry (see Davis et al. 2014 for details). We find that ≈17 per cent of the early-type galaxies in ATLAS^{3D} lie $>2\sigma$ from the median relation in EAGLE, pointing to a possible conflict between EAGLE and the observations. However, this cannot be confirmed here because we would need to make a thorough comparison with ATLAS^{3D} by morphologically selecting galaxies to be early-type in EAGLE. This is beyond the scope of this paper. If the discrepancy was real, it could be due to the star formation law used in EAGLE, which is constructed from the observed Kennicutt (1998) relation between the surface densities of SFR and gas. Galaxies in ATLAS^{3D} as well as the upper limits from COLD GASS (shown in Fig. 7) are offset from the Kennicutt (1998) relation in the sense that their efficiency of star formation is inferred to be lower (see also conclusions of Davis et al. 2014).

Fig. 8 shows the H_2 depletion time-scale, $\tau_{H_2} \equiv M_{H_2}/\mathrm{SFR}$, as a function of the stellar surface density at $z = 0$ for the simulation Ref-L100N1504 with the GK11 prescription and the median and

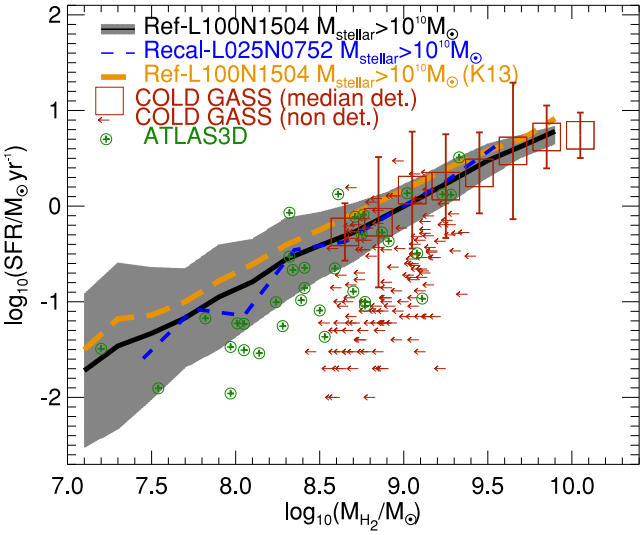

Figure 7. As Fig. 6 but for galaxies with $M_{\rm H_2}/M_{\rm stellar} > 0$ and $M_{\rm stellar} > 10^{10}\,{\rm M}_\odot$. Here, the filled region corresponds to the 2.5th and 97.5th percentiles for the simulation Ref-L100N1504 with the GK11 prescription applied. To the observations shown in Fig. 6, we add those of the ATLAS$^{\rm 3D}$ survey (Young et al. 2011; Davis et al. 2014), as labelled, which sample a similar range of stellar mass as COLD GASS, but go much deeper in the CO(1–0) emission line. The main difference is that ATLAS$^{\rm 3D}$ only includes early-type galaxies, while COLD GASS does not distinguish between different morphological types. Thus, ATLAS$^{\rm 3D}$ cannot be directly compared to EAGLE (given the selection) but one can conclude that about 17 per cent of the early-type galaxies in ATLAS$^{\rm 3D}$ are $>2\sigma$ from the median relation predicted by EAGLE.

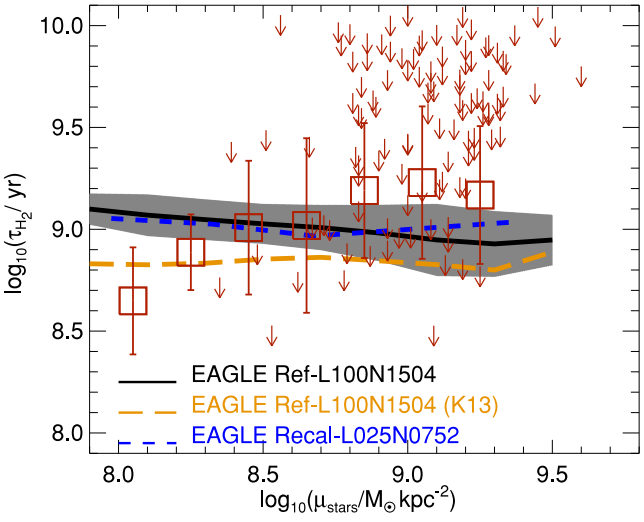

Figure 8. The molecular hydrogen depletion time-scale, $\tau_{\rm H_2} \equiv M_{\rm H_2}/{\rm SFR}$, as a function of the stellar mass surface density, $\mu_\star = M_{\rm stellar}/2\pi\, r_{50}^2$ at $z = 0$ (see the text for details on how $\mu_\star$ is calculated in observations and EAGLE), for galaxies in EAGLE with $M_{\rm stellar} > 10^{10}{\rm M}_\odot$ and $\rm H_2$ gas fractions above the sensitivity limit defined for COLD GASS). Simulations and observations shown are as in Fig. 6.

16th and 84th percentiles of the observations of COLD GASS. The $\rm H_2$ depletion time-scale is defined as the time that would be required to exhaust the current $\rm H_2$ mass if the SFR were constant. The stellar surface density is defined in EAGLE and in the observations as $\mu_\star = (M_{\rm stellar}/2)/(\pi r_{50}^2)$, where $M_{\rm stellar}$ and r_{50} in EAGLE are the stellar mass in the subhalo and the projected radius that contains

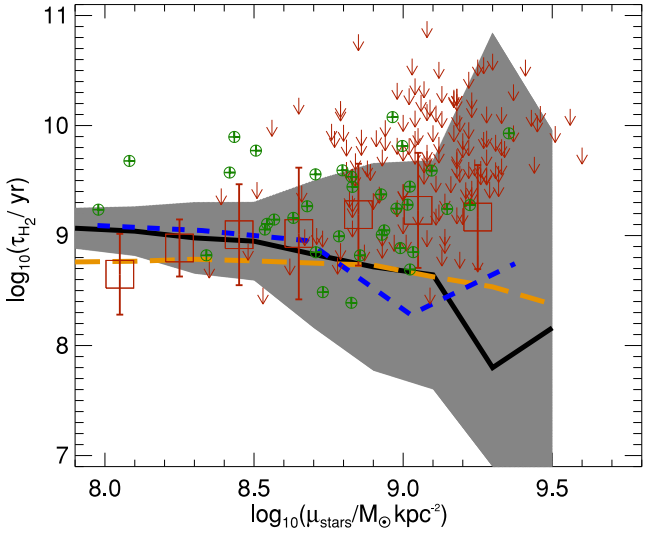

Figure 9. As Fig. 8 but for galaxies with $M_{\rm H_2}/M_{\rm stellar} > 0$. Here, the filled region corresponds to the 2.5th and 97.5th percentiles of the simulation Ref-L100N1504 with the GK11 prescription applied. As in Fig. 7, we add the observations of ATLAS$^{\rm 3D}$, for which $\mu_\star$ is calculated using the r-band half-mass radius. ATLAS$^{\rm 3D}$ cannot be directly compared to EAGLE (given the selection) but one can conclude that about 7 per cent of the early-type galaxies in ATLAS$^{\rm 3D}$ are $>2\sigma$ from the median relation predicted by EAGLE.

half of that stellar mass, respectively. In COLD GASS, the stellar mass is derived from the broad-band photometry and r_{50} is the radius encompassing half of the Petrosian flux in the z band, which is measured in circular apertures (see Saintonge et al. 2012 for details). We show the median and 16th and 84th percentiles of galaxies in COLD GASS for which the CO(1–0) emission line is detected as squares with error bars, while upper limits are shown as arrows.

Even though the EAGLE predictions lie in a similar region to the observations in Fig. 8, there are some interesting discrepancies. It has been suggested from the observations shown in Fig. 8 that there is a tendency for $\tau_{\rm H_2}$ to increase with increasing $\mu_\star$ (Saintonge et al. 2012; Martig et al. 2013; Davis et al. 2014). This trend is apparent if upper limits are considered, but less clear if only detections in COLD GASS are considered. It has been suggested that such a trend can arise if the central stellar surface density increases the surface density threshold for gas instability, therefore lowering the SFR (Saintonge et al. 2012; Martig et al. 2013; Davis et al. 2014), in other words, as dynamically driven star formation suppression. Such a trend is not seen in the reference simulation, Ref-L100N1504, when the GK11 prescription is applied, but instead the predicted relation is consistent with a constant $\tau_{\rm H_2}$.

To gain more insight into the relation between $\tau_{\rm H_2}$ and $\mu_\star$, we again relax the limit imposed on $M_{\rm H_2}/M_{\rm stellar}$ to $M_{\rm H_2}/M_{\rm stellar} > 0$, while maintaining the cut in stellar mass, $M_{\rm stellar} > 10^{10}\,{\rm M}_\odot$. The results are shown in Fig. 9 where we also include the observations of the ATLAS$^{\rm 3D}$ survey, in which $\mu_\star$ is measured from the K band derived stellar masses and the half-light radius in the r band. Since most early-type galaxies in ATLAS$^{\rm 3D}$ do not have significant ongoing star formation, measuring r_{50} in the r-band is similar to measuring it in the infrared (as was done for COLD GASS).

The observations of ATLAS$^{\rm 3D}$ confirm the existence of a population of galaxies with larger values of $\tau_{\rm H_2}$ than the galaxies in COLD GASS with CO(1–0) line detections. The trends predicted by EAGLE are as in Fig. 8, although the dispersion around the

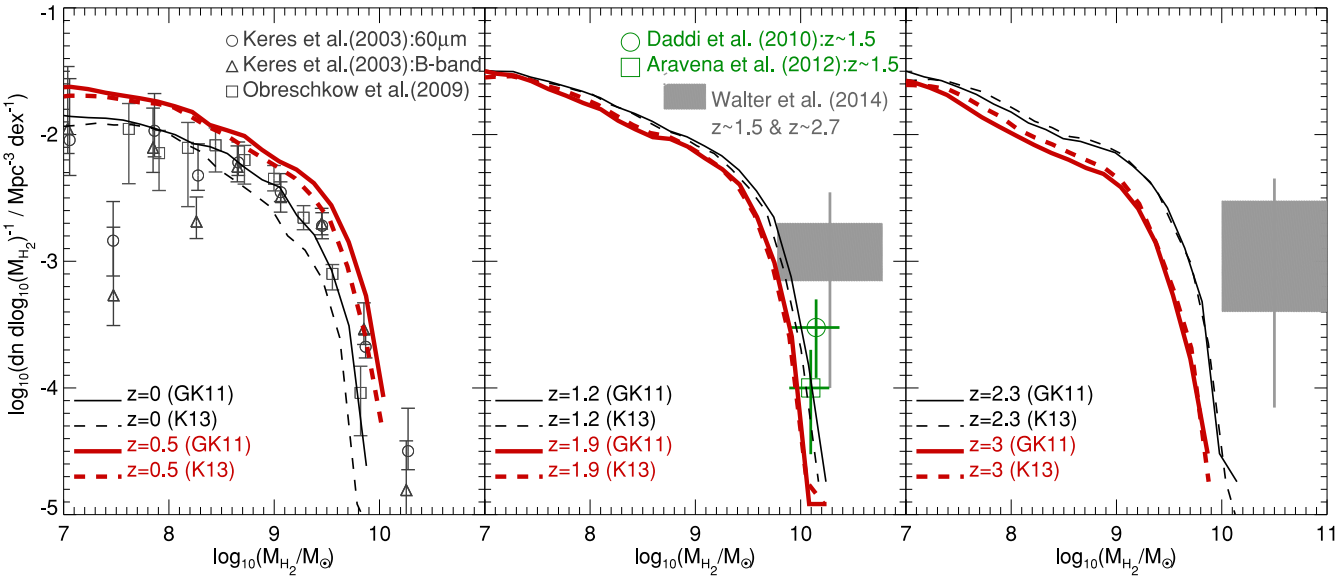

Figure 10. The H_2 mass functions for the reference simulation Ref-L100N1504 using the GK11 (solid lines) and K13 (dashed lines) prescriptions applied, at $z = 0$ and 0.5 in the left-hand panel; $z = 1.2$ and 1.9 in the middle panel; and $z = 2.3$ and 3 in the right-hand panel. The part affected by resolution is approximately $M_{H_2} \lesssim 5 \times 10^7\,M_\odot$ (see Appendix B). Observations shown in the left-hand panel are as in Fig. 3. The solid grey regions in the middle and right-hand panels correspond to the blind CO measurements presented of Walter et al. (2014) and Decarli et al. (2014), at $z \approx 1.5$ and $z \approx 2.7$, respectively. The upper side of the grey regions show the number density obtained if all the high-quality line candidates are considered real, while the bottom side of the boxes show the number density if only secure detections are considered. The error bars in the boxes indicate the Poisson errors on the number densities. In the middle panel, we also show the observational inferences from Daddi et al. (2010) and Aravena et al. (2012), as labelled, which are based on CO emission line follow-up of $z \approx 1.5$ galaxies or fields with spectroscopic or photometric redshifts.

medians becomes much larger at $\mu_\star > 10^{8.5}\,M_\odot kpc^{-2}$. This means that some galaxies in EAGLE that have large values of $\mu_\star$ also have high values for τ_{H_2}, consistent with the observations, but not the majority of them. However, we find that only 7 per cent of early-type galaxies in ATLAS^{3D} are $>2\sigma$ from the median relation predicted by EAGLE, and again, we cannot confirm this is a real issue in the simulation given that we are not selecting galaxies in EAGLE to be early type. Although we have not applied the same definition of $\mu_\star$ used in the observational data, we find that the trends of Fig. 8 are not strongly sensitive to the definition of $\mu_\star$. For example, similar trends and values for $\mu_\star$ and τ_{H_2} are found if we take the half-mass radius measured in apertures of 30 and 20 pkpc.

We conclude that EAGLE captures the main scaling relations between stellar mass, SFR, stellar mass surface density and H_2 mass of galaxies that are on the main sequence of star formation. Some discrepancies are present when we consider galaxies with high values of τ_{H_2}, which in EAGLE are rare but observationally represent a large fraction of the early-type galaxy population in the local Universe. These tests are important since none of these observations were used to fix the free parameters in EAGLE. A study of the radial profiles of these quantities is ongoing and will be presented in a future paper (Lagos et al., in preparation).

5 THE EVOLUTION OF THE H_2 ABUNDANCE OF GALAXIES

In this section, we present the analysis of the evolution of the H_2 mass function (Section 5.1), H_2 scaling relations (Section 5.2), the stellar masses and SFRs of galaxies selected by their H_2 mass (Section 5.3), and the global H_2 density (Section 5.4), over the redshift range $0 \leq z \leq 5$.

5.1 The evolution of the H_2 mass function

At $z = 0$, the H_2 mass function has been measured by Keres et al. (2003) and Obreschkow et al. (2009a). At $z = 1.5$ and 2.7 constraints on the H_2 mass function have been provided from a blind molecular emission line survey by Decarli et al. (2014) and Walter et al. (2014). Although the uncertainties in these measurements are significant (see Fig. 10), these are currently the best available measurements of the H_2 mass function at these redshifts. Other constraints on the H_2 mass functions at $z \approx 1$–2 have been obtained by following-up galaxies at mm wavelengths that have spectroscopic or photometric redshifts. This is the case for Aravena et al. (2012), who performed a Jansky Very Large Array survey towards a candidate cluster at $z \sim 1.5$. Aravena et al. (2012) obtained two detections and were able to place constraints on the number density of bright CO(1–0) galaxies. Another interesting measurement at $z \approx 1.5$ was provided by Daddi et al. (2010), who detected CO(2–1) emission lines for six galaxies from a sample of colour-selected star-forming galaxies.

Fig. 10 shows the H_2 mass function of the simulation Ref-L100N1504 using both the GK11 and K13 prescriptions for several redshift bins ranging from $z = 0$ to 3. The H_2 mass function shows an increase in the number density of galaxies of all masses from $z = 0$ to $z \approx 1.2$–1.5. At $z > 1.5$, the trend reverses and the number density of galaxies starts to decrease with redshift. The driver of this reversal is that at $z \gtrsim 1$ the gas metallicity decreases faster than at $z \lesssim 1$, which hampers H_2 formation. In addition, the median ISRF in galaxies increases monotonically with redshift (at least up to $z = 4.5$), hampering H_2 formation even further. We come back to this point in Section 5.4. Interestingly, the differences at $z = 0$ between the GK11 and K13 prescriptions applied to EAGLE disappear at $z \approx 1$–2, and at $z \approx 2$–3 the K13 prescription gives slightly higher abundances of galaxies with $M_{H_2} < 10^9\,M_\odot$ than the GK11 prescription. This is because the GK11 prescription gives

lower H_2 fractions than the K13 prescription in the case of intense ISRF ($G_0^{'} \gtrsim 100$). The latter values of $G_0^{'}$ are typical at $z \gtrsim 1.5$ (see bottom panel of Fig. 12 and Appendix A for an analysis of the differences between the H_2 masses predicted when applying the GK11 and K13 prescriptions).

We find that EAGLE predictions at $z \approx 1.2$ (middle panel of Fig. 10), with either the GK11 and K13 prescriptions applied, are within the uncertainties of the observations of Daddi et al. (2010), Aravena et al. (2012) and Walter et al. (2014). At $z \approx 2.5$ (right-hand panel of Fig. 10), the predicted abundance of galaxies with $M_{H_2} > 10^{10}\,M_\odot$ is low compared to the constraints of Walter et al. (2014), particularly if we compare with the observational constraints at $z = 3$. However, the redshift range of that data point in the right-hand panel of Fig. 10 is large, $z \approx 2$–3.1, which can partially explain the offset with the predictions. Another important uncertainty is the CO-H_2 conversion factor: $X = 2$ was assumed in Walter et al. (2014) and Decarli et al. (2014). If this conversation factor were on average lower in galaxies with large abundances of H_2, as suggested by the theoretical studies of Lagos et al. (2012), Narayanan et al. (2012) and Popping et al. (2014), the H_2 masses would be overestimated by Walter et al. (2014), improving the consistency with the EAGLE results.

Note that the small number of galaxies with $M_{H_2} \gtrsim 10^{10}\,M_\odot$ is not an effect of the limited box size, as we find from comparing the Ref-L100N1504 and Ref-L050N0752 simulations which produce a very similar number density of galaxies with $M_{H_2} \gtrsim 10^{10}\,M_\odot$ and a very similar value of the cosmic mean H_2 density. (Fewer than 10 objects per dex bin in Fig. 10 corresponds to a number density of $10^{-4.3}$ in the units of the y-axis.)

The EAGLE results in Fig. 10 are slightly different from those given by semi-analytic models. The latter predict the peak of the number density of galaxies with $M_{H_2} \gtrsim 10^{10}\,M_\odot$ to be at $z \approx 2$–2.5 (Obreschkow et al. 2009b; Lagos et al. 2012). A possible explanation for this is that the models that have been preferred in semi-analytic work to calculate the ratio of H_2-to-H I gas mass (for example the empirical model of Blitz & Rosolowsky 2006) do not consider gas metallicity. Future observations can distinguish between these different predictions, for example with telescopes such as the Atacama Large Millimeter Array (ALMA).

5.2 H_2 scaling relations at high redshift

Fig. 11 shows the ratio $M_{H_2}/(M_{H_2} + M_{stellar})$ and the H_2 depletion time-scale, $\tau_{H_2} \equiv M_{H_2}/\mathrm{SFR}$, as a function of redshift, calculated for galaxies on the main sequence of star formation and with stellar masses $M_{stellar} > 5 \times 10^9\,M_\odot$ in the Ref-L100N1504 simulation using the GK11 and K13 prescriptions. We define the main sequence of galaxies as ±0.5 dex around the median $\log_{10}(\mathrm{SFR})$ at a given stellar mass, calculated from the simulation outputs. Observational data are from Saintonge et al. (2011), Geach et al. (2011), Magdis et al. (2012), Saintonge et al. (2013) and Tacconi et al. (2013), and also refer to quantities measured for main-sequence galaxies over a range of stellar masses similar to the one chosen here. The definition of the main sequence of galaxies in the observational data is also based on the median SFR of galaxies of a given stellar mass. However, since observations are flux limited, the possibility exists that the observational data are biased towards higher SFRs.

EAGLE predicts the ratio $M_{H_2}/(M_{H_2} + M_{stellar})$ to increase by 0.5–0.7 dex from $z = 0$ to 1 and to evolve very little from $z \approx 1$ to 4. These trends are also observed. The GK11 prescription applied to the Ref-L100N1504 simulation predicts a slightly lower median value of $M_{H_2}/(M_{H_2} + M_{stellar})$ at $z \gtrsim 2.5$ than is the case for the

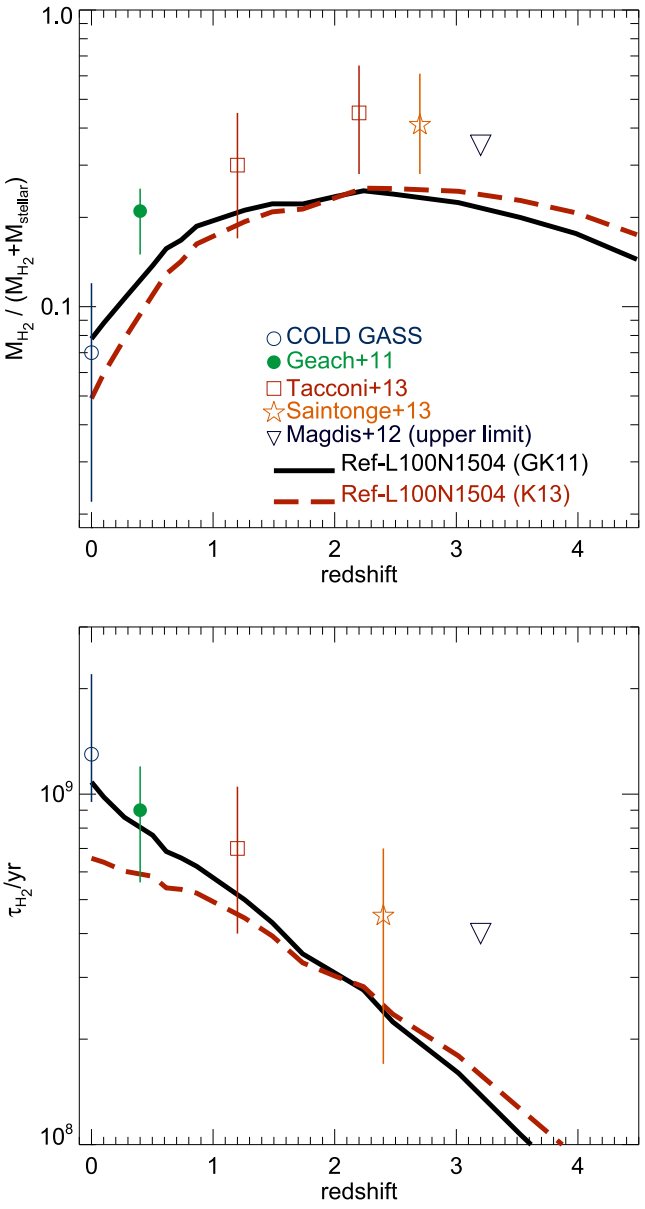

Figure 11. Top panel: the mean H_2 gas fraction, $M_{H_2}/(M_{H_2} + M_{stellar})$ as a function of redshift for the Ref-L100N1504 simulation using the GK11 (solid line) and the K13 (dashed line) prescriptions. The mean H_2 gas fraction was calculated in EAGLE using only main-sequence galaxies with $M_{stellar} > 5 \times 10^9\,M_\odot$. The observational data also correspond to the median H_2 gas fraction for main-sequence galaxies at different redshifts. Data was taken from Saintonge et al. (2011), Geach et al. (2011), Magdis et al. (2012), Saintonge et al. (2013) and Tacconi et al. (2013). Magdis et al. (2012) presented a detection and a non-detection of the CO(3–2) emission line of galaxies at $z \approx 3$, and therefore the combined estimate presented here is only an upper limit. Bottom panel: the H_2 depletion time-scale, M_{H_2}/SFR, as a function of redshift. The observational data are as in the top panel except for the measurement at $z \approx 2.4$ which is a combination of the Tacconi et al. (2013) and the Saintonge et al. (2013) measurements, as was presented by Saintonge et al. (2013).

K13 prescription. The reason is that the GK11 prescription is more sensitive to intense ISRF, $G_0^{'} \gtrsim 100$, than the K13 prescription, and the latter values are typical for $z \gtrsim 2$ (see bottom panel of Fig. 12).

There is a slight discrepancy of a factor of ≈1.5 between the predicted ratio $M_{H_2}/(M_{H_2} + M_{stellar})$ and the observationally

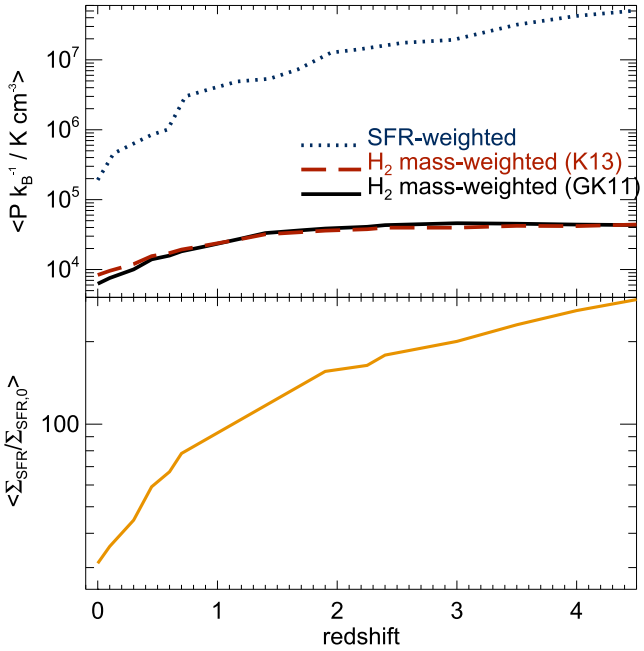

Figure 12. Top panel: the median of the M_{H_2}-weighted pressure, P, when the GK11 (solid line) and the K13 (long dashed line) prescriptions are applied to the simulation Ref-L100N1504, as a function of redshift. The SFR-weighted P (dotted line) is also shown as a function of redshift in the same simulation. Pressure is expressed in terms of $P\,k_B^{-1}$, where k_B is Boltzmann's constant. Bottom panel: the median $\Sigma_{SFR}/\Sigma_{SFR,0}$, where $\Sigma_{SFR,0} = 10^{-3}$ M$_\odot$ yr^{-1} kpc^{-2} is the SFR surface density in the solar neighbourhood (Bonatto & Bica 2011), calculated using all the star-forming particles at each time, as a function of redshift.

inferred ones at $z \approx 2$–3, regardless of whether the GK11 or the K13 prescription is applied. It is unclear at this stage whether this offset is a problem for EAGLE, since the observations have so far been biased towards gas-rich systems (as they only quote detections, except for the values in the local Universe given by COLD GASS). However, the offset between the predicted and observed ratio $M_{H_2}/(M_{H_2} + M_{stellar})$ is, in any case, within the scatter seen in observations.

The increase of the H_2 mass fraction with redshift is driven by the greater fraction of gas particles that reach the regions of high H_2 fractions ($T < 10^{4.5}$ K and $P/k_B \gtrsim 5 \times 10^3$ cm^{-3}). Such a trend can be seen from the evolution of the M_{H_2}-weighted pressure in the top panel of Fig. 12. The trend shows that at higher redshift more H_2 is locked up in particles of higher pressure than at $z = 0$.

The bottom panel of Fig. 11 shows the evolution of the depletion time-scale, τ_{H_2}, measured for the same main-sequence galaxies used to calculate the evolution of $M_{H_2}/(M_{H_2} + M_{stellar})$ shown in the top panel. We find that τ_{H_2} decreases with redshift by a factor of $\sim$10 between $z = 0$ and 4, with the GK11 prescription predicting a slightly larger decrease than the K13 prescription. Such evolution indicates that EAGLE galaxies at higher redshift are more efficient at forming stars. The non-monotonic behaviour of the H_2 fraction with redshift, seen at the top panel of Fig. 11, affects τ_{H2} by accelerating the decrease with redshift at $z \gtrsim 2.3$. This can be seen from fitting a power-law relation before and after $z = 2.3$: at $z \leq 2.3$, $\tau_{H2} \propto (1+z)^{-1.2}$, while at $z > 2.3$, $\tau_{H2} \propto (1+z)^{-3}$.

To help visualize the reasons behind the increase in the efficiency of star formation with redshift, we show in the top panel of Fig. 12 the SFR-weighted and M_{H_2}-weighted pressures for the Ref-L100N1504 simulation, which we refer to as $\langle P \rangle_{SFR}$ and $\langle P \rangle_{M_{H_2}}$, respectively. Applying the GK11 or K13 prescriptions in this simulation does not make any significant difference to $\langle P \rangle_{M_{H_2}}$. Another important observation is that both $\langle P \rangle_{SFR}$ and $\langle P \rangle_{M_{H_2}}$ increase with redshift, but $\langle P \rangle_{SFR}$ does it to a greater extent. Since the star formation law adopted in EAGLE has a power-law, $n = 1.4$ (see equation 2), forming stars at higher pressure implies having higher gas surface density and therefore higher $\Sigma_{SFR}/\Sigma_{gas}$. Although $\langle P \rangle_{M_{H_2}}$ also increases, its weaker evolution compared to $\langle P \rangle_{SFR}$ helps τ_{H_2} to decrease faster with redshift.

Lilly et al. (2013) have suggested that the observations shown in Fig. 11 are consistent with the 'equilibrium model', where the evolution in the specific SFR and the H_2 depletion time-scale are mainly driven by the evolution of the gas reservoir (neutral gas in the galaxy). We find that besides the gas reservoir, an additional parameter plays an important role: the typical pressures at which star formation takes place and at which H_2 forms. In other words, galaxies move along the Kennicutt–Schmidt relation towards higher gas surface densities as redshift increases. The change in the H_2 depletion time-scale is therefore a result of adopting a superlinear star formation law (i.e. power-law index >1 in equation 2) and the transition from H I- to H_2-dominated gas evolving towards higher mean ISM pressures as the redshift increases. Note that our conclusion is subject to the assumption that the Kennicutt–Schmidt relation holds at redshifts higher than 0. This is a reasonable assumption since observations so far show that galaxies at $z > 0$ follow the same Kennicutt–Schmidt relation (e.g. Genzel et al. 2010, 2013).

The trends obtained in EAGLE for the evolution of H_2 gas fractions and H_2 depletion time-scales are qualitatively the same as those inferred from observations. However, observational results have so far presented contradictory conclusions. For example, Genzel et al. (2015) presented CO and dust mass observations for ≈500 galaxies from $z \approx 0$ to $z \approx 3$, and concluded that the H_2 depletion time-scale is only a weak function of redshift, $\tau_{H_2} \propto (1+z)^{-0.3}$, while EAGLE favours a stronger dependence, $\tau_{H_2} \propto (1+z)^{-1.1}$ for the Ref-L100N1504 with the K13 prescription applied and $\tau_{H_2} \propto (1+z)^{-1.4}$ with the GK11 prescription applied. However, recent ALMA observations by Scoville et al. (2015) favour a much stronger redshift dependence of τ_{H_2}[3], with $z \approx 1$ galaxies having τ_{H_2} lower by an order of magnitude than $z = 0$ galaxies, in good agreement with the EAGLE results. Moreover, Scoville et al. (2015) found no difference in τ_{H_2} between galaxies on and off the main sequence, while Genzel et al. (2015) found an order of magnitude difference. This argues for further observations and the need for galaxy samples with simpler selection functions that can then be applied to simulations to make a one-to-one comparison (as we did in Section 4.2 for COLD GASS).

5.3 The properties of H_2 mass-selected samples

An interesting question is what are the stellar masses and SFRs of galaxies selected by their H_2 mass? This question is relevant for planned blind CO surveys as an important aspect of these surveys is finding optical and near-IR counterparts (see for example Decarli et al. 2014). We show in Fig. 13 the galaxy stellar mass functions for all galaxies (dashed lines) and for the subsample that have $M_{H_2} > 10^9$ M$_\odot$ (solid lines) at $0 < z < 3$. Typically, galaxies with

[3] Scoville et al. (2015) measure ISM masses rather than H_2 masses. However, from the very high H_2 gas fractions measured in high-z galaxies, one can safely assume that the ISM masses of high-z galaxies are dominated by H_2, on average.

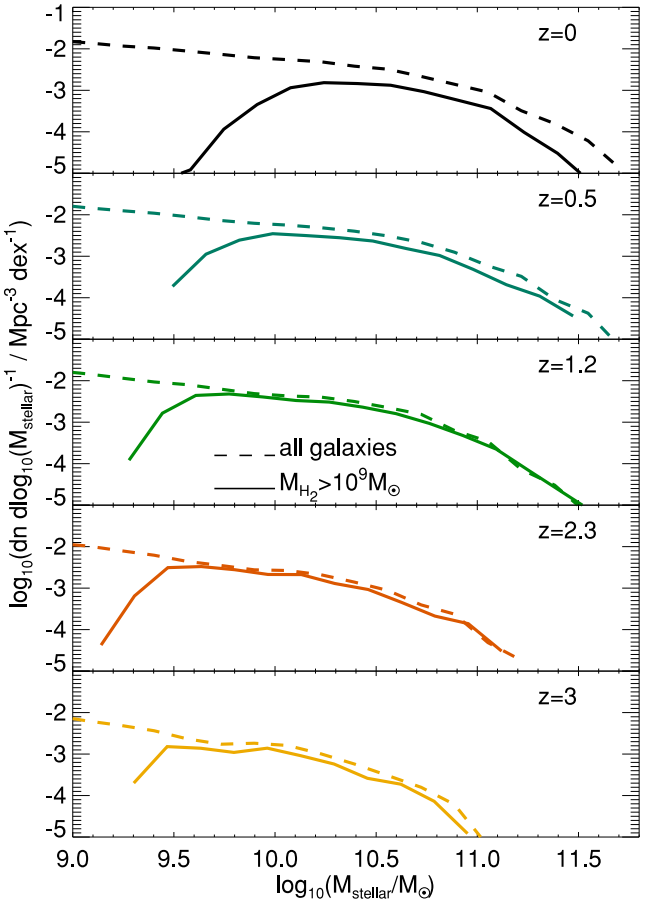

Figure 13. Galaxy stellar mass functions of all galaxies (dashed lines) and those having $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ (solid lines) at different redshifts, as labelled in each panel, for the simulation Ref-L100N1504 using the GK11 prescription. The H_2 mass-selected sample traces galaxies with stellar masses, $M_{\rm stellar} \gtrsim 10^{10}\,{\rm M}_\odot$, but at $z \lesssim 0.5$ about half of those massive galaxies have $M_{\rm H_2} < 10^9\,{\rm M}_\odot$.

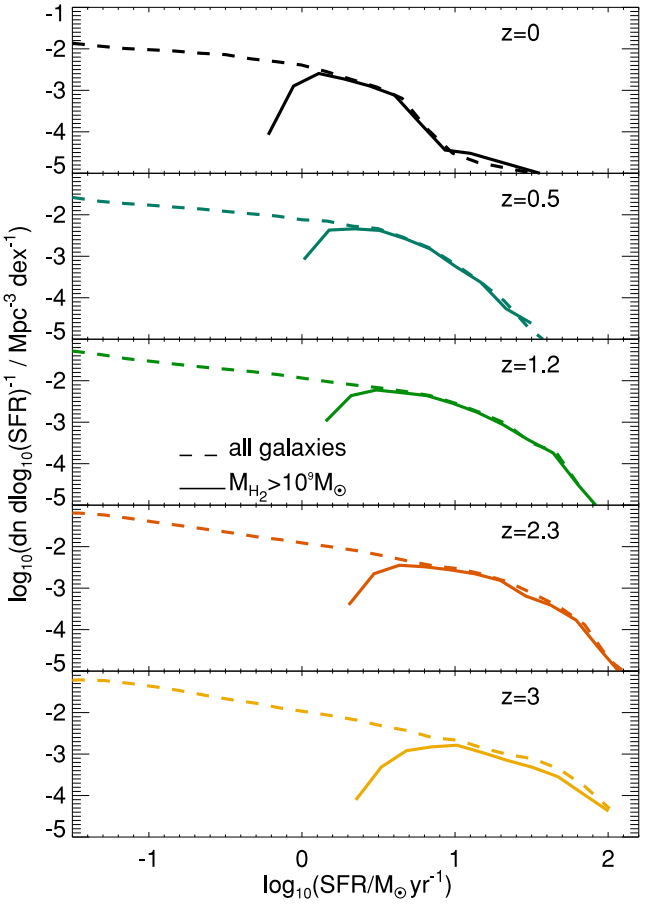

Figure 14. SFR functions of all galaxies (dashed lines) and those having $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ (solid lines) at different redshifts, as labelled in each panel, for the simulation Ref-L100N1504 using the GK11 prescription. The H_2 mass-selected sample traces galaxies with SFR$\gtrsim 1\,{\rm M}_\odot\,{\rm yr}^{-1}$ at $z = 0$, while at $z = 3$ the typical galaxies present in H_2 mass-selected samples have SFR $\gtrsim 10\,{\rm M}_\odot\,{\rm yr}^{-1}$ at $z = 3$.

large reservoirs of H_2 also have high stellar masses. However, at $z \lesssim 0.5$ we observe an increasing number of massive galaxies with $M_{\rm H_2} < 10^9\,{\rm M}_\odot$, which is clear from the difference in normalization between the galaxy stellar mass function of all galaxies and of those with $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ in the top two panels of Fig. 13. At $z \gtrsim 1$, galaxies with $M_{\rm stellar} > 10^{10}\,{\rm M}_\odot$ almost always have $M_{\rm H_2} > 10^9\,{\rm M}_\odot$. For a comparison between the predicted galaxy stellar mass function in EAGLE and observations see Furlong et al. (2015).

Fig. 14 shows the SFR function of all galaxies (dashed lines) and of those with $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ (solid lines). The H_2 mass-selected sample is a very good tracer of the most actively star-forming galaxies at any time. For example, at $z = 0$ galaxies with $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ have SFR $\gtrsim 1\,{\rm M}_\odot\,{\rm yr}^{-1}$. The average SFR of the H_2 mass-selected sample increases with redshift, not only because galaxies have on average higher SFRs at fixed stellar mass, but also because the ratio between the H_2 mass and the SFR decreases with redshift (see bottom panel of Fig. 11). This means that the H_2 mass at a fixed SFR decreases with redshift. The outcome of this is that the typical SFRs of galaxies with $M_{\rm H_2} > 10^9\,{\rm M}_\odot$ at $z = 3$ are SFR $\gtrsim 10\,{\rm M}_\odot\,{\rm yr}^{-1}$.

5.4 The evolution of $\Omega_{\rm H_2}$

Fig. 15 shows the evolution of $\Omega_{\rm H_2} \equiv \rho_{\rm H_2}/\rho_{\rm crit}$, where $\rho_{\rm crit}$ is the critical density of the Universe at the given redshift, for Ref-L100N1504 using both the GK11 and K13 prescriptions. The global density of H_2 was calculated using the H_2 abundance in galaxies measured in a 30 pkpc aperture. If we instead use all the gas particles that are bound to the subhalo of a galaxy, we find a value higher by 1.5–3 per cent, which shows that most of the H_2 in the universe resides close to galaxy centres. We also show for reference the evolution of the SFR cosmic density (dotted line) in the Ref-L100N1504 simulation.

The GK11 prescription applied to EAGLE predicts a peak at a slightly lower redshift, $z \approx 1.2$, compared with $z \approx 1.4$ for the K13 prescription. The latter results in an increase of a factor of $\approx$4.3 between $z = 0$ and $z \approx 1.4$, while GK11 yields an increase of a factor of $\approx$3.5 over the same redshift range. The difference is driven by the GK11 prescription being more sensitive to intense ISRF, as discussed in Section 5.2. The peak of $\Omega_{\rm H_2}$ is located at a redshift which in both cases is lower than the peak of the SFR density in EAGLE, at $z \approx 2$ (see dotted line in Fig. 15 and Furlong et al. 2015 for a detailed analysis on the SFR density evolution). The reason for this offset is the faster decrease in gas metallicity at $z > 1$ and the monotonic increase of the median $\Sigma_{\rm SFR}$ with redshift (see the bottom panel of Fig. 12). At $z < 1$, the H_2 mass-weighted metallicity decreases as $\log_{10}(Z) \propto (1+z)^{-0.6}$, while at $z > 1$ the decrease accelerates and scales as $\log_{10}(Z) \propto (1+z)^{-1.8}$. Note that

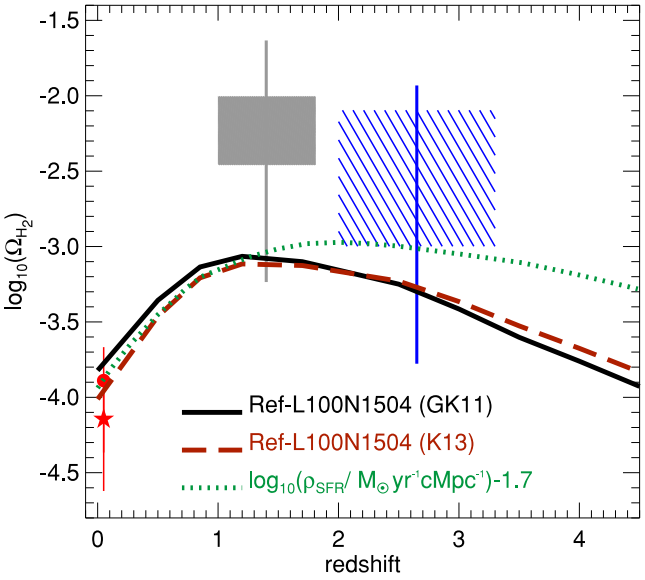

Figure 15. Ω_{H_2}, defined as the ratio of the H_2 density to the critical density of the universe, $\Omega_{H_2} \equiv \rho_{H_2}/\rho_{crit}$, as a function of redshift for the simulation Ref-L100N1504 using the GK11 (solid line) and the K13 (dashed line) prescriptions. For reference, we show the SFR cosmic density evolution as dotted line, offset by 1.7 dex so that it lies in a similar range than Ω_{H_2}. Observations at $z = 0$ by Keres et al. (2003) and Obreschkow & Rawlings (2009) are shown as a filled circle and star, respectively, with 1σ error bars, while the observations from the blind CO survey of Walter et al. (2014) are shown as solid and hatched regions. The boxes show the contribution of the CO blind detections without extrapolating to account for the undetected sources below the detection limits in the millimetre (which would translate into H_2 masses $< 5 \times 10^9\,M_\odot$). The error bars in the boxes indicate 1σ Poisson errors.

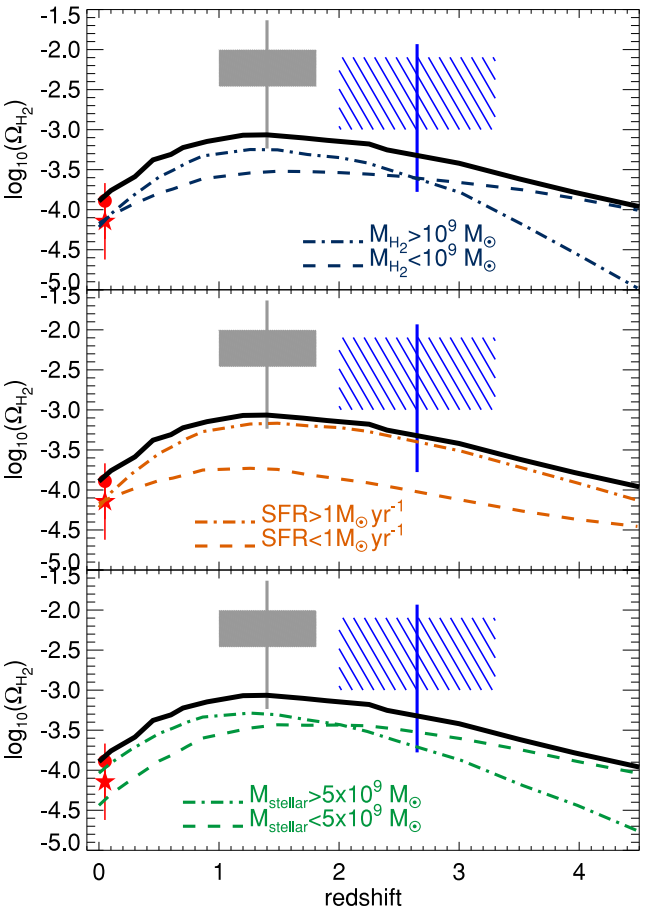

Figure 16. Ω_{H_2} as a function of redshift for the simulation Ref-L100N1504 using the GK11 prescription separated into the contribution from galaxies with different H_2 masses (top panel), SFRs (middle panel) and stellar masses (bottom panel), as labelled. The solid line in each panel corresponds to the total H_2 density also shown as a solid line in Fig. 15. Observations are as in Fig. 15 and are shown in all panels.

we still see a rapid increase of Ω_{H_2} between 0 and 1 even though some of the conditions for H_2 formation become less favourable (lower gas metallicities and higher ISRF). This is due to an increase in the pressure of the ISM in galaxies and to an increase in the total neutral gas content.

We also show in Fig. 15 the observational estimates of Keres et al. (2003), Obreschkow & Rawlings (2009) and Walter et al. (2014). The error bars on the Walter et al. (2014) inferences are as described in Section 5.1. Compared to current observational estimates, EAGLE captures the increase in the H_2 content of galaxies, but predicts a lower Ω_{H_2} in the redshift range $z = 1$–3, although within the error bars of the observations. Some possibilities to explain this tension are (i) EAGLE is truly predicting a lower value of Ω_{H_2} than in the real Universe, and (ii) the observational bias Ω_{H_2} towards higher values. Concerning (i), this could be related to the SFR cosmic density in EAGLE being below the observations by a factor of $\approx$2 (Furlong et al. 2015). Related to (ii), Walter et al. (2014) adopted a value of X to convert their CO measurements to H_2 similar to the MW value. However, theoretical evidence points to a lower value of X being more reasonable for these high-redshift galaxies. For example, Narayanan et al. (2012) show that the value of X decreases with increasing velocity width of the CO line. Since the CO velocity widths of these high-redshift galaxies are much larger than for $z = 0$ galaxies, one would expect lower values of X. These systematics in the observations need to be carefully addressed before we can conclude that the tension between EAGLE and the observational inferences of Ω_{H_2} is a true shortcoming of the simulation.

Walter et al. (2014) argue that the tension seen between several predictions of Ω_{H_2} in the literature and their observational estimates are, in reality, worse than it seems from comparisons such as the one presented here. This is because their estimate of Ω_{H_2} accounts only for detections (secure and candidates) and it does not include the contribution from galaxies with $M_{H_2} \lesssim 5 \times 10^9\,M_\odot$. However, to assess whether this really worsens the tension with theoretical predictions, one needs to estimate how much of Ω_{H_2} is contributed by galaxies with different values of M_{H_2} or other properties. We show in Fig. 16 the value of Ω_{H_2} for subsamples of galaxies in EAGLE selected by their H_2 mass, SFR and stellar mass. Galaxies with $M_{H_2} > 10^9\,M_\odot$ dominate the global Ω_{H_2} at $z < 2.5$, which means that we expect the H_2 mass function integrated over the massive end only (above the break) to recover most of the H_2 in the universe for these redshifts.

In terms of stellar mass and SFR, Ω_{H_2} is dominated by galaxies with

(i) SFR $> 10\,M_\odot\,yr^{-1}$ at $z \gtrsim 0.1$,
(ii) $M_{stellar} > 5 \times 10^9\,M_\odot$ at $z \lesssim 2$, and
(iii) $M_{stellar} < 5 \times 10^9\,M_\odot$ at $z \gtrsim 2$.

The dominant contribution of galaxies with $M_{stellar} < 5 \times 10^9\,M_\odot$ at $z \gtrsim 2$ is unlikely to be a consequence of the limited

simulated volume (100 cMpc box size). In fact, $\Omega_{\rm H_2}$ convergences to better than 15 per cent in the simulations Ref-L100N1504 and Ref-L050N0752.

These results are slightly different to those found in semi-analytic models. For example Lagos et al. (2014a) find that galaxies with $1\,{\rm M_\odot\,yr^{-1}} < {\rm SFR} < 10\,{\rm M_\odot\,yr^{-1}}$ make a larger contribution to $\Omega_{\rm H_2}$ than those with ${\rm SFR} > 10\,{\rm M_\odot\,yr^{-1}}$. These differences are interesting and will be tested by future observations.

The galaxies that dominate $\Omega_{\rm H_2}$ in EAGLE are readily observable in current optical and near-IR surveys (for example the CANDELS survey; Grogin et al. 2011). This suggests that a cheap and effective observing strategy to unveil $\Omega_{\rm H_2}$ is to follow up existing optical and near-infrared surveys to measure CO, rather than blind CO surveys that are more expensive and harder to characterize. In follow-up strategies, molecular emission line surveys can be guided by well-known UV, optical, IR and/or radio surveys, with well-identified positions and spectroscopic or photometric redshifts. Even in the case that only galaxies with ${\rm SFR} > 10\,{\rm M_\odot\,yr^{-1}}$ are followed up, we expect them to uncover 50 per cent or more of $\Omega_{\rm H_2}$, placing strong constraints on theoretical models of galaxy formation.

6 CONCLUSIONS

We have presented a comprehensive comparison between two hydrodynamic simulations of the EAGLE suite and surveys of molecular gas in galaxies at low and high redshift. To determine $\rm H_2$ masses in simulated galaxies, we first extracted the fraction of neutral gas (atomic and molecular) in each gas particle using the fitting functions of Rahmati et al. (2013a), which were derived applying radiative transfer algorithms to cosmological simulations in order to calculate local neutral fractions as a function of the collisional and photoionization rates, temperature and gas density. After estimating the fraction of the gas that is neutral, we proceeded to calculate the fraction of that gas that is molecular. We applied two schemes to EAGLE to calculate the $\rm H_2$ fraction of individual gas particles, those advanced by Gnedin & Kravtsov (2011) and by Krumholz (2013). The first one, GK11, provides fitting functions for the $\rm H_2$ fraction that depend on the dust-to-gas mass ratio and the local ISRF. These fitting functions were calculated from a set of small, high-resolution cosmological simulations, which included on-the-fly radiative transfer, a simple model for $\rm H_2$ formation and dissociation, where the main catalyst for $\rm H_2$ formation are dust grains. The second prescription, K13, is a theoretical model where the $\rm H_2$ fraction is calculated assuming pressure balance between the warm and cold neutral media, and depends again on the dust-to-gas mass ratio and the local ISRF. After computing the $\rm H_2$ fraction in each gas particle, we used apertures of 30 pkpc to calculate the $\rm H_2$ mass in individual galaxies.

We compared the predicted $\rm H_2$ masses with observations, and our main conclusions are as follows.

(i) The $\rm H_2$ galaxy mass functions at $z = 0$ predicted by EAGLE for the two prescriptions tested here are very similar to each other and in good agreement with the observational estimates of Keres et al. (2003) and Obreschkow & Rawlings (2009, see Fig. 3). We showed that the total star-forming gas mass in EAGLE becomes a poor proxy for the $\rm H_2$ mass for $M_{\rm H_2} \gtrsim 5 \times 10^8\,{\rm M_\odot}$. We find that $\rm H_2$ masses become affected by resolution below $M_{\rm H_2} \approx 5 \times 10^7\,{\rm M_\odot}$, and for $M_{\rm stellar} < 10^9\,{\rm M_\odot}$.

(ii) We made an extensive comparison with the COLD GASS survey (Saintonge et al. 2011), applying the same selection criteria as were applied to the observations: $M_{\rm stellar} > 10^{10}\,{\rm M_\odot}$ and $M_{\rm H_2}/M_{\rm stellar} > 0.015$ in galaxies with $M_{\rm stellar} > 10^{10.6}\,{\rm M_\odot}$ or $\rm H_2$ masses $> 10^{8.8}\,{\rm M_\odot}$ in galaxies with $10^{10}\,{\rm M_\odot} < {\rm M_{stellar}} < 10^{10.6}\,{\rm M_\odot}$. We found generally good agreement in the scaling between the $\rm H_2$ mass and other properties such as stellar mass (Fig. 5), SFR (Fig. 6) and stellar surface density (Fig. 8). An interesting discrepancy was seen in the scaling between the SFR and the $\rm H_2$ mass. Observations reveal a population of galaxies that have low SFRs, but a considerable amount of $\rm H_2$ (for example the early-type galaxies in Davis et al. 2014; see Fig. 7). These galaxies are offset from the main relation traced by spiral galaxies. In EAGLE, we find that all galaxies follow the latter relation regardless of their morphology, with little scatter, and therefore no galaxies lie in the region of low SFR for a fixed $\rm H_2$ mass. This is because at small scales the star formation law in EAGLE follows the standard Kennicutt (1998) star formation relation, which has a higher normalization than the star formation relation inferred for early-type galaxies (Davis et al. 2014).

(iii) We presented the SFRs and stellar masses of $\rm H_2$ mass-selected samples of galaxies in EAGLE. The most $\rm H_2$-rich galaxies are also the most actively star-forming galaxies at any redshift (Fig. 14), while also being the most massive at $z \gtrsim 0.5$ (Fig. 13). At $z \lesssim 0.5$, a significant fraction of galaxies with $M_{\rm stellar} > 10^{10}\,{\rm M_\odot}$ are $\rm H_2$-poor ($M_{\rm H_2} < 10^9\,{\rm M_\odot}$).

(iv) EAGLE predicts (for both the GK11 or K13 prescriptions applied) that $\Omega_{\rm H_2}$ increases by a factor of $\approx$4 from $z = 0$ to $z \approx 1.3$–1.5, followed by a decrease towards higher redshifts (Fig. 15). In EAGLE, the peak redshift of $\Omega_{\rm H_2}$ is lower than the peak redshift of the SFR density, $z \approx 2$. We find that this is due to the gas metallicity decreasing faster at $z > 1$ compared to the decrease at $z < 1$, and the SFR surface density increasing with increasing redshift, on average. Both trends hamper the formation of $\rm H_2$ at $z \gtrsim 1$. We find that the EAGLE $\rm H_2$ mass functions and $\Omega_{\rm H_2}$ agree well with the observational inferences in the redshift range $0 < z < 1.5$, but underestimate the $\rm H_2$ abundance at $2 < z < 3$ (see Fig. 10). However, it is unclear how much of the latter discrepancy is driven by systematic errors in the observations. We also find that $\Omega_{\rm H_2}$ in EAGLE is dominated by galaxies with $M_{\rm H_2} > 10^9\,{\rm M_\odot}$ at $z < 2.5$, ${\rm SFR} > 10\,{\rm M_\odot\,yr^{-1}}$ at $0.1 < z < 5$ and $M_{\rm stellar} > 5 \times 10^9\,{\rm M_\odot}$ at $z \lesssim 2$ (Fig. 16). The latter implies that an efficient strategy for unveiling $\Omega_{\rm H_2}$ is to use current optical and near-IR surveys with spectroscopic or good photometric redshifts and accurate galaxy positions and to follow them up using millimetre telescopes targeting the CO emission.

The agreement between the properties of molecular gas in the EAGLE simulations and observations at low and high redshift is remarkable, particularly when one considers that all the free parameters of the subgrid physics models in the simulation were fixed by requiring a match to the observed galaxy stellar mass function and sizes at the present day. No parameters were adjusted to match any of the $\rm H_2$ observations discussed in this paper. We now plan to study the distribution of $\rm H_2$ within galaxies in EAGLE and to compare with resolved observations, in an effort to further our understanding of how the physical processes modelled in EAGLE determine the properties of dense gas in galaxies.

ACKNOWLEDGEMENTS

We thank Matt Bothwell, Paola Santini and Tim Davis for providing observational data sets, and Thorsten Naab, Alessandro Romeo and Padelis Papadopoulosfor useful discussions. CL is funded by a Discovery Early Career Researcher Award (DE150100618). RAC is a Royal Society University Research Fellow. The research

leading to these results has received funding from the European Community's Seventh Framework Programme (/FP7/2007-2013/) under grant agreement no. 229517 and from the ERC grants: 278594-GasAroundGalaxies. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. Support was also provided by the European Research Council (grant numbers GA 267291 'Cosmiway' and GA 238356 'Cosmocomp'), the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office ([AP P7/08 CHARM]), the National Science Foundation under Grant no. NSF PHY11-25915, and the UK Science and Technology Facilities Council (grant numbers ST/F001166/1 and ST/I000976/1) through rolling and consolidating grants awarded to the ICC.

REFERENCES

Allende Prieto C., Lambert D. L., Asplund M., 2001, ApJ, 556, L63
Aravena M. et al., 2012, MNRAS, 426, 258
Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, ApJS, 196, 22
Bayet E., Viti S., Williams D. A., Rawlings J. M. C., Bell T., 2009, ApJ, 696, 1466
Bell T. A., Viti S., Williams D. A., 2007, MNRAS, 378, 983
Béthermin M. et al., 2015, A&A, 573, A113
Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
Bigiel F., Leroy A., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, AJ, 140, 1194
Bisbas T. G., Papadopoulos P. P., Viti S., 2015, ApJ, 803, 37
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Bonatto C., Bica E., 2011, MNRAS, 415, 2827
Boselli A., Lequeux J., Gavazzi G., 2002, A&A, 384, 33
Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, A&A, 564, A66
Bothwell M. S., Kennicutt R. C., Lee J. C., 2009, MNRAS, 400, 154
Bothwell M. S. et al., 2014, MNRAS, 445, 2599
Caldú-Primo A., Schruba A., Walter F., Leroy A., Sandstrom K., de Blok W. J. G., Ianjamasimanana R., Mogotsi K. M., 2013, AJ, 146, 150
Cappellari M. et al., 2011, MNRAS, 413, 813
Carilli C. L., Walter F., 2013, ARA&A, 51, 105
Catinella B. et al., 2010, MNRAS, 403, 683
Chabrier G., 2003, PASP, 115, 763
Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, MNRAS, 425, 3058
Crain R. A. et al., 2015, MNRAS, 450, 1937
Cullen L., Dehnen W., 2010, MNRAS, 408, 669
Daddi E. et al., 2010, ApJ, 713, 686
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davis T. A. et al., 2011, MNRAS, 414, 968
Davis T. A. et al., 2014, MNRAS, 444, 3427
Decarli R. et al., 2014, ApJ, 782, 78
Dessauges-Zavadsky M. et al., 2015, A&A, 577, A50
Duffy A. R., Kay S. T., Battye R. A., Booth C. M., Dalla Vecchia C., Schaye J., 2012a, MNRAS, 420, 2799
Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012b, MNRAS, 426, 3385
Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465
Feldmann R., Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 732, 115
Feldmann R., Gnedin N. Y., Kravtsov A. V., 2012, ApJ, 758, 127
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515
Furlong M. et al., 2015, MNRAS, 450, 4486
Geach J. E., Smail I., Moran S. M., MacArthur L. A., Lagos C. d. P., Edge A. C., 2011, ApJ, 730, L19
Genel S. et al., 2014, MNRAS, 445, 175
Genzel R. et al., 2010, MNRAS, 407, 2091
Genzel R. et al., 2013, ApJ, 773, 68
Genzel R. et al., 2015, ApJ, 800, 20
Glover S. C. O., Clark P. C., 2012, MNRAS, 421, 9
Glover S. C. O., Jappsen A.-K., 2007, ApJ, 666, 1
Gnedin N. Y., Draine B. T., 2014, ApJ, 795, 37 (GD14)
Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88 (GK11)
Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55
Grogin N. A. et al., 2011, ApJS, 197, 35
Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. CEA, Saclay, p. 64
Habing H. J., 1968, Bull. Astron. Inst. Neth., 19, 421
Hopkins P. F., 2013, MNRAS, 428, 2840
Huang M.-L., Kauffmann G., 2014, MNRAS, 443, 1329
Kennicutt R. C., Jr, 1998, ApJ, 498, 541
Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659
Krumholz M. R., 2013, MNRAS, 436, 2747 (K13)
Krumholz M. R., McKee C. F., 2005, ApJ, 630, 250
Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011a, MNRAS, 416, 1566
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011b, MNRAS, 418, 1649
Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, MNRAS, 426, 2142
Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014a, MNRAS, 440, 920
Lagos C. d. P., Davis T. A., Lacey C. G., Zwaan M. A., Baugh C. M., Gonzalez-Perez V., Padilla N. D., 2014b, MNRAS, 443, 1002
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119
Magdis G. E. et al., 2012, ApJ, 758, L9
Marigo P., 2001, A&A, 370, 194
Martig M. et al., 2013, MNRAS, 432, 1914
Meijerink R., Spaans M., Israel F. P., 2007, A&A, 461, 793
Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, MNRAS, 421, 3127
Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857
Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009a, ApJ, 698, 1467
Obreschkow D., Heywood I., Klöckner H.-R., Rawlings S., 2009b, ApJ, 702, 1321
Ostriker E. C., McKee C. F., Leroy A. K., 2010, ApJ, 721, 975
Pelupessy F. I., Papadopoulos P. P., 2009, ApJ, 707, 954
Pelupessy F. I., Papadopoulos P. P., van der Werf P., 2006, ApJ, 645, 1024
Pety J. et al., 2013, ApJ, 779, 43
Planck Collaboration XVI, 2014, A&A, 571, A16
Popping G., Pérez-Beaupuits J. P., Spaans M., Trager S. C., Somerville R. S., 2014, MNRAS, 444, 1301
Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505
Price D. J., 2008, J. Comput. Phys., 227, 10040
Putman M. E., Peek J. E. G., Joung M. R., 2012, ARA&A, 50, 491
Rahmati A., Pawlik A. H., Raicevic M., Schaye J., 2013a, MNRAS, 430, 2427
Rahmati A., Schaye J., Pawlik A. H., Raicevic M., 2013b, MNRAS, 431, 2261
Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, preprint (arXiv:1503.05553)

Richings A. J., Schaye J., Oppenheimer B. D., 2014a, MNRAS, 440, 3349
Richings A. J., Schaye J., Oppenheimer B. D., 2014b, MNRAS, 442, 2780
Robertson B. E., Kravtsov A. V., 2008, ApJ, 680, 1083
Rosas-Guevara Y. M. et al., 2013, preprint (arXiv:1312.0598)
Saintonge A. et al., 2011, MNRAS, 415, 32
Saintonge A. et al., 2012, ApJ, 758, 73
Saintonge A. et al., 2013, ApJ, 778, 2
Santini P. et al., 2014, A&A, 562, A30
Schaye J., 2001, ApJ, 562, L95
Schaye J., 2004, ApJ, 609, 667
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Schruba A. et al., 2011, AJ, 142, 37
Scoville N. et al., 2015, preprint (arXiv:1505.02159)
Shetty R., Glover S. C., Dullemond C. P., Klessen R. S., 2011, MNRAS, 412, 1686
Springel V., 2005, MNRAS, 364, 1105
Springel V. et al., 2008, MNRAS, 391, 1685
Swinbank A. M. et al., 2014, MNRAS, 438, 1267
Tacconi L. J. et al., 2013, ApJ, 768, 74
Thielemann F.-K. et al., 2003, in Hillebrandt W., Leibundgut B., eds, From Twilight to Highlight: The Physics of Supernovae. Springer, Heidelberg, p. 331
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Walch S. K. et al., 2015, MNRAS, 451, 2757
Walter F. et al., 2014, ApJ, 782, 79
Wendland H., 1995, Adv. Comput. Math., 4, 389 396
Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
Wolfire M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, ApJ, 587, 278
Young L. M. et al., 2011, MNRAS, 414, 940
Young L. M. et al., 2014, MNRAS, 444, 3408
Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014, ApJ, 791, 130

APPENDIX A: THE TRANSITION FROM NEUTRAL TO MOLECULAR HYDROGEN

In this appendix, we provide the equations used to apply the GK11 and K13 prescriptions to EAGLE.

A1 The prescription of Gnedin & Kravtsov (2011) applied to EAGLE

In this section, we describe the analytic formulae that Gnedin & Kravtsov (2011) obtained for computing the H_2 fraction and the way we apply them to EAGLE. The H_2 fraction is calculated as

$$f_{H_2} \equiv \frac{\Sigma_{H_2}}{\Sigma_n} \approx \left(1 + \frac{\Sigma_c}{\Sigma_n}\right)^2, \tag{A1}$$

where, Σ_{H_2} is the surface density of H_2 and Σ_n is the surface density of neutral hydrogen. The parameter Σ_c is defined as

$$\Sigma_c = 20\,M_\odot\,pc^{-2}\,\frac{\Lambda^{4/7}}{D_{MW}}\,\frac{1}{\sqrt{1 + G_0^{'}\,D_{MW}^2}}, \tag{A2}$$

where D_{MW} is the dust-to-gas mass ratio in units of the MW dust-to-gas mass ratio, which we take to be $D_{MW} \equiv Z/Z_\odot$ under the assumption that the dust-to-gas ratio scales linearly with gas metallicity, Z, and that this relation is universal, $Z_\odot = 0.0127$ is the solar metallicity (Allende Prieto, Lambert & Asplund 2001), and $G_0^{'}$ is the ISRF (defined in the wavelength range 912–2400 Å) in units of the Habing radiation field, 1.6×10^{-3} erg cm^{-2} s^{-1} (Habing 1968). Radiation at these frequencies is produced mainly by OB stars. In EAGLE, we adopt a universal IMF and therefore the rate at which OB stars are formed is proportional to the SFR, such that we can express $G_0^{'}$ as $G_0^{'} = \Sigma_{SFR}/\Sigma_{SFR,0}$, where $\Sigma_{SFR,0} = 10^{-3}\,M_\odot\,yr^{-1}\,kpc^{-2}$ is the SFR surface density in the solar neighbourhood (Bonatto & Bica 2011). Finally, Λ is defined as

$$\Lambda \equiv \ln\left[1 + g\,D_{MW}^{3/7}\left(\frac{G_0^{'}}{15}\right)^{4/7}\right]. \tag{A3}$$

Here, the function g is

$$g = \frac{1 + \alpha s + s^2}{1 + s}, \tag{A4}$$

where α and s are defined as

$$s \equiv \frac{0.04}{D_* + D_{MW}}, \tag{A5}$$

$$\alpha \equiv 5\,\frac{G_0^{'}/2}{1 + (G_0^{'}/2)^2}. \tag{A6}$$

Here D_* is defined as

$$D_* = 1.5 \times 10^{-3}\ln\left(1 + [3\,G_0^{'}]^{1.7}\right). \tag{A7}$$

Following Schaye (2001), we estimate the local neutral hydrogen surface density, Σ_n, by multiplying the volume density of hydrogen by the neutral fraction and the Jeans length, λ_J:

$$\Sigma_n = \eta\rho_H\,\lambda_J = \eta\rho_H\,\frac{c_s}{\sqrt{G\,\rho_H}}. \tag{A8}$$

Here, c_s is the effective sound speed which is a function of the pressure, P, and the gas density, ρ (see Schaye & Dalla Vecchia 2008 for details). From equations A1 and A8, we can estimate Σ_{H_2}.

We apply the above prescriptions to individual gas particles, where $\Sigma_{SFR} = \rho_{SFR}\,\lambda_J$, ρ_{SFR} is the SFR density (computed from the pressure using equation 12 of Schaye & Dalla Vecchia 2008), λ_J is calculated as in equation (A8), n_n is the particle density of neutral hydrogen and Z the particle smoothed metallicity.

For non-star-forming particles (those that have densities below the density threshold for star formation, see equation 1), $\rho_{SFR} = 0$, so we adopt the UV photoionization background of Haardt & Madau (2001), which is a function of redshift, to determine $G_0^{'}$. Rahmati et al. (2013b) showed that star-forming galaxies produce a galactic scale photoionization rate of $\approx 1.3 \times 10^{-13}$ s^{-1}, which is of a similar magnitude than the UV background at $z = 0$, and smaller than it at $z > 0$. This favours our approximation, which uses the UV photoionization background as the relevant photoionization rate for non-star-forming particles.

Note that, since equations A1–A7 are fits to the simulation set of Gnedin & Kravtsov (2011), there are no free parameters we can adjust. Instead, we apply the prescription as presented in Gnedin & Kravtsov directly to EAGLE. We also implemented the updated version of the GK11 prescription, described in Gnedin & Draine (2014) and found that H_2 masses were typically lower when the updated prescription is applied (which we refer to as GD14) than with the GK11 prescription. The top panel of Fig. A1 shows the H_2 masses predicted by the GD14 prescription as a function of the H_2 masses resulting from applying the GK11 prescription, at different redshifts. The differences seen in the H_2 masses with the GK11 and GD14 prescriptions depends on redshift, because the correction introduced by GD14 to the GK11 prescriptions changes

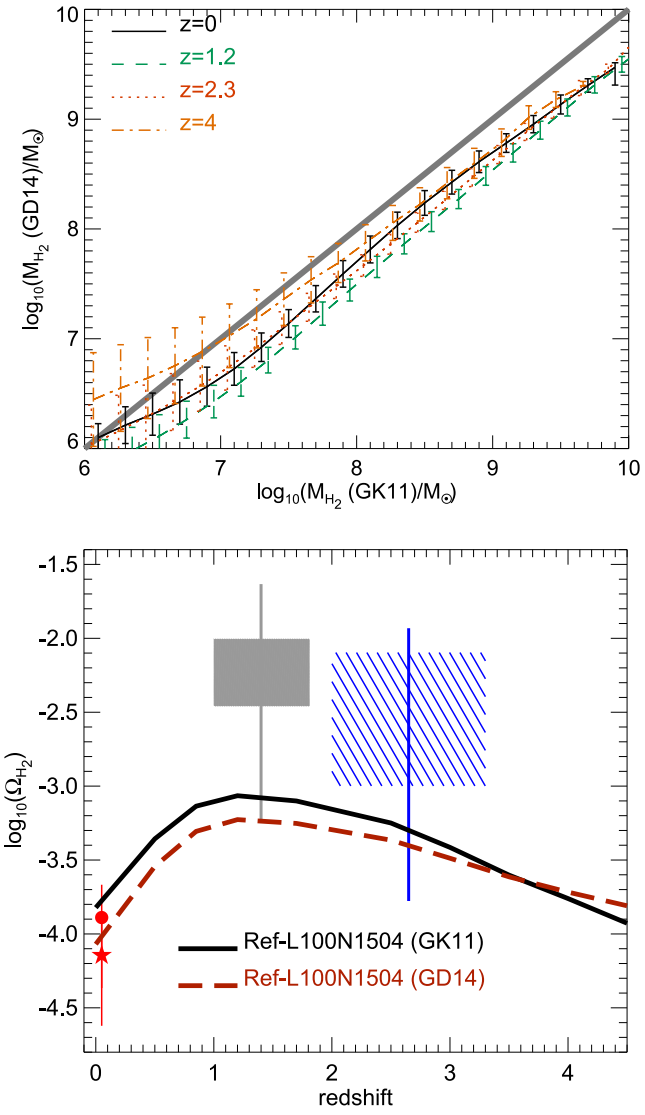

Figure A1. Top panel: the H_2 masses in the Ref-L100N1504 simulation when the GD14 prescription is applied compared to when the GK11 is applied, at different redshifts, as labelled. The diagonal solid line shows the relation M_{H_2}(GD14) $\equiv$ M_{H_2}(GK11). Lines with error bars show the median and 16th and 84th percentiles. Bottom panel: the evolution of Ω_{H_2} in the Ref-L100N1504 simulation with the GK11 and GD14 prescriptions applied, as labelled. Symbols are as in Fig. 15.

the dependency of f_{H_2} on the ISRF. In the bottom panel of Fig. A1, we show the effect the change in M_{H_2} has on Ω_{H_2}. Although there is a change in overall normalization, the behaviour of the two models is quite similar: there is a peak in Ω_{H_2} at $z \approx 1.2$ followed by a decline at higher redshift.

A2 The prescription of Krumholz (2013) applied to EAGLE

Krumholz (2013) considered an ISM composed of a WNM and a CNM which are in pressure equilibrium. This implies a minimum density of the CNM, first described by Wolfire et al. (2003), which is

$$n_{\rm CNM,min} \approx 31\, G_0^{'} \left(\frac{Z_{\rm d}/Z}{1 + 3.1 (G_0^{'}\, Z_{\rm d}/\zeta_{\rm t})^{0.365}} \right) \ {\rm cm}^{-3}, \quad \text{(A9)}$$

where $G_0^{'}$ is the ISRF in units of the Habing radiation field, $Z_{\rm d}$ and Z are the dust and gas phase metallicities, respectively, and $\zeta_{\rm t}$ is the ionization rate due to cosmic rays and X-rays. Assuming that the cosmic rays density and the UV intensity scale with the local SFR, Krumholz (2013) approximates $G_0^{'}/\zeta_{\rm t} = 1$. Similarly, the metallicity of the dust and gas phase are assumed to be equal, $Z_{\rm d}/Z \equiv 1$. In two-phase equilibrium, the CNM can have densities between $n_{\rm CNM,\,min}$ and $\approx 5\, n_{\rm CNM,\,min}$. With this in mind, Krumholz (2013) writes a fiducial CNM density for two-phase equilibrium

$$n_{\rm CNM,2p} \approx 23\, G_0^{'} \left(\frac{1 + 3.1\, D_{\rm MW}^{0.365}}{4.1} \right)^{-1} \ {\rm cm}^{-3}. \quad \text{(A10)}$$

Note that we slightly changed the notation of Krumholz (2013) to make it consistent with the notation used in Section A1.

In the regime where $G_0^{'} \rightarrow 0$, $n_{\rm CNM,\,2p}$ in equation (A10) also tends to 0, which would imply that the pressure of the CNM tends to 0 too. This regime is unphysical, and therefore Krumholz (2013) argues that the thermal pressure of the gas becomes relevant in this regime and sets a minimum CNM density to maintain hydrostatic balance, $n_{\rm CNM,\,hydro}$ (see also Ostriker, McKee & Leroy 2010). The latter depends on the maximum temperature at which the CNM can exist ($\approx$243 K; Wolfire et al. 2003), $T_{\rm CNM,\,max}$, the matter density (of stars and dark matter), $\rho_{\rm sd}$, and the neutral gas surface density, $\Sigma_{\rm n}$ (which we compute using equation A8). Krumholz (2013) writes

$$n_{\rm CNM,hydro} = \frac{P_{\rm th}}{1.1\, k_{\rm B}\, T_{\rm CNM,max}}, \quad \text{(A11)}$$

where

$$P_{\rm th} \approx \frac{\pi\, G\, \Sigma_{\rm n}^2}{4\,\alpha} \left[1 + \sqrt{1 + \frac{32\, \zeta_{\rm d}\, \alpha\, f_{\rm w}\, c_{\rm w}^2\, \rho_{\rm sd}}{\pi\, G\, \Sigma_{\rm n}^2}} \right]. \quad \text{(A12)}$$

Here, $\alpha \approx 5$ represents how much of the mid-plane pressure support comes from turbulence, magnetic fields and cosmic rays, compared to the thermal pressure (Ostriker et al. 2010), $\zeta_{\rm d} \approx 0.33$ is a numerical factor that depends on the shape of the gas surface isodensity contour, $f_{\rm w} = 0.5$ is the ratio between the mass-weighted mean square thermal velocity dispersion and the square of the sound speed of the warm gas (the value adopted here originally comes from Ostriker et al. 2010) and $c_{\rm w} = 8\,{\rm km\,s^{-1}}$ is the sound's speed in the WNM (e.g. Leroy et al. 2008). The value of the gas density in the CNM is then taken to be $n_{\rm CNM} = \max(n_{\rm CNM,\,2p}, n_{\rm CNM,\,hydro})$.

Krumholz (2013) defines a dimensionless radiation field parameter

$$\chi = 7.2\, G_0^{'} \left(\frac{n_{\rm CNM}}{10\,{\rm cm^{-3}}} \right)^{-1}, \quad \text{(A13)}$$

and writes the ratio between the H_2 density and the total neutral hydrogen density as

$$f_{\rm H_2} = \begin{cases} 1 - 0.75\, s/(1 + 0.25\, s), & s < 2 \\ 0, & s \geq 2 \end{cases} \quad \text{(A14)}$$

where

$$s \approx \frac{\ln(1 + 0.6\,\chi + 0.01\,\chi^2)}{0.6\,\tau_{\rm c}}, \quad \text{(A15)}$$

$$\tau_{\rm c} = 0.066\, f_{\rm c}\, D_{\rm MW} \left(\frac{\Sigma_{\rm n}}{{\rm M_\odot pc^{-2}}} \right). \quad \text{(A16)}$$

Here, $f_{\rm c}$ is a clumping factor representing the ratio between the surface density of a gas complex and the surface density of the diffuse medium. This value is suggested to be $f_{\rm c} \approx 5$ on scales of $\approx$1 kpc by Krumholz & McKee (2005) based on observations of

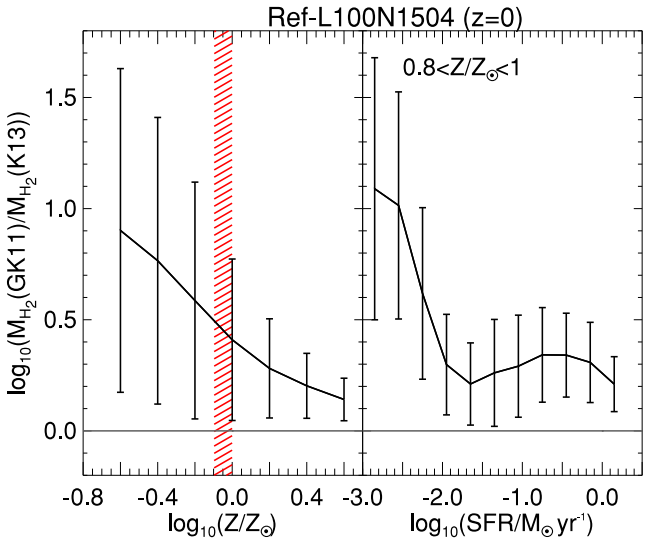

Figure A2. The ratio between the H_2 mass calculated using the GK11 prescription and the mass calculated using the K13 prescription, $M_{H_2}(GK11)/M_{H_2}(K13)$, as a function of the neutral gas metallicity at $z = 0$ in the simulation Ref-L100N1504 (left-hand panel). There is a trend of increasing $M_{H_2}(GK11)/M_{H_2}(K13)$ with decreasing metallicity. Lines with error bars show the median and the 16th and 84th percentiles. We select galaxies in a narrow range in metallicity, shown by the hatched strip in the left-hand panel, and show in the right-hand panel the ratio $M_{H_2}(GK11)/M_{H_2}(K13)$ as a function of the SFR. For reference, we show in both panels the ratio $M_{H_2}(GK11)/M_{H_2}(K13) = 1$ as horizontal line. The major driver of the dispersion at a fixed metallicity is the SFR.

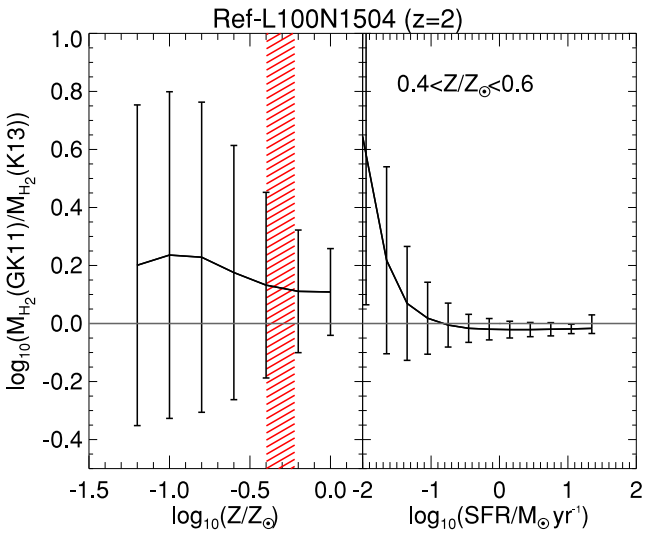

Figure A3. As Fig. A2, but for galaxies at $z = 2$. Note that to produce the right-hand panel, we select galaxies with lower metallicities than in the right-hand panel of Fig. A2 to increase the number of galaxies selected.

nearby galaxies. Since the spatial resolution of EAGLE is of the order of 1 kpc at low redshift (see Table 1), we adopt $f_c = 5$.

As in Section A1, we apply equations A10–A16 to individual gas particles to obtain the H_2 mass per particle, and then calculate the total H_2 mass within 30 pkpc of the centre of individual subhalos.

A3 Differences between the GK11 and K13 prescriptions

Fig. A2 shows the ratio between the H_2 mass calculated using the GK11 prescription and the mass calculated using the K13 prescription, $M_{H_2}(GK11)/M_{H_2}(K13)$, as a function of the neutral gas metallicity at $z = 0$ in the simulation Ref-L100N1504. The largest differences between $M_{H_2}(GK11)$ and $M_{H_2}(K13)$ are obtained in the low-metallicity regime, $Z \lesssim 0.5\,Z_\odot$, where $M_{H_2}(GK11)/M_{H_2}(K13) > 5$. However, the variation around the median value of $M_{H_2}(GK11)/M_{H_2}(K13)$ seen at a fixed gas metallicity is large. Fow example, at $Z \approx 0.25\,Z_\odot$, the median value of $M_{H_2}(GK11)/M_{H_2}(K13)$ is ≈8, but the 16th and 84th percentiles are ≈1.5 and ≈38, respectively, revealing that there is another important parameter driving the differences seen in the H_2 mass resulting from applying the GK11 and K13 prescriptions.

To gain insight into what is driving the large dispersion in $M_{H_2}(GK11)/M_{H_2}(K13)$ at a fixed gas metallicity, we select galaxies in a narrow range of metallicities (shown by the hatched strip in the left-hand panel of Fig. A2) and plot for those galaxies the ratio $M_{H_2}(GK11)/M_{H_2}(K13)$ as a function of SFR in the right-hand panel of Fig. A2. It can be seen that galaxies with SFR$< 0.01\,M_\odot\,yr^{-1}$ exhibit the largest differences between $M_{H_2}(GK11)$ and $M_{H_2}(K13)$. We conclude that large deviations between H_2 mass calculated using the GK11 or the K13 prescriptions appear in regimes of low metallicity and weak ISRF. Note there is still a large scatter in the distribution shown at the right-hand panel of Fig. A2, which is due to the SFR entering in the H_2 fraction equations as SFR surface density.

At $z \gtrsim 1$, an interesting regime appears, which is the one of intense ISRF, $G_0' \gtrsim 100$ in units of the Habing field (see Fig. 12). In this regime, we observe that the K13 prescription gives similar or slightly higher H_2 fractions than the GK11. Fig. A3 is analogous to Fig. A2 but for $z \approx 2$. As can be seen from the right-hand panel of Fig. A3, galaxies with SFR $> 1\,M_\odot\,yr^{-1}$, have a ratio $M_{H_2}(GK11)/M_{H_2}(K13) \sim 1$. This is the regime of intense ISRF. Note that for the same metallicity, galaxies at $z = 0$ have a higher ratio $M_{H_2}(GK11)/M_{H_2}(K13)$ than galaxies at $z = 2$, on average.

A3.1 Comparison with previous work

In previous works, the empirical prescription of Blitz & Rosolowsky (2006) has been applied to gas particles in hydrodynamic simulations (e.g. Duffy et al. 2012b), but restricting the calculation to gas particles that are star forming. Duffy et al. (2012b) adopted the following prescription (following the observations presented in Leroy et al. 2008),

$$f_{H_2} = \left(\frac{P/k_b}{10^{4.23}\,K\,cm^{-3}}\right)^{0.8}, \quad \text{(A17)}$$

where k_b is Boltzmann's constant. To get a better insight into the main differences between the theoretical prescriptions adopted here and the empirical prescription adopted in Duffy et al. (2012b), we first estimate how much of the H_2 in the Ref-L100N1504 simulation with the prescriptions GK11 and K13 applied is locked up in particles that are star forming. We find that on average 82 per cent of the H_2 is located in gas particles that are star forming when the GK11 prescription is applied. This percentage rises to 99 per cent when the K13 prescription is applied. However, we find that variations around that median are large, and galaxies can have as little as 20 per cent of their H_2 mass in gas particles that are star forming.

Secondly, we study the dependency of f_{H_2}, calculated using the GK11 and K13 prescriptions, on pressure in the simulation Ref-L100N1504 and compare it with equation (A17). This is shown in Fig. A4. The relations resulting from applying the GK11 and K13 prescriptions deviate significantly from equation (A17) particularly

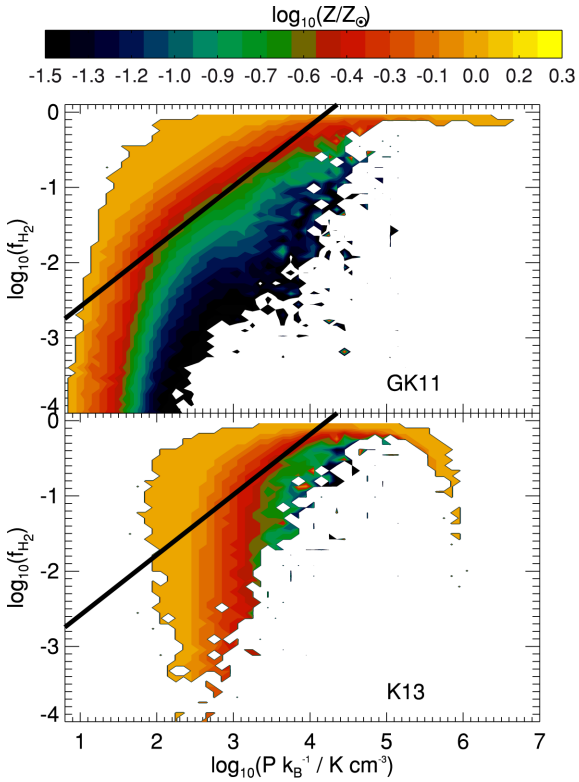

Figure A4. The fraction f_{H_2} as a function of pressure for the simulation Ref-L100N1504 when the GK11 (top panel) and the K13 (bottom panel) prescriptions are applied. Pixels are coloured by the median gas metallicity, as labelled at the top. The solid line shows the relation of equation (A17).

in regions of low metallicity. This is a good example of the differences seen between applying the empirical relation of equation (A17) and more physically motivated prescriptions that include a dependence on gas metallicity.

APPENDIX B: STRONG AND WEAK CONVERGENCE TESTS

Here, we compare the simulations Ref-L025N0376, Ref-L025N0752 and Recal-L025N0752 to address the effect of resolution on the H_2 masses of galaxies in a fixed cosmological volume. The comparison at a fixed volume is needed as simulations with different cosmological volumes sample different haloes, which has the problem that differences can come from either resolution or halo sampling. We remind the reader that the simulation Recal-L025N0752 has four parameters that are modified slightly compared to Ref-L025N0376 and Ref-L025N0752 (see Section 2 for details).

Fig. B1 shows the H_2 mass as a function of total subhalo mass (dark plus baryonic) at $z = 0$ for the simulations Ref-L025N0376 (solid lines), Ref-L025N0752 (dotted lines) and Recal-L025N0752 (dashed line) using either the GK11 (top panel) or K13 (bottom panel) prescriptions. The H_2 mass is measured in an aperture of 30 pkpc, while $M_{\rm Tot}$ is the total subhalo mass. We decide to use $M_{\rm Tot}$ to characterize when the resolution becomes important in the H_2 mass of galaxies in the intermediate-resolution simulations because the subhalo mass is a well converged quantity. For the strong convergence test, we need to compare Ref-L025N0376 with Ref-L025N0752 (i.e. fixed subgrid parameters). We see that the simulation Ref-L025N0752 shows H_2 masses for a fixed $M_{\rm Tot}$ that are

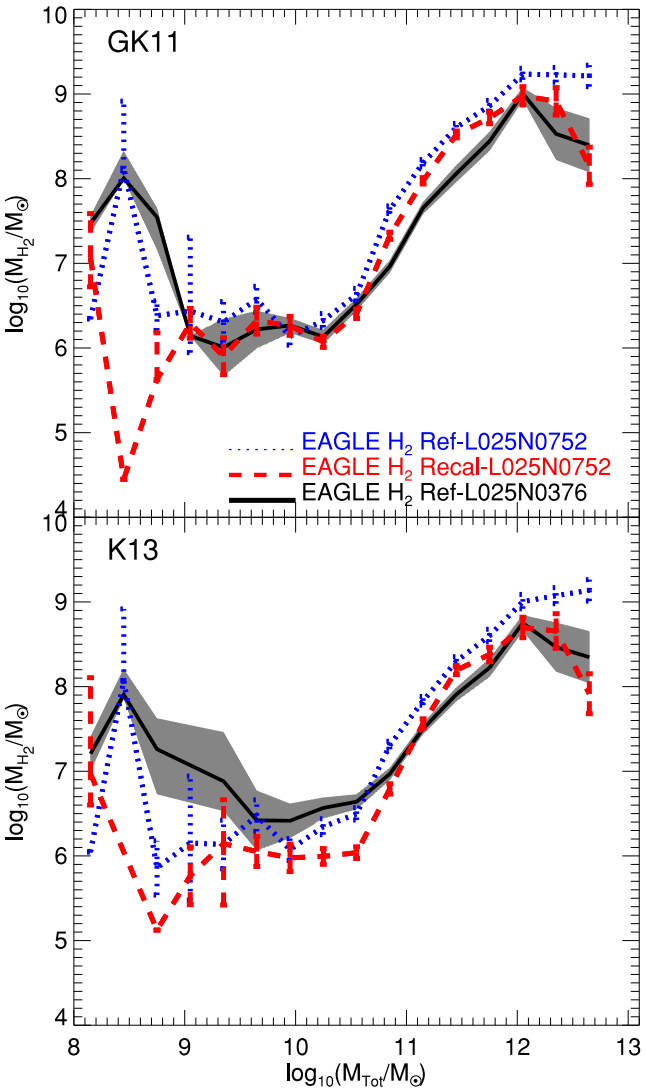

Figure B1. The H_2 mass as a function of the total matter, $M_{\rm Tot}$ (dark plus baryonic matter), at $z = 0$ for the simulations Ref-L025N0376, Ref-L025N0752 and Recal-L025N0752, as labelled, using the GK11 (top panel) and K13 (bottom panel) prescriptions. The H_2 mass here corresponds to the mass measured in an aperture of 30 pkpc, while $M_{\rm Tot}$ is the total subhalo mass. Lines show the medians, while the shaded region and error bars show the uncertainty on the median (e.g. the 16th and 84th percentiles of the distributions divided by $\sqrt{(N)}$, where N is the number of objects at each bin). The simulations Ref-L025N0376 and Recal-L025N0752 agree very well at $M_{\rm Tot} > 10^{10}\,{\rm M}_{\odot}$ regardless of the prescription used to calculate H_2. At $M_{\rm Tot} < 10^{10}\,{\rm M}_{\odot}$, the simulation Ref-L025N0376 starts to differ significantly from the higher resolution simulations which implies that resolution problems become important.

a factor of 2–3 larger than the H_2 masses in Ref-L025N0376 for $M_{\rm Tot} > 5 \times 10^{10}\,{\rm M}_{\odot}$. This difference is similar to the difference seen by S15 and Furlong et al. (2015) in the $z = 0.1$ galactic stellar mass function in the same set of simulations.

For the weak convergence test, we compare Ref-L025N0376 with Recal-L025N0752. We find that the medians are closer to each other than in the strong convergence test. The difference seen between the median relations in the simulations Ref-L025N0376 and Recal-L025N0752 is a factor of $\lesssim$1.5–2 for $M_{\rm Tot} > 5 \times 10^{10}\,{\rm M}_{\odot}$, which again is consistent with the differences seen in the galactic stellar mass function in S15 and Furlong et al. (2015). Note that the exact

prescription we use to calculate H_2 makes little difference to the offset obtained between the simulations we are comparing here.

The higher resolution of Ref-L025N0752 and Recal-L025N0752 enables us to establish that, at the resolutions of the simulations Ref-L025N0376 and Ref-L100N1504, we can trust the H_2 content of subhalos with $M_{\rm Tot} > 10^{10}\,{\rm M}_\odot$. In the three simulations analysed here, there are ≈1500 subhalos with $M_{\rm Tot} > 10^{10}\,{\rm M}_\odot$.

Note that it is difficult to establish a H_2 mass below which resolution effects become significant. This is because each gas particle has a weight, the H_2 fraction, and therefore a low H_2 mass can be obtained by either having a small H_2 fraction but a large number of gas particles, or by having a larger H_2 fraction but few gas particles. Thus, we find it more appropriate to express the resolution limit in terms of $M_{\rm Tot}$. We adopt $M_{\rm Tot} = 10^{10}\,{\rm M}_\odot$ as our resolution limit, and show unresolved results in Fig. 3 as a dotted line. In terms of H_2 mass, this resolution limit corresponds to $M_{\rm H_2} \approx 5 \times 10^7\,{\rm M}_\odot$ on average for the intermediate-resolution simulations, both for the GK11 and K13 prescriptions.

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