LETTER

Stellar populations dominated by massive stars in dusty starburst galaxies across cosmic time

Zhi-Yu Zhang[1,2], D. Romano[3], R. J. Ivison[2,1], P. P. Papadopoulos[4,5] & F. Matteucci[6,7,8]

[1]Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK [2]ESO, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany [3]INAF, Astrophysics and Space Science Observatory, Bologna via Piero Gobetti 93/3, 40129, Bologna, Italy [4]Department of Physics, Section of Astrophysics, Astronomy and Mechanics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece [5]Research Center for Astronomy, Academy of Athens, Soranou Efesiou 4, GR-115 27 Athens, Greece [6]Department of Physics, Section of Astronomy, University of Trieste, Trieste, Italy [7]INAF, Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy [8]INFN, Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy

All measurements of cosmic star formation must assume an initial distribution of stellar masses – the stellar initial mass function – in order to extrapolate from the star-formation rate measured for typically rare, massive stars ($M_\star \geq 8$ M$_\odot$) to the total star-formation rate across the full stellar mass spectrum[1]. The shape of the stellar initial mass function in various galaxy populations underpins our understanding of the formation and evolution of galaxies across cosmic time[2]. Classical determinations of the stellar initial mass function in local galaxies are traditionally made at ultraviolet, optical and near-infrared wavelengths, which cannot be probed in dust-obscured galaxies[2,3], especially in distant starbursts, whose apparent star-formation rates are hundreds to thousands of times higher than in our Milky Way, selected at submillimetre (rest-frame far-infrared) wavelengths[4,5]. The ^{13}C/^{18}O abundance ratio in the cold molecular gas – which can be probed via the rotational transitions of the ^{13}CO and C^{18}O isotopologues – is a very sensitive index of the stellar initial mass function, with its determination immune to the pernicious effects of dust. Here we report observations of ^{13}CO and C^{18}O emission for a sample of four dust-enshrouded starbursts at redshifts of approximately two to three, and find unambiguous evidence for a top-heavy stellar initial mass function in all of them. A low ^{13}CO/C^{18}O ratio for all our targets – alongside a well-tested, detailed chemical evolution model benchmarked on the Milky Way[6] – implies that there are considerably more massive stars in starburst events than in ordinary star-forming spiral galaxies. This can bring these extraordinary starbursts closer to the 'main sequence' of star-forming galaxies[7], though such main-sequence galaxies may not be immune to changes in initial stellar mass function, depending upon their star-formation densities.

Oxygen, carbon and their stable isotopes are produced solely by nucleosynthesis in stars[8]. The minor isotopes, ^{13}C and ^{18}O, are released mainly by low- and intermediate-mass stars (LIMSs; $M_\star < 8$ M$_\odot$) and massive stars ($M_\star > 8$ M$_\odot$), respectively[9], owing to their differing energy barriers in nuclear reactions and evolution of stars[10]. These isotopes then mix with the interstellar medium (ISM) such that the ^{13}C/^{18}O abundance ratio measured in ISM becomes a 'fossil', imprinted by evolutionary history and the stellar initial mass function (IMF)[6]. The abundances of the ^{13}CO and C^{18}O isotopologues in the molecular ISM, whose measurements are immune to the pernicious effects of dust, are therefore a very sensitive index of the IMF in galaxies.

Galaxies in the early Universe, having had much less cosmological time available for prior episodes of evolution, are expected to have simpler star-formation histories than local galaxies. Our sample comprises the strongest carbon monoxide (CO) emitters in the early Universe, selected from the literature (see Methods): four gravitational lensed submillimetre galaxies (SMGs) at $z \sim 2$–3, with look-back times, $\gtrsim 10$ Gyr.

Using the Atacama Large Millimeter Array (ALMA), we have robustly ($> 5\sigma$, where σ is the standard deviation) detected multiple transitions of ^{13}CO and C^{18}O in most of our target galaxies. The $J = 3 \rightarrow 2$ line from SDP.17b and the $J = 5 \rightarrow 4$ line from SPT 0103−45 are marginally detected at $\sim 4\sigma$ levels. But the $J = 4 \rightarrow 3$ transitions of SDP.17b are detected at high signal-to-noise so we can be confident that emission features seen at the expected velocities of the weaker transitions are also real. We also detected ^{12}CO $J = 4 \rightarrow 3$ and $J = 5 \rightarrow 4$ for SPT 0125−47 and SPT 0103−45, respectively.

As shown in Fig. 1, there is a decreasing trend in the ratio of velocity-integrated line intensities, $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$, with increasing infrared luminosity, $L_{\rm IR}$ (or, the apparent star-formation rate, SFR, traced by massive stars). For all the galaxies in our observed sample (see Methods), the line ratios of $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ are close to unity, similar to those found[11,12] in three nearby ultraluminous infrared galaxies (ULIRGs; $L_{\rm IR} > 10^{12}$ L$_\odot$) – Arp 220, Mrk 231 and IRAS 13120−5453, as well as in the strongly lensed SMG, SMM J2135−0102 at $z \approx 2.3$[13]. Galactic disks of nearby spiral galaxies have $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ ratios similar to the representative ratio[8,14] of our Milky Way's disk, $\sim$ 7–10. In the central nuclear regions of these spiral galaxies, where the star-formation activity is more intense than in the disks, $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ ratios are lower[15], though remain restricted to $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O}) \geq 4$. The Magellanic clouds – our nearest dwarf galaxies – show the highest $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ ratios, $\gtrsim$ 30–60.

For representative Galactic abundance ratios of ^{13}CO/C^{18}O $\sim$ 7–10, the $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ line intensity ratio can approach values near unity – which is what we measure for all the galaxies in our sample – only if even the rarest of our three isotopologue lines, C^{18}O, were to acquire substantial optical depths on galactic scales (see Methods). On the other hand, to reach line ratios, $I(^{12}\mathrm{CO})/I(^{13}\mathrm{CO})$ and $I(^{12}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ in excess of 30 – as found in our sample – the optical depth of ^{13}CO and C^{18}O have to be $\ll 1$ for either type of conditions, namely local thermodynamic equilibrium (LTE) or non-LTE excitation (see Methods), assuming the typical abundance ratios of ^{12}CO/^{13}CO $\sim$ 40–100 found in the Milky Way[6,8].

The magnification factors of gravitational lensing in our objects are modest ($\mu \sim 5$, see Extended Data Table 1), with the notable exception of the Cloverleaf ($\mu \sim 10$). It is unlikely that differential lensing could skew the measured $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ line ratio away from the value intrinsic to the galaxy, even in cases of much stronger lensing[13].

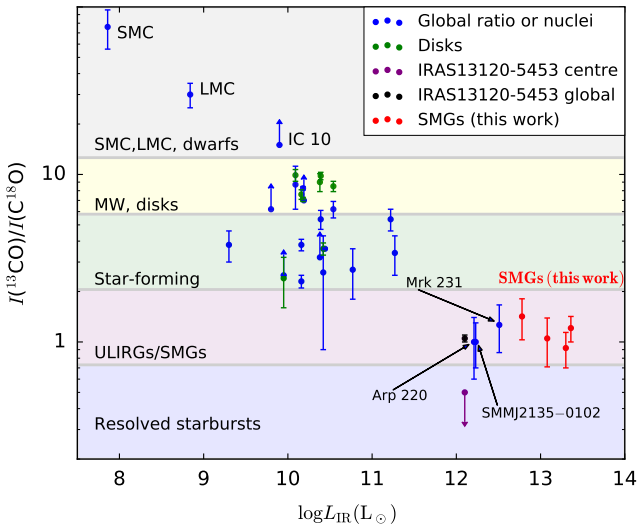


Figure 1 | $I(^{13}CO)/I(C^{18}O)$ as a function of IR luminosity (L_{IR}, rest-frame 8–1000 μm, corrected for gravitational amplification when appropriate). Red symbols refer to submillimetre galaxies (SMGs) in our sample. We include $I(^{13}CO)/I(C^{18}O)$ measurements of nearby star-forming spiral galaxies, disks, a few resolved nuclei [15], three local ULIRGs [11], Arp 220, Mrk 231, and IRAS 13120−5453, and a SMG, SMM J2135−0102 [13] at $z \approx 2.3$. The ratios of the Small and Large Magellanic Clouds (SMC and LMC) are averaged from multiple positions [16,17]. $I(^{13}CO)/I(C^{18}O)$ in a dwarf galaxy IC 10 is reported as a lower limit [18]. A decreasing trend of $I(^{13}CO)/I(C^{18}O)$ as a function of L_{IR} is clearly evident, indicating that $^{13}CO/C^{18}O$ abundance ratios are varying systematically in galaxies with different rates of apparent star formation. Error bars stand for an one σ uncertainty.

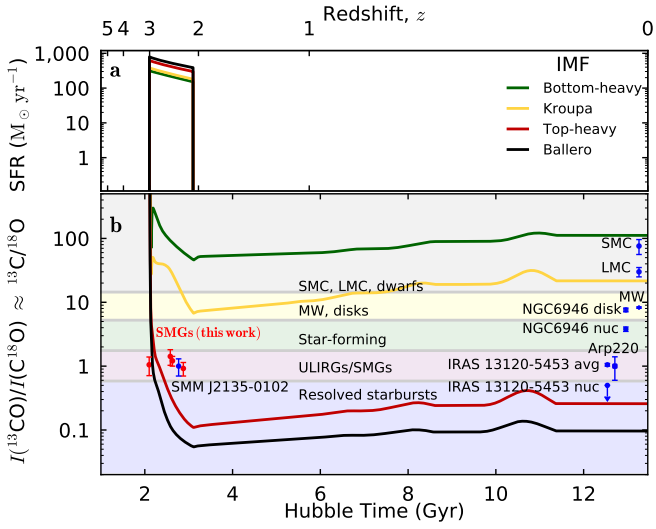


Figure 2 | Theoretical ^{13}C and ^{18}O isotopic abundance ratios in the ISM for different evolutionary tracks, predicted using various IMFs. **a:** Star-formation history for a delayed starburst, starting two Gyr after the Big Bang, with a total cessation of subsequent star formation. Coloured lines correspond to different IMFs. **b:** Theoretical $^{13}C/^{18}O$ abundance ratio in the ISM as a function of time, following the different IMFs (shown in colour). Red symbols refer to SMGs in our sample. Blue symbols show the $^{13}CO/C^{18}O$ ratios measured in SMM J2135−0102 and a few representative local galaxies. Table 1 lists the detailed definitions of the IMFs adopted here.

For differential lensing to operate in this way, the global $I(^{13}CO)$ and $I(C^{18}O)$ distributions over the galaxies must be very different – which is improbable given that these two isotopologue lines have almost identical excitation requirements and any differences in their distribution are expected to be confined within individual molecular clouds (see Methods). Finally, the isotopologue lines have been observed simultaneously, making the uncertainties from pointing and calibration negligible. Furthermore, it was recently shown that known photo-chemical effects, such as selective photodissociation and fractionation, cannot induce global isotopologue abundances to differ from the intrinsic, IMF-determined, isotopic abundances in star-forming galaxies [6,12].

We thus conclude that emission in both ^{13}CO and $C^{18}O$ is optically thin for the bulk of the molecular gas mass in these galaxies. The systematically low $I(^{13}CO)/I(C^{18}O)$ ratios found in all our high-redshift starbursts – as well as in local ULIRGs – reflect intrinsic isotopologue abundance ratios over galaxy-sized molecular hydrogen reservoirs, i.e. $I(^{13}CO)/I(C^{18}O) \approx {}^{13}CO/C^{18}O \approx {}^{13}C/^{18}O$. Fig. 1 thus reflects a strong decrease of the $^{13}C/^{18}O$ abundance ratio in starburst galaxies, compared to local spiral galaxies, galactic disks and dwarf galaxies.

The only plausible basis for such systematic variations of isotopologue abundance ratios over galaxy-sized molecular hydrogen reservoirs is a change of the stellar IMF, which must cause the intrinsic abundances of isotopic elements to deviate significantly from those found in ordinary star-forming systems where the standard IMF prevails. The robustness of this conclusion is made possible by tremendous advances in chemical evolution modeling, which now takes into account nucleosynthesis, isotopic yields across stellar mass, time differentials for their release into the ISM, and the dependence of stars' initial metallicity on prior galactic evolution. Benchmarked against the rich isotopic datasets of the Milky Way, these models can now follow the chemical evolution of various isotopes and their abundance ratios, uniquely identifying the effects of different IMFs upon them [6] (see Methods).

In Fig. 2, we present chemical evolution models that show how the isotopologue abundance ratios are altered by IMF types, and how they evolve as a function of cosmic time. The models show a massive galaxy that started an intense burst of star formation at $z \sim 3$, reached a stellar mass of 10^{11} $M_\odot$ one Gyr later, ceased forming stars after the burst, then evolved passively to the present day. This represents an extreme case for the evolution of the $^{13}C/^{18}O$ abundance ratio for a pure starburst. The $^{13}C/^{18}O$ abundance ratio starts from a high value set by the first generations of metal-poor massive stars, because ^{13}C is released from its primary and secondary nucleosynthesis channels earlier than ^{18}O, which is purely a secondary element (see Methods). The ratio drops quickly during the starburst, then slowly increases with time, varying by a factor of 2–3× depending on the adopted IMF and the time interval. The late increase of the ratio is due mostly to the slow but continuous release of ^{13}C from LIMS (see Methods), which keep releasing ^{13}C for a long time after the star formation – and, at the same time, the ^{18}O pollution from massive stars – has ceased; wiggles in the ratio correspond to the lifetimes of roughly Solar-mass stars.

It is not possible to reproduce the observed $I(^{13}CO)/I(C^{18}O)$ ratio in SMGs with a Kroupa IMF (or similar IMFs; see Table 1), i.e. with an IMF that can reproduce the ratios found in the Milky Way and in the disks of local spiral galaxies [3]. The top-heavy IMF and the Ballero [19] IMF (which can reproduce the chemical abundances of stars in the Galactic bulge) under-produce the $I(^{13}CO)/I(C^{18}O)$ ratios observed in $z \sim 2$–3 starburst galaxies and local ULIRGs, but they can reproduce the extremely low $I(^{13}CO)/I(C^{18}O)$ ratio recently measured [12] in the central 500-pc region of the starburst ULIRG, IRAS 13120−5453, which has been until now the lowest value reported in literature. Note that the average star-formation event may have a less top-heavy IMF over galactic scales, or a mix of both top-heavy and canonical IMFs that produces a galaxy-sized average $^{13}C/^{18}O$ ratio ≤ 1, which also applies for the resolved studies of IRAS 13120−5453.

A clear trend is shown in Fig. 2: the more top-heavy the IMF, the lower the $I(^{13}CO)/I(C^{18}O)$ ratio, which is also compatible with the ratios found in local ULIRGs and the exceptionally small $I(^{13}CO)/I(C^{18}O)$ ratio found in the centre of IRAS 13120−5453 [12]. This paints a consistent picture in which a top-heavy IMF operates within both local ULIRGs and the much more numerous, distant starburst galaxies, where starburst events can quickly enrich the ^{18}O abundance, pushing the $^{13}C/^{18}O$ ratio to (or below) unity. A canonical IMF can never produce a $I(^{13}CO)/I(C^{18}O)$ ratio close to unity, no matter what type of star-formation history or at what time along a galaxy's evolutionary track the measurement is made.

Table 1 | Details of the IMFs used in this work.

IMF Name	$\alpha 0$	$\alpha 1$	$\alpha 2$	$m0$ $M_\odot$	$m1$ $M_\odot$	$m2$ $M_\odot$	$m3$ $M_\odot$	$M_\star^{(8-100)M_\odot}/M_\star^{\rm total}$ (%)
Bottom heavy	−1.7	−1.7	−1.7	0.1	0.5	1.0	100	3.9
Kroupa [3,6]	−0.3	−1.2	−1.7	0.1	0.5	1.0	100	6.9
Top heavy	−0.3	−1.1	−1.1	0.1	0.5	1.0	100	33.3
Ballero [19]	−0.3	−0.95	−0.95	0.1	0.5	1.0	100	44.0

The slopes quoted in the table are for IMFs in mass, where $m0$ (= 0.1 $M_\odot$) and $m3$ (= 100 $M_\odot$) are, respectively, the lower and upper limits of stellar masses assumed in the models, i.e. the IMF is normalised to unity in the 0.1–100 $M_\odot$ range. $m1$ and $m2$ indicate the masses at which there is a change in the IMF slope, if any. For instance, for the Kroupa IMF, the slope changes at both $m1$ and $m2$; for the Ballero and top-heavy IMFs, the slope changes only at $m1$; finally, the bottom-heavy IMF has a single slope. The Kroupa IMF slopes are adopted for reproducing typical Milky Way values in chemical evolution models [6], which are within the error bars of the original reported values [3].

Multiple evidence in the local Universe has shown that the stellar IMF in galaxies with very high SFR densities are likely biased to massive stars, such as ultra-compact dwarf galaxies [20], ULIRGs [21], and progenitors of early-type galaxies [22]. A top-heavy stellar IMF was recently also found in compact stellar associations in the LMC [23,24,25], whose high-density star-formation events may closely replicate what happens over galactic scales in distant starbursts. Our results – for the most intensive star-forming systems in the distant Universe, where classical UV and optical methods cannot be applied – are in line with these findings. We also note that metal-poor dwarf galaxies likely have an IMF biased towards low-mass stars, which is predicted by the integrated galaxy-wide IMF theory and is consistent with the results found in dwarf galaxies [26] and the outer regions of disk galaxies, using Hα and UV observations [27].

An IMF biased to massive stars implies that SFRs determined for SMGs must be reduced considerably, since they are based on extrapolations of observables related to massive stars [1]. Moving from the Kroupa IMF to the Ballero IMF, the relative mass fraction of massive stars increases by a factor of 6–7× (see Table 1), meaning that SFRs derived from most classical tracers [1] (e.g. $L_{\rm IR}$, radio continuum, etc.) must decrease by a similar factor. As a result, dusty starburst galaxies probably lie much closer to the so-called 'main sequence' of star-forming galaxies [7] than previously thought. Classical ideas about the evolutionary tracks of galaxies [28] and our understanding of cosmic star-formation history [29] are challenged. Fundamental parameters governing galaxy formation and evolution – SFRs, stellar masses, gas-depletion and dust-formation timescales, dust extinction laws, and more [30] – must be re-addressed, exploiting recent advances in stellar physics.

1. Kennicutt, R. C., Jr. Star Formation in Galaxies Along the Hubble Sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).
2. Bastian, N., Covey, K. R. & Meyer, M. R. A Universal Stellar Initial Mass Function? A Critical Look at Variations. *Annu. Rev. Astron. Astrophys.* **48**, 339–389 (2010).
3. Kroupa, P. *et al.* *The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations,* 115 (2013).
4. Smail, I., Ivison, R. J. & Blain, A. W. A Deep Sub-millimeter Survey of Lensing Clusters: A New Window on Galaxy Formation and Evolution. *Astrophys. J. Let.* **490**, L5–L8 (1997).
5. Hughes, D. H. *et al.* High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey. *Nature* **394**, 241–247 (1998).
6. Romano, D., Matteucci, F., Zhang, Z.-Y., Papadopoulos, P. P. & Ivison, R. J. The evolution of CNO isotopes: a new window on cosmic star formation history and the stellar IMF in the age of ALMA. *Mon. Not. R. Astron. Soc.* **470**, 401–415 (2017).
7. Noeske, K. G. *et al.* Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. *Astrophys. J. Let.* **660**, L43–L46 (2007).
8. Wilson, T. L. & Matteucci, F. Abundances in the interstellar medium. *A&ARv* **4**, 1–33 (1992).
9. Romano, D., Karakas, A. I., Tosi, M. & Matteucci, F. Quantifying the uncertainties of chemical evolution studies. II. Stellar yields. *Astron. Astrophys.* **522**, A32 (2010).
10. Pagel, B. E. J. *Nucleosynthesis and Chemical Evolution of Galaxies* (2009).
11. Henkel, C. *et al.* Carbon and oxygen isotope ratios in starburst galaxies: New data from NGC 253 and Mrk 231 and their implications. *Astron. Astrophys.* **565**, A3 (2014).
12. Sliwa, K., Wilson, C. D., Aalto, S. & Privon, G. C. Extreme CO Isotopic Abundances in the ULIRG IRAS 13120-5453: An Extremely Young Starburst or Top-heavy Initial Mass Function. *Astrophys. J. Let.* **840**, L11 (2017).
13. Danielson, A. L. R. *et al.* ^{13}CO and $C^{18}O$ emission from a dense gas disc at z = 2.3: abundance variations, cosmic rays and the initial conditions for star formation. *Mon. Not. R. Astron. Soc.* **436**, 2793–2809 (2013).
14. Barnes, P. J. *et al.* The Three-mm Ultimate Mopra Milky Way Survey. I. Survey Overview, Initial Data Releases, and First Results. *Astrophys. J.* **812**, 6 (2015).
15. Jiménez-Donaire, M. J. *et al.* $^{13}CO/C^{18}O$ Gradients across the Disks of Nearby Spiral Galaxies. *Astrophys. J. Let.* **836**, L29 (2017).
16. Heikkila, A., Johansson, L. E. B. & Olofsson, H. The C^18O/C^17O ratio in the Large Magellanic Cloud. *Astron. Astrophys.* **332**, 493–502 (1998).
17. Muraoka, K. *et al.* ALMA Observations of N83C in the Early Stage of Star Formation in the Small Magellanic Cloud. *Astrophys. J.* **844**, 98 (2017).

18. Nishimura, Y. *et al.* Spectral Line Survey toward a Molecular Cloud in IC10. *Astrophys. J.* **829**, 94 (2016).

19. Ballero, S. K., Matteucci, F., Origlia, L. & Rich, R. M. Formation and evolution of the Galactic bulge: constraints from stellar abundances. *Astron. Astrophys.* **467**, 123–136 (2007).

20. Dabringhausen, J., Kroupa, P. & Baumgardt, H. A top-heavy stellar initial mass function in starbursts as an explanation for the high mass-to-light ratios of ultra-compact dwarf galaxies. *Mon. Not. R. Astron. Soc.* **394**, 1529–1543 (2009).

21. Dabringhausen, J., Kroupa, P., Pflamm-Altenburg, J. & Mieske, S. Low-mass X-Ray Binaries Indicate a Top-heavy Stellar Initial Mass Function in Ultracompact Dwarf Galaxies. *Astrophys. J.* **747**, 72 (2012).

22. Peacock, M. B. *et al.* Further Constraints on Variations in the Initial Mass Function from Low-mass X-ray Binary Populations. *Astrophys. J.* **841**, 28 (2017).

23. Schneider, F. R. N. *et al.* An excess of massive stars in the local 30 Doradus starburst. *Science* **359**, 69–71 (2018).

24. Banerjee, S. & Kroupa, P. On the true shape of the upper end of the stellar initial mass function. The case of R136. *Astron. Astrophys.* **547**, A23 (2012).

25. Kalari, V. M., Carraro, G., Evans, C. J. & Rubio, M. The Magellanic Bridge cluster NGC 796: Deep optical AO imaging reveals the stellar content and initial mass function of a massive open cluster. *ArXiv e-prints* (2018).

26. Lee, J. C. *et al.* Comparison of Hα and UV Star Formation Rates in the Local Volume: Systematic Discrepancies for Dwarf Galaxies. *Astrophys. J.* **706**, 599–613 (2009).

27. Pflamm-Altenburg, J. & Kroupa, P. Clustered star formation as a natural explanation for the Hα cut-off in disk galaxies. *Nature* **455**, 641–643 (2008).

28. Speagle, J. S., Steinhardt, C. L., Capak, P. L. & Silverman, J. D. A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ˜ 0-6. *Astrophys. J. Supp.* **214**, 15 (2014).

29. Madau, P. *et al.* High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z˜4. *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).

30. Pflamm-Altenburg, J. & Kroupa, P. The Fundamental Gas Depletion and Stellar-Mass Buildup Times of Star-Forming Galaxies. *Astrophys. J.* **706**, 516–524 (2009).

METHODS

1 Sample

Our sample comprises the strongest CO emitters in the early Universe, taken from the literature[31,32,33]: four strongly-lensed submillimetre galaxies (SMGs) at $z \sim 2$–3, with look-back times, $\gtrsim 10$ Gyr. Two of these galaxies, SPT-S J010312−4538.8 ($z = 3.09$, also known as SPT 0103−45) and SPT-S J012506−4723.7 ($z = 2.51$, also known as SPT 0125−47), were selected[32] using the South Pole Telescope at $\lambda = 1.4$ and 2 mm; another, HATLAS J090302.9−014127 ($z = 2.31$, also known as SDP.17b), was discovered using the *Herschel Space Observatory*[34] at far-IR wavelengths; the last, H1413+117 ($z = 2.56$, the 'Cloverleaf' quasar), was discovered as a result of its rare quadruple-spot optical morphology, and was later found to be bright in CO and in dust continuum[31,35]. We list the basic characteristics of the sample in Extended Data Table 1.

2 Observations and data reduction

We have performed simultaneous observations of ^{13}CO and C^{18}O using ALMA in its relatively compact array configurations (C36-1 and C36-2), with two 2-GHz-wide spectral windows (SPWs) in bands 3 and 4. We used the remaining two SPWs to cover continuum emission. Between 10 and 30 min were spent on target for each transition. For SDP.17b, we observed both $J = 3 \rightarrow 2$ and $J = 4 \rightarrow 3$ transitions of ^{13}CO and C^{18}O, in order to have a redundant measurement for their line ratios as well as constraints on the relative excitation in these rare isotopic lines. We also observed ^{12}CO $J = 5 \rightarrow 4$ for SPT 0103−45 and ^{12}CO $J = 4 \rightarrow 3$ for SPT 0125−47, with similar configurations. Calibrators, integration time, atmospheric conditions are listed in Extended Data Table 2.

All the data were calibrated manually using CASA v4.7.1[36], using standard procedures. We subtracted the continuum using the CASA task, UVCONTSUB, by fitting a linear slope to the line-free channels. We cleaned the visibility data with a channel width of ~ 20–30 km s^{-1}, using a Briggs weighting with ROBUST = 1.5 to optimise sensitivity. We applied a primary beam correction to all the cleaned data. Our target galaxies are mostly unresolved, or only marginally resolved. We assume that line widths of ^{13}CO and C^{18}O are the same as those of ^{12}CO lines, to minimise uncertainties in the line flux fitting. Extended Data Figs 1, 2 and 3 present the velocity-integrated flux (moment-0) maps of ^{13}CO and C^{18}O, overlaid with contours of high-resolution submm continuum. In Extended Data Fig. 4 we present the spectra. SPT 0103−45 has two velocity components that cover a very large velocity span. The overall line profile of ^{13}CO is consistent with ^{12}CO, but limited by the noise level. We adopt only the narrow (stronger) component, seen for the yellow shadow region in the ^{12}CO $J = 5 \rightarrow 4$ spectrum, to avoid confusion from the broad (weaker) component (see Extended Data Fig. 4). Our synthesised beamsizes are mostly larger than, or at least comparable to, the apparent sizes revealed by the high-resolution submm continuum images, so any missing flux is expected to be negligible. We extract spectra of ^{13}CO and C^{18}O using circular apertures ≈ 4–$6''$ in diameter, as shown in Extended Data Figs 1, 2 and 3.

To measure velocity-integrated line fluxes and the associated errors, we performed three independent methods: we first fit one-dimensional Gaussian profiles to the extracted spectra with a fixed linewidth from ^{12}CO, and fixed the frequency interval between ^{13}CO and C^{18}O, since we can assume confidently that these two lines are emitted from the same excitation component, such that their line centres do not shift relative to one another. Second, we made moment-0 maps and fitted two-dimensional Gaussian profiles, as an independent check of line flux. Third, to better estimate the noise level, we also calculated the theoretical noise using the ALMA sensitivity calculator[37], given the integration time, precipitable water vapour, linewidth, and array configuration. The measured line fluxes and properties are listed in Extended Data Table 3. To be conservative, in Figs 1 and 2 we adopt the largest error among the results from the three methods in the analysis.

3 Line ratios and optical depths

3.1 Conditions of local thermodynamic equilibrium (LTE) We first analyse the observed line ratios of ^{13}CO to C^{18}O, to constrain the molecular line optical depths, assuming local thermodynamic equilibrium (LTE; $T_{\rm ex}^{\rm line} = T_{\rm kin}^{\rm line}$). The line brightness temperature ratios of ^{12}CO to ^{13}CO and ^{13}CO to C^{18}O can be expressed as:

$$\frac{T_{\rm b}^{^{12}{\rm CO}}}{T_{\rm b}^{^{13}{\rm CO}}} = \frac{J_\nu(T_{\rm ex}^{^{12}{\rm CO}}) - J_\nu(T_{\rm bg}^{^{12}{\rm CO}})}{J_\nu(T_{\rm ex}^{^{13}{\rm CO}}) - J_\nu(T_{\rm bg}^{^{13}{\rm CO}})} \cdot \frac{1-\exp(-\tau^{^{12}{\rm CO}})}{1-\exp(-\tau^{^{13}{\rm CO}})} \quad (1)$$

and

$$\frac{T_{\rm b}^{^{13}{\rm CO}}}{T_{\rm b}^{{\rm C}^{18}{\rm O}}} = \frac{J_\nu(T_{\rm ex}^{^{13}{\rm CO}}) - J_\nu(T_{\rm bg}^{^{13}{\rm CO}})}{J_\nu(T_{\rm ex}^{{\rm C}^{18}{\rm O}}) - J_\nu(T_{\rm bg}^{{\rm C}^{18}{\rm O}})} \cdot \frac{1-\exp(-\tau^{^{13}{\rm CO}})}{1-\exp(-\tau^{{\rm C}^{18}{\rm O}})} \quad (2)$$

where $T_{\rm b}$ is the brightness temperature of the molecular transition, $T_{\rm ex}$ is the excitation temperature, J is the rotational quantum number, and $T_{\rm bg}$ is the radiation temperature of the background emission field, which is dominated by the cosmic microwave background (CMB) at $T_{\rm CMB} \sim 2.73 \times (1+z)$ K, where z is the redshift. $\tau^{\rm line}$ is the optical depth of the given transition. $J(T) = (h\nu/k_{\rm B})/[\exp(h\nu/k_{\rm B}T) - 1]$ is the Planck radiation temperature at the rest frequency of the line emission, $\nu^{\rm line}$. $k_{\rm B}$ is the Boltzmann constant, h is the Planck constant, and T is the considered temperature. For optically thick lines (e.g. ^{12}CO for most conditions), $1 - \exp(-\tau^{\rm line}) \sim 1$; for optically thin lines (e.g. ^{13}CO and C^{18}O), $1 - \exp(-\tau^{\rm line}) \sim \tau^{\rm line}$.

In Extended Data Fig. 5 we present the $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ and $I(^{12}{\rm CO})/I(^{13}{\rm CO})$ velocity-integrated line intensity ratios as a function of the optical depths of ^{13}CO and C^{18}O, under LTE conditions. We also calculated the corresponding H_2 column densities, assuming Galactic abundances[38,81]. A representative Galactic abundance ratio of ^{12}CO/^{13}CO = 70 is assumed, the optical depth for ^{13}CO needs to be < 0.03, which is around 10× lower than the Galactic average values[14] for producing the observed high $I(^{12}{\rm CO})/I(^{13}{\rm CO})$ ratios, $\gtrsim 30$. The corresponding optical depth of ^{12}CO is ~ 2, much lower than the typical values for ^{12}CO $J = 1 \rightarrow 0$ found in typical Galactic molecular clouds[14], but consistent with a moderate optical depth of ^{12}CO found in local starburst galaxies[39].

Only when the optical depth of C^{18}O $\gg 1$ (corresponding to $\tau^{^{13}{\rm CO}} \gg 7$, which leads to an H_2 column density of $\gg 10^{25}$ cm^{-2}), does the $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ line ratio approach unity. In this case the line ratios of $I(^{12}{\rm CO})/I(^{13}{\rm CO})$ and $I(^{12}{\rm CO})/I({\rm C}^{18}{\rm O})$ would also move towards unity, in conflict with our observed ratios. Even for moderate $\tau^{^{13}{\rm CO}} \sim 0.2$–0.5, the line ratio of $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ stays at ~ 6–7, not strongly biased by $\tau^{^{13}{\rm CO}}$.

3.2 Non-LTE conditions In Extended Data Fig. 6, we present non-LTE models derived with a non-LTE radiative transfer code, RADEX[40], showing the optical depth, $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ and $I(^{12}{\rm CO})/I(^{13}{\rm CO})$ line ratios as a function of ^{13}CO column density and $N_{\rm H_2}$ column density, $N_{\rm H_2}$. We calculate different H_2 volume densities, $n_{\rm H_2}$, of 10^3 cm^{-3}, 10^4 cm^{-3} and 10^5 cm^{-3}, covering the most common $n_{\rm H_2}$ range – typical values from normal molecular clouds to dense cores. We assume that the same abundance ratios as we assumed for the LTE conditions, i.e. ^{12}CO/^{13}CO = 70, and ^{13}CO/C^{18}O = 7, which are representative values of the Milky Way disk. The velocity width (full-width at half-maximum, FWHM) is set to 300 km s^{-1}, as the typical (indeed, at the lower end) linewidth found in ULIRGs and SMGs[41]. In the panels **b** and **c** of Extended Data Fig. 6 we overlay the LTE results, for comparison.

For all models, we set the kinetic temperature, $T_{\rm kin}$, to be 30 K, which is a typical dust temperature for the SMG population[42], and is also the lower limit of the kinetic temperate of the H_2 gas, as the minimum temperature powered by the cosmic-ray heating for such starburst conditions[43]. Higher $T_{\rm kin}$ would bring the CO energy population towards higher-J transitions, making optical depths even smaller.

Extended Data Fig. 6 shows that, for $n_{\rm H_2} = 10^3$ cm^{-3}, which is a typical value for normal Galactic molecular cloud conditions, only when the H_2 column density $N_{\rm H_2} \gg 10^{26}$ cm^{-2}, the line ratio of $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ can approach unity (< 1.5, considering uncertainties in line ratios). The required high column densities are a few orders of magnitude higher than the typical values measured in SMGs: $\sim 10^{23}$–10^{24} cm^{-2}, obtained using X-rays[44], CO LVG modeling[45], and dust[42]. This is especially supported by the Cloverleaf quasar, whose X-ray emission has been clearly detected[46], given the Compton limit of $\sim 10^{24}$ cm^{-2}.

On the other hand, the high-density results, i.e. $n_{\rm H_2} = 10^4$ and 10^5 cm^{-3}, are very similar to those under LTE conditions. For all conditions, a moderate ^{13}CO optical depth does not strongly vary the $I(^{13}{\rm CO})/I({\rm C}^{18}{\rm O})$ ratio

Extended Data Table. 1 | Target properties

Short name	IAU name	R.A. J2000	Dec. J2000	Redshift z	Lensing amplification, μ	$L_{\rm IR}/\mu$ 10^{13} L$_\odot$	$M_\star/\mu$ 10^{10} M$_\odot$
SPT 0103−45	SPT-S J010312−4538.8	01:03:11.50	−45:38:53.9	3.0917	5.3 ± 0.11	1.2	$5.5^{+6.1}_{-2.9}$
SPT 0125−47	SPT-S J012506−4723.7	01:25:07.08	−47:23:56.0	2.5148	5.5 ± 0.1★	2.2	–
SDP.17b	HATLAS J090302.9−014127	09:03:03.02	−01:41:26.9	2.3051	$3.56^{+0.19}_{-0.17}$	2.0	$24.2^{+8.6}_{-4.0}$
Cloverleaf	H1413+117	14:15:46.23	+11:29:44.0	2.5585	11†	6.0	–

Data for SDP.17b is from ref. [73,74].
★ We adopt an amplification factor, μ, derived from lens modelling using 850-μm ALMA visibility data in the literature [71,72].
† We adopt the best-determined amplification factor, μ, reported in the literature [35,75], whose lensing model is derived from CO $J = 7 \rightarrow 6$ line emission, and could better reproduce a single source on the lens plane [75]. Unfortunately the uncertainty of this amplification factor was not reported, but the uncertainty in μ does not jeopardise our conclusions.

from the abundance ratio. So, it is highly unlikely that the unity value of the $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ ratio can be due to high optical depths.

4 Possible HNCO contamination of C^{18}O lines

HNCO $5_{05} \rightarrow 4_{04}$ ($J = 5 \rightarrow 4$) has a rest frequency of 109.9058 GHz, close to the rest frequency of C^{18}O $J = 1 \rightarrow 0$ (109.7822 GHz), with a velocity offset of $\sim$ 370 km s^{-1}. When the linewidths are broad, these two lines are sometimes blended, which leads to a possible contamination of the C^{18}O measurements [12,47,48]. In these observations it has been found that HNCO $J = 5 \rightarrow 4$ could contribute up to $\sim$ 30% of the total flux (C^{18}O + HNCO), for the most extreme cases in the local Universe, e.g. Arp 220 [48] and IRAS 13120−543 [12].

HNCO $J = 5 \rightarrow 4$ has a critical density, $n_{\rm crit} \sim 10^6$ cm^{-3} and is regarded as a dense-gas tracer that could be excited in slow-shock regions [49,50]. The Einstein A coefficient increases as $A \propto (J+1)^3$, thus $n_{\rm crit}$ increases quickly for high-J transitions, such as HNCO $J = 15 \rightarrow 14$ ($\nu_{\rm rest}$ = 329.66 GHz) and $J = 20 \rightarrow 19$ ($\nu_{\rm rest}$ = 439.62 GHz), making much less contribution to the C^{18}O $J = 3 \rightarrow 2$ and $J = 4 \rightarrow 3$ transitions involved in our study.

To better estimate how much HNCO lines may contaminate the C^{18}O lines, in Extended Data Fig. 7 we show the theoretical line ratios between HNCO and C^{18}O and high-J $I(\mathrm{HNCO})/I(\mathrm{C}^{18}\mathrm{O})$ ratios normalised with $I(\mathrm{HNCO}\ J = 5 \rightarrow 4)/I(\mathrm{C}^{18}\mathrm{O}\ J = 1 \rightarrow 0)$, using RADEX [40]. We assume the same abundances measured in Arp 220 [47], and use molecular data from the Leiden Atomic and Molecular Database [51] (LAMBDA). We assume $T_{\rm kin}$ = 30 K as the representative temperature of the H_2 gas.

Extended Data Fig. 7**a** shows that the $I(\mathrm{HNCO})/I(\mathrm{C}^{18}\mathrm{O})$ ratio increases with the volume density of H_2 gas, $n_{\rm H_2}$. Moreover, the ratio decreases quickly with J transitions, meaning that the contamination from HNCO to C^{18}O is much less severe for the high-J transitions, compared to the C^{18}O $J = 1 \rightarrow 0$ line. Extended Data Fig. 7**b** shows $I(\mathrm{HNCO})/I(\mathrm{C}^{18}\mathrm{O})$ line brightness ratios normalised by $I(\mathrm{HNCO}\ J = 5 \rightarrow 4)/I(\mathrm{C}^{18}\mathrm{O}\ J = 1 \rightarrow 0)$. With a weak dependency on $n_{\rm H_2}$, the ratios of $I(\mathrm{HNCO}\ J = 10 \rightarrow 9)/I(\mathrm{HNCO}\ J = 5 \rightarrow 4)$ and $I(\mathrm{HNCO}\ J = 15 \rightarrow 14)/I(\mathrm{HNCO}\ J = 5 \rightarrow 4)$ are one order of magnitude lower than unity. Unfortunately the LAMBDA database does not have the data for the transition of HNCO $J = 20 \rightarrow 19$, whose HNCO/C^{18}O ratio is expected to be even lower.

Even if we take the highest HNCO $J = 5 \rightarrow 4$ contamination found in local galaxies, i.e. 30%, the corresponding contamination for the high-J C^{18}O transitions will be at most 3%, which can be regarded as negligible for the lines in our study.

5 Chemical evolution model

We adopt a single-zone chemical evolution model for our analysis, originally developed to describe the evolution of the Milky Way [52], then further extended to other galaxies [53]. The model computes the evolution of abundances of multiple elements, including ^{12}C, ^{16}O, ^{13}C and ^{18}O in the ISM of galaxies. We use detailed numerical models to solve the classical set of equations of chemical evolution [10,52,53,54,55,56], with the following assumptions:

- Gas inflow with primordial chemical composition provides raw material for star formation. The gas is accreted at an exponentially fading rate and the timescale of the process is a free parameter of the model;
- Galactic outflows remove both the stellar ejecta and a fraction of the ambient ISM;
- Star formation follows the canonical Kennicutt-Schmidt law [57]; the masses of the newly-formed stars follow the input IMF;
- Finite stellar lifetimes for different stars need to be considered (i.e. no instantaneous recycling approximation (non-IRA) is adopted) [59];
- Stars release the elements they have synthesised during their lifetime, as well as those already present when they were born that are left unaltered by the nucleosynthesis processes, when they die;
- Stellar ejecta are mixed with the ISM homogeneously.

The adopted yields account for the dependence of several stellar processes on the initial metallicity of the stars, and have been calibrated with the best fit using the Milky Way data, which are relevant to a range of metallicity and evolution timescales [6]. The time-delay effect is considered in the chemical evolution, namely the differences between the lifetimes of massive stars and low-mass stars [58]. We used an analytical formula for the stellar lifetimes that linearly interpolate stellar lifetime tables [59]. The time lag in producing and releasing primary (those synthesised directly from H and He; i.e. ^{12}C, ^{16}O) and secondary elements (those derived from metals already present in the star at birth; i.e. ^{13}C, ^{18}O – but note that a fraction of ^{13}C is also synthesised as a primary element) is also considered [52]. These two effects, corroborated by the star-formation history and the adopted IMF, determine the amount of different isotopes released to the ISM on different timescales. In particular, different isotopes are released as a function of time, e.g. the bulk of ^{13}C is released later than ^{12}C, and the bulk of ^{18}O is released later than ^{16}O [6,8,60]. Chemical evolution models can now follow the evolution of various isotopic ratios, tracing abundance ratios not only between the isotopes of each element [6], but also between different elements.

The most important aspect regarding this work is that such models can now compare the effects of a young starburst (with a regular stellar IMF, e.g. the Kroupa IMF) against those due to different stellar IMFs via carefully chosen isotope (and thus isotopologue) ratios. This critical advance was made by no longer assuming instantaneous element enrichment of the ISM by the stars, but incorporating the different timescales of their release into the ISM. It should be noted that these timescales, and the relative delays between the release of various isotopes into the ISM, are set by stellar physics, i.e. they are not free parameters. With the Kroupa IMF, only an unphysical combination of star-formation history, i.e. $\tau \lesssim 10$ Myr with an SFR $\gtrsim 20\,000$ M$_\odot$ yr^{-1} could approach the observed ^{13}C/^{18}O ratios of near unity.

6 Origins of carbon and oxygen isotopes

The $I(^{12}\mathrm{CO})/I(^{13}\mathrm{CO})$ line ratios have been found to vary systematically in galaxies with different SFRs and Hubble types [39,61]. Owning to the differences in the origins of ^{12}C and ^{13}C, it has been proposed that the

$I(^{12}\mathrm{CO})/I(^{13}\mathrm{CO})$ line ratio can be used to derive their abundance ratio, which can further probe the stellar IMF, or different star-formation modes [12,43,60,62].

The ^{12}C element is primarily produced by helium burning (the classical triple-α process), and multiple channels can produce ^{12}C in nucleosynthesis [63]. In the Milky Way, ^{12}C is primarily produced by LIMSs, revealed by data for stars in the Solar vicinity: in fact, the [C/Fe]/[Fe/H] ratio is almost a constant, with [C/Fe] being Solar, indicating that C and Fe are produced in the same proportions by the same stars [9,64]. If mass loss from massive stars is considered, ^{12}C released from massive stars still accounts for $< 50\%$ of the total [65,66]. On the other hand, ^{13}C is released from LIMSs largely as a secondary element, because ^{13}C production needs a pre-existing seed, namely, the primary element, ^{12}C [63,67]. ^{13}C also has a primary component in nucleosynthesis but it can only occur in red AGB stars, where periodic dredge-up episodes convect ^{12}C to the stellar surface and form ^{13}C.

Since both ^{12}C and ^{13}C are mostly produced by LIMSs, the ^{12}C/^{13}C ratio cannot discriminate between IMFs unambiguously. In previous work, however, we show that by switching from the Ballero IMF [19] (a very top-heavy IMF, which can reproduce the chemical abundances of stars in the Galactic bulge) to the Kroupa IMF, the ^{12}C/^{13}C ratio only varies [6] by a factor of 2×, indicating that carbon isotopologue ratios are not very sensitive to IMF. Furthermore, the ^{12}C abundance is very difficult to obtain because the ^{12}C-bearing major isotopologue lines are mostly optically thick.

The origin of oxygen isotopes is rather different. As earlier work suggested [68], the stellar yields of ^{16}O and ^{18}O are sensitive to different stellar masses, due to their temperature sensitivity in the stellar nucleosynthesis [69]. Production of ^{16}O is dominated by massive stars [54], as revealed by chemical evolution models in the Milky Way using detailed stellar yields [6]. Only a tiny fraction of ^{16}O is contributed by AGB stars [67]. Massive stars also dominate the production of ^{18}O [70], predominantly in the early stages of helium burning [69]. As a secondary element, the ^{18}O yield relies strongly on the pre-existence of ^{16}O, so the metallicity in oxygen also plays a major role in producing ^{18}O. The production of ^{18}O is biased to more massive stars compared to ^{13}C which is more biased to LIMSs. So, the abundance ratio of ^{13}C and ^{18}O does indeed reflect different IMFs (see Fig. 2).

The abundance ratios of ^{12}C/^{13}C and ^{16}O/^{18}O can trace star-formation timing and IMF [6], respectively. The ^{12}C and ^{16}O abundances are compromised by the optical depths of molecular lines, which are difficult to estimate accurately. However, the combination of both carbon and oxygen isotopologues – the abundance ratio of ^{13}C and ^{18}O – can be obtained easily from the ratio of two optically thin lines, ^{13}CO and C^{18}O, in the same J transition. Moreover, these two lines can be obtained simultaneously using current facilities, due to the close spacing of their rest frequencies. They have almost identical critical densities and upper energy levels, essentially free from excitation differences. Even for strongly lensed galaxies, it is safe to assume that a differential lensing effect among the two lines is negligible.

31. Magain, P., Surdej, J., Swings, J.-P., Borgeest, U. & Kayser, R. Discovery of a quadruply lensed quasar - The 'clover leaf' H1413 + 117. *Nature* **334**, 325–327 (1988).

32. Weiß, A. *et al.* ALMA Redshifts of Millimeter-selected Galaxies from the SPT Survey: The Redshift Distribution of Dusty Star-forming Galaxies. *Astrophys. J.* **767**, 88 (2013).

33. Negrello, M. *et al.* The Detection of a Population of Submillimeter-Bright, Strongly Lensed Galaxies. *Science* **330**, 800 (2010).

34. Griffin, M. J. *et al.* The Herschel-SPIRE instrument and its in-flight performance. *Astron. Astrophys.* **518**, L3 (2010).

35. Solomon, P., Vanden Bout, P., Carilli, C. & Guelin, M. The essential signature of a massive starburst in a distant quasar. *Nature* **426**, 636–638 (2003).

36. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA Architecture and Applications. In Shaw, R. A., Hill, F. & Bell, D. J. (eds.) *Astronomical Data Analysis Software and Systems XVI,* vol. 376 of *Astronomical Society of the Pacific Conference Series,* 127 (2007).

37. Proposal & observing preparation system team within the ALMA Computing IPT. *The ALMA Sensitivity Calculator (ASC)* (2017). URL: `https://almascience.eso.org/proposing/sensitivity-calculator`.

38. Mangum, J. G. & Shirley, Y. L. How to Calculate Molecular Column Density. *Publ. Astron. Soc. Pac.* **127**, 266 (2015).

39. Aalto, S., Booth, R. S., Black, J. H. & Johansson, L. E. B. Molecular gas in starburst galaxies: line intensities and physical conditions. *Astron. Astrophys.* **300**, 369 (1995).

40. van der Tak, F. F. S., Black, J. H., Schöier, F. L., Jansen, D. J. & van Dishoeck, E. F. A computer program for fast non-LTE analysis of interstellar line spectra. With diagnostic plots to interpret observed line intensity ratios. *Astron. Astrophys.* **468**, 627–635 (2007).

41. Yang, C. *et al.* Molecular gas in the Herschel-selected strongly lensed submillimeter galaxies at z 2-4 as probed by multi-J CO lines. *Astron. Astrophys.* **608**, A144 (2017).

42. Simpson, J. M. *et al.* The SCUBA-2 Cosmology Legacy Survey: Multi-wavelength Properties of ALMA-identified Submillimeter Galaxies in UKIDSS UDS. *Astrophys. J.* **839**, 58 (2017).

43. Papadopoulos, P. P. *et al.* Molecular Gas Heating Mechanisms, and Star Formation Feedback in Merger/Starbursts: NGC 6240 and Arp 193 as Case Studies. *Astrophys. J.* **788**, 153 (2014).

44. Wang, S. X. *et al.* An ALMA Survey of Submillimeter Galaxies in the Extended Chandra Deep Field-South: The AGN Fraction and X-Ray Properties of Submillimeter Galaxies. *Astrophys. J.* **778**, 179 (2013).

45. Spilker, J. S. *et al.* The Rest-frame Submillimeter Spectrum of High-redshift, Dusty, Star-forming Galaxies. *Astrophys. J.* **785**, 149 (2014).

46. Chartas, G., Eracleous, M., Agol, E. & Gallagher, S. C. Chandra Observations of the Cloverleaf Quasar H1413+117: A Unique Laboratory for Microlensing Studies of a LoBAL Quasar. *Astrophys. J.* **606**, 78–84 (2004).

47. Martín, S., Martín-Pintado, J. & Mauersberger, R. HNCO Abundances in Galaxies: Tracing the Evolutionary State of Starbursts. *Astrophys. J.* **694**, 610–617 (2009).

48. Greve, T. R., Papadopoulos, P. P., Gao, Y. & Radford, S. J. E. Molecular Gas in Extreme Star-Forming Environments: The Starbursts Arp 220 and NGC 6240 as Case Studies. *Astrophys. J.* **692**, 1432–1446 (2009).

49. Zinchenko, I., Henkel, C. & Mao, R. Q. HNCO in massive galactic dense cores. *Astron. Astrophys.* **361**, 1079–1094 (2000).

50. Li, J., Wang, J. Z., Gu, Q. S. & Zheng, X. W. Distribution of HNCO 5_{05}-4_{04} in massive star-forming regions. *Astron. Astrophys.* **555**, A18 (2013).

51. Schöier, F. L., van der Tak, F. F. S., van Dishoeck, E. F. & Black, J. H. An atomic and molecular database for analysis of submillimetre line observations. *Astron. Astrophys.* **432**, 369 (2005).

52. Matteucci, F. *Chemical Evolution of Galaxies* (2012).

53. Romano, D., Bellazzini, M., Starkenburg, E. & Leaman, R. Chemical enrichment in very low metallicity environments: Boötes I. *Mon. Not. R. Astron. Soc.* **446**, 4220–4231 (2015).

54. Tinsley, B. M. Evolution of the Stars and Gas in Galaxies. *Fundamentals of Cosmic Physics* **5**, 287–388 (1980).

55. Pagel, B. E. J. *Nucleosynthesis and Chemical Evolution of Galaxies* (1997).

56. Matteucci, F. (ed.). *The chemical evolution of the Galaxy,* vol. 253 of *Astrophysics and Space Science Library* (2001).

57. Kennicutt, R. C., Jr. The Global Schmidt Law in Star-forming Galaxies. *Astrophys. J.* **498**, 541–552 (1998).

58. Matteucci, F. & Greggio, L. Relative roles of type I and II supernovae in the chemical enrichment of the interstellar gas. *Astron. Astrophys.* **154**, 279–287 (1986).

59. Schaller, G., Schaerer, D., Meynet, G. & Maeder, A. New grids of stellar models from 0.8 to 120 solar masses at Z = 0.020 and Z = 0.001. *A&AS* **96**, 269–331 (1992).

60. Henkel, C. & Mauersberger, R. C and O Nucleosynthesis in Starbursts - the Connection Between Distant Mergers the Galaxy and the Solar System. *Astron. Astrophys.* **274**, 730 (1993).

61. Davis, T. A. Systematic variation of the ^{12}CO/^{13}CO ratio as a function of star formation rate surface density. *Mon. Not. R. Astron. Soc.* **445**, 2378–2384 (2014).

62. Henkel, C., Downes, D., Weiß, A., Riechers, D. & Walter, F. Weak ^{13}CO in the Cloverleaf quasar: evidence for a young, early generation starburst. *Astron. Astrophys.* **516**, A111 (2010).

63. Hughes, G. L. *et al.* The evolution of carbon, sulphur and titanium isotopes from high redshift to the local Universe. *Mon. Not. R. Astron. Soc.* **390**, 1710–1718 (2008).

64. Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C. & Maeda, K. Nucleosynthesis yields of core-collapse supernovae and hypernovae, and galactic chemical evolution. *Nuclear Physics A* **777**, 424–458 (2006).

65. Cescutti, G., Matteucci, F., McWilliam, A. & Chiappini, C. The evolution of carbon and oxygen in the bulge and disk of the Milky Way. *Astron. Astrophys.* **505**, 605–612 (2009).

66. Carigi, L., Peimbert, M., Esteban, C. & García-Rojas, J. Carbon, Nitrogen, and Oxygen Galactic Gradients: A Solution to the Carbon Enrichment Problem. *Astrophys. J.* **623**, 213–224 (2005).

67. Meyer, B., L.R., N., N., N. A. & S., M. Nucleosynthesis and Chemical Evolution of Oxygen. *Reviews in Mineralogy and Geochemistry* **68**, 1 (2008).

68. Sage, L. J., Henkel, C. & Mauersberger, R. Extragalactic O-18/O-17 ratios and star formation - High-mass stars preferred in starburst systems? *Astron. Astrophys.* **249**, 31–35 (1991).

69. Kobayashi, C., Karakas, A. I. & Umeda, H. The evolution of isotope ratios in the Milky Way Galaxy. *Mon. Not. R. Astron. Soc.* **414**, 3231–3250 (2011).

70. Timmes, F. X., Woosley, S. E. & Weaver, T. A. Galactic chemical evolution: Hydrogen through zinc. *Astrophys. J. Supp.* **98**, 617–658 (1995).

71. Ma, J. *et al.* Stellar Masses and Star Formation Rates of Lensed, Dusty, Star-forming Galaxies from the SPT Survey. *Astrophys. J.* **812**, 88 (2015).

72. Aravena, M. *et al.* A survey of the cold molecular gas in gravitationally lensed star-forming galaxies at z $\geq$ 2. *Mon. Not. R. Astron. Soc.* **457**, 4406–4420 (2016).

73. Dye, S. *et al.* Herschel-ATLAS: modelling the first strong gravitational lenses. *Mon. Not. R. Astron. Soc.* **440**, 2013–2025 (2014).

74. Negrello, M. *et al.* Herschel *-ATLAS: deep HST/WFC3 imaging of strongly lensed submillimetre galaxies. *Mon. Not. R. Astron. Soc.* **440**, 1999–2012 (2014).

75. Venturini, S. & Solomon, P. M. The Molecular Disk in the Cloverleaf Quasar. *Astrophys. J.* **590**, 740–745 (2003).

76. Falgarone, E. *et al.* Large turbulent reservoirs of cold molecular gas around high-redshift starburst galaxies. *Nature* **548**, 430–433 (2017).

77. Vieira, J. D. *et al.* Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature* **495**, 344–347 (2013).

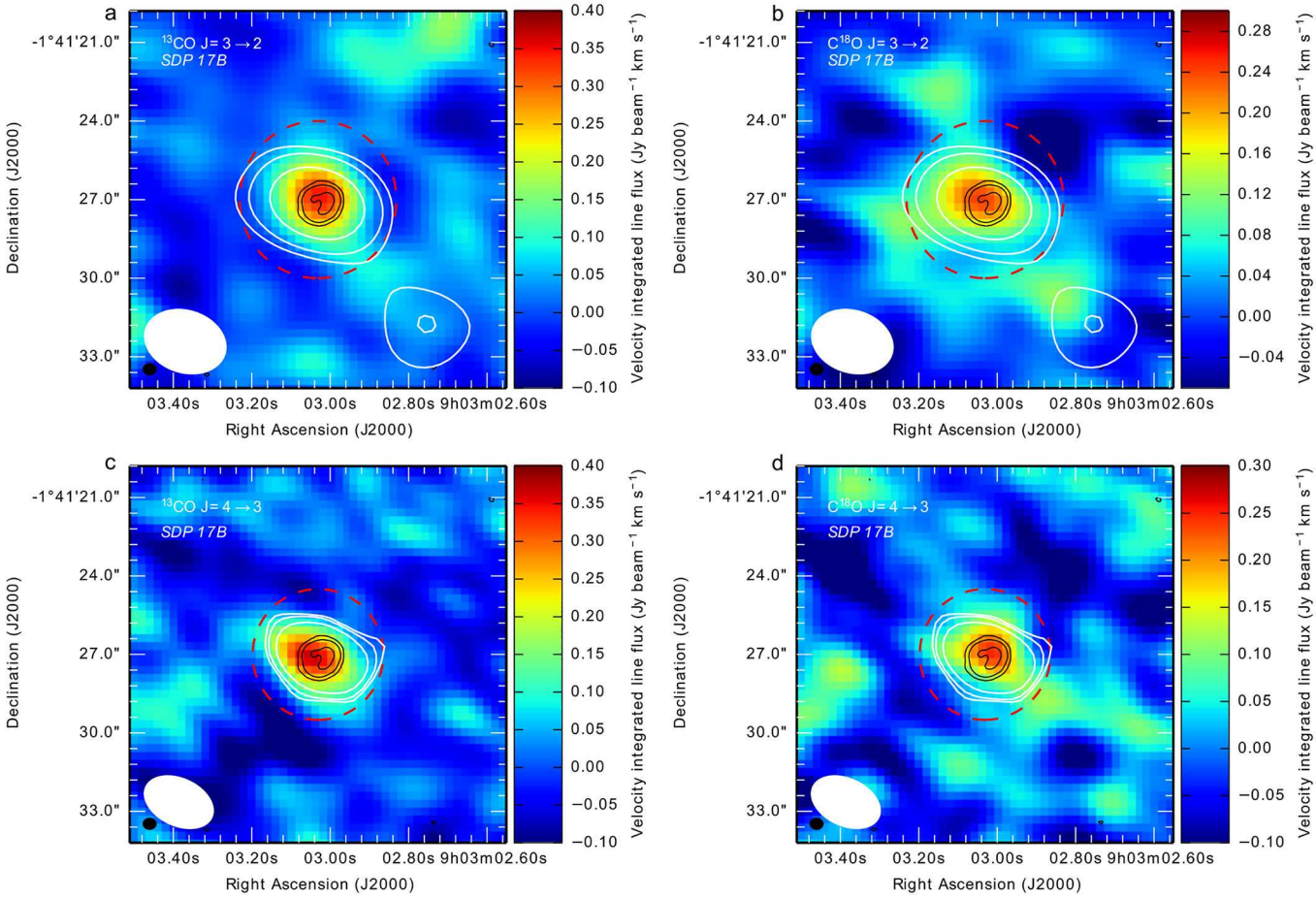
78. Ferkinhoff, C. *et al.* Band-9 ALMA Observations of the [N II] 122 μm Line and FIR Continuum in Two High-z galaxies. *Astrophys. J.* **806**, 260 (2015).

79. Omont, A. *et al.* H_2O emission in high-z ultra-luminous infrared galaxies. *Astron. Astrophys.* **551**, A115 (2013).

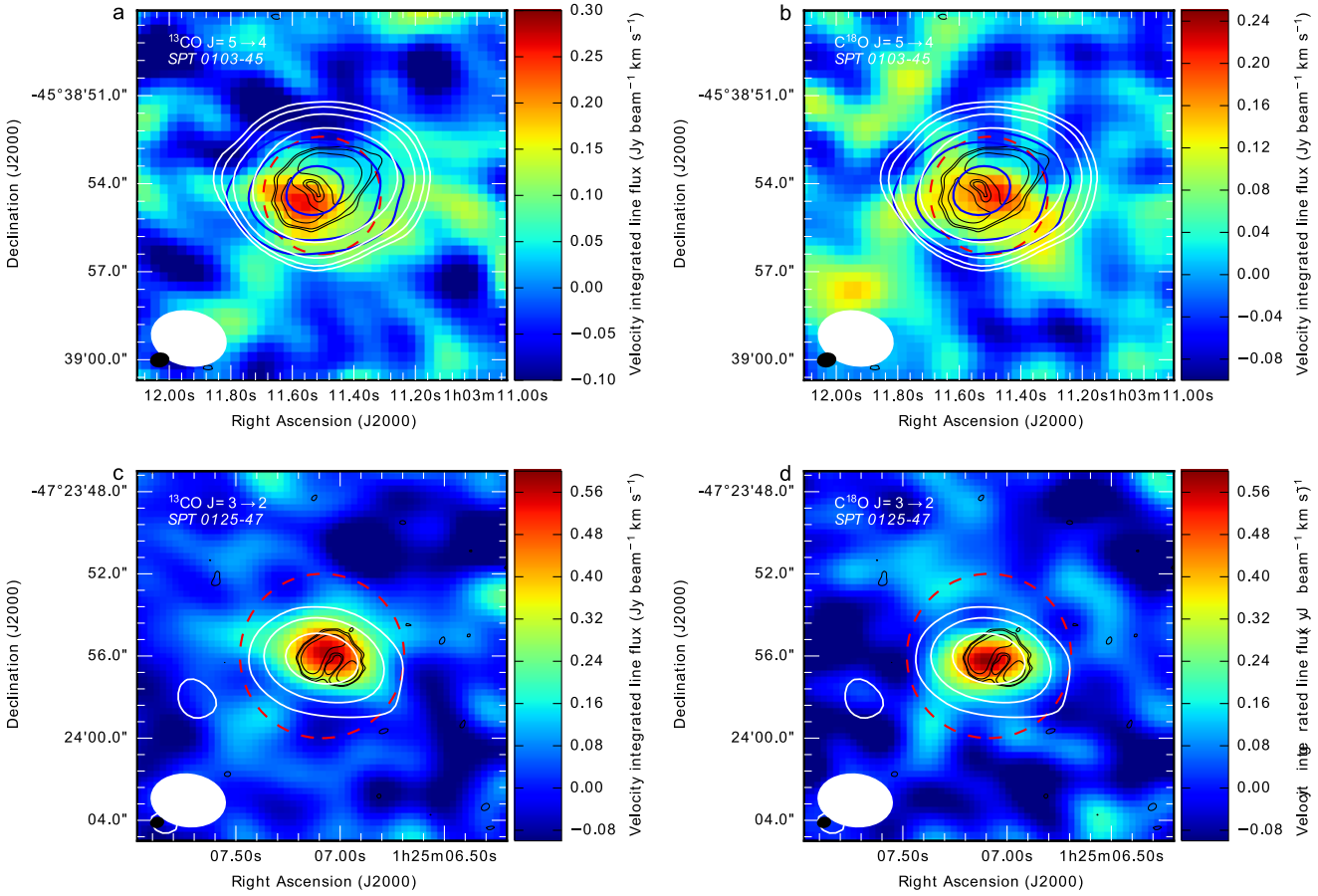
80. Weiß, A., Henkel, C., Downes, D. & Walter, F. Gas and dust in the Cloverleaf quasar at redshift 2.5. *Astron. Astrophys.* **409**, L41–L45 (2003).

81. Frerking, M. A., Langer, W. D. & Wilson, R. W. The relationship between carbon monoxide abundance and visual extinction in interstellar clouds. *Astrophys. J.* **262**, 590–605 (1982).

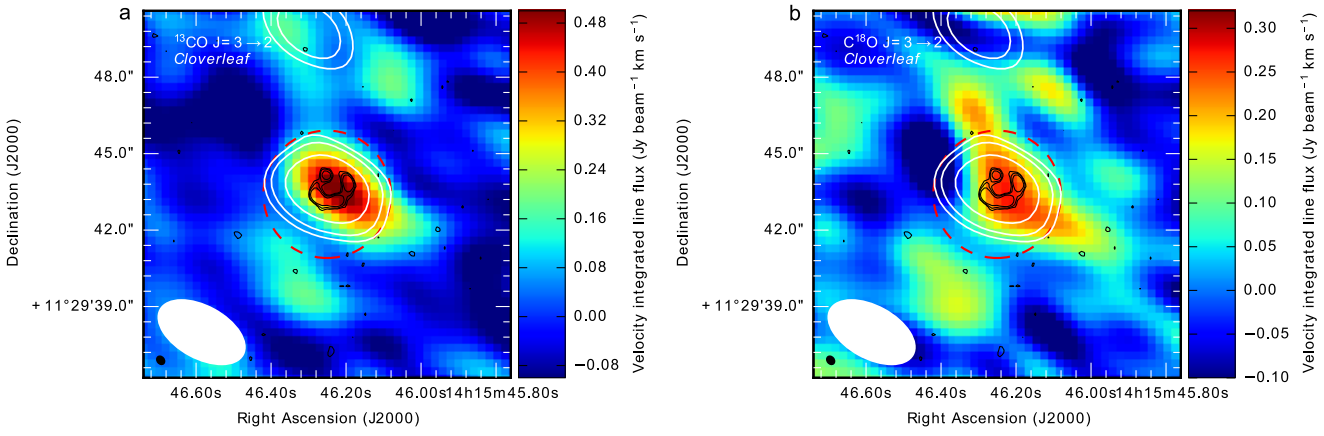
EXTENDED DATA



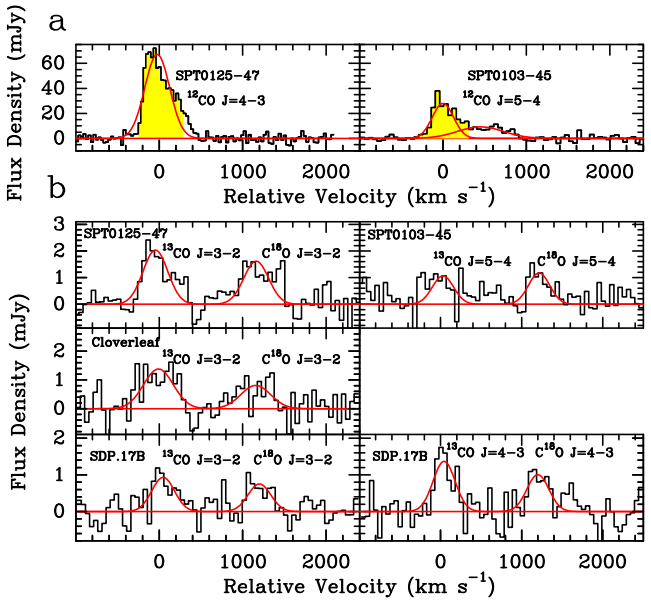
Extended Data Fig. 1 | Velocity-integrated flux maps (Moment-zero) of SDP 17b. Black contours show the high resolution 250-GHz continuum image, obtained from the ALMA archive [75], with levels of 3, 10 and 50 σ ($\sigma = 0.6 \times 10^{-1}$ mJy beam^{-1}). Dashed red circles show the adopted apertures for extracting spectra. **a & b:** Images of ^{13}CO and C^{18}O for the $J = 3 \rightarrow 2$ transition. White contours show the 95-GHz continuum, with levels of 3, 5 and 10 σ ($\sigma = 1.7 \times 10^{-2}$ mJy beam^{-1}). **c & d:** Images of ^{13}CO and C^{18}O for the $J = 4 \rightarrow 3$ transition. White contours show the 133-GHz continuum, with levels of 3, 5 and 10 σ ($\sigma = 2.3 \times 10^{-2}$ mJy beam^{-1}).



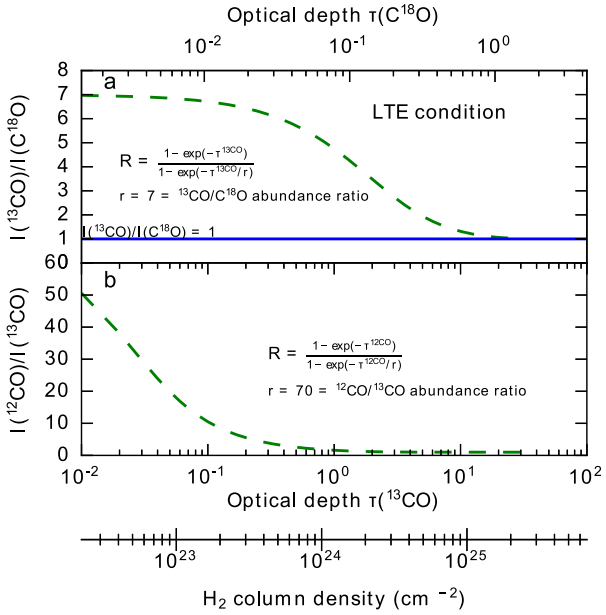
Extended Data Fig. 2 | Velocity-integrated flux maps (Moment-zero) of SPT 0103−45 and SPT 0125−47. Black contours show the high-resolution 336-GHz continuum image, obtained from the ALMA archive [76], with levels of 3, 10 and 30 σ ($\sigma = 2.3 \times 10^{-2}$ mJy beam^{-1}). Dashed red circles show the adopted apertures for extracting spectra. **a & b:** Images of ^{13}CO and C^{18}O J =5→ 4 for SPT 0103−45. Blue contours show the narrow ^{12}CO J = 4→ 3 emission, with levels of 3, 10 and 30 σ ($\sigma = 0.14$ Jy beam^{-1} km s^{-1}). White contours show the 135-GHz continuum, with levels of 3, 10 and 30 σ ($\sigma = 2 \times 10^{-2}$ mJy beam^{-1}). **c & d:** Images of ^{13}CO and C^{18}O for the J = 3→ 2 transition in SPT 0125−47. White contours show the 94-GHz continuum, with levels of 3, 5 and 10 σ ($\sigma = 2.2 \times 10^{-2}$ mJy beam^{-1}).



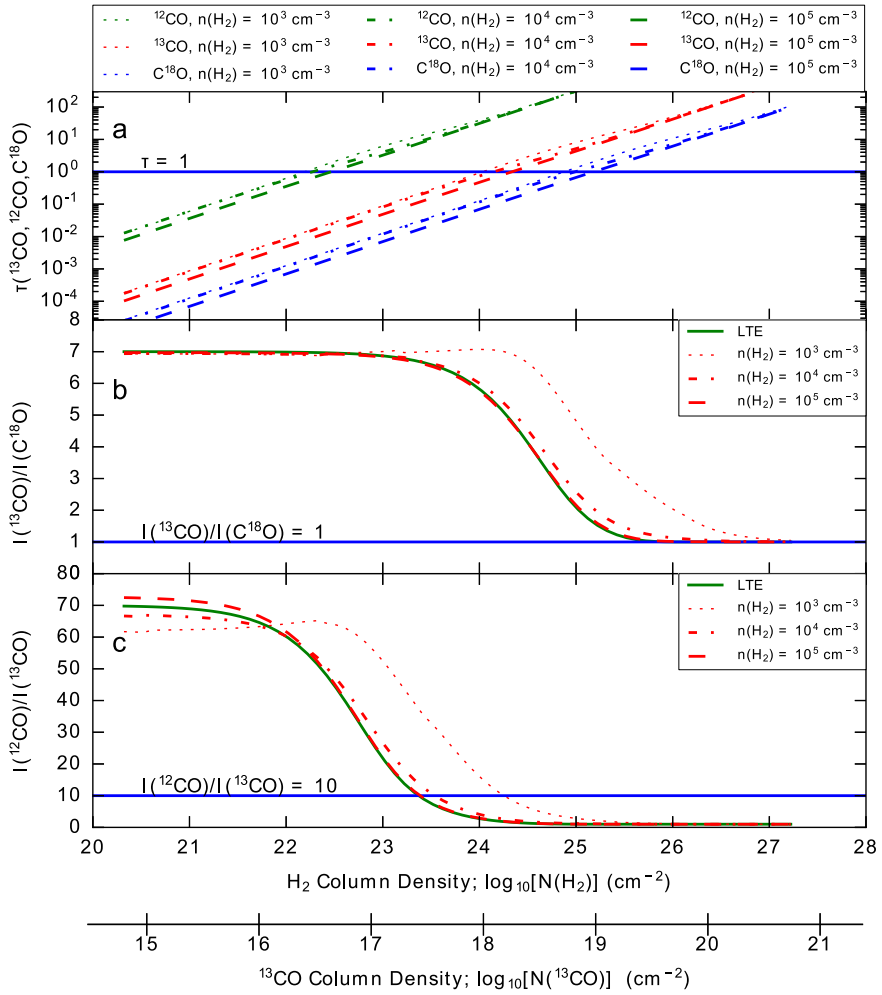
Extended Data Fig. 3 | Velocity-integrated flux maps (Moment-zero) of Cloverleaf. **a:** Image of the ^{13}CO J =3→ 2 transition. **b:** Image of the C^{18}O J =3→ 2 transition. Black contours show the high-resolution 690-GHz continuum image, obtained from the ALMA archive [77], with levels of 3, 5 and 10 σ (σ= 0.8 mJy beam^{-1}). Dashed red circles show the adopted apertures for extracting spectra. White contours show the 92-GHz continuum, with levels of 3, 5 and 10 σ ($\sigma = 2 \times 10^{-2}$ mJy beam^{-1}).



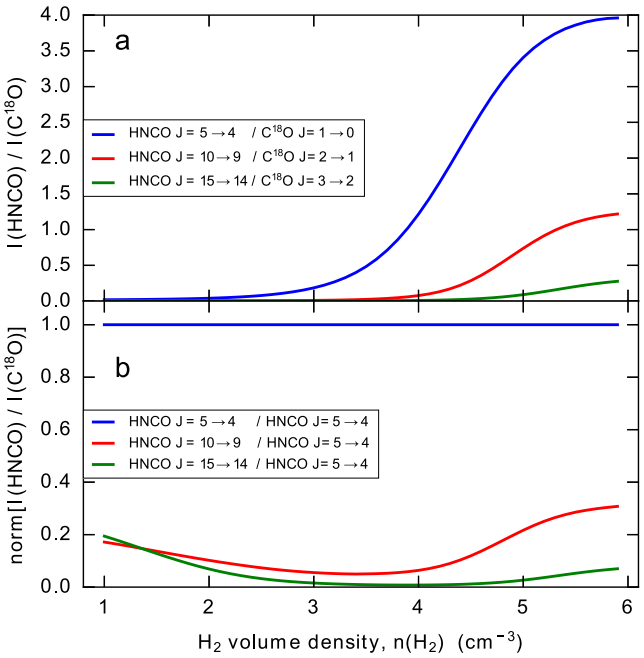
Extended Data Fig. 4 | **a:** ALMA spectra of ^{12}CO in SPT 0125−47 and SPT 0103−45. Yellow shadows show the velocity range adopted from ^{12}CO in the analysis. **b:** ALMA spectra of ^{13}CO and C^{18}O for all targets. All spectra are in black. Red lines show Gaussian fits to the observed lines. Velocities are labelled relative to their ^{12}CO or ^{13}CO transitions.



Extended Data Fig. 5 | **a:** $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ line ratio as a function of optical depth of ^{13}CO. **b:** $I(^{12}\mathrm{CO})/I(^{13}\mathrm{CO})$ line ratio as a function of optical depth of ^{13}CO. Both ratios assume local thermal equilibrium conditions. We assume abundance ratios of ^{13}CO/C^{18}O and ^{12}CO/^{13}CO are 7 and 70, respectively, which are representative values found in the Milky Way. This shows that the $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ line ratio approaches unity (blue line) only when the optical depth of C^{18}O is ≥ 1 (and the corresponding $\tau^{^{13}\mathrm{CO}} = 7$.). The bottom scale bar shows the corresponding column density of H_2 gas, assuming a CO/H_2 abundance of 8.5×10^{-5} [80].



Extended Data Fig. 6 | a: Optical depths of ^{12}CO, ^{13}CO and C^{18}O, for the J=3→2 transition; **b:** $I(^{13}\mathrm{CO})/I(\mathrm{C}^{18}\mathrm{O})$ line ratio, and **c:** $I(^{12}\mathrm{CO})/I(^{13}\mathrm{CO})$ line ratio as a function of H_2 column density, n_{H_2}, and ^{13}CO column density in various physical conditions, for non-LTE models calculated with RADEX[40]. For all models, we set the abundance ratios of ^{12}CO, ^{13}CO and C^{18}O to be Galactic: $^{12}\mathrm{CO}/^{13}\mathrm{CO}$ = 70 and $^{13}\mathrm{CO}/\mathrm{C}^{18}\mathrm{O}$ = 7, which are representative values of the Milky Way disk. Different line styles show the gas conditions of $n_{H_2} = 10^3$, 10^4 and 10^5 cm^{-3}. The T_{kin} for all models are set to 30 K, which is a typical dust temperature for the SMG population, and the lowest T_{kin} that H_2 gas can reach for such intensive starburst conditions, due to cosmic ray heating[43]. In panels **b** and **c**, we also overlay the line ratio (in thick green lines) with the LTE assumption for comparison. All three plots show that for Galactic abundances the line ratio of $^{13}\mathrm{CO}/\mathrm{C}^{18}\mathrm{O}$ can approach unity only when the ^{13}CO column density is higher than 10^{19}–10^{20} cm^{-2} (i.e. H_2 column density $N_{H_2} > 10^{25}$–10^{26} cm^{-2}).



Extended Data Fig. 7 | a: I(HNCO)/I(C^{18}O) line ratio as a function of H_2 volume density. **b:** I(HNCO)/I(C^{18}O) line ratio as a function of H_2 volume density, normalised with I(HNCO $J = 5 \rightarrow 4$)/I(C^{18}O $J = 1 \rightarrow 0$). Both ratios are calculated using RADEX[40], in which we assume the same abundances as measured in Arp 220[47]. We assume $T_{\rm kin} = 30$ K as the representative kinetic temperature of the H_2 gas.

Extended Data Table. 2 | ALMA Observational information

Target	SPT 0103−45	SPT 0103−45	SPT 0125−47	SPT 0125−47	SDP.17B	SDP.17B	Cloverleaf
Isotopologue	^{13}CO, C^{18}O	^{12}CO	^{13}CO, C^{18}O	^{12}CO	^{13}CO, C^{18}O	^{13}CO, C^{18}O	^{13}CO, C^{18}O
Transition	$J = 5 \rightarrow 4$	$J = 5 \rightarrow 4$	$J = 3 \rightarrow 2$	$J = 4 \rightarrow 3$	$J = 3 \rightarrow 2$	$J = 4 \rightarrow 3$	$J = 3 \rightarrow 2$
Observing Date	21-Jan-2016	21-Jan-2016	21-Jan-2016	21-Jan-2016	17-Jan-2016	16-Jan-2016	08-Apr-2016
Bandpass Calibrator	J2357−5311	J2357−5311	J2357−5311	J2357−5311	J0854+2006	J0854+2006	J1337−1257
Flux Calibrator	Neptune	J2357−5311	Neptune	Neptune	J0854+2006	J0854+2006	Callisto
Gain Calibrator	J0056−4451	J0051−4226	J0124−5113	J0124−5113	J0909+0121	J0909+0121	J1415+1320
Integration Time (s)	1300	120	605	120	816	726	1753
Median PWV (mm)	6.1	5.6	6.2	6.0	2.0	3.2	3.0
Median T_{sys} (K)	86	85	93	88	52	60	70
Angular Resolution	2.5″×1.8″	2.7″×1.6″	3.6″×2.5″	2.5″×1.8″	3.1″×2.3″	2.7″×1.8″	3.7″×2.0″

PWV, precipitable water vapour.

Extended Data Table. 3 | Observed targets, lines, frequencies, line widths and fluxes.

Target	Transition	ν_{obs}	I_{line}	I_{line}^{mom0}	σ^{theo} ⋆	ΔV_{line}	F_{peak}
	$J \rightarrow J-1$	GHz	Jy km s^{-1}	Jy km s^{-1}	Jy km s^{-1}	km s^{-1}	mJy
SDP.17b	^{12}CO $J = 4 \rightarrow 3$	139.49	9.1 ± 0.3	–	–	320	∼40
SDP.17b	^{13}CO $J = 3 \rightarrow 2$	100.02	0.32 ± 0.05	0.34 ± 0.08	0.08		0.9 ± 0.3
SDP.17b	C^{18}O $J = 3 \rightarrow 2$	99.64	0.26 ± 0.05	0.32 ± 0.08	0.08		0.8 ± 0.3
SDP.17b	^{13}CO $J = 4 \rightarrow 3$	133.36	0.47 ± 0.07	0.46 ± 0.08	0.07		1.3 ± 0.4
SDP.17b	C^{18}O $J = 4 \rightarrow 3$	132.85	0.34 ± 0.06	0.50 ± 0.08	0.07		1.0 ± 0.4
Cloverleaf	^{12}CO $J = 3 \rightarrow 2$	97.17	13.2 ± 0.2	–	–	400	30 ± 1.7
Cloverleaf	^{13}CO $J = 3 \rightarrow 2$	92.90	0.65 ± 0.09	0.61 ± 0.06	0.07		1.4 ± 0.4
Cloverleaf	C^{18}O $J = 3 \rightarrow 2$	92.55	0.40 ± 0.10	0.43 ± 0.06	0.07		0.8 ± 0.4
SPT 0103−45	^{12}CO $J = 4 \rightarrow 3$	112.68	8.2 ± 0.6	–	–		32 ± 0.6
SPT 0103−45	^{12}CO $J = 5 \rightarrow 4$	140.91	8.8 ± 0.5†	8.8 ± 0.6†	–	300†	27.8 ± 0.2
SPT 0103−45	^{13}CO $J = 5 \rightarrow 4$	134.65	0.37 ± 0.07	0.38 ± 0.05	0.07		1.2 ± 0.4
SPT 0103−45	C^{18}O $J = 5 \rightarrow 4$	134.13	0.35 ± 0.09	0.39 ± 0.07	0.07		1.2 ± 0.4
SPT 0125−47	^{12}CO $J = 3 \rightarrow 2$	98.38	18.1 ± 0.5	18.0 ± 0.5	–		43 ± 4
SPT 0125−47	^{12}CO $J = 4 \rightarrow 3$	131.21	26.9 ± 0.7	26.8 ± 0.7	–	400	69 ± 3
SPT 0125−47	^{13}CO $J = 3 \rightarrow 2$	94.06	0.78 ± 0.09	0.86 ± 0.07	0.1		2.0 ± 0.4
SPT 0125−47	C^{18}O $J = 3 \rightarrow 2$	93.70	0.63 ± 0.07	0.71 ± 0.1	0.1		1.6 ± 0.4

Literature data are from refs [32,78,79]. T_{sys} is the system temperature. ⋆ Theoretical noise level calculated using the ALMA sensitivity calculator [37]. † There are two velocity components in the ^{12}CO spectrum. We adopt only the narrow component, seen for ^{12}CO $J = 5 \rightarrow 4$, to avoid the broad and weaker component (see Extended Data Fig. 4).

Acknowledgements The authors are grateful to the referees for their constructive suggestions and comments. Z.-Y.Z. is grateful to Xiaoting Fu, Hau-Yu Baobab Liu, Yancy Shirley, and Peter Barnes for helpful discussions. Z.-Y.Z., R.J.I. and P.P.P. acknowledge support from the European Research Council in the form of the Advanced Investigator Programme, 321302, COSMICISM. F.M. acknowledges financial funds from Trieste University, FRA2016. This research was supported by the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence *"Origin and Structure of the Universe"*. This work also benefited from the International Space Science Institute (ISSI) in Bern, thanks to the funding of the team *"The Formation and Evolution of the Galactic Halo"* (PI D. Romano) This paper makes use of the ALMA data. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

Author Contributions Z.-Y.Z. is the Principal Investigator of the ALMA observing project. Z.-Y.Z. reduced the data and wrote of the initial manuscript. R.J.I. and P.P.P. provided ideas to initialise the project and helped write the manuscript. Z.-Y.Z. and P.P.P. worked on molecular line modeling of isotopologue ratios and chemical/thermal effects on the abundances. D.R. and F.M. ran the chemical evolution models and provided theoretical interpretation of the data. All authors discussed and commented on the manuscript.

Competing interests The authors declare no competing financial interests.

Code Availability We opt not to make the code used for the chemical evolution modeling publicly available because it is an important asset of the researchers' toolkits. The code for analysing the line ratios and optical depths of ^{12}CO, ^{13}CO, C^{18}O are based on the publicly available non-LTE radiative transfer code, RADEX[40].

Data Availability The dataset that supports the findings of this study is available in the ALMA archive (http://almascience.eso.org/aq/) under the observing project #2015.1.01309.S, #2013.1.00164.S, #2011.0.00958.S, and #2011.0.00747.S. Additional requests can be directed to the corresponding author upon reasonable request.