Astronomy & Astrophysics

The MUSE *Hubble* Ultra Deep Field Survey

XIV. Evolution of the Lyα emitter fraction from *z* = 3 to *z* = 6

Haruka Kusakabe[1], Jérémy Blaizot[2], Thibault Garel[1,2], Anne Verhamme[1,2], Roland Bacon[2], Johan Richard[2],
Takuya Hashimoto[3], Hanae Inami[4], Simon Conseil[2,5], Bruno Guiderdoni[2], Alyssa B. Drake[6],
Edmund Christian Herenz[7], Joop Schaye[8], Pascal Oesch[1], Jorryt Matthee[9], Raffaella Anna Marino[9],
Kasper Borello Schmidt[10], Roser Pelló[11], Michael Maseda[8], Floriane Leclercq[1,2],
Josephine Kerutt[1,10], and Guillaume Mahler[12]

[1] Observatoire de Genève, Université de Genève, 51 chemin de Pégase, 1290 Versoix, Switzerland
  e-mail: haruka.kusakabe@unige.ch
[2] Univ. Lyon, Univ. Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon, UMR5574, 69230 Saint-Genis-Laval, France
[3] Tomonaga Center for the History of the Universe (TCHoU), Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan
[4] Hiroshima Astrophysical Science Center, Hiroshima University 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
[5] Gemini Observatory/NSF's OIR Lab, Casilla 603, La Serena, Chile
[6] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[7] ESO Vitacura, Alonso de Córdova 3107,Vitacura, Casilla 19001, Santiago de Chile, Chile
[8] Leiden Observatory, Leiden University, PO Box 9513, 2300, RA Leiden, The Netherlands
[9] Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland
[10] Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany
[11] Aix Marseille Université, CNRS, CNES, LAM (Laboratoire d'Astrophysique de Marseille), UMR 7326, 13388 Marseille, France
[12] Department of Astronomy, University of Michigan, 1085 South University Ave, Ann Arbor, MI 48109, USA

ABSTRACT

Context. The Lyα emitter (LAE) fraction, $X_{\rm LAE}$, is a potentially powerful probe of the evolution of the intergalactic neutral hydrogen gas fraction. However, uncertainties in the measurement of $X_{\rm LAE}$ are still under debate.

Aims. Thanks to deep data obtained with the integral field spectrograph Multi Unit Spectroscopic Explorer (MUSE), we can measure the evolution of the LAE fraction homogeneously over a wide redshift range of $z \approx 3$–6 for UV-faint galaxies (down to UV magnitudes of $M_{1500} \approx -17.75$). This is a significantly fainter range than in former studies ($M_{1500} \leq -18.75$) and it allows us to probe the bulk of the population of high-redshift star-forming galaxies.

Methods. We constructed a UV-complete photometric-redshift sample following UV luminosity functions and measured the Lyα emission with MUSE using the latest (second) data release from the MUSE *Hubble* Ultra Deep Field Survey.

Results. We derived the redshift evolution of $X_{\rm LAE}$ for $M_{1500} \in [-21.75; -17.75]$ for the first time with a equivalent width range $EW(\rm Ly\alpha) \geq 65\,\text{\AA}$ and found low values of $X_{\rm LAE} \lesssim 30\%$ at $z \lesssim 6$. The best-fit linear relation is $X_{\rm LAE} = 0.07^{+0.06}_{-0.03}z - 0.22^{+0.12}_{-0.24}$. For $M_{1500} \in [-20.25; -18.75]$ and $EW(\rm Ly\alpha) \geq 25\,\text{\AA}$, our $X_{\rm LAE}$ values are consistent with those in the literature within 1σ at $z \lesssim 5$, but our median values are systematically lower than reported values over the whole redshift range. In addition, we do not find a significant dependence of $X_{\rm LAE}$ on M_{1500} for $EW(\rm Ly\alpha) \geq 50\,\text{\AA}$ at $z \approx 3$–4, in contrast with previous work. The differences in $X_{\rm LAE}$ mainly arise from selection biases for Lyman Break Galaxies (LBGs) in the literature: UV-faint LBGs are more easily selected if they have strong Lyα emission, hence $X_{\rm LAE}$ is biased towards higher values when those samples are used.

Conclusions. Our results suggest either a lower increase of $X_{\rm LAE}$ towards $z \approx 6$ than previously suggested, or even a turnover of $X_{\rm LAE}$ at $z \approx 5.5$, which may be the signature of a late or patchy reionization process. We compared our results with predictions from a cosmological galaxy evolution model. We find that a model with a bursty star formation (SF) can reproduce our observed LAE fractions much better than models where SF is a smooth function of time.

Key words. dark ages, reionization, first stars – early Universe – cosmology: observations – galaxies: evolution – galaxies: high-redshift – intergalactic medium

1. Introduction

In the early Universe, the first objects formed and filled the Universe with light. They ionized the neutral gas in the intergalactic medium (IGM) via a phenomenon called "cosmic reionization".

One of the candidates for the main source of reionization is star-forming galaxies, whose ionizing radiation, which is called the "Lyman Continuum" (LyC, $\lambda < 912\,\text{\AA}$) and is emitted from massive stars, its ionizing radiation is expected to leak into the IGM (e.g., Bouwens et al. 2015a,b; Finkelstein et al. 2015

Robertson et al. 2015; Livermore et al. 2017). Another candidate is active galactic nuclei (AGNs, e.g., Madau & Haardt 2015). However, they have recently been reported to contribute less than $\approx 10\%$ of the ionizing photons needed to keep the IGM ionized (over a UV magnitude range of -18 to -30 mag; Matsuoka et al. 2018, see also Parsa et al. 2018). Previous studies using the Gunn-Peterson absorption trough seen in quasar spectra (e.g., Gunn & Peterson 1965; Fan et al. 2006; McGreer et al. 2015), and in gamma-ray burst spectra (e.g., Totani et al. 2006, 2014) suggest that cosmic reionization was completed by $z \approx 6$ (see however, Bosman et al. 2018). The Thomson optical depth of the cosmic microwave background measured by *Planck* suggests that the midpoint redshift of reionization (i.e., when half the IGM had been reionized) is at $z \approx 7.7 \pm 0.7$ (1σ confidence interval, Planck Collaboration VI 2020).

Lyα emission is intrinsically the strongest UV spectral feature of young star-forming galaxies, and galaxies with mostly detectable Lyα emission or with Lyα equivalent widths higher than ≈ 25 Å are called "Lyα emitters (LAEs)". Lyα emission is scattered by neutral hydrogen gas (H I) in the IGM, and, therefore, the detectability of LAEs is affected by the H I gas fraction in the IGM. The redshift evolution of Lyα luminosity functions has thus been used to investigate the history of the neutral hydrogen gas fraction of the IGM (e.g., Malhotra & Rhoads 2004; Kashikawa et al. 2006; Hu et al. 2010; Ouchi et al. 2010; Santos et al. 2016; Drake et al. 2017a; Ota et al. 2017; Zheng et al. 2017; Konno et al. 2018; Itoh et al. 2018). Lyα luminosity functions can be used to compute the evolution of the Lyα luminosity density, and its rapid decline at $z \gtrsim 5.7$ compared with that of the cosmic star formation rate density derived from UV luminosity functions is interpreted to be caused by IGM absorption (e.g., Ouchi et al. 2010; Konno et al. 2018).

Similarly, the fraction of LAEs among UV selected galaxies, $X_{\rm LAE}$, can also be used to probe the evolution of the H I gas fraction of the IGM (e.g., Fontana et al. 2010; Pentericci et al. 2011; Stark et al. 2011; Ota et al. 2012; Treu et al. 2013; Caruana et al. 2014; Faisst et al. 2014; Schenker et al. 2014). $X_{\rm LAE}$ has been reported to increase from $z \approx 3$ to 6 and then to drop at $z > 6$. This has again been interpreted as a signature of the IGM becoming more neutral at $z > 6$ (e.g., Dijkstra et al. 2011; Jensen et al. 2013; Mason et al. 2018). The LAE fraction is complementary to the test of Lyα luminosity functions (LFs) and has some advantages: efficient spectroscopic observations as a follow-up of continuum-selected galaxies, which is insensitive to the declining number density of star forming galaxies, and rich information obtained from the spectra such as spectroscopic redshifts and kinematics of the interstellar medium (e.g., Stark et al. 2010; Hashimoto et al. 2015). It also enables us to solve the degeneracy between the Lyα escape fraction among star forming galaxies with different UV magnitudes and the comparison between luminosity densities of Lyα emission and UV continuum, which are obtained from the integration of UV and Lyα LFs. In addition, Kakiichi et al. (2016) recently, suggested that the UV magnitude-dependent evolution of the LAE fraction combined with the Lyα luminosity function can be used to constrain the ionization topology of the IGM and the history of reionization.

Using $X_{\rm LAE}$ to set quantitative constraints on the evolution of the neutral content of the IGM remains challenging. In particular, we need to understand whether observed variations of $X_{\rm LAE}$ are exclusively due to variations in the IGM properties, or whether they can be attributed to galaxy evolution. Following the Lyα spectroscopic observations of Lyman

break galaxies (LBGs) at $z \approx 3$ by Steidel et al. (2000) and Shapley et al. (2003), Stark et al. (2010, 2011) have found that $X_{\rm LAE}$ among LBGs evolves with redshift and depends on the rest-frame Lyα equivalent width ($EW({\rm Ly}\alpha)$) cut. They also show that $X_{\rm LAE}$ depends on the absolute rest-frame UV magnitude (M_{1500}), so that UV-faint galaxies are more likely to show Lyα than UV-bright galaxies (see also Schaerer et al. 2011a; Forero-Romero et al. 2012; Garel et al. 2012). One conclusion from these studies is that the evolution of $X_{\rm LAE}$ with redshift is more prominent for UV-faint galaxies and low $EW({\rm Ly}\alpha)$ cuts.

However, several recent studies show lower values of $X_{\rm LAE}$ for UV-faint galaxies ($-20.25 < M_{1500} < -18.75$ mag) than those in the pioneering work of Stark et al. (2011). At $z \approx 4$ and $z \approx 5$, Arrabal Haro et al. (2018) show more than 1σ lower $X_{\rm LAE}$ for the faint M_{1500} and low EW cut (25 Å), though their result at $z \approx 6$ is consistent with that in Stark et al. (2011). De Barros et al. (2017) also investigate $X_{\rm LAE}$ for UV-faint galaxies with a low EW cut, at $z \approx 6$. They obtained a low median value of $X_{\rm LAE}$, which is even slightly lower than the value previously found at $z \approx 5$, though their $X_{\rm LAE}$ is consistent within 1σ. They concluded that the drop at $z > 6$ is less dramatic than what was previously found (see also Pentericci et al. 2018, for their recent study at $z \approx 7$). De Barros et al. (2017) and Pentericci et al. (2018) also suggest the possibility that the effect of an increase of the H I gas fraction in the IGM is observed between $5 < z < 6$. This would be consistent with a later and more inhomogeneous reionization process than previously thought, as has also been recently suggested by observations and simulations of fluctuations in Lyα forest (e.g., Bosman et al. 2018; Kulkarni et al. 2019; Keating et al. 2019). The parent LBG sample in De Barros et al. (2017) is selected with an additional UV magnitude cut on a normal LBG selection, while the parent sample in Arrabal Haro et al. (2018) is mostly based on photometric redshift (photo-z) even though it is regarded as an LBG sample in their paper. Therefore, the results of $X_{\rm LAE}$ for the faint M_{1500} are not yet conclusive, possibly due to different parent sample selections. Moreover, for the UV-bright galaxies, the redshift evolution of $X_{\rm LAE}$ for a 25 Å EW cut has not been confirmed yet (e.g., Stark et al. 2011, 2017; Curtis-Lake et al. 2012; Ono et al. 2012; Schenker et al. 2014; Cassata et al. 2015; Mason et al. 2019). De Barros et al. (2017) and Pentericci et al. (2018) suggest that some previous results are affected by an LBG selection bias. As strong Lyα emission affects the red band, strong LAEs can be selected more easily compared to galaxies without Lyα emission at faint UV magnitudes. It results in a high LAE fraction of LBGs (see also Stanway et al. 2008; Inami et al. 2017). Arrabal Haro et al. (2018) assess UV completeness of their parent sample using UV luminosity functions and find that their 90% completeness magnitude is ≈ -20 and -21 mag at $z \approx 4$ and $z \approx 5$, respectively.

To summarize, it is important to obtain a firm conclusion about the evolution of $X_{\rm LAE}$ in the post-reionization epoch in order to quantify the drop of $X_{\rm LAE}$ at $z > 6$ and to assess the reliability of using $X_{\rm LAE}$ as a good probe of reionization. However, although there are a number of observational studies of $X_{\rm LAE}$, uncertainties in the measurement and interpretation of $X_{\rm LAE}$ are still a matter of debate (e.g., Stark et al. 2011; Garel et al. 2015; De Barros et al. 2017; Caruana et al. 2018; Mason et al. 2018; Hoag et al. 2019a,b). One of the biggest problems is the LBG selection bias due to the different depths of selected bands in previous studies. It is worth pointing out that none of the previous studies were based on complete parent samples of UV faint galaxies ($-20.25 < M_{1500} < -18.75$ mag). Completeness in terms of UV magnitudes, as well as homogeneously selected

samples over a wide redshift range are essential for the determination of $X_{\rm LAE}$. In addition, we also need deep and homogeneous spectroscopic observations of Lyα emission over a wide redshift range.

In this study, we use a combination of deep *Hubble* Space Telescope (HST) and Very Large Telescope (VLT)/Multi-Unit Spectroscopic Explorer (MUSE) data, to overcome these limitations and improve our knowledge of the evolution with redshift of $X_{\rm LAE}$ among a homogeneous parent sample of UV faint galaxies. We use HST bands that are not contaminated by Lyα emission to measure UV magnitudes to avoid a selection bias. Deep and homogeneous spectroscopic Lyα observations at a wide redshift range have been achieved in the *Hubble* Ultra Deep field (HUDF) by VLT/MUSE (Bacon et al. 2010) in the guaranteed time observations (GTO), MUSE-HUDF survey (e.g., Bacon et al. 2017). The LAE fraction has already been investigated with MUSE and HST data, using the MUSE-Wide GTO survey in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) Deep region in Caruana et al. (2018). Their sample are constructed with an apparent magnitude cut of $F775W \leq 26.5$ mag for an HST catalog, which is roughly converted to $M_{1500} \approx -19$ to -20 mag at $z \approx 3$–6. However, the MUSE HUDF data enable us to measure faint Lyα emission even for faint UV sources (-17.75 mag) in existing HUDF photometric catalogs.

The paper is organized as follows. In Sect. 2, we describe the data, methods, and samples: our UV-selected samples, the MUSE data, the detection and measurement of Lyα emission, the calculation of the LAE fraction, and its uncertainties. Section 3 presents the LAE fraction as a function of redshift and UV magnitude. In Sect. 4, we discuss our results: the differences in LAE fraction from previous results, a comparison with predictions from a model of galaxy formation, and implications for reionization. Finally, the summary and conclusions are given in Sect. 5.

Throughout this paper, we assume a flat cosmological model with a matter density of $\Omega_{\rm m} = 0.3$, a cosmological constant of $\Omega_{\Lambda} = 0.7$, and a *Hubble* constant of $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$ ($h_{100} = 0.7$). Magnitudes are given in the AB system (Oke & Gunn 1983).

2. Data, methods, and samples

In Sect. 2.1, we first discuss the construction of a volume-limited UV-selected sample of galaxies from the HUDF catalog of Rafelski et al. (2015, hereafter R15)[1]. In Sect. 2.2 we explain how we use MUSE data to detect and measure Lyα emission from galaxies of our UV sample. In Sect. 2.3 we lay out our calculations of $X_{\rm LAE}$ and discuss the error budget. In Sect. 2.4 we present our measurement of slopes of $X_{\rm LAE}$ as a function of z and M_{1500}. We briefly discuss in Sect. 2.5 the effects of extended Lyα emission.

2.1. UV-selected samples

We built our parent sample of high-redshift UV-selected galaxies using the latest HUDF catalog from R15. Sources in this catalog are detected in the average-stacked image of eight HST bands: four optical bands from *ACS/WFC* ($F435W$, $F606W$, $F775W$, and $F850LP$), and four near infrared (NIR) bands from *WFC3/IR* ($F105W$, $F125W$, $F140W$, and $F160W$). In total, out of the 9969 sources in R15's catalog, 1095 and 7904 objects are within the footprints of the udf-10 and mosaic regions

of the MUSE HUDF Survey[2] (Bacon et al. 2017), respectively (the duplicated region in udf-10 is removed). We note that the $F140W$ image only covers 6.8 arcmin^2 of the 11.4 arcmin^2 footprint of the R15's catalog. The $F140W$ photometry is only used when it is available in R15. As discussed in footnote 2 of Hashimoto et al. (2017a), the lack of $F140W$ may affect the detection of sources in R15's catalog. Moreover, the footprint of $F140W$ is also covered by deeper NIR images ($F105W$, $F125W$, and $F160W$). Indeed, the fraction of sources identified by R15 at $z \gtrsim 6$ within the footprint of the $F140W$ image is higher than where there is no $F140W$ data.

In order to avoid contamination from neighboring objects, we followed Inami et al. (2017) and discarded all HST sources which have at least one neighbour within $0\rlap{.}{''}6$. Such associations cannot be resolved in our MUSE observations where the full width at half maximum (FWHM) of the average seeing of DR1 data is $\approx 0\rlap{.}{''}6$ at 7750 Å. This procedure excludes $\approx 20\%$ of the sources. We assume that this does not result in a significant bias as it is effectively only decreasing the survey area. This assumption is true if interacting systems are not more often LAEs than isolated systems.

R15 provide photometric redshifts and associated errors for all objects. These are obtained via spectral energy distribution (SED) fitting of photometric data in 11 HST bands using either the Bayesian photometric redshift (BPZ) algorithm (Benitez 2000; Benitez et al. 2004; Coe et al. 2006) or the EAZY software (Brammer et al. 2008). In the present paper, we choose to use the results from BPZ because they are found to be more accurate (Rafelski et al. 2015; Inami et al. 2017; Brinchmann et al. 2017). We note that R15 do not include *Spitzer*/IRAC data in their SED fitting. We state in Appendix A.1 that this addition would not improve the photometric redshifts of the faint galaxies studied here. Below, $z_{\rm p}$ denotes the photometric redshift given in R15[3], and we use it to define redshift selections of all our UV samples (see Table 1). In R15's catalog, the 95% lower and upper limits of $z_{\rm p}$ are provided as uncertainties on $z_{\rm p}$. We use these to construct an "inclusive parent sample", that we will use for Lyα searches. This sample includes all sources with photometric redshift estimates (95% confidence interval) overlapping with the redshift range (2.91–6.12) where Lyα can be observed with MUSE. We note that we remove sources at $z > 6.12$ from our sample because the parent photometric-redshift sample may be affected by selection bias (see Sect. 2.1), and because Lyα detectability in MUSE spectra is strongly reduced by sky lines (Drake et al. 2017a). In the end, this inclusive parent sample consists of 3233 and 402 sources in the mosaic and udf-10, respectively (without duplication).

We derived the absolute rest-frame UV magnitude using two or three HST photometric points to fit a power-law to the UV continuum. The power-law describes the spectral flux density f_ν as $f_\nu = f_0(\lambda/\lambda_0)^{\beta+2}$, where $\lambda_0 = 1500$ Å, f_0 is the spectral flux density at 1500 Å (in erg s^{-1} cm^{-2} Hz^{-1}), and β is the continuum slope. We then simply have $M_{1500} = -2.5 \log (f_0) - 48.6 - 5 \log (d_{\rm L}/10) + 2.5 \log (1 + z_{\rm p})$, where $d_{\rm L}$ (pc) is the luminosity distance. We chose the HST filters following Hashimoto et al. (2017a) so that Lyα emission and IGM absorption are not included in the photometry: we used HST/$F775W$, $F850LP$, and $F105W$ for objects at $2.91 \leq z_{\rm p} \leq 4.44$; $F105W$,

---

[1] https://archive.stsci.edu/hlsp/uvudf

[2] The survey consists of two layers of different depths: a shallower area with 9 MUSE pointings (mosaic) and a deeper area with 1 MUSE pointing (udf-10) within the mosaic. More details are given in Sect. 2.2.1.

[3] The photometric redshift value is the one which maximizes the likelihood estimate in BPZ.

Table 1. Subsample criteria.

z range	Mean z	M_{1500} range (mag)	$N_{1500}(z_p, M_{1500})$	$EW(\mathrm{Ly}\alpha)$ cut (Å)
Subsamples for Fig. 7: X_{LAE} vs. z				
$2.91 < z \leq 3.68$	3.3	$-21.75 \leq M_{1500} \leq -17.75$ [a]	228	45 [b], 65
$3.68 < z \leq 4.44$	4.1	$-21.75 \leq M_{1500} \leq -17.75$ [a]	119	45 [b], 65
$4.44 < z \leq 5.01$	4.7	$-21.75 \leq M_{1500} \leq -17.75$ [a]	98	45 [b], 65
$5.01 < z \leq 6.12$	5.6	$-21.75 \leq M_{1500} \leq -17.75$ [a]	89	45 [b], 65
Subsamples for Figs. 8, 11, and 12: X_{LAE} vs. z for comparison with previous work				
$2.91 < z \leq 3.68$	3.3	$-20.25 \leq M_{1500} \leq -18.75$	87	25, 55
$3.68 < z \leq 4.44$	4.1	$-20.25 \leq M_{1500} \leq -18.75$	40	25, 55
$4.44 < z \leq 5.01$	4.7	$-20.25 \leq M_{1500} \leq -18.75$	28	25, 55
$5.01 < z \leq 6.12$	5.6	$-20.25 \leq M_{1500} \leq -18.75$	35	25, 55
Subsamples for Figs. 5 and 10: X_{LAE} vs. M_{1500}				
$2.91 < z \leq 3.68$	3.3	$-21.5 \leq M_{1500} \leq -20.0$	31	25, 45, 65, 85
$2.91 < z \leq 3.68$	3.3	$-20.0 < M_{1500} \leq -19.0$	58	25, 45, 65, 85
$2.91 < z \leq 3.68$	3.3	$-19.0 < M_{1500} \leq -18.0$	106	45, 65, 85
$2.91 < z \leq 3.68$	3.3	$-18.0 < M_{1500} \leq -17.0$	197	85
Subsamples for Fig. 6: X_{LAE} and M_{1500} for comparison with previous work				
$2.91 < z \leq 3.68$	3.3	$-21.5 \leq M_{1500} \leq -20.0$	31	50
$2.91 < z \leq 3.68$	3.3	$-20.0 < M_{1500} \leq -19.0$	58	50
$2.91 < z \leq 3.68$	3.3	$-19.0 < M_{1500} \leq -18.0$	106	50
$3.68 < z \leq 4.44$	4.1	$-21.5 \leq M_{1500} \leq -20.0$	8	50
$3.68 < z \leq 4.44$	4.1	$-20.0 < M_{1500} \leq -19.0$	23	50
$3.68 < z \leq 4.44$	4.1	$-19.0 < M_{1500} \leq -18.0$	56	50

Notes. Subsample criteria of redshift and M_{1500} for the continuum-selected parent sample. The mean redshift, sample size ($N_{1500}(z_p, M_{1500})$), and $EW(\mathrm{Ly}\alpha)$ cut are also shown. To increase the sample size used in Figs. 7, 8, 11, and 12, we combine the two highest redshift bins used to compute the UV magnitude and Lyα completeness. [a]We also calculate X_{LAE} for the $-21.75 \leq M_{1500} \leq -18.75$ mag and $-18.75 < M_{1500} \leq -17.75$ mag subsamples. [b]The 45 Å $EW(\mathrm{Ly}\alpha)$ cut is only applied to the $-21.75 < M_{1500} \leq -18.75$ mag subsamples.

$F125W$, and $F140W$ for objects at $4.44 < z_p \leq 5.58$; and $F125W$, $F140W$, and $F160W$ for objects at $5.58 < z_p \leq 6.12$. While Hashimoto et al. (2017a) use the MUSE spectroscopic redshifts, we used the values of the photometric redshifts from Rafelski et al. (2015). Our derived M_{1500} values are consistent with those of Hashimoto et al. (2017a) for the sources which we have in common (LAEs). The standard deviation of the relative difference in M_{1500} for the sources included in both studies is $\approx 3\%$ without a systematic offset.

In Fig. 1, we show the distribution of M_{1500} as a function of z_p for sources in our parent sample that have $2.91 < z_p \leq 6.12$. In order to construct a complete parent sample in terms of M_{1500}, we define a limiting magnitude M_{1500}^{lim} so that objects brighter than M_{1500}^{lim} are detected with a signal-to-noise ratio (S/N) larger than two in at least two HST bands among the rest-frame UV HST bands. To compute M_{1500}^{lim}, we again use a power-law continuum model, this time with a fixed UV slope $\beta = -2$, which is commonly used as a fixed value (e.g., Caruana et al. 2018). At each redshift, we derive the normalisation M_{1500}^{lim} such that the flux can be detected in at least two HST UV bands with $S/N > 2$. The resulting M_{1500}^{lim} is shown with the thick black curve overplotted to the data in magenta in the upper panel of Fig. 1. From this panel, we see that we can build a complete UV-selected sample at redshifts $z \approx 3$–4 down to $M_{1500}^{\mathrm{lim}} \approx -16.3$ mag, and even at $z \approx 6$ we can achieve completeness down to $M_{1500}^{\mathrm{lim}} \approx -17.7$ mag.

In the upper panel of Fig. 1, we highlight two regions of parameter space that we select to do two complementary studies: X_{LAE} vs. M_{1500} with z in the range [2.91; 4.44] (the polygon marked with a solid black line) and X_{LAE} vs. z with z in the range [2.91; 6.12] (the dashed-line rectangle). To define the criteria for X_{LAE} vs. M_{1500} plots, completeness simulations for

Lyα emission (see Sect. 2.2.4) and z_p bins for the M_{1500}^{lim} calculation are also taken into account. In the lower panel of Fig. 1, we show for comparison the locus of previous studies in the $(M_{1500} - z)$ plane. The faint galaxies in Stark et al. (2011) are shown by the light-grey shaded area. De Barros et al. (2017) use a UV magnitude cut of $F160W \leq 27.5$ mag at $z \approx 6$ in their sample, which is shown with a dark-grey arrow. The recent sample of Arrabal Haro et al. (2018) is shown with dark-grey crosses which indicate the UV magnitudes at which they reach $\approx 90\%$ completeness. The LAE fraction with the MUSE-Wide GTO survey data (Caruana et al. 2018) adopt an apparent magnitude cut of $F775W \leq 26.5$ mag for an HST catalog (Guo et al. 2013), which we roughly convert to M_{1500} for illustrative purposes, and show with the solid grey line in the lower panel of Fig. 1. Our MUSE-Deep data combined with the HST catalog from R15 allow us to probe deeper than all previous work, and to extend our study to UV-faint galaxies (i.e., $M_{1500} \geq -18.75$ mag).

In Fig. 2, we demonstrate the completeness of our UV-selected sample at different redshifts by comparing our UV number counts to what we would expect from the UV luminosity function (UVLF) of Bouwens et al. (2015b). We find that the distribution of M_{1500} for our samples (magenta) follow well those expected from the UVLFs for the same survey area at similar redshifts (black solid lines) within 1σ error bars. For comparison, we also show in the bottom panels of Fig. 2 the distribution of magnitudes of the sample of Stark et al. (2010). Clearly, their samples are incomplete even at relatively bright magnitudes (≈ -20.25 mag), most likely because of the shallow depth of their data and the LBG selection. We note that the LBG samples in Stark et al. (2010, 2011) consist of two LBG samples. One is from their own sample (Stark et al. 2009), and the other
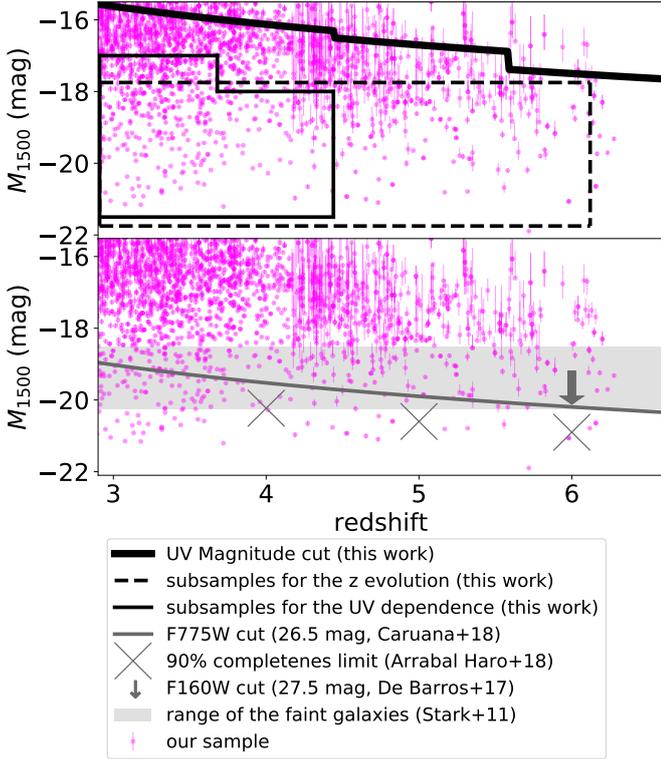
Fig. 1. M_{1500} versus z_p for our sample and the literature. The M_{1500} and z_p of our parent sample from Rafelski et al. (2015) are shown by magenta filed circles (identical in the two panels). *Upper panel*: the M_{1500} cut (N_{1500}^{lim}) for our sample defined from the rest-frame UV HST bands is indicated by a black thick solid line. The parameter space studied here, X_{LAE} vs. M_{1500} at $z \approx 3$–4 and X_{LAE} vs. z at $z \approx 3$–6, is shown by a black solid polygon and a black dashed rectangle, respectively. *Lower panel*: the UV magnitude cut of $F160W \leq 27.5$ mag at $z \approx 6$ in De Barros et al. (2017) and UV magnitudes corresponding to $\approx 90\%$ completeness at $z \approx 4$, 5, and 6 in Arrabal Haro et al. (2018) are represented by a dark-grey arrow and dark-grey crosses, respectively. The apparent magnitude cut of $F775W \leq 26.5$ mag in Caruana et al. (2018) is shown by a grey solid line. The parameter space for the faint galaxies studied in Stark et al. (2011) is indicated by a light-grey shaded region.

Fig. 2. Histograms of M_{1500} for our parent samples at $z \approx 3.7$ (=2.91–4.44, *left*), 5.0 (= 4.44–5.58, *middle*), and 5.9 (= 5.58–6.12, *right*). *Upper panels*: magenta histograms and black dashed lines represent the number distribution of our parent sample and that expected from the UV luminosity functions (UVLF) in Bouwens et al. (2015b) for the same effective survey area, respectively. The uncertainty of the number distribution of our parent sample is given by the Poisson error. Grey hashed areas indicate M_{1500} ranges that are not used in this work. *Lower panels*: light-grey histograms shows the number distribution in Stark et al. (2010) at $z \approx 3.75$, $z \approx 5.0$ and $z \approx 6.0$. Dark-grey dashed and dotted lines indicate the M_{1500} for 90% completeness at $z \approx 4$, 5, and 6 in Arrabal Haro et al. (2018) and the magnitude cut at $z \approx 6$ in De Barros et al. (2017), respectively.

is a sample from the literature, which is biased towards bright objects including few B and V dropouts with magnitudes fainter than 26 mag in $F850LP$ (z) according to Stark et al. (2010). Generally, the M_{1500} ranges used in LAE fraction studies in the literature are close to those in Stark et al. (2010, 2011). Recently, Arrabal Haro et al. (2018) tested the UVLF of their LBG sample (mainly constructed from photo-z samples) used in their LAE fraction study. At $z \approx 4$ and $z \approx 5$, their LBG samples follow the UVLF in Bouwens et al. (2015b) at $M_{1500} \lesssim -20.5$ mag (dark-grey dashed lines in Fig. 2). However, it is still not complete in the UV for the faint sample ($-20.25 \leq M_{1500} \leq -18.75$ mag) just as Stark et al. (2011).These comparisons illustrate the methodological improvement of our study in terms of the M_{1500} completeness of the sample of galaxies for which we estimate the LAE fraction.

For future reference, $N_{1500}(z_p, M_{1500})$ denotes the number of galaxies with a photometric redshift z_p and absolute magnitude M_{1500} within the given ranges (see Table 1 which summarizes different samples). As discussed above, our UV-selected samples are volume-limited and N_{1500} is directly measured from the catalog with no need for incompleteness corrections.
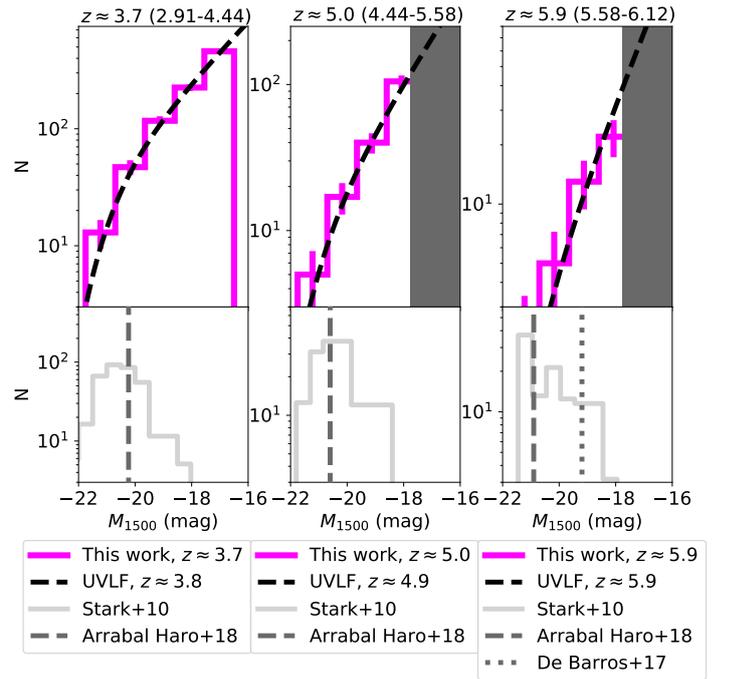
2.2. Counting LAEs within the UV-selected sample

2.2.1. MUSE data

The data of the MUSE HUDF Survey were obtained as part of the MUSE GTO program (PI: R. Bacon). The MUSE HUDF Survey design is presented in Bacon et al. (2017). It consists of two layers of different depths: the mosaic is composed of nine MUSE pointings that cover a $3' \times 3'$ area (9.92 arcmin^2) with an integration time of approximately 10 h; the udf-10 is a deeper integration at a $1' \times 1'$ sub field within the mosaic, with an integration time of ≈ 31 hours. MUSE covers a wide optical wavelength range, from 4750 Å to 9300 Å, which allows the observation of the Lyα line from $z \approx 2.9$ to $z \approx 6.6$. The typical spectral resolving power is $R = 3000$, with a spectral sampling of 1.25 Å. The spatial resolution (pixel size) is $0\rlap{.}''2 \times 0\rlap{.}''2$ per pixel.

In the present paper, we use the latest data release (the second data release, hereafter DR2) from the MUSE HUDF (Bacon et al., in prep.). The improved data reduction process results in data cubes with fewer systematics and a better sky subtraction. The FWHM of the Moffat point spread function (PSF) is $0\rlap{.}''65$ at 7000 Å in the MUSE HUDF. The estimated 1σ surface brightness limits are 2.8×10^{-20} erg s^{-1} cm^{-2} Å^{-1} arcsec^{-2} and 5.5×10^{-20} erg s^{-1} cm^{-2} Å^{-1} arcsec^{-2} in udf-10 and mosaic, respectively, in the wavelength range of 7000–8500 Å (excluding

regions of OH sky emission, see Inami et al. 2017; Bacon et al., in prep., for more details). For instance, the estimated 3σ flux limits are 1.5×10^{-19} erg s^{-1} cm^{-2} and 3.1×10^{-19} erg s^{-1} cm^{-2} in udf-10 and mosaic, respectively, for a point-like source extracted over three spectral channels (i.e., 3.75 Å) around 7000 Å (see Fig. 20 in Bacon et al. 2017). The PSF and noise characteristics are similar to the DR1 data, except in the reddest part of the wavelength range.

In order to measure the fraction of galaxies which have a strong Lyα line, we first extract a 1D spectrum from the MUSE cube for each HST source in our parent sample. We proceed as follows. First, we convolve the HST segmentation map of the R15 catalog with the MUSE PSF, which is normalized to 1. To obtain a spatial mask applicable to MUSE observations for each object, we apply a threshold value of 0.2 to the normalized convolved segmentation map. The median value of the radius of the normalized mask is $\approx$1″0 to 0″8 arcsecond at $z \approx 3$ to 6, which is not affected by Lyα halo flux (Leclercq et al. 2017, see Sect. 2.5 for more detailed discussion of our choice). Second, we integrate the cube spatially over the extent of the mask. We note that PSF weighted or white-light weighted integrations are used to extract spectra in the DR1 catalog of Inami et al. (2017). These provide a higher S/N in the extracted spectrum. However, in the present paper, we do not use a spatial weighting. This results in slightly lower S/N values, but more accurate estimates of the fluxes (i.e., conserved flux), which are needed to assess the completeness of our Lyα detections. Third, we subtract local residual background emission from the extracted spectrum for the 1D spectra as in Inami et al. (2017). The local background is defined in 5″ × 5″ subcubes avoiding the masks of any source.

2.2.2. Search for Lyα emission

In order to detect Lyα emission lines in the 1D spectra extracted above, we use a customized version of the MARZ software[4] (Hinton et al. 2016) described in Inami et al. (2017). MARZ compares 1D spectra to a list of templates and returns the best-fitting spectroscopic redshift, the best-fitting 1D template, and a confidence level for the result (called the quality operator, QOP). In our customized MARZ version, the list of templates consists of templates made using MUSE data, and the interface is improved (Inami et al. 2017). We use our version of MARZ in a similar manner to Inami et al. (2017), except for the two following changes. First, we do not activate cosmic ray replacement in MARZ because (1) it affects the detectability of bright and spectrally peaky Lyα emission and (2) cosmic rays are efficiently removed in the data reduction. Second, we only use template spectra with Lyα emission: those of IDs= 10, 18, and 19, which are used in Inami et al. (2017), and those of IDs= 25, 26, 27, and 30, which are newly built from MUSE data in Bacon et al. (in prep.) and show single-peaked Lyα (see Appendix B for the template spectra). As in Inami et al. (2017), we use the 1D spectra and source files including the subcubes and cutouts of HST UV to NIR images for the parent sources as input for MARZ.

To select robust Lyα detections, we keep only galaxies which MARZ identifies as LAEs with a high confidence level ("Great" and "Good" shown in Fig. 1 in Inami et al. 2017)[5]. Sources

with lower confidence levels are not regarded as detected LAEs. According to this selection, among the 3233 (402) sources in mosaic (udf-10), 374 (70) are LAEs. However, some of these LAE candidates are in fact [O II] emitters or non-LAEs polluted by extended Lyα emission from LAE neighbors. We visually inspect all the LAE candidates as in Inami et al. (2017): we check the entire MUSE spectra, Lyα line profiles, MUSE white-light images, MUSE narrow-band images of Lyα emission, all the existing HST UV to NIR images, and HST colors by eye using the customized MARZ. The MUSE white-light image is created by collapsing the 5″×5″ MUSE subcubes in wavelength direction (see Inami et al. 2017), while the narrow-band image for the Lyα emission is extracted from the wavelength range around the Lyα emission in the subcubes for the sources (see Drake et al. 2017a,b). In contrast to Inami et al. (2017), we also use a consistent photometric redshift (95% uncertainty range) as an evidence of Lyα emission. As a result, we have 276 (58) LAEs at $z \approx 2.9$–6.1 among the parent sample in the mosaic (udf-10) field. Most of the removed sources ($\approx$80%) have a 1D spectrum contaminated by (extended) Lyα emission from neighboring objects, which can be distinguished using the Lyα narrow-band images, MUSE white-light images, and HST images. We show an example of Lyα contamination in Appendix C.

2.2.3. Measurement of Lyα fluxes

For our LAEs, we measured the Lyα fluxes from the 1D spectra used for Lyα detection and described in Sect. 2.2.1. The aperture size is defined by the R15's segmentation map for each source convolved with the MUSE PSF (see Sects. 2.2 and 2.5 for more details). It has been reported that Lyα emission is often spatially offset from the stellar UV continuum (e.g., Erb et al. 2019; Hoag et al. 2019b). The typical offset values for our LAEs are however measured to be less than 0″1 (Leclercq et al. 2017), significantly less than the PSF scale of our observations. We therefore assume that all the flux of the central Lyα component can be captured by our apertures, centered on the continuum emission peak, and with median radius ranging from $\approx$1″0 to 0″8 at $z \approx 3$ to 6. We fit the Lyα emission either with an asymmetric Gaussian or a double Gaussian profile. The choice between these two solutions is made after visual inspection. In practice, we use the gauss_asymfit and gauss_dfit methods of the publicly available software MPDAF (Piqueras et al. 2017)[6]. We use the spectroscopic redshift from MARZ as information for the center wavelength of the fit, with a fitting range of rest-frame 1190 Å to 1270 Å. We inspect all the Lyα spectral profile fits to confirm their validity.

Once we had measured the Lyα fluxes, we computed the rest-frame Lyα equivalent width using M_{1500} and β defined in Sect. 2.1 to estimate the continuum. The distribution of EW(Lyα) as a function of redshift is shown in Fig. 3.

---

[4] The original MARZ in Hinton et al. (2016) is based on a cross-correlation algorithm (AUTOZ, Baldry et al. 2014) and is publicly available at: https://github.com/Samreay/Marz

[5] The confidence level is given as QOP, which is calculated from the peak values of the cross-correlation function (figure of merit, hereafter
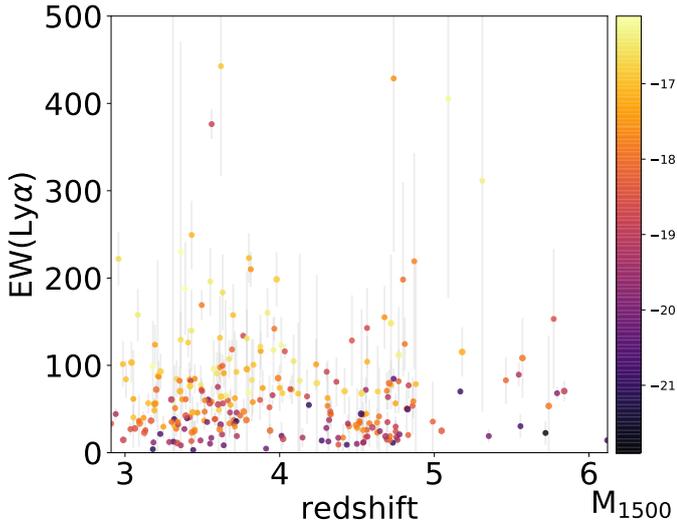
FOM, see Sect. 5.3 in Hinton et al. 2016, for more details). In our version of MARZ (Inami et al. 2017), QOP = 3 (QOP = 2) is regarded as "great" ("good") and corresponds to 99.55% (95%) confidence in the original MARZ (Hinton et al. 2016). However, the original relation between QOP and confidence percentage is calibrated with SED templates different from ours, and the confidence percentage may not be directly applicable to our data. We note that the FOM criterion for QOP = 3 (QOP = 2) in our version of MARZ is the same as that for QOP $\geq$ 4 (QOP = 3) in the original MARZ (see Fig. 1 in Inami et al. 2017 and Fig. 12 in Hinton et al. 2016).

[6] MPDAF, the MUSE Python Data Analysis Framework, is publicly available from the following link: https://mpdaf.readthedocs.io/en/latest/

Fig. 3. $EW(Ly\alpha)$ versus z_z for our final LAE sample. Colors show the M_{1500}.

2.2.4. Completeness estimate and correction

In order to estimate the detection completeness of Lyα emission for the MUSE HUDF data with MARZ, we inserted fake Lyα emission lines into 1D spectra and try to detect them as explained in Sect. 2.2.2.

We took realistic noise into account by creating 1D sky-background spectra from MUSE sub-cubes extracted for different continuum-selected sources, as detailed in Sect. 2.2.1. We chose a sample of spectra which show a clear Lyα emission line (detected with a very high confidence level of "great" shown in Fig. 1 in Inami et al. 2017), and which do not have continuum or other spectroscopic features. We masked the spectral pixels covered by the Lyα emission in these 1D spectra (including ±20 pixels $\approx \pm 25$ Å around the line center). We note that we did not insert fake Lyα lines in the masked regions. With this procedure, we obtained 131 (35) 1D sky-background spectra in the mosaic (udf-10) field.

We added fake emission lines with fluxes taking 18 values that are regularly spaced in the log between 6×10^{-19} erg s^{-1} cm^{-2} and 50×10^{-18} erg s^{-1} cm^{-2}, and 3102 redshift values regularly distributed between $z = 2.93$ and $z = 6.12$. For each flux-redshift pair, we drew 4 lines (yielding a total of 223 344 lines) and added each of them to one of our 131 (35) 1D spectra chosen at random. Each fake Lyα line, has line-shape parameter values (total FWHM and FWHM ratio of red wing to blue wing) randomly drawn from the measured distribution of Lyα emission line shapes of LAEs used in Bacon et al. (2015) and Hashimoto et al. (2017a). We used the add_asym_gaussian method of MPDAF to generate the fake lines and added them to our test spectra.

We then repeated the detection procedure of Sect. 2.2.2, applying the same cut at a confidence level of "good". In each field (udf-10 or mosaic), we computed the completeness of Lyα detection as a function of the Lyα flux $f_{Ly\alpha}$ for five redshift bins: $2.91 \leq z < 3.68$ ($z \approx 3.3$), $3.68 \leq z < 4.44$ ($z \approx 4.1$), $4.44 \leq z < 5.01$ ($z \approx 4.7$), $5.01 \leq z < 5.58$ ($z \approx 5.3$), and $5.58 \leq z \leq 6.12$ ($z \approx 5.9$), which are defined from the redshift bins used to derive M_{1500}. We fit each simulated completeness curve with a formula based on the error function, (e.g., Rykoff et al. 2015):

$$C(f_{Ly\alpha}) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{-2.5\log_{10} f_{Ly\alpha} - a}{\sqrt{2b}}\right)\right], \quad (1)$$
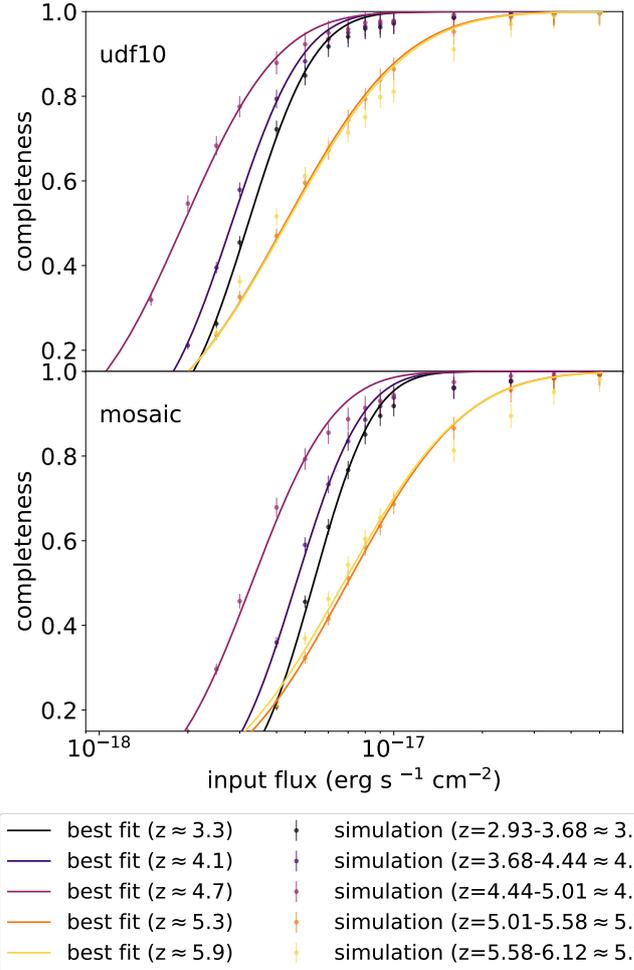

Fig. 4. Completeness of Lyα detection as a function of Lyα flux in the udf-10 (*upper panel*) and mosaic (*lower panel*) fields. The simulated data points and their best-fit completeness functions are indicated by circles and lines, respectively. Black, purple, violet, orange and yellow colors represent redshifts $z \approx 3.3$, 4.1, 4.7, 5.3, and 5.9, respectively. Error bars are calculated from the Poisson errors of the numbers of the detected fake emission lines.

where a and b are two free parameters for fitting (see Fig. 4). We used the function curve_fit from scipy.optimize to perform the fit. The best-fit parameters for the completeness curve in the mosaic and udf-10 are summarized in Table 2.

At completeness above ≈0.8, our best-fit relations slightly overestimate the measured completeness. The analytic fit is however at most $\approx5\%$ above the 1σ upper errors, and this does not have a noticeable effect on the calculation of X_{LAE}. We nevertheless took this into account in the error propagation of X_{LAE} in Sect. 2.3.

Theoretically, completeness functions should just scale with S/N and thus be applicable throughout the wavelength (or redshift) range of the instrument. For the MUSE-WIDE survey, Herenz et al. (2019) indeed find that the shape of their completeness function is independent of redshift. As expected, we also find very similar behavior of completeness at $z \approx 3.3$ and 4.1, in both fields. At these redshifts, the noise is well behaved and there are only few accidents due to sky-line removal in the spectra. At $z \approx 4.7$, the shape of the completeness curve is still well described by Eq. (1), but the curve is shifted to fainter flux with a shallower slope. At $z \approx 5.3$ and 5.9, the shapes of the best-fitting completeness are different from those at lower

Table 2. Best-fit parameters of completeness functions.

Mean z	a in udf-10	b in udf-10	a in mosaic	b in mosaic
$z \approx 3.3$	43.2	0.163	43.7	0.202
$z \approx 4.1$	43.3	0.197	43.9	0.228
$z \approx 4.7$	43.7	0.303	44.3	0.405
$z \approx 5.3$	42.9	0.614	43.4	0.629
$z \approx 5.9$	42.9	0.653	43.4	0.647

Notes. Best-fit parameter of a and b of Eq. (1) in each redshift bin for the mosaic and udf-10 fields.

redshifts: they have a shallower slope, the data points with completeness above 0.8 are not well fit, and the completeness at a given flux is much lower than that at lower redshifts. The lower normalisation and distorted shape of the completeness may be caused by the many sky emission lines at high redshifts (at $z \gtrsim 5$, Drake et al. 2017a). Because MARZ is not a local line detector, as opposed, e.g., to the matched-filtering approach implemented in the tool LSDCat utilized in Herenz et al. (2019) where the filter has a compact support in spectral space, MARZ is affected by relatively long-range noise or distant spectroscopic features in the spectra. Thus, MARZ often does not return a high confidence level ("great" and "good") for LAEs at $z \gtrsim 5$, and our completeness goes down at $z \gtrsim 5$ even for relatively bright Lyα fluxes. We note that the LAE template spectra that we use all have a single-peaked Lyα profile (see Fig. B.1). However, the exact shape of the line profile is shown to have little impact on the detectability with cross-correlation function in general (for instance, see Sect. 4.3 in Herenz & Wisotzki 2017). The shape of the Lyα line of MARZ's template should not affect significantly the detection rate of LAEs (see also Appendix B). Indeed, for example, our sample contains LAEs with double-peaked lines even though none of our templates have such features. We experimented using all the templates in MARZ, including templates of different galaxy populations such as [O II] and [O III] emitters, and we found only a small impact on our LAE sample, which is well within the error bars. Finally, we checked the dependence of completeness on the FWHM of fake Lyα emission lines at fixed flux, and again found no significant trend.

In the following, $N_{\mathrm{LAE}}^{\mathrm{det}}(z_s, M_{1500}, EW)$ denotes the number of galaxies with detected Lyα emission and with spectroscopic redshift z_s, absolute magnitude M_{1500}, and rest-frame equivalent width EW in given ranges. We estimated the true number of LAEs with a corrected value $N_{\mathrm{LAE}}^{\mathrm{corr}}$ defined as follows. For a given field and redshift bin, we used the fits to the completeness function above to define four Lyα flux bins which correspond to regularly spaced bins in the logarithm of the completeness (C), ranging from $C = 0.1$ to $C = 0.9$. We then counted the number of detected LAEs (within a given z_s, M_{1500}, and EW bin) in each flux bin and divide it by the mean completeness (in log) in each of the flux bins. We computed $N_{\mathrm{LAE}}^{\mathrm{corr}}$ as the sum of these over the four flux bins. When the uncertainties of the LAE fraction is calculated, the completeness correction value in each flux bin is propagated, and the uncertainty of completeness correction itself is taken into account as described in the next section. The number of flux bins is defined through a test described in Appendix D. Four to six bins are a sweet spot where the error bars are small and they appear converged. For a larger number of bins, we often get flux-bins with no object at all, and these bins contribute to a large error bar. For a smaller number of bins, we introduce a larger error on the completeness correction (averaged over the bin). Here we adopted four bins.

2.3. X_{LAE} and its error budget

Knowing the number of UV-selected galaxies in a volume-limited sample ($N_{1500}(z_p, M_{1500})$, Sect. 2.1), and the number of LAE among the "inclusive parent sample" ($N_{\mathrm{LAE}}^{\mathrm{corr}}(z_s, M_{1500}, EW)$, Sect. 2.2.4), the fraction of UV-selected galaxies with a Lyα line is simply given as:

$$X_{\mathrm{LAE}} = \frac{N_{\mathrm{LAE}}^{\mathrm{corr}}(z_s, M_{1500}, EW)}{N_{1500}(z_p, M_{1500})}. \qquad (2)$$

The uncertainties on X_{LAE} arise from four components: (1) the uncertainty due to contaminants (type II error) and missed objects (type I error) in $N_{1500}(z_p, M_{1500})$, (2) the uncertainty due to the completeness correction of $N_{\mathrm{LAE}}^{\mathrm{corr}}(z_s, M_{1500}, EW)$, (3) the uncertainty for Bernoulli trials (i.e., fraction of $N_{\mathrm{LAE}}^{\mathrm{corr}}(z_s, M_{1500}, EW)$ over $N_{1500}(z_p, M_{1500})$) measured by a binomial proportion confidence interval and (4) the uncertainty due to cosmic variance. We note that there is no obvious sample selection bias in our UV galaxies as shown in Fig. 2, and our Lyα measurements are homogeneous over the whole redshift range (see discussion in Sect. 4.1). We estimated the relative uncertainty of X_{LAE} from error propagation and discuss each of these contributions below.

Contaminants and missed objects in $N_{1500}(z_p, M_{1500})$. Sources with z_p in a given z range that are truly located outside of the z range are contaminants in the parent continuum-selected sample, while sources with z_p outside of the z that are truly located in the z range are missed sources. These mismatches of z_p can happen because of confusion between Lyman and 4000 Å breaks in the SED fitting. In addition, IGM absorption modeling has been suggested to affect z_p estimation (Brinchmann et al. 2017). As discussed in Inami et al. (2017) and Brinchmann et al. (2017), the fraction of contaminants is very low for high confidence level objects (with secure redshift, see Fig. 20 in Inami et al. 2017). The fraction of missed objects with a relative redshift difference of more than 15%, $|z_s$-$z_p|/(1+z_s)>$ 0.15, is suggested to be $\approx 10\%$ (outlier fraction, Brinchmann et al. 2017). Since the missed objects whose 95% uncertainty range for z_p is outside the z range for MUSE-LAEs are not included in our parent continuum-selected sample, they are also not included in our LAE sample. With an assumption that the fractions of missed objects are the same for the parent and LAE samples, the uncertainties due to missed objects can be neglected as well as those due to contaminants. We note that we do not find any significant relation between the z_p-z_s difference and Lyα EW as well as Lyα flux.

Completeness correction of $N_{\mathrm{LAE}}^{\mathrm{corr}}(z_s, M_{1500}, EW)$. Simulated data of completeness are fitted well with Eq. (1) at $z \approx 3.3$, 4.1, and 4.7. Even at $z \approx 5.3$ and 5.9, the differences in completeness between the simulated data and the best-fitting functions are at most $\approx 5\%$, which is much smaller than the uncertainty due to the flux binning. The flux binning for the completeness correction described in Sect. 2.2.4 causes an uncertainty of at most $\approx \pm 32\%$. The completeness bins, spaced regularly in log from 0.1 to 0.9, correspond to steps of a factor 1.73, which has a square-root of ≈ 1.32. We used this very conservative estimate of 32% for the completeness correction error in our error budget. The completeness correction error is smaller than the error component (3) as described later and does not change the total uncertainty of X_{LAE}. We note that the completeness correction value is also taken into account according to error propagation, when the error component (3) is calculated.

Uncertainties for Bernoulli trials. Measuring the fraction of a sub-sample among a parent sample is a kind of experiment of Bernoulli trials. An uncertainty for a Bernoulli trial is given by a binomial proportion confidence interval (BPCI). We used the python module binom_conf_interval from astropy.stats and provided an approximate uncertainty for a given confidence interval (=68%, 1σ, in this work), the number of trials, and the number of successes. Here, the number of trials and the number of successful experiments are $N_{1500}(z_p, M_{1500})$ and $N_{LAE}^{rcorr}(z_s, M_{1500}, EW)$, respectively. However, we cannot obtain $N_{LAE}^{rcorr}(z_s, M_{1500}, EW)$ directly from the observations. To include the effect of the completeness correction in each Lyα flux bin (described in Sect. 2.2.4) in the error propagation, we calculated the uncertainty of the LAE fraction in each Lyα flux bin without applying a completeness correction from binom_conf_interval and multiplied the uncertainties by the correction value in the flux bin. We chose as an approximation formula the Wilson score interval (Wilson 1927), which is known to return an appropriate output even for a small number of trials and/or success experiments. Our method is confirmed to be accurate by numerical tests described in Appendix E. For the flux-bins with no LAEs, the average completeness value among the bins and the number of LAEs (= 0) are used to derive the uncertainties conservatively. When we summed over all the uncertainties in a flux bin to derive the total uncertainties for the component (3), a python module, add_asym developed in Laursen et al. (2019), is used to treat asymmetric errors by BPCI. We note that the Poisson errors of $N_{LAE}^{corr}(z_s, M_{1500}, EW)$ and $N_{1500}(z_p, M_{1500})$ are commonly used in the literature. However, the error for the LAE fraction should be derived by BPCI like in other fraction studies (e.g., the galaxy merger fraction, Ventou et al. 2017) to obtain statistically correct errors. The BPCI method is reviewed for astronomical uses in Cameron (2011).

Cosmic variance. The survey volume in each redshift range is limited to $\approx 1.5-2.5 \times 10^4$ cMpc^3. However, we find that the uncertainty due to cosmic variance is less than the BPCI error and thus not affecting our X_{LAE} significantly (see Sect. 4.2.3 for more details). Since the uncertainty due to cosmic variance cannot be included in our MUSE measurements, we neglected this error component (4).

In addition to (1)–(4), uncertainties in photo-z estimations and flux measurements are also potential error components. In Appendix A, we discuss the impact of uncertainties on z_p and conclude that it does not affect the error bar of X_{LAE} significantly. Uncertainties of M_{1500}, β, and the Lyα flux, combine into uncertainties on $EW(Ly\alpha)$. As shown in Figs. 5 and 6, X_{LAE} shows a slight dependence on M_{1500} and $EW(Ly\alpha)$. We thus expect that a small error on these quantities will translate into an even smaller error on X_{LAE}. We note that although some objects have large errors in M_{1500} in Fig. 2, they are not included in our analysis because they are very faint[7]. Thus, we ignored these two uncertainties, as is commonly done in the rest of the literature.

---

[7] Among the subsamples shown in Table 1, fainter-M_{1500} and higher-z_p objects have greater uncertainties in M_{1500}. The medians (standard deviations) of the uncertainties for the subsamples with $-18.75 \leq M_{1500} \leq -17.75$ mag are 0.05 mag (0.02 mag) at $z \approx 3.3$ and 0.15 mag (0.12 mag) at $z \approx 5.6$. Those for the subsamples with $-19.0 \leq M_{1500} \leq -18.0$ mag is 0.04 mag (0.01 mag) at $z \approx 3.3$ and 0.08 mag (0.11 mag) at $z \approx 4.1$. These uncertainties are much smaller than the width of M_{1500} bins. With regard to S/N cuts of Lyα fluxes corresponding to $EW(Ly\alpha)$ cuts, those in the mosaic field for $M_{1500} = -17.75$ mag and $EW(Ly\alpha) = 65$ Å are estimated to be ≈ 17.4 at $z \approx 3.3$ and ≈ 4.4 at $z \approx 5.6$, if we assume $\beta = -2$. Similarly, the S/N cuts for $M_{1500} = -18.0$ mag and $EW(Ly\alpha) = 50$ Å are ≈ 16.9 at $z \approx 3.3$ and ≈ 11.1 at $z \approx 4.1$. Here, the noise is the median of those shown in Fig. 8 in each redshift bin.
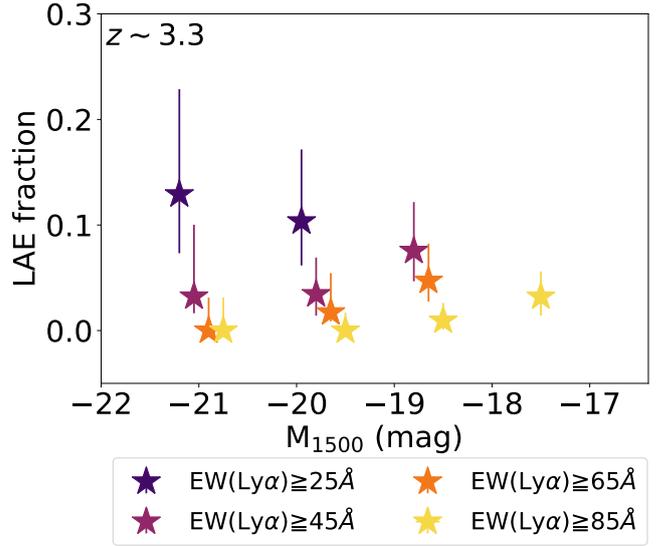

Fig. 5. X_{LAE} vs. M_{1500} at $z \approx 3.3$ (= 2.91–3.68) for $M_{1500} \in [-21.5; -17.0]$. Purple, violet, orange and yellow stars indicate our MUSE results with $EW(Ly\alpha) \geq 25$ Å, $EW(Ly\alpha) \geq 45$ Å, $EW(Ly\alpha) \geq 65$ Å, and $EW(Ly\alpha) \geq 85$ Å, respectively. For visualization purposes, we show the width of M_{1500} only for the violet stars and slightly shift the other points along the abscissa.
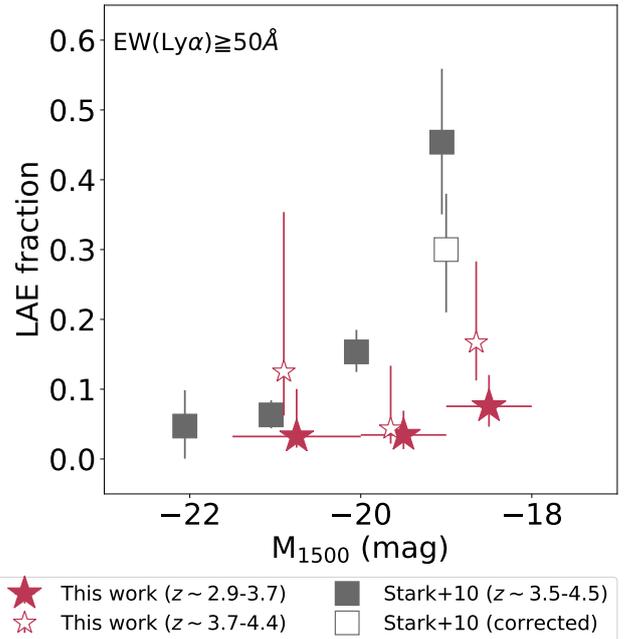

Fig. 6. X_{LAE} vs. M_{1500} for $EW(Ly\alpha) \geq 50$ Å at $z \approx 3$-4. Our X_{LAE} at $z \approx 3.3$ ($\approx$ 2.9–3.7) and $z \approx 4.1$ ($\approx$3.7–4.4), are indicated by filled and open magenta stars, respectively. Stark et al. (2010)'s X_{LAE} at $z \approx 3.5$–4.5 is shown by filled grey squares. The open grey square indicates the corrected X_{LAE} value of Stark et al. (2010) at $M_{1500} = -19$ mag (see Sect. 4.1.1). For visualization purposes, we slightly shift the magenta open stars and the grey open square along the abscissa and show the width of M_{1500} only for the red filled stars.

With these considerations, the uncertainty of the LAE fraction is the quadratic sum of the uncertainty terms (2) and (3). Below, the error bars on X_{LAE} represent the 68% confidence intervals around the values calculated by Eq. (2). We note that the dominant error for X_{LAE} derived in this work is

component (3), the uncertainties for a Bernoulli trial, which are, for instance, 38% and 78% of X_{LAE} for $-21.75 \leq M_{1500} \leq -18.75$ mag and $EW(\mathrm{Ly}\alpha) \geq 65$ Å at $z \approx 3$ and 5.6, respectively.

2.4. Measurement of the slope of X_{LAE} as a function of z and M_{1500}

We measured linear slopes of X_{LAE} as a function of z and M_{1500} using a python package for orthogonal distance regression (ODR) fitting, scipy.odr, to account for widths of bins in x-axis and uncertainties of X_{LAE} in y-axis. The ODR fitting minimizes the sum of squared perpendicular distances from the data to the fitting model. Since uncertainties of X_{LAE} are not symmetric, a Monte Carlo simulation with 10 000 trials is used. We assumed an asymmetric Gaussian profile as a probability distribution function for X_{LAE} with each of the upper and lower uncertainties at a given bin (z or M_{1500}) in x-axis. We fit a linear relation ($y = ax + b$) in each trial drawing X_{LAE} randomly with scipy.odr. The best-fit values of a and b and their error bars are derived from the median values and the 68% confidence intervals around the median values. The results of the fitting are shown in Sect. 3.

2.5. Extended Lyα emission

Our aim is to measure how the fraction of UV-selected galaxies showing a strong Lyα line varies with redshift. We made the choice to discard possible significant contributions to the Lyα luminosities by extended Lyα haloes (LAH, e.g., Wisotzki et al. 2016; Drake et al. 2017a; Leclercq et al. 2017, Leclercq et al. in prep.) in our study, though the contribution of LAHs to the total Lyα fluxes is typically more than ≈50% (e.g., Momose et al. 2016; Leclercq et al. 2017). There are a number of reasons for this choice. First, it is largely motivated by our lack of understanding of the physical processes lighting up these halos. In particular, it is not clear how they relate to the UV luminosities of their associated galaxies (e.g., Leclercq et al. 2017; Kusakabe et al. 2019), to what extent they are associated to star formation (e.g., Yajima et al. 2012, their Fig. 12), or whether the nature of this association could vary with redshift. In order to assess the evolution of X_{LAE} with redshift, it thus appears more conservative to limit our measurement of the Lyα emission from galaxies to the part which is most likely to have the same origin as the continuum UV light. Any evolution is then more likely to be related to the evolution of the ionisation state of the IGM. With the above procedure, our 1D spectra include as little as possible of the extended Lyα emission that is found around LAEs (Wisotzki et al. 2016; Leclercq et al. 2017). Second, our choice has the advantage of following a similar methodology without including halo fluxes used in other studies (e.g., Stark et al. 2011; De Barros et al. 2017; Arrabal Haro et al. 2018), and thus allows for a fair comparison. Third, it is difficult to measure the faintest halos. If we demanded that LAHs were detected around galaxies in our sample, we would limit our sample to the brightest or most compact halos only (see Sect. 2.2 and Fig. 8 of Leclercq et al. 2017). So even though an IFU enables us in principle to separate the central and halo component more clearly than slit and fiber spectrometers, the S/N required to do so is still prohibitive for statistical studies such as ours.

We note that because of our choice, the Lyα fluxes and EWs in the present paper are smaller than the total Lyα fluxes and EWs reported by e.g., Hashimoto et al. (2017a), Drake et al. (2017a), and Leclercq et al. (2017).

3. Results

In order to measure the variation of X_{LAE} with redshift or UV absolute magnitude, we designed several sub-samples shown in Table 1. We used $EW(\mathrm{Ly}\alpha)$ cuts starting at 25 Å, which is a common limit in the literature, and then increase in steps of 20 Å to 45 Å, 65 Å, and 85 Å. We also used 50 Å and 55 Å cuts, for comparison to Stark et al. (2010, 2011). In Sect. 3.1, we present our results for the X_{LAE}-z relation, going as faint as $M_{1500} = -17.75$ mag for the first time in a homogeneous way over the redshift range $z \approx 3$ to $z \approx 6$. We discuss how these results compare to existing measurements. In Sect. 3.2, we present the first measurement of the $X_{\mathrm{LAE}}-M_{1500}$ relation for galaxies as faint as $M_{1500} = -17.00$ mag at $z \approx 3$, and compare our findings to other studies. The numerical values of X_{LAE} are summarized in Tables 3 and 4. The slopes of the best-fit linear relations of X_{LAE} as a function of z and M_{1500} are shown in Fig. F.1, and summarised in Tables 3 and 4.

3.1. Redshift evolution of X_{LAE}

We derive the redshift evolution of X_{LAE} for $EW(\mathrm{Ly}\alpha) \geq 65$ Å from $M_{1500} = -21.75$ mag to the faint UV magnitude of -17.75 mag. We show the results in the upper panel of Fig. 7 with the large filled hexagons. We find a weak rise of X_{LAE} from $z \approx 3$ to $z \approx 6$, even though poor statistics do not allow us to set a firm constraint at $z \approx 5.6$. Breaking our sample into a bright end (with $-21.75 \leq M_{1500} \leq -18.75$ mag, purple circles), and a faint end (with $-18.75 \leq M_{1500} \leq -17.75$ mag, purple squares), we find similar trends that are consistent within the 1σ error bars, suggesting that the X_{LAE}-z relation does not depend strongly on the rest-frame UV absolute magnitude. The best-fit linear relations are $X_{\mathrm{LAE}} = 0.07^{+0.06}_{-0.03}z - 0.22^{+0.12}_{-0.24}$, $X_{\mathrm{LAE}} = 0.05^{+0.05}_{-0.03}z - 0.13^{+0.12}_{-0.18}$, and $X_{\mathrm{LAE}} = 0.09^{+0.09}_{-0.04}z - 0.28^{+0.18}_{-0.36}$ for $-21.75 \leq M_{1500} \leq -17.75$ mag, $-21.75 \leq M_{1500} \leq -18.75$ mag, and $-18.75 \leq M_{1500} \leq -17.75$ mag, respectively. We note that the bright sample here is dominated by the more numerous sub-L* galaxies, which are fainter than the bright samples in the literature.

If we select the brighter part of our sample ($-21.75 \leq M_{1500} \leq -18.75$ mag), we can estimate X_{LAE} down to lower equivalent widths. In the lower panel of Fig. 7, the orange circles show X_{LAE} for galaxies with $EW(\mathrm{Ly}\alpha) \geq 45$ Å. This is also found to increase from $z \approx 3$ to $z \approx 4$–6, and is above the relation obtained for $EW(\mathrm{Ly}\alpha) \geq 65$ Å also shown in the lower panel, as expected. The best-fit linear relation is $X_{\mathrm{LAE}} = 0.05^{+0.05}_{-0.03}z - 0.07^{+0.14}_{-0.20}$. These results are qualitatively consistent with the trend of increasing X_{LAE} with increasing z based on bright samples in the literature (see below).

The fact that only ≈0–30% of galaxies within the UV magnitude range at $z \leq 5$ are observed as LAEs with $EW(\mathrm{Ly}\alpha) \geq 65$ Å requires some explanation of the physical mechanisms. We discuss this further in Sect. 4.2.

Next, we compare our MUSE results of X_{LAE} with previous studies. For this purpose, we derive X_{LAE} for M_{1500} in the range $[-20.25; -18.75]$ (which corresponds to the faint UV magnitude range of Stark et al. 2011, see Fig. 1), and for $EW(\mathrm{Ly}\alpha) \geq 25$ Å and $EW(\mathrm{Ly}\alpha) \geq 55$ Å. Figure 8 shows our results and those of other studies as a function of redshift. At $z \lesssim 5$, we confirm the low values from Arrabal Haro et al. (2018) (grey crosses at $z \approx 4$ and $z \approx 5$) for $EW(\mathrm{Ly}\alpha) \geq 25$ Å. Our median values of X_{LAE} at $z \approx 4.1$ and $z \approx 4.7$ are somewhat smaller than those at $z \approx 4$ and $z \approx 5$ from Stark et al. (2011), although they are compatible within the large error bars. At $z \approx 5.6$, our value for

Table 3. LAE fraction as a function of redshift.

Mean z	X_{LAE} or slope	1σ upper error	1σ lower error
$-21.75 \leq M_{1500} \leq -17.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 65\,\text{Å}$			
3.3	0.04	0.02	0.01
4.1	0.07	0.04	0.02
4.7	0.11	0.07	0.04
5.6	0.20	0.16	0.06
Modest positive correlation:			
Slope	0.07	0.06	0.03
$-21.75 \leq M_{1500} \leq -18.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 45\,\text{Å}$			
3.3	0.05	0.03	0.02
4.1	0.16	0.10	0.06
4.7	0.21	0.13	0.08
5.6	0.13	0.13	0.05
Modest positive correlation:			
Slope	0.05	0.05	0.03
$-21.75 \leq M_{1500} \leq -18.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 65\,\text{Å}$			
3.3	0.02	0.02	0.01
4.1	0.09	0.07	0.04
4.7	0.12	0.09	0.05
5.6	0.13	0.13	0.05
Modest positive correlation:			
Slope	0.05	0.05	0.03
$-18.75 \leq M_{1500} \leq -17.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 65\,\text{Å}$			
3.3	0.05	0.04	0.02
4.1	0.05	0.05	0.02
4.7	0.10	0.09	0.03
5.6	0.25	0.24	0.09
Modest positive correlation:			
Slope	0.09	0.09	0.04
$-20.25 \leq M_{1500} \leq -18.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 25\,\text{Å}$			
3.3	0.13	0.07	0.05
4.1	0.25	0.14	0.09
4.7	0.32	0.22	0.11
5.6	0.13	0.13	0.05
No correlation (flat):			
Slope	0.01	0.05	0.05
$-20.25 \leq M_{1500} \leq -18.75\,\text{mag}$, $EW(\text{Ly}\alpha) \geq 55\,\text{Å}$			
3.3	0.06	0.04	0.03
4.1	0.10	0.09	0.04
4.7	0.18	0.13	0.07
5.6	0.13	0.12	0.05
No correlation (almost flat):			
Slope	0.04	0.05	0.03

Notes. The values and 1σ uncertainties of the LAE fraction as a function of z and the values of the slope are summarized.

$EW(\text{Ly}\alpha) \geq 25\,\text{Å}$ appears to be significantly lower than those reported in the literature. We note that the discrepancy between our work and De Barros et al. (2017) is however only 1.14σ, and might thus be caused by the statistical error. However, our result is more than 2σ away from those of Arrabal Haro et al. (2018) and Stark et al. (2011), which is less likely to be statistical fluctuation. We discuss in Sect. 4.1.1 potential biases which may explain why these two latter references find large values of X_{LAE}. We discuss the effect of cosmic variance in Sect. 4.2.3, and find that it cannot explain such large differences. Another possibility is that the low value we find reflects a late and/or patchy reionization process, and we discuss that further in Sect. 4.3.

Table 4. LAE fraction as a function of UV magnitude.

Mean M_{1500}	X_{LAE} or slope	1σ upper error	1σ lower error
$z \approx 3.3$, $EW(\text{Ly}\alpha) \geq 25\,\text{Å}$			
-20.75	0.13	0.10	0.06
-19.50	0.10	0.07	0.04
No correlation (flat):			
Slope	-0.02	0.07	0.08
$z \approx 3.3$, $EW(\text{Ly}\alpha) \geq 45\,\text{Å}$			
-20.75	0.03	0.07	0.02
-19.50	0.03	0.03	0.02
-18.50	0.08	0.05	0.03
No correlation (flat):			
Slope	0.01	0.02	0.03
$z \approx 3.3$, $EW(\text{Ly}\alpha) \geq 65\,\text{Å}$			
-20.75	0.00	0.03	0.00
-19.50	0.02	0.04	0.01
-18.50	0.05	0.04	0.02
Modest positive correlation:			
Slope	0.02	0.01	0.01
$z \approx 3.3$, $EW(\text{Ly}\alpha) \geq 85\,\text{Å}$			
-20.75	0.00	0.03	0.00
-19.50	0.00	0.02	0.00
-18.50	0.01	0.02	0.00
-17.50	0.03	0.03	0.01
No correlation (flat):			
Slope	0.01	0.01	0.01
$z \approx 3.3$, $EW(\text{Ly}\alpha) \geq 50\,\text{Å}$			
-20.75	0.03	0.07	0.02
-19.50	0.03	0.03	0.02
-18.50	0.08	0.04	0.03
No correlation (flat):			
Slope	0.01	0.02	0.03
$z \approx 4.1$, $EW(\text{Ly}\alpha) \geq 50\,\text{Å}$			
-20.75	0.12	0.23	0.06
-19.50	0.04	0.09	0.02
-18.50	0.17	0.12	0.05
No correlation (flat):			
Slope	0.01	0.06	0.10

Notes. The values and 1σ uncertainties of the LAE fraction as a function of M_{1500} and the values of the slope are summarized.

We also check the best-fit linear relation of X_{LAE} as a function of z for $EW(\text{Ly}\alpha) \geq 25\,\text{Å}$, which is $X_{\text{LAE}} = 0.01^{+0.05}_{-0.05}z + 0.16^{+0.20}_{-0.22}$. The best-fit slope is lowered by the point at $z \gtrsim 5.6$, and is shallower than 0.11 ± 0.04 in Stark et al. (2011) and $0.18^{+0.06}_{-0.06}$ in Arrabal Haro et al. (2018). It is consistent with the flat relation reported by Hoag et al. (2019b) within the 1σ error bars (0.014 ± 0.02) and that by Caruana et al. (2018). We note that Caruana et al. (2018) discuss the slope of X_{LAE} against z using a sample of the MUSE-Wide GTO survey with an apparent magnitude cut of $F775 < 26.5\,\text{mag}$, which is shown by a grey solid line in Fig. 1. They also include the contribution of extended Lyα halos to their Lyα fluxes, which enhances the values, contrary to us. With regard to $EW(\text{Ly}\alpha) \geq 55\,\text{Å}$ the best-fit relation is $X_{\text{LAE}} = 0.04^{+0.05}_{-0.03}z - 0.05^{+0.14}_{-0.19}$, whose slope is consistent with that in Stark et al. (2011), 0.018 ± 0.036.

3.2. UV magnitude dependence of X_{LAE}

Figure 5 shows a diagram of X_{LAE} and M_{1500} at $z = 2.91$–3.68 (≈ 3.3) for our MUSE sample. This is the first time that
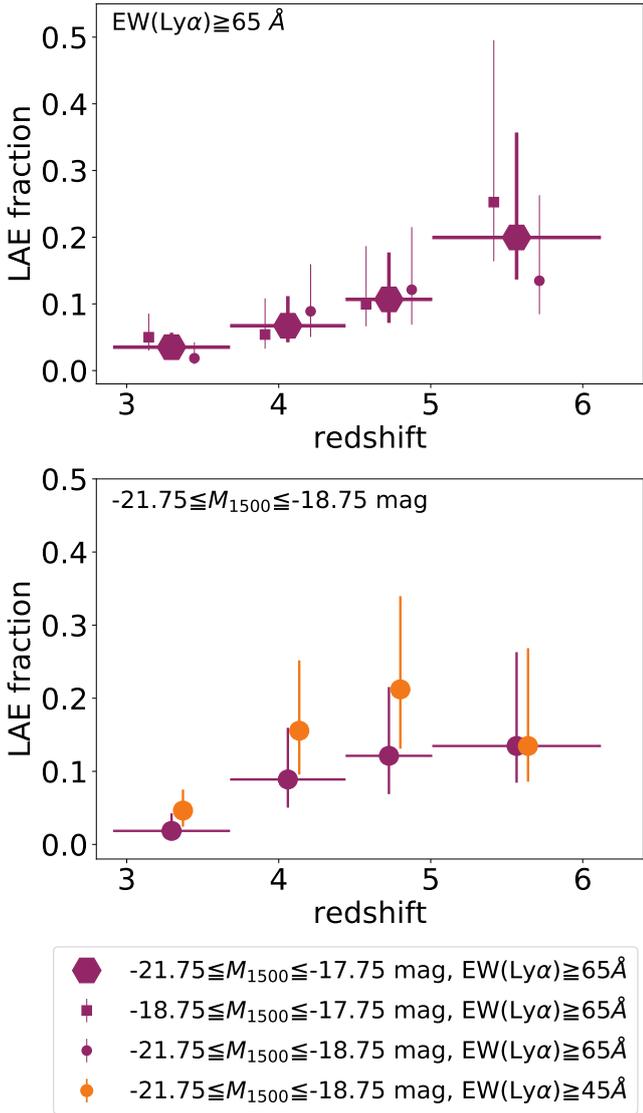
Fig. 7. $X_{\rm LAE}$ vs. z. *Upper panel*: big purple hexagons, small purple circles, and small purple squares indicate LAE fractions for $EW({\rm Ly}\alpha) \geq 65$ Å derived with MUSE at $-21.75 \leq M_{1500} \leq -17.75$ mag, $-21.75 \leq M_{1500} \leq -18.75$ mag, and $-18.75 \leq M_{1500} \leq -17.75$ mag, respectively. *Lower panel*: purple and orange circles represent $X_{\rm LAE}$ for $EW({\rm Ly}\alpha) \geq 65$ Å and $EW({\rm Ly}\alpha) \geq 45$ Å at $-21.75 \leq M_{1500} \leq -18.75$ mag, respectively. For visualization purposes, we show the width of z only for one symbol in each panel and slightly shift the other points along the abscissa.

the dependence of the LAE fraction on M_{1500} is studied at $M_{1500} \geq -18.5$ mag. The LAE fractions for $EW({\rm Ly}\alpha) \geq 25$ Å (45 Å, 65 Å, and 85 Å) are shown with the purple (violet, orange, and yellow) stars. The best-fit linear relations are $X_{\rm LAE} = -0.02^{+0.07}_{-0.08}M_{1500} - 0.30^{+1.39}_{-1.62}$, $X_{\rm LAE} = 0.01^{+0.02}_{-0.03}M_{1500} + 0.33^{+0.45}_{-0.54}$, $X_{\rm LAE} = 0.02^{+0.01}_{-0.01}M_{1500} + 0.41^{+0.18}_{-0.30}$, and $X_{\rm LAE} = 0.01^{+0.01}_{-0.01}M_{1500} + 0.12^{+0.13}_{-0.14}$ for EW cuts of 25 Å, 45 Å, 65 Å, and 85 Å, respectively. We find no clear dependence of $X_{\rm LAE}$ on M_{1500} in tension with the clear rise of $X_{\rm LAE}$ to faint UV magnitude for an EW cut of 50 Å reported in Stark et al. (2010).

Our results also show that the LAE fraction is sensitive to the equivalent width selection, as expected e.g., from Hashimoto et al. (2017a). Although this means that the LAE fraction is useful in itself to test cosmological galaxy

evolution models (see Sect. 4.2 and Forero-Romero et al. 2012; Inoue et al. 2018), it also raises concern for the usage of $X_{\rm LAE}$ as a probe of the IGM neutral fraction at the end of reionization (see also Mason et al. 2018), since homogeneous measurements of Lyα emission over a wide redshift range are required for a fair comparison.

In Fig. 6, our results for the relation between $X_{\rm LAE}$ and M_{1500} for $EW({\rm Ly}\alpha) \geq 50$ Å at $z \approx 3$–4 (filled and open red stars) are compared with those in Stark et al. (2010) (filled grey squares). The best-fit linear relations for our results at $z \approx 2.9$–3.7 and at ≈ 3.7–4.4 are $X_{\rm LAE} = 0.01^{+0.02}_{-0.03}M_{1500} + 0.34^{+0.44}_{-0.55}$ and $X_{\rm LAE} = 0.01^{+0.06}_{-0.10}M_{1500} + 0.33^{+1.12}_{-1.80}$, respectively. We find no dependence of $X_{\rm LAE}$ on M_{1500} as opposed to the claim in Stark et al. (2010), whose best-fit relation is $X_{\rm LAE} = 0.13^{+0.03}_{-0.03}M_{1500} + 2.87^{+0.74}_{-0.72}$. Our $X_{\rm LAE}$ is lower than that in Stark et al. (2010) at UV magnitude fainter than $M_{1500} \approx -19$ mag, and possibly at $M_{1500} \approx -20$ mag. We discuss the difference in $X_{\rm LAE}$ between this work and Stark et al. (2011) in Sect. 4.1.

4. Discussion

In this section, we assess the cause of the differences between our results and previous results, compare our results with predictions from a cosmological galaxy formation model, and discuss the evolution of the LAE fraction and implication for reionization.

4.1. Possible causes of the differences between our MUSE results and previous results

In Fig. 8, we find that our measurements of $X_{\rm LAE}$ are systematically lower than those of Stark et al. (2011), although consistent within the error bars. This tension supports the results of Arrabal Haro et al. (2018), who also find low median values of $X_{\rm LAE}$ at $z \approx 4$–5. It is worth discussing the potential origins of this tension, since the median values of $X_{\rm LAE}$ have been used to assess cosmic reionization in theoretical studies (e.g., Dijkstra 2014). The difference between our study and that of Stark et al. (2010) is best illustrated in Fig. 6, which shows $X_{\rm LAE}$ as a function of M_{1500}. Here, our results are inconsistent with theirs at a faint UV magnitude, even when taking into account the large error bars. Below we discuss two possible origins of this discrepancy in the plot of $X_{\rm LAE}$ as a function of M_{1500}: the LBG selection bias, and systematics due to different observing methods.

4.1.1. LBG selection bias

There is a possibility that the LBG sample of Stark et al. (2010) is biased towards bright Lyα emission that is, higher $X_{\rm LAE}$, as pointed out in previous studies (LBG selection bias, for example, Stanway et al. 2008; De Barros et al. 2017; Inami et al. 2017, see also Cooke et al. 2014 for another potential bias of LBGs due to LyC leakers). In other words, LBG selections could be biased towards having higher $X_{\rm LAE}$ if they preferentially miss low-EW sources. The LBG selection consists of a set of color-color criteria and S/N cuts (e.g., Stark et al. 2009). De Barros et al. (2017) obtain a relatively low median $X_{\rm LAE}$ at $z \approx 6$ and discuss the causes. They use common selection criteria for the i-dropout (Bouwens et al. 2015b), but add an additional criterion of the H-band ($F160W$, rest-frame UV at ≈ 6) magnitude cut. When Lyα emission is located in the red band, strong Lyα emission in a UV spectrum can significantly enhance the Lyman break. The
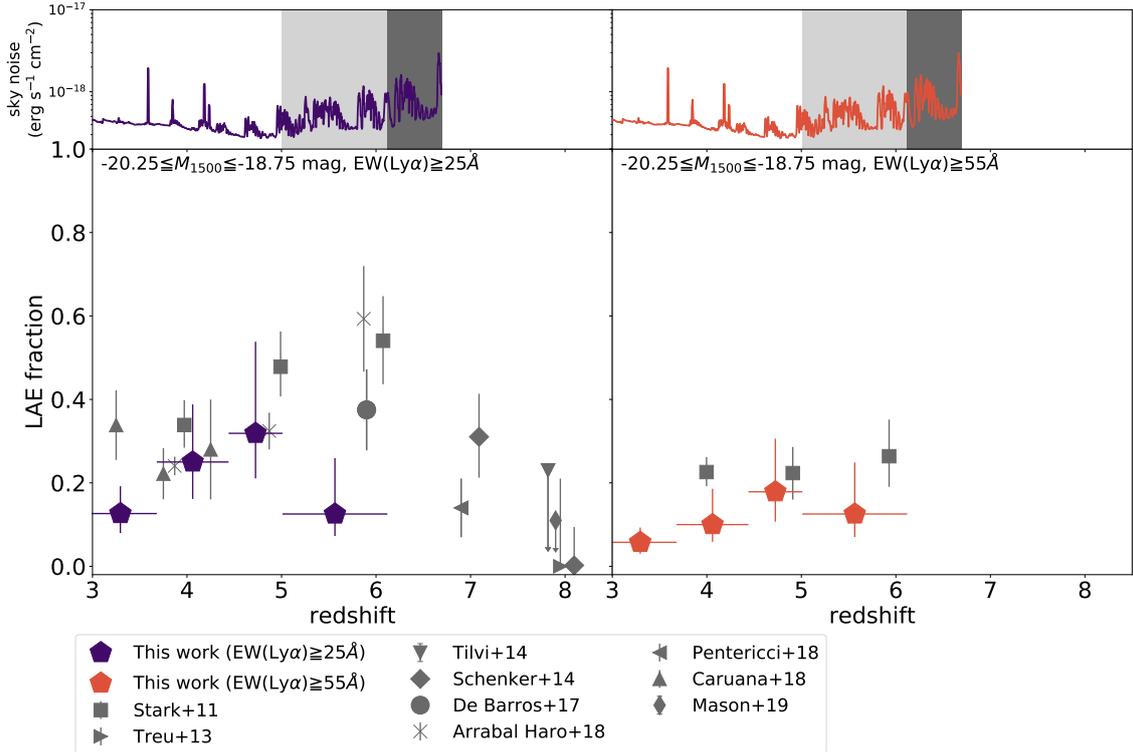
Fig. 8. X_{LAE} vs. z compared with previous results. *Upper left and right panels*: sky noise vs. z. The purple and orange lines show the 1σ flux of the sky noise in the mosaic field (for a wavelength width of $600\,\mathrm{km\,s^{-1}}$). The sky noise in the udf-10 field is ≈ 1.7 times lower than in the mosaic field. The dark-grey and light-grey hashed areas show redshift ranges that are not included in our sample and that are highly affected by sky lines, respectively. *Lower left (right) panel*: X_{LAE} vs. z for $-20.25 \leq M_{1500} \leq -18.75$ mag and $EW(Ly\alpha) \geq 25$ Å ($EW(Ly\alpha) \geq 55$ Å). Purple (orange) pentagons indicate our MUSE results. A grey square, triangle (*right*), diamond, inverted triangle, circle, triangle (*left*), cross, triangle, and thin diamond represent the results by Stark et al. (2011), Treu et al. (2013), Schenker et al. (2014), Tilvi et al. (2014), De Barros et al. (2017), Pentericci et al. (2018), Arrabal Haro et al. (2018), Caruana et al. (2018), and Mason et al. (2019), respectively. The upper limits in Tilvi et al. (2014) and Mason et al. (2019) show the 86% and 68% confidence levels of X_{LAE}, respectively, while other error bars show 1σ uncertainties of X_{LAE}. We note that the original M_{1500} ranges in Treu et al. (2013), Schenker et al. (2014), Tilvi et al. (2014), De Barros et al. (2017), Arrabal Haro et al. (2018), and Mason et al. (2019) are $M_{1500} \geq -20.25$ mag. In Caruana et al. (2018), the original M_{1500} range can be roughly estimated from the apparent $F775W$ cut (see Fig. 1), and they include $Ly\alpha$ halos in the flux measurement. For visualization purposes, we slightly shift the data points of Tilvi et al. (2014) and Arrabal Haro et al. (2018) along the abscissae.

additional criterion in De Barros et al. (2017) can suppress the LBG selection bias, which increases X_{LAE} for faint UV sources as will be discussed below.

Here we estimate the effect on X_{LAE} of two aspects of the LBG selection bias for B ($F435W$)-dropouts: the impact of $Ly\alpha$ contamination on the signal-to-noise ratio cut in the V ($F606W$) band, and on the color-color criteria in a diagram of $F435W$–$F606W$ vs. $F606W$–$F850LP$ (black solid line in Fig. 9). We estimate $F606W$ magnitudes assuming a power-law spectrum with a UV slope of -2 ($\lambda \geq 912$ Å in rest frame) and IGM transmission following Madau (1995). As shown in the upper panel in Fig. 9, strong $Ly\alpha$ emission can increase the flux in $F606W$ noticeably, especially at $z \gtrsim 4$. The magnitude shift becomes larger at higher redshifts because of the increasingly smaller rest-frame wavelength range of the UV continuum that is covered by $F606W$ as the redshift goes up. Even at $z \approx 4$, however, a moderate $Ly\alpha$ emission line with $EW(Ly\alpha) = 50$ Å can cause a ≈ 0.16 mag difference in $F606W$. This affects both the S/N and the colors.

We first illustrate the effect on the signal-to-noise ratio cut of $F606W$ by considering an object with $M_{1500} = -19$ mag. At $z = 4.5$ ($z = 4.0$), a source without $Ly\alpha$ emission has $F606W = 28.3$ mag ($F606W = 27.5$ mag), while a source with $EW(Ly\alpha) = 50$ Å has a 0.27 (0.16) brighter magnitude. Since these $F606W$

magnitudes are close to the 5σ limiting magnitude of 28.0 mag in Stark et al. (2009), which corresponds to a completeness of 50% in the case of a $S/N \geq 5$ cut, the completeness for their B-dropout changes drastically around 28.0 mag.

Second, the effect on the two colors for the B-dropout selection is shown in the lower panel of Fig. 9. The green, red and violet arrows show color-color shifts in the diagram of $F435W$–$F606W$ and $F606W$–$F850LP$ due to the magnitude shift of $F606W$ in the case of $EW(Ly\alpha) = 25$ Å, 50 Å, and 100 Å, respectively. The shift of $F606W$ enhances the possibility of meeting the dropout criteria. Indeed, at $z = 3.5$–4.5, all of our LAEs with an absolute magnitude of $F775W = -20.25$–-18.75 mag are located in the dropout selection region (upper left region of bottom panel in Fig. 9). However, $\approx 10\%$ of continuum-selected galaxies do not meet the dropout criteria. Therefore, strong $Ly\alpha$ emission can enhance the probability to meet LBG-selection criteria both in terms of the signal-to-noise cut and of the color-color criteria.

We estimate the completeness of the B-dropout galaxies in Stark et al. (2009) using a plot of surface number density as a function of UV magnitude for B-dropouts in Bouwens et al. (2007). Stark et al. (2009) use the same color-color criteria as Bouwens et al. (2007) but with ≈ 0.6 mag shallower data sets

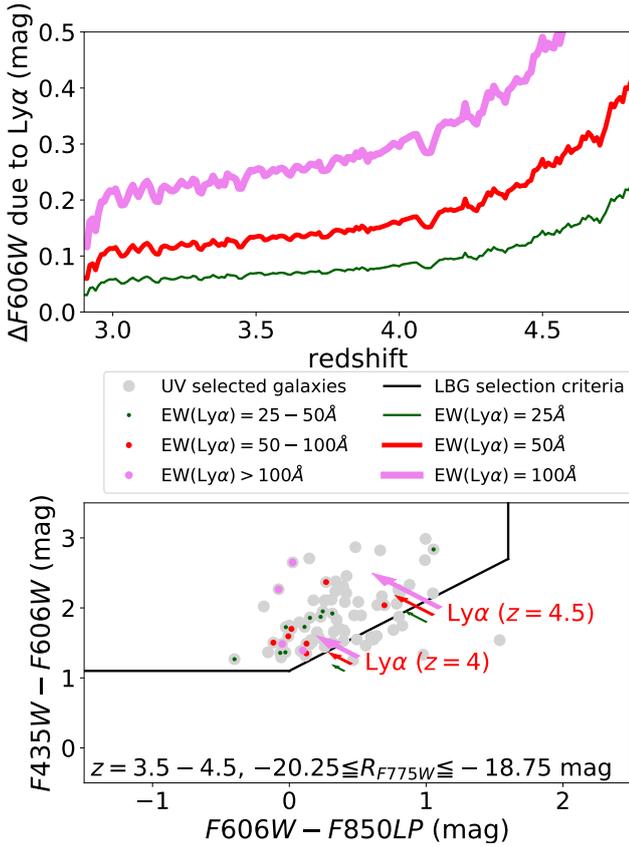**Fig. 9.** Tests of a possible LBG selection bias. *Upper panel*: shift of *F606W* magnitude due to contamination of Lyα emission as a function of redshift. Green, red, and violet lines show the shifts for $EW(Ly\alpha) = 25$ Å, $EW(Ly\alpha) = 50$ Å, and $EW(Ly\alpha) = 100$ Å, respectively. *Lower panel*: color-color diagram for *B* (*F435W*)-dropouts: *F435W*–*F606W* as a function of *F606W*–*F850LP*. The grey, green, red, and violet points indicate UV selected galaxies with $-20.25 \leq M_{1500} \leq -18.75$ mag at $z_p = 3.5$–4.5, those with $EW(Ly\alpha) = 20$–50 Å, $EW(Ly\alpha) = 50$–100 Å, and $EW(Ly\alpha) \geq 100$ Å, respectively. The black line represents the color-color criteria for *B*-dropout. Green, red, and violet arrows show the shifts of colors due to contamination of Lyα emission to *F606W* at $z \approx 4$ and $z \approx 4.5$ for $EW(Ly\alpha) = 25$ Å, $EW(Ly\alpha) = 50$ Å, and $EW(Ly\alpha) = 100$ Å, respectively.

than those in Bouwens et al. (2007). Figure 1 in Bouwens et al. (2007) shows the surface number density of the *B*-dropouts as a function of apparent *F775W* magnitude (i.e., apparent rest-frame UV magnitude). At absolute UV magnitudes of ≈ -18.8 and -18.3 at $z \approx 4$, the completeness values are $\approx 25\%$ and $\approx 5\%$, respectively. The completeness values for ≈ 0.6 mag shallower data in Stark et al. (2009) are estimated to be $\approx 25\%$ and $\approx 5\%$ at $M_{1500} \approx -19.4$ mag and $M_{1500} \approx -18.9$ mag, respectively, if the behavior of completeness as a function of S/N is similar. As shown in the lower panels of Fig. 2, the *B*-dropout galaxies in Stark et al. (2009) are not complete at $M_{1500} \approx -19.0$ mag. Moreover, Stark et al. (2009) adopt stricter S/N cuts than Bouwens et al. (2007)'s criteria for *B*-dropouts, and the completeness values in Stark et al. (2009) may be lower than estimated here, especially for faint sources.

Following the discussions above, at $z \approx 3.5$–4.5, the observed number (the denominator of X_{LAE}) for their B-dropouts with $M_{1500} \approx -19$ mag is estimated to be less than 25% of the true value, while the observed number (the numerator of X_{LAE}) for LAEs with $EW(Ly\alpha) = 50$ Å is estimated to be larger than

50% of the true value under the assumption that all the LAEs meet the color-color criteria. This means that X_{LAE} for their B-dropout sample may be more than ≈ 1.5 times larger than the X_{LAE} for a complete sample, $0.3^{+0.08}_{-0.09}$. If the overestimate would be corrected for X_{LAE} at $M_{1500} \approx -19$ mag, their X_{LAE} would be consistent with ours within the 1σ error bars as shown by the open grey square in Fig. 6.

Therefore, the *B*-dropout selection bias may be the dominant cause of the difference in X_{LAE} between LBGs and photo-*z* selected galaxies at $z \approx 4$ at faint UV magnitudes. This may also have an effect on the difference in X_{LAE} at $z \approx 4$ shown in Fig. 8. Indeed, the difference for the 55 Å cut is more pronounced than that for the 25 Å cut. Although we do not discuss quantitatively biases of dropout selections at other redshifts, strong Lyα emission will cause similar effects, as discussed in the references. We note that the LBG sample in Arrabal Haro et al. (2018) consists of $\approx 70\%$ photometric-redshift selected objects and of $\approx 30\%$ dropout selected objects based on dropout selection criteria for their medium-band filters. Since we measure X_{LAE} for a photometric-redshift selected sample, this may result in the similarity of X_{LAE} to ours at $z \lesssim 5$, though their sample is not complete in UV at $M_{1500} \gtrsim -20$ mag.

We note that Oyarzún et al. (2017) mention yet another potential bias for LBG samples which are incomplete in UV, due to a correlation between the $EW(Ly\alpha)$ and M_{1500}. This bias leads to an underestimate of X_{LAE} because large equivalent width objects are preferentially missed when faint-UV galaxies drop out of the sample. Our results are not affected by this bias, but it could affect the results of other work shown in Fig. 8, and could have compensated the LBG selection bias we discussed above. In Fig. 6, the bias from Oyarzún et al. (2017) has no impact because we are looking at X_{LAE} as a function of UV magnitude.

4.1.2. Different observational methods

The Lyα emission in our sample is measured with an IFU (without including the Lyα halo) and is thus less affected by uncertainties due to slit-loss and aperture corrections. Hoag et al. (2019b) measure the spatial offset between the Lyα emission and the UV continuum. They find a typical standard deviation for the offset which decreases towards higher redshifts ($2.17^{+0.19}_{-0.14}$ kpc ($\approx 0\farcs3$) at $z \approx 3.25$ to $1.19^{+1.29}_{-0.33}$ kpc ($\approx 0\farcs2$) at $z \approx 5.25$). They argue that the evolution of the spatial offset contributes to the increasing trend of X_{LAE} with z measured with slit spectroscopy with $1''$ slits such as in Stark et al. (2011). According to Hoag et al. (2019b) Fig. 7, the simulated cumulative distribution function (CDF) of slit-loss is similar from $z \approx 3.5$ to 5.5 but is shifted at $z \approx 3$–3.5 to a larger slit-loss for a $1''$ slit. At $z \approx 3.5$–4.5, the CDF reaches $\approx 90\%$ at a slitloss of $\approx 10\%$. Moreover, their measured offsets are much larger than that for a lensed LAE at $z \approx 1.8$ (0.65 kpc, Erb et al. 2019) and typical values for LAEs at $z \approx 3$–6 ($\lesssim 0\farcs1$, Leclercq et al. 2017). Hence, this is probably not the dominant cause of the high X_{LAE} values of Stark et al. (2010, 2011) at $z \approx 3.5$–4.5. We note that the aperture diameters (convolved mask diameters) for our Lyα measurement with IFU data are typically larger than $1''$ (see Sect. 2.2.1). Our measurements are less affected by the spatial offset between the Lyα emission and the UV continuum. Meanwhile, Hoag et al. (2019b) estimate slitlossess based on the spatial component of slit spectra.

4.1.3. Summary

We find indications that the dominant cause of the difference in X_{LAE} measured in Stark et al. (2010, 2011) and presented here

is the LBG selection bias. Strong Lyα emission can enhance the probability to meet the LBG-selection criteria both in terms of the signal-to-noise ratio and color. The LBG selection bias has a strong effect on X_{LAE} especially for faint UV magnitudes, where LBG samples are not complete. Possible discrepancies arising from different observational methods probably affect X_{LAE} to a lesser extent. Thanks to the MUSE observations and to the HST photo-z sample, our X_{LAE} are derived from the most homogeneous and complete sample to date.

4.2. Comparison with the GALICS model

Using a homogeneous and complete UV sample and MUSE spectroscopic data, we have measured the LAE fraction for the first time at very faint magnitudes ($M_{1500} \leq -17.0$ mag). While we confirm a weak increase of X_{LAE} as a function of redshift at $3 \lesssim z \lesssim 5$, we find that LAEs with $EW(\text{Ly}\alpha) \geq 45$ Å make up a relatively low fraction of the underlying rest-frame UV-detected galaxy population, ≈ 0–20%. This implies the existence of a duty cycle either for the star formation activity or for the escape and/or production of Lyα photons. Another possibility is that only a small fraction of all galaxies can evolve into LAEs or can be observed as LAEs in a limited range of inclinations (e.g., Verhamme et al. 2012). Our results suggest no dependence of X_{LAE} on M_{1500} at $z \approx 3$. Keeping in mind that we want to assess the merits of the redshift evolution of X_{LAE} to probe the IGM neutral fraction at $z \gtrsim 6$, it is essential to understand these trends after reionization. To do so, we compare our results with predictions from the semi-analytic model of Garel et al. (2015).

4.2.1. Description of the model

Garel et al. (2015) present an updated version of the GAL-ICS hybrid model (Galaxies In Cosmological Simulations, Hatton et al. 2003) which is designed to study the formation and evolution of galaxies in the high redshift Universe. GAL-ICS relies on an N-body cosmological simulation to follow the hierarchical growth of dark matter structures and on semi-analytic prescriptions to describe the physics of the baryonic component. The box size of the simulation is $100\,h^{-1}$ cMpc on a side, and the dark-matter particle mass is $\approx 8.5 \times 10^7\,M_\odot$ (with 1024^3 particles) (Garel et al. 2012). In Garel et al. (2015), stars are formed according to a Kennicutt-Schmidt law when the galaxy's gas surface density Σ_{gas} is larger than a threshold value, $\Sigma_{\text{gas}}^{\text{thresh}}$ (e.g., Schmidt 1959; Kennicutt 1998), and the intrinsic Lyα emission from galaxies is computed assuming case B recombination as $L_{\text{Ly}\alpha}^{\text{intr}} \propto 0.67Q(H)$. Here, $Q(H)$ is the production rate of hydrogen-ionising photons estimated from the stellar spectral energy distributions. In order to predict the observed Lyα properties of galaxies, Garel et al. (2015) combine GALICS with the library of radiative transfer (RT) simulations of Schaerer et al. (2011b) which predict the escape fraction of Lyα photons through galactic winds (f_{esc}; see Verhamme et al. 2006, 2008; Garel et al. 2012, for more details). f_{esc} depends on the wind parameters (the wind expansion velocity, velocity dispersion, dust opacity, neutral hydrogen column density) which are computed by GALICS. The Lyα luminosity emerging from each individual galaxy is then given by $L_{\text{Ly}\alpha} = L_{\text{Ly}\alpha}^{\text{intr}} \times f_{\text{esc}}$.

The GALICS model was tuned to reproduce the UV and Lyα luminosity functions at $z \approx 3$–6 in Garel et al. (2012). This model was then shown to also accurately reproduce the observed stellar mass functions and star-formation-rate to stellar mass relations at $z \approx 3$–6 (Garel et al. 2016). While their fiducial

model can match these observational constraints at $3 \lesssim z \lesssim 6$, it fails to reproduce the wide distribution of Lyα EWs, in particular the high EW values (at $EW(\text{Ly}\alpha) \gtrsim 50$ Å). Garel et al. (2015) discuss the possibility that this mismatch is linked to the lack of "burstiness" for star formation in GALICS given that the Lyα EW is primarily set by the combination of (i) the production rate of Lyα photons, dominated by short-lived stars, and (ii) the stellar UV emission which traces star formation over longer timescales (see also Charlot & Fall 1993; Madau et al. 1998). In the fiducial model, most galaxies keep forming stars at a rather constant rate because the surface gas density threshold is almost always met. In an alternative model (labeled bursty SF), they increase this threshold by a factor of 10, such that gas needs to accrete onto galaxies for longer periods before reaching the required surface density. This naturally gives rise to a star formation duty cycle and Garel et al. (2015) show that their bursty SF model predicts EW distributions in much better agreement with observations than the fiducial model does.

Following the procedure of Garel et al. (2016), we created mock surveys for both the fiducial and bursty SF models using the Mock Map Facility (MOMAF) tool (Blaizot et al. 2005). In practice, for each model, we generated 100 lightcones that mimic the geometry and redshift range of the MUSE HUDF survey, that is, a square field of ≈ 10 arcmin^2 and $2.8 \lesssim z \lesssim 6.7$, and we computed X_{LAE} from the mocks in the same bins of redshift and UV magnitude as for the observational measurements.

4.2.2. Measured LAE fraction vs. GALICS predictions

In Fig. 10, we show our MUSE measurement of X_{LAE} as a function of M_{1500} at $z \approx 3.3$ for $EW(\text{Ly}\alpha) \geq 45$ Å and $EW(\text{Ly}\alpha) \geq 65$ Å. These EW cuts correspond to our most secure measurements. We also show in this figure the predictions from the GAL-ICS model. Both fiducial and bursty SF GALICS models (dashed and solid lines, respectively) show an increase of X_{LAE} towards faint UV magnitudes with a slope which is in good agreement with the data (star symbols). As discussed in Verhamme et al. (2008) and Garel et al. (2015), this trend may be the result of two factors. First, UV bright sources have intrinsically smaller $EW(\text{Ly}\alpha)$ due to less significant or less recent bursts of star formation. Second, these galaxies are often more massive with higher H I and dust contents which can dramatically reduce the escape fraction of Lyα photons and therefore the observed EW. Hence, few bright UV galaxies display a strong Lyα emission line.

For a more detailed comparison, we see that the fiducial GALICS model (dashed lines) does not reproduce the observed X_{LAE}. This model overestimates (underestimates) the LAE fraction at almost all UV magnitudes for $EW(\text{Ly}\alpha) \geq 45$ Å ($EW(\text{Ly}\alpha) \geq 65$ Å). This is a consequence of the too narrow EW distribution predicted by this model (see Sect. 4.2.1), which Garel et al. (2015) attribute to overly smooth star formation histories. In the bursty SF model however (solid lines), galaxies have more diverse recent star formation histories which result in wider EW distributions, and consequently this model is able to reproduce our measured X_{LAE} much better.

In Fig. 11, we compare the GALICS predictions of X_{LAE} as a function of z with X_{LAE} from the MUSE data at $z \lesssim 5$, i.e., where our observational measurements are most robust. For the same reasons as above, we find that the bursty SF model is much more successful at reproducing the observations than the fiducial model. This is particularly true for the lowest EW cut (i.e., 45 Å) where the agreement is quite good (solid orange curve).
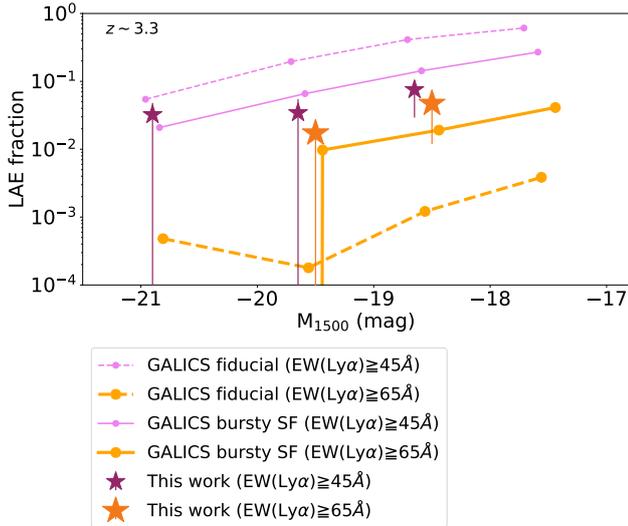
Fig. 10. X_{LAE} vs. M_{1500} at $z \approx 3.3$ from our MUSE results compared to predictions from the GALICS mocks. The MUSE results at $z \approx 3.3$ for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å are indicated by violet and orange stars, respectively. The magenta and orange dashed lines with dots show the average X_{LAE} computed from 100 mocks of the fiducial GALICS model (Garel et al. 2015) at the same redshift for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å, respectively. Those from the bursty SF model are shown by solid lines with dots. For visualization purposes, we slightly shift the points along the abscissa. We note that X_{LAE} for $EW(Ly\alpha) \geq 65$ Å from MUSE and from the bursty SF model is 0 at $M_{1500} \approx -21$ mag.

For $EW(Ly\alpha) \geq 65$ Å however, we note that the bursty model slightly underpredicts the observed LAE fraction (solid magenta curve), especially at $z \gtrsim 4.5$. Additional ingredients could possibly be missing from this model that would help produce more galaxies with large EWs (in particular in the higher redshift bin) such as radiative transfer in asymmetric geometries, or Lyα production from other channels like collisions (gravitational cooling) or fluorescence (see e.g., Verhamme et al. 2012; Garel et al. 2015; Dijkstra 2017, for a more detailed discussion on these aspects). Also, the assumed IMF or the metallicity evolution of model galaxies may not be realistic and lead to low EWs (e.g., Hashimoto et al. 2017b,a).

Overall, these comparisons suggest that the measurements of X_{LAE} by MUSE in the post-reionization epoch can be reasonably well interpreted with current models of high-z galaxies such as GALICS. We find that the observed trends between X_{LAE} and redshift/UV magnitude are mainly shaped by the burstiness of star formation in GALICS. It is also caused by the variation of f_{esc} with respect to the physical properties of the galaxies as discussed in Garel et al. (2015). In GALICS, these two aspects modulate the observed Lyα EWs of galaxies and therefore the LAE fraction at $z \lesssim 5$.

4.2.3. Cosmic variance

The area of the MUSE HUDF survey is limited to 9.92 arcmin^2 which translates into comoving volumes of $\approx 1.5 - 2.5 \times 10^4$ cMpc^3 for the redshift ranges we are considering here. As explained in Sect. 2.3, we have accounted for several sources of uncertainty to compute the error on X_{LAE} but so far we ignored cosmic variance (see Sect. 2.3 for discussion). To assess the significance of this effect, we can estimate the cosmic variance
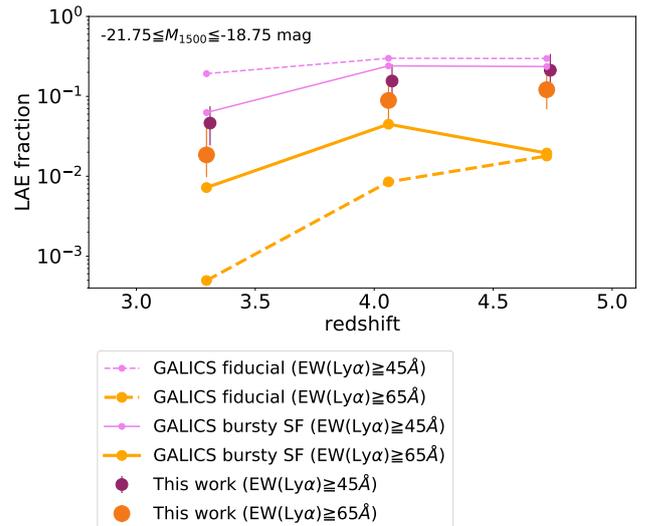

Fig. 11. X_{LAE} vs. z for $M_{1500} \in [-21.75; -18.75]$, at $z < 5$, from our MUSE results compared to predictions from the GALICS model. The MUSE results for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å are indicated by violet and orange circles, respectively. The magenta and orange dashed (solid) lines with dots show the average X_{LAE} computed from 100 mocks of the fiducial (bursty SF) GALICS model (Garel et al. 2015) for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å, respectively. For visualization purposes, we slightly shift the points along the x-axis.

from the GALICS mock lightcones, which are cut out from the $\approx 3 \times 10^6$ cMpc^3 simulation box (Garel et al. 2016). We compute the 1σ standard deviation as the field-to-field variation, which includes the effects of both cosmic variance and the binomial proportion confidence interval. We note that our estimate of cosmic variance with GALICS only accounts for the clustering of galaxies and not for the possible contribution of large-scale variations in the IGM transmissivity due to a patchy reionization.

In the middle and bottom panels of Fig. 12, we compare the relative uncertainty of our MUSE X_{LAE} estimates (circles) with the relative uncertainties due to the field-to-field variation of X_{LAE} (solid lines with dots), which is estimated from the 100 mocks based on the bursty SF model. For a fair comparison, we match X_{LAE} of GALICS to that of MUSE for $EW(Ly\alpha) \geq 45$ Å ($EW(Ly\alpha) \geq 65$ Å), by adopting slightly different $EW(Ly\alpha)$ cuts in the model catalogs, of 46 Å, 48 Å, and 46Å (53 Å, 52 Å, and 49 Å) at $z \approx 3.3$, 4.1, and 4.7, respectively (see the top panel of Fig. 12). We note that the total uncertainty for the MUSE X_{LAE} is calculated by summing the statistical error (binomial proportional confidence interval) multiplied by completeness corrections in each flux bin for completeness correction (see Sect. 2.3 for more details). For both EW cuts, the relative upper errors of our MUSE X_{LAE} are much larger than those of the field-to-field variance for the bursty GALICS model. The relative lower errors of our MUSE X_{LAE} are in the same level of those of the field-to-field variance at $z \approx 4$–5. The contribution of cosmic variance to the field-to-field variance in the GALICS mock is negligible, since statistical errors (crosses) are dominant. It suggests that the cosmic variance is a subdominant source of uncertainties in our measurement of X_{LAE}. Therefore, we conclude that our MUSE results are not strongly affected by cosmic variance. We note that the field-to-field variance may be slightly underestimated in the mock catalogs because the fluctuations on scales larger than the simulated box (150 cMpc) are not sampled. The size of the simulated
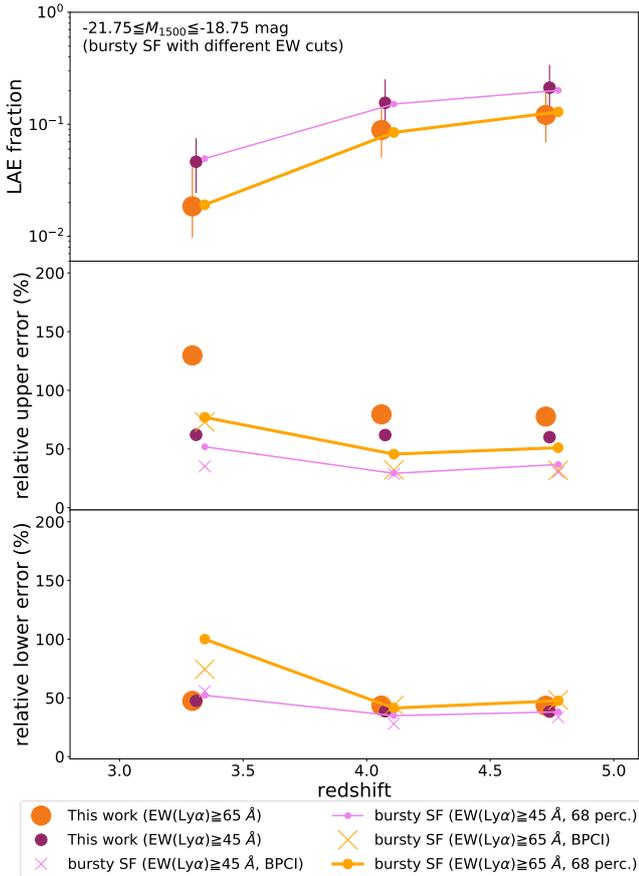
Fig. 12. Test of cosmic variance and uncertainties of X_{LAE} for our MUSE observations using GALICS mocks of the bursty SF model. *Top panel*: X_{LAE} vs. z for $M_{1500} \in [-21.75; -18.75]$, at $z < 5$. In order to better compare GALICS with our observations and to provide a more accurate estimate of cosmic variance, we use slightly different EW cuts for the model. We replace the 45 Å cut with 46 Å, 48 Å, and 46 Å cuts at $z \approx$ 3.3, 4.1, and 4.7, and we replace the 65 Å cut with 53 Å, 52 Å, and 49 Å at the same redshifts. With these cuts, the values of X_{LAE} from MUSE (violet and orange circles at for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å) and GALICS (solid lines) match. *Middle panel*: relative upper 1σ uncertainties of X_{LAE} vs. z for $M_{1500} \in [-21.75; -18.75]$, at $z < 5$. The relative 68% percentiles of X_{LAE} (field-to-field variance) measured among 100 GALICS mocks are indicated by violet and orange circles with soloid lines for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å, respectively. The 68% percentile includes both of the cosmic variance and statistical error. The statistical errors estimated from BPCI are shown by violet and orange crosses. The MUSE uncertainties (estimated from BPCI including completeness correction effects) for $EW(Ly\alpha) \geq 45$ Å and $EW(Ly\alpha) \geq 65$ Å are indicated by violet and orange circles, respectively. *Bottom panel*: relative lower 1σ uncertainties of X_{LAE} vs. z for $M_{1500} \in [-21.75; -18.75]$, at $z < 5$. The symbols are the same as those in the middle panel. For visualization purposes, we slightly shift the points along the x-axis.

volume is however significantly larger than our MUSE survey volume ($\approx 1.5-2.5 \times 10^4$ cMpc^3) and so this underestimate should be weak.

4.3. The redshift evolution of the LAE fraction and implications for reionization

The main purpose of this paper is to measure the evolution of X_{LAE} in the post-reionization epoch, at $z \lesssim 6$ (as shown in Figs. 7,

8, and 11). Our results confirm the rise of X_{LAE} with redshift found in the literature between $z \approx 3$ and 6 for $-21.75 \leq M_{1500} \leq -17.75$ mag, $X_{LAE} = 0.07^{+0.06}_{-0.03}z - 0.22^{+0.12}_{-0.24}$, in Fig. 7. Meanwhile, the trend stopps at $z \sim 5.6$ for $-20.25 \leq M_{1500} \leq -18.75$ mag in Fig. 8. As discussed in Sect. 4.2.3, this evolution at $z \approx 3-5$ is not caused by the cosmic variance of the limited survey field of the MUSE HUDF Survey. Instead, it is probably caused by higher intrinsic $EW(Ly\alpha)$ and/or higher Lyα escape fractions at higher redshift in a given M_{1500} range, due to less massive and less dusty galaxies at higher redshift (e.g., Speagle et al. 2014; Bouwens et al. 2016; Santini et al. 2017). It is also very important to understand the co-evolution of the Lyα and UV luminosity functions (e.g., Ouchi et al. 2008; Dunlop 2013; Konno et al. 2016, 2018; Ono et al. 2018).

De Barros et al. (2017) obtain a relatively low X_{LAE} ($\approx 40\%$) at $z \approx 6$, which implies a less dramatic turn-over at $z > 6$ than previously found (e.g., Stark et al. 2011; Pentericci et al. 2014; Schenker et al. 2014; Tilvi et al. 2014). If we interpret our point at $z \approx 5.5$ as a statistical fluctuation 1σ below a true value ≈ 0.35 as found by De Barros et al. (2017), we confirm this shallower increase of the LAE fraction towards $z \approx 5$ and $z \approx 6$. This could indicate the stop of the evolution of the Lyα escape fraction, possibly related to the plateau evolution of the star formation main sequence at $z \approx 5-6$ suggested by Speagle et al. (2014) and Salmon et al. (2015), and implying a constant stellar mass at a given M_{1500} (see however, Santini et al. 2017).

Another possibility is that our low value of X_{LAE} at $z \approx 5.5$ is genuine and indeed indicates a late transition in the ionisation state of the IGM. However, our data at $z \leq 5$ combined the work of De Barros et al. (2017) and Pentericci et al. (2018), which is not affected either by the LBG selection bias discussed above, suggest an earlier reionization, at $z \approx 6-7$. Our measurement at $z \approx 5.5$ may thus indicate a patchy reionization process. Bosman et al. (2018) measure the mean and scatter of the IGM Lyα opacity with the largest sample of quasars so far. They confirm the existence of tails towards high values in the Lyα opacity distributions, which may persist down to $z \approx 5.2$. They find a linear increase in the mean Lyα opacity from ≈ 1.8 at $z \approx 5$ to ≈ 3.8 at $z \approx 6$. These results also imply a late or patchy reionization scenario, in which reionization ends at $z \approx 5.2-5.3$ (e.g., Kulkarni et al. 2019; Keating et al. 2019, see also Kashino et al. 2019). The Gunn-Peterson absorption trough in quasar spectra is only sensitive to a low Lyα opacity (very low H I gas fraction) and the LAE fraction is therefore a complementary tool. In the near future, the *James Webb* Space Telescope (JWST)/Near Infrared Spectrograph (NIRspec) will enable us to observe Lyα emission at $z \approx 5$ to $z \gtrsim 10$ homogeneously and help make significant progress. We can also use the WST/Near Infrared Imager and Slitless Spectrograph (NIRISS) for Lyα spectroscopy. Most importantly, one can measure Hα emission at $z \approx 0$ to $z \approx 7$ with JWST/NIRspec and, subsequently, the line ratio of Lyα to Hα to disentangle between intrinsic evolution and escape fraction.

Another important point raised in this paper is that X_{LAE} estimates are sensitive to $EW(Ly\alpha)$ selections (see Figs. 7 and 8). Although a general consensus seems to emerge from previous work, a quantitative interpretation of the evolution of X_{LAE} with redshift requires more accurately constructed samples. In addition, the contribution of extended Lyα emission to the total Lyα budget is typically large and has a large scatter (typically more than $\approx 50\%$, Momose et al. 2016; Leclercq et al. 2017). Methods for measuring Lyα flux have a large effect on $EW(Ly\alpha)$ and then X_{LAE}. It means that accurate and homogeneous measurements of Lyα emission are required to use X_{LAE} as a tracer of the H I gas fraction of the IGM. In Figs. 5 and 6, the $X_{LAE}-z$

relation does not depend strongly on the rest-frame UV absolute magnitude. This suggests that at $z < 6$, combining UV-bright and faint samples can give us better statistics for $X_{\rm LAE}$ measurements. In addition, as discussed in Sect. 4.1.1, a firm definition of parent samples avoiding a selection bias is also required to assess the evolution of $X_{\rm LAE}$. The uncertainties of $X_{\rm LAE}$ have to be calculated with BPCI (Bernoulli trials, see Sect. 2.3). Moreover, Mason et al. (2018) warn of the interpretation of the evolution of $X_{\rm LAE}$ with the same UV magnitude, since galaxies with the same UV magnitude have very different stellar and halo masses at different redshifts (e.g., Speagle et al. 2014; Behroozi et al. 2013). Because of such effects, sophisticated models of galaxy formation are needed to robustly interpret variations of $X_{\rm LAE}$ with cosmic time. We propose a method using a UV complete sample including faint galaxies, based on a photo-z selection and an absolute magnitude cut, together with Lyα measurements by an IFU with a high sensitivity and a wide-wavelength coverage in a large field-of-view like VLT/MUSE and VLT/BlueMUSE (Richard et al. 2019) at $z \approx 2$–6.6.

5. Summary and conclusions

We investigated the LAE fraction at $z \approx 3$–6 using the second data release of the MUSE *Hubble* Ultra Deep Field Survey and the HST catalog of the UVUDF. Thanks to the unprecedented depth of the MUSE and HST data for Lyα and UV, respectively, we studied the LAE fraction for galaxies as faint as $M_{1500} = -17.0$ mag at $z \approx 3$ for the first time with a UV-complete sample. We also derived the LAE fraction as a function of redshift homogeneously from $z = 3$ to 6, down to $M_{1500} = -17.75$ mag. Our results are summarized as follows:

- We derived the redshift evolution of $X_{\rm LAE}$ for a number of EW and UV magnitude selections, including the first estimate down to -17.75 mag. These results are summarized in Table 3. For all selections, we find low values of $X_{\rm LAE} \approx 0.04$–0.3, and a weak rise of $X_{\rm LAE}$ with z, qualitatively consistent with the trend reported for brighter samples in the literature.

- We compared our MUSE results with those in the literature for $M_{1500} \in [-20.25; -18.75]$. At $z \lesssim 5$, our values of $X_{\rm LAE}$ are consistent with those in Arrabal Haro et al. (2018) and Stark et al. (2011) within 1σ error bars for $EW(\mathrm{Ly}\alpha) \geq 25$ Å (see left panel of Fig. 8). Our $X_{\rm LAE}$ at $z \approx 5.6$ is lower than those in the literature, which may be caused by statistical errors, or a late and/or patchy reionization process.

- We measured the dependence of $X_{\rm LAE}$ on M_{1500} at $z = 2.9$–3.7 for $EW(\mathrm{Ly}\alpha) \geq 25$ Å, 45 Å, 65 Å, and 85 Å (see Fig. 5). This is the first time this has been measured down to $M_{1500} = -17.0$ mag (for the largest EWs of our sample), and for a volume-limited sample. We found no clear dependence of $X_{\rm LAE}$ on M_{1500}, in contrast to previous reports.

- We compared the dependence of $X_{\rm LAE}$ on M_{1500} for $EW(\mathrm{Ly}\alpha) \geq 50$ Å at $z \approx 3$–4 derived from MUSE with results from the literature (Fig. 6). Again we found no dependence of $X_{\rm LAE}$ on M_{1500}. Our slopes of $0.01^{+0.02}_{-0.03}$ and $0.01^{+0.06}_{-0.10}$ at $z \approx 2.9$–3.7 and at ≈ 3.7–4.4, respectively, are shallower than that in Stark et al. (2010), $0.13^{+0.03}_{-0.03}$ at ≈ 3.5–4.5. We also found lower values of our $X_{\rm LAE}$ at a faint UV magnitude of $M_{1500} \gtrsim -19$ mag.

- The dominant causes of the difference of $X_{\rm LAE}$ in our work and previous studies appear to be LBG selection biases in those studies. We showed how these can lead to an overestimate by a factor ≈ 1.5 of $X_{\rm LAE}$ at $z \approx 4$ for galaxies with $M_{1500} = -19$ mag and $EW(\mathrm{Ly}\alpha) \geq 50$ Å.

- We compared our MUSE results with predictions from a cosmological semi-analytic galaxy evolution model (GALICS, Garel et al. 2015). When GALICS uses a bursty star formation model, it can reproduce our measurement of $X_{\rm LAE}$ as a function of M_{1500} at $z \approx 3$. The fiducial GALICS model however cannot. The bursty model can also reproduce $X_{\rm LAE}$ as a function of z at $z \lesssim 4$. We assessed cosmic variance for our MUSE results using the bursty SF model and found that it does not have a significant effect on our results.

- Overall, we found that $X_{\rm LAE}$ is lower than $\approx 30\%$. This implies a low duty cycle of LAEs, suggesting bursty star formation or strong time variations in the production of Lyα photons or in their escape fraction.

Despite the difficulties of the method, the dominant source of uncertainties in our work is the Poisson noise due to the small number of objects in our samples. This is encouraging and suggests that future deep surveys with for example, MUSE and JWST will enable us to produce accurate measurements of $X_{\rm LAE}$ with secure samples and to extend our understanding of the evolution of $X_{\rm LAE}$ at all redshifts, after and during the epoch of reionization.

References

Arrabal Haro, P., Rodríguez Espinosa, J. M., Muñoz-Tuñón, C., et al. 2018, MNRAS, 478, 3740
Astropy Collaboration (Tollerud, E. J., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Bacon, R., Accardo, M., Adjali, L., et al. 2010, in The MUSE second-generation VLT instrument, eds. I. S. McLean, S. K. Ramsay, H. Takami, Proc. SPIE, 7735, 773508
Bacon, R., Brinchmann, J., Richard, J., et al. 2015, A&A, 575, A75
Bacon, R., Conseil, S., Mary, D., et al. 2017, A&A, 608, A1
Baldry, I. K., Alpaslan, M., Bauer, A. E., et al. 2014, MNRAS, 441, 2440
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Benitez, N. 2000, ApJ, 536, 571
Benitez, N., Ford, H., Bouwens, R., et al. 2004, ApJS, 150, 1
Blaizot, J., Wadadekar, Y., Guiderdoni, B., et al. 2005, MNRAS, 360, 159
Bosman, S. E., Fan, X., Jiang, L., et al. 2018, MNRAS, 479, 1055
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015a, ApJ, 811, 140
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015b, ApJ, 803, 1
Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, ApJ, 833, 72
Bradač, M., Huang, K.-H., Fontana, A., et al. 2019, MNRAS, 489, 99
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brinchmann, J., Inami, H., Bacon, R., et al. 2017, A&A, 608, A3
Cameron, E. 2011, PASA, 28, 128
Caruana, J., Bunker, A. J., Wilkins, S. M., et al. 2014, MNRAS, 443, 2831
Caruana, J., Wisotzki, L., Herenz, E. C., et al. 2018, MNRAS, 473, 30
Cassata, P., Tasca, L. A. M., Le Fèvre, O., et al. 2015, A&A, 573, A24
Charlot, S., & Fall, S. M. 1993, ApJ, 415, 580

Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, ApJ, 132, 926
Cooke, J., Ryan-Weber, E. V., Garel, T., & Díaz, C. G. 2014, MNRAS, 441, 837
Curtis-Lake, E., McLure, R. J., Pearce, H. J., et al. 2012, MNRAS, 422, 1425
De Barros, S., Pentericci, L., Vanzella, E., et al. 2017, A&A, 608, A123
Dijkstra, M. 2014, PASA, 31
Dijkstra, M. 2017, Lecture notes for the 46th Saas-Fee winterschool, 13
Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, MNRAS, 414, 2139
Drake, A. B., Garel, T., Wisotzki, L., et al. 2017a, A&A, 608, A6
Drake, A. B., Guiderdoni, B., Blaizot, J., et al. 2017b, MNRAS, 471, 267
Dunlop, J. S. 2013, Astrophys. Space Sci. Lib. (Berlin, Heidelberg: Springer-Verlag), 223
Erb, D. K., Berg, D. A., Auger, M. W., et al. 2019, ApJ, 884, 7
Faisst, A. L., Capak, P., Carollo, C. M., Scarlata, C., & Scoville, N. 2014, ApJ, 788, 87
Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, ApJ, 132, 117
Finkelstein, S. L., Ryan, R. E., Papovich, C., et al. 2015, ApJ, 810, 71
Fontana, A., Vanzella, E., Pentericci, L., et al. 2010, ApJ, 725, L205
Forero-Romero, J. E., Yepes, G., Gottlöber, S., & Prada, F. 2012, MNRAS, 419, 952
Garel, T., Blaizot, J., Guiderdoni, B., et al. 2012, MNRAS, 422, 310
Garel, T., Blaizot, J., Guiderdoni, B., et al. 2015, MNRAS, 450, 1279
Garel, T., Guiderdoni, B., & Blaizot, J. 2016, MNRAS, 455, 3436
Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Hashimoto, T., Verhamme, A., Ouchi, M., et al. 2015, ApJ, 812, 157
Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017a, A&A, 608, A10
Hashimoto, T., Ouchi, M., Shimasaku, K., et al. 2017b, MNRAS, 465, 1543
Hatton, S., Devriendt, J. E. G., Ninin, S., et al. 2003, MNRAS, 343, 75
Herenz, E. C., & Wisotzki, L. 2017, A&A, 602, A111
Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, A&A, 621, A107
Hinton, S. R., Davis, T. M., Lidman, C., Glazebrook, K., & Lewis, G. F. 2016, Astron. Comput., 15, 61
Hoag, A., Bradač, M., Huang, K., et al. 2019a, ApJ, 878, 12
Hoag, A., Treu, T., Pentericci, L., et al. 2019b, MNRAS, 488, 706
Hu, E. M., Cowie, L. L., Barger, A. J., et al. 2010, ApJ, 725, 394
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Inoue, A. K., Hasegawa, K., Ishiyama, T., et al. 2018, PASJ, 70, 1
Itoh, R., Ouchi, M., Zhang, H., et al. 2018, ApJ, 867, 46
Jensen, H., Laursen, P., Mellema, G., et al. 2013, MNRAS, 428, 1366
Kakiichi, K., Dijkstra, M., Ciardi, B., & Graziani, L. 2016, MNRAS, 463, 4019
Kashikawa, N., Shimasaku, K., Malkan, M. A., et al. 2006, ApJ, 2, 7
Kashino, D., Lilly, S. J., Shibuya, T., Ouchi, M., & Kashikawa, N. 2019, ApJ, 888, 6
Keating, L. C., Weinberger, L. H., Kulkarni, G., et al. 2019, MNRAS, 491, 1736
Kennicutt, R. C. 1998, ARA&A, 36, 189
Konno, A., Ouchi, M., Nakajima, K., et al. 2016, ApJ, 823, 20
Konno, A., Ouchi, M., Shibuya, T., et al. 2018, PASJ, 70
Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, MNRAS, 485, L24
Kusakabe, H., Shimasaku, K., Momose, R., et al. 2019, PASJ, 71, 1
Laursen, P., Sommer-Larsen, J., Milvang-Jensen, B., Fynbo, J. P. U., & Razoumov, A. O. 2019, A&A, 627, A84
Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, A&A, 608, A8
Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113
Madau, P. 1995, ApJ, 441, 18
Madau, P., & Haardt, F. 2015, ApJ, 813, L8
Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106
Malhotra, S., & Rhoads, J. E. 2004, ApJ, 617, L5
Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, ApJ, 856, 2

Mason, C. A., Fontana, A., Treu, T., et al. 2019, MNRAS, 485, 3947
Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. 2018, ApJ, 869, 150
McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, MNRAS, 447, 499
Momose, R., Ouchi, M., Nakajima, K., et al. 2016, MNRAS, 457, 2318
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, ApJ, 744, 83
Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, PASJ, 70, 2
Ota, K., Richard, J., Iye, M., et al. 2012, MNRAS, 423, 2829
Ota, K., Iye, M., Kashikawa, N., et al. 2017, ApJ, 844, 85
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301
Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, ApJ, 723, 869
Oyarzún, G. A., Blanc, G. A., González, V., Mateo, M., & Bailey, J. I. 2017, ApJ, 843, 133
Parsa, S., Dunlop, J. S., & McLure, R. J. 2018, MNRAS, 474, 2904
Pentericci, L., Fontana, A., Vanzella, E., et al. 2011, ApJ, 743, 132
Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, ApJ, 793, 113
Pentericci, L., Vanzella, E., Castellano, M., et al. 2018, A&A, 619, A147
Piqueras, L., Conseil, S., Shepherd, M., et al. 2017, ArXiv eprints [arXiv:1710.03554]
Planck Collaboration VI. 2020, A&A, in press, https://doi.org/10.1051/0004-6361/201833910
Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31
Richard, J., Bacon, R., Blaizot, J., et al. 2019, ArXiv eprints [arXiv:1906.01657]
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJ, 802, L19
Rykoff, E. S., Rozo, E., & Keisler, R. 2015, ArXiv e-prints [arXiv:1509.00870]
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Santini, P., Fontana, A., Castellano, M., et al. 2017, ApJ, 847, 76
Santos, S., Sobral, D., & Matthee, J. 2016, MNRAS, 463, 1678
Schaerer, D., De Barros, S., & Stark, D. P. 2011a, A&A, 536, A72
Schaerer, D., Hayes, M., Verhamme, A., & Teyssier, R. 2011b, A&A, 531, A12
Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2014, ApJ, 795, 20
Schmidt, M. 1959, ApJ, 129, 243
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Stanway, E. R., Bremer, M. N., & Lehnert, M. D. 2008, MNRAS, 385, 493
Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, MNRAS, 408, 1628
Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, ApJ, 728
Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, MNRAS, 464, 469
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, ApJ, 532, 170
Tilvi, V., Papovich, C., Finkelstein, S. L., et al. 2014, ApJ, 794, 1
Totani, T., Aoki, K., Hattori, T., et al. 2014, PASJ, 66, 63
Totani, T., Kawai, N., Kosugi, G., et al. 2006, Proc. Int. Astron. Union, 2, 265
Treu, T., Schmidt, K. B., Trenti, M., Bradley, L. D., & Stiavelli, M. 2013, ApJ, 775, 5
Ventou, E., Contini, T., Bouché, N., et al. 2017, A&A, 608, A9
Verhamme, A., Schaerer, D., & Maselli, A. 2006, A&A, 460, 397
Verhamme, A., Schaerer, D., Atek, H., & Tapken, C. 2008, A&A, 491, 89
Verhamme, A., Dubois, Y., Blaizot, J., et al. 2012, A&A, 546, A111
Wilson, E. B. 1927, J. Am. Stat. Assoc., 22, 209
Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98
Yajima, H., Li, Y., Zhu, Q., et al. 2012, ApJ, 754, 118
Zheng, Z.-Y., Wang, J., Rhoads, J., et al. 2017, ApJ, 842, L22

Appendix A: Uncertainties of $z_{\rm p}$

Table A.1. Comparison of the median of 95% upper and lower errors of $z_{\rm p}$ to the half-width of redshift bins and the fraction of galaxies with a large lower error.

Mean $z_{\rm p}$	Half-width of z bins	Median of 95% upper error	Median of 95% lower error	$f_{\rm large\ error}$
		$-20.25 \leq M_{1500} \leq -18.75$ mag		
$z \approx 3.3$	0.39	0.19	0.20	0.02
$z \approx 4.1$	0.38	0.24	0.25	0.05
$z \approx 4.7$	0.29	0.28	0.27	0.0
$z \approx 5.6$	0.56	0.30	0.32	0.09
		$-18.75 \leq M_{1500} \leq -17.75$ mag		
$z \approx 3.3$	0.39	0.21	0.23	0.05
$z \approx 4.1$	0.38	0.30	0.34	0.2
$z \approx 4.7$	0.29	0.33	0.34	0.13
$z \approx 5.6$	0.56	0.35	0.44	0.35

Notes. The mean redshift, the half-width of the redshift bin, the median of upper 95% errors of $z_{\rm p}$, the median of lower 95% errors, and the fraction of galaxies with a large lower error suggesting a bimodal $z_{\rm p}$ PDF are shown.

A.1. Impact of lacking IRAC data on z_p estimation in Rafelski et al. (2015)

It is well known that LBG samples and photo-z samples at $z \approx 3$–7 can be contaminated by lower-z galaxies with a 3646 Å Balmer or 4000 Å break at $z \approx 0$–1, especially at faint magnitude. *Spitzer*/IRAC data can provide rest-frame optical data which are useful to break the degeneracy of $z_{\rm p}$ (e.g., Bradač et al. 2019). In this work, we used the catalog from R15, where $z_{\rm p}$ are derived using HST data alone. The advantage of this choice is discussed in Brinchmann et al. (2017), where they show that IRAC data can in fact worsen photo-z performance for faint galaxies in the MUSE HUDF sample (see their Appendix A). It might be a reflection of the difficulty of providing reliable IRAC photometry for sources as faint as most of our sample.

As discussed in Sect. 2.3, the fraction of low-z contaminants among R15 galaxies is found to be low within MUSE samples (Fig. 20 in Inami et al. 2017). To avoid contaminants due to poor photo-z estimations, we apply an $S/N > 2$ cut for our sample and then applied a stricter cut on M_{1500} (see Sect. 2.1 and Fig. 1 for more details).

A.2. Effects of z_p errors on redshift binning

To check the effect of the error on $z_{\rm p}$ on redshift binning, we compare the median of upper and lower 95% errors of $z_{\rm p}$ to the half-width of redshift bins for $-20.25 \leq M_{1500} \leq -18.75$ mag and $-18.75 \leq M_{1500} \leq -17.75$ mag shown in Figs. 7 and 8. The upper (lower) 95% errors are calculated from a difference between 95% upper (lower) limit of photo-z and photo-z where likelihood is maximized in BPZ. The results are summarized in Table A.1. For $-20.25 \leq M_{1500} \leq -18.75$ mag, the median values of the 95% $z_{\rm p}$ errors are smaller than the width of redshift bins at $z \approx 3.3$ to 5.6. For $-18.75 \leq M_{1500} \leq -17.75$ mag, the errors are larger than those for the brighter M_{1500}, but the median of 95% errors are still smaller than or comparable to the width of redshift bins at the all over redshift range. Therefore, the 95% errors of $z_{\rm p}$ are typically smaller than the width of redshift bins

in Figs. 7 and 8. We note that we included widths of the redshift bins in the linear relation fitting (see Sect. 2.4).

Next, we checked the fraction of possible low-z contaminants, $f_{\rm large\ error}$. Although the probability distribution functions (PDFs) of $z_{\rm p}$ in R15 catalog are not published, galaxies with a bimodal PDF of $z_{\rm p}$ show a large lower 95% error, which reaches at $z \approx 0$–1 for a sample at $z \approx 3$–7. We calculated $f_{\rm large\ error}$ from the wavelengths of breaks, $z_{\rm p}$, and its 95% errors and summarize the results in Table A.1. For $-20.25 \leq M_{1500} \leq -18.75$ mag, $f_{\rm large\ error}$ is 0.02 to 0.09, implying that our $X_{\rm LAE}$ in Fig. 8 is not affected significantly by contaminants. Meanwhile, for $-18.75 \leq M_{1500} \leq -17.75$ mag, $f_{\rm large\ error}$ is 0.05 at $z \approx 3.3$ so that our $X_{\rm LAE}$ is not suppressed significantly by contaminants in Fig. 6. At $z \approx 4.1$ to 5.6, $f_{\rm large\ error}$ is relatively high, 0.13 to 0.35. However, these are conservative estimations of the upper limit of low-z contaminated fraction, since not all of the galaxies with a large photo-z error locate at $z \approx 0$–1. In fact, our sample shows the maximum likelihood at $z \approx 3$–6. Our $X_{\rm LAE}$ at $z \gtrsim 4.1$ and 5.6 in Fig. 7 should not be affected significantly by low-z contaminants significantly.

Appendix B: The template spectra used in MARZ

Figure B.1 shows the template spectra of LAEs in MARZ used in this work. The continuum are subtracted in the templates and the bright Lyα lines are clipped to lie between -30 and $+30$ times the mean absolute deviation in the similar manner to those used in the original MARZ (Hinton et al. 2016) and AUTOZ (Baldry et al. 2014) as shown in the right panels in Fig. B.1. Cross-correlation functions indicate locations of lines in the fitting range and are not affected significantly by the line shape of templates in general (see Sect. 4.3 in Herenz & Wisotzki 2017). In fact, the completeness of MARZ does not depend on the FWHM of fake lines in completeness simulations as we mention in Sect. 2.2.4. Although our templates do not cover various types of Lyα lines, it does not have an effect on detection of Lyα emission.
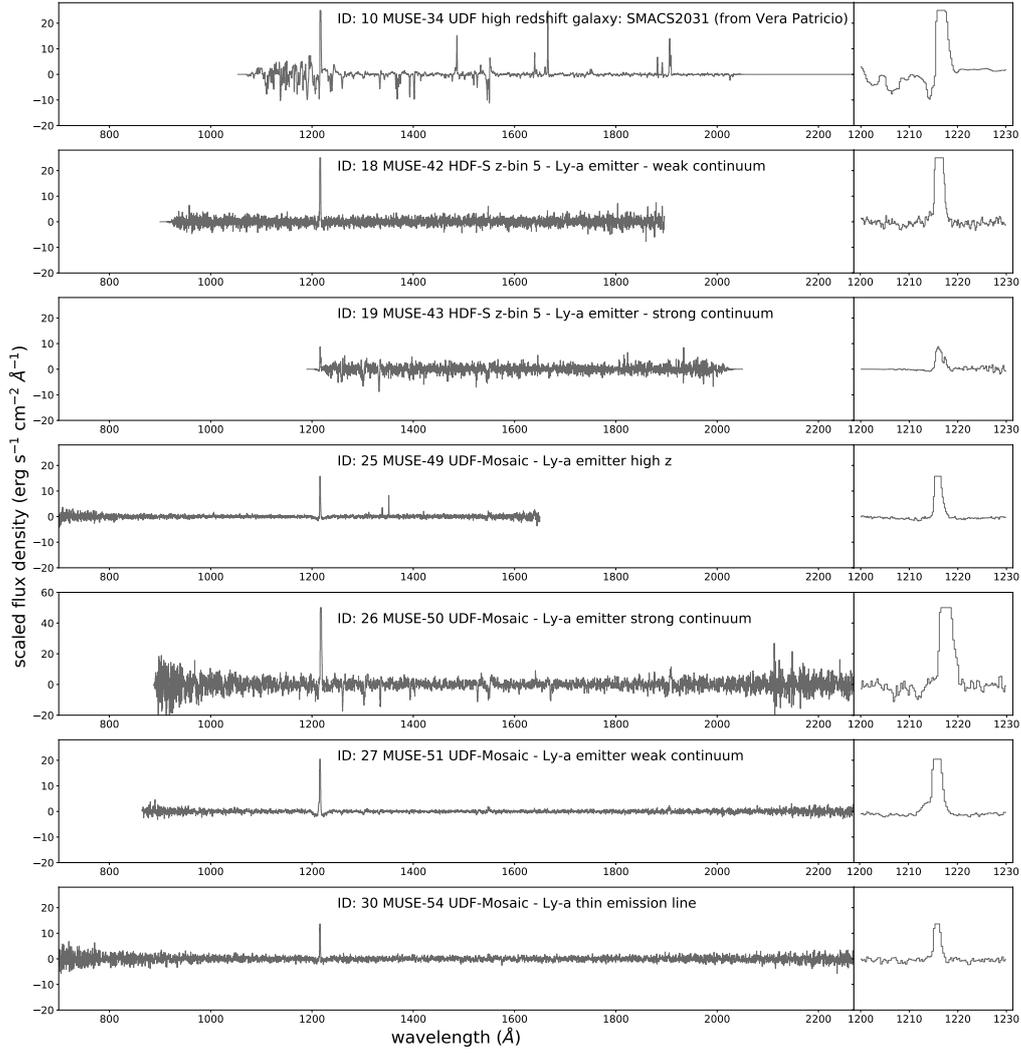
Fig. B.1. LAE template spectra of MARZ used in this work: those of ID= 10, 18, and 19 are used in Inami et al. (2017), while those of ID= 25, 26, 27, and 30 are newly created from MUSE data (Bacon et al. in prep.). *Left panels*: scaled spectra of the templates in the rest frame. *Right panels*: zooms of the Lyα emission line in each left panel.

Appendix C: An example of contamination of Lyα detection

Figure C.1 shows an example of contamination of Lyα emission from a neighboring object. Panels a–d show sub-panels in MARZ for a UV-selected source (see Inami et al. 2017, for more details of MARZ's screen). The HST cutout around a UV-galaxy with Rafelski et al. (2015) ID = 628 is shown in panel a. The 1D spectrum shown in panel d clearly shows a strong Lyα emission line with the highest confidence level by MARZ. The spectrum is extracted from the object mask (panel b), and clearly contaminated by diffuse Lyα emission from a neighboring object (MUSE ID = 1185 in the DR1 catalog) as shown in the narrow band in panel c. We can remove these contaminated objects from our sample of Lyα emitter candidates in visual inspection.
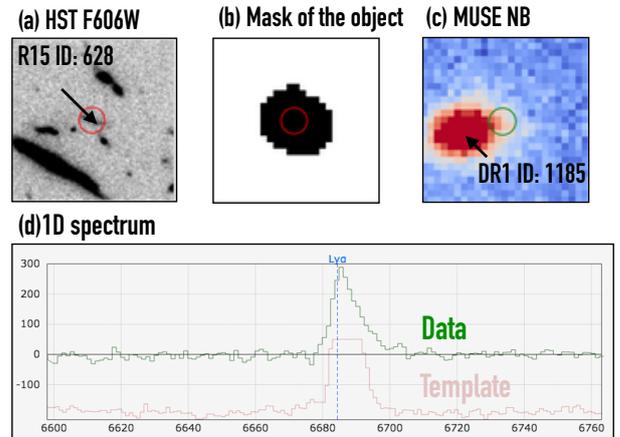

Fig. C.1. Example of contamination of Lyα emission from a neighboring object. *Panels a–d*: sub-panels in MARZ's screen (Inami et al. 2017) for a UV-selected source with Rafelski et al. (2015) ID = 628: (*a*) HST F606W cutout, (*b*) mask of the object for the extraction of the 1D spectrum, (*c*) MUSE NB cutout, and (*d*) 1D spectrum. The red and green circle in the images show the position of the UV-selected galaxy. The green and red lines in *panel d* indicate observational data and the best-fit template.

Appendix D: The number of flux bins used to correct incompleteness of Lyα detection
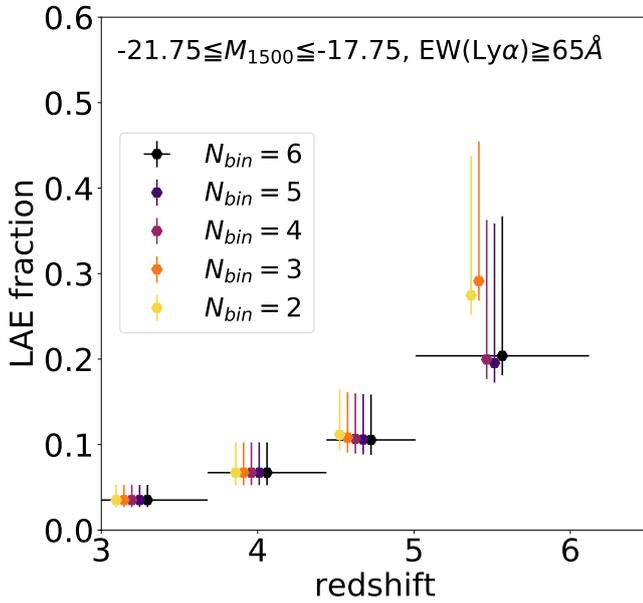

Fig. D.1. Test for the effect of binning of Lyα flux to correct incompleteness of the number of LAEs. The median values and error bars of X_{LAE} for the plot of the evolution of X_{LAE} (Fig. 7) are shown. The black, purple, violet, orange, and yellow hexagons indicate $N_{\mathrm{bin}} = 6, 5, 4, 3,$ and 2, respectively. For visualization purposes, we slightly shift the points along the x-axis and show the width of z only for $N_{\mathrm{bin}} = 6$.

We also examine the effect of binning to correct incompleteness of the number of LAEs. Because the number of objects we have is small, somewhat arbitrary binning may affect X_{LAE}. For a large number of bins, we often get flux-bins with no object at all, and these bins increase the error bars. For a small number of bins, we introduce a large error on the completeness correction as described above. To test the effect of binning, we vary the number of bins (N_{bin}) we use, from 2 to 6, and see how the median values and error bars change. In Fig. D.1, we show the median values and error bars of X_{LAE} for $N_{\mathrm{bin}} = 2$–6 for the plot of the evolution of X_{LAE} (Fig. 7). The uncertainties of the completeness correction are 20%, 25%, 32%, 44%, and 73% for $N_{\mathrm{bin}} = 6, 5, 4, 3,$ and 2 (see Sect. 2.3). We find that 4 to 6 bins are a sweet spot where the error bars are small and appear converged and adopt $N_{\mathrm{bin}} = 4$ in Sect. 2.3.

Appendix E: Error propagation of completeness correction values in a binomial proportion confidence interval

We test the applicability of a binomial proportion confidence interval (BPCI) for the case of a completeness correction to calculate error bars of the LAE fraction. First, we examine how to propagate completeness correction values in the error calculation with BPCI numerically. To do mock observations, we generate $N_{\mathrm{LAE}}^{\mathrm{true}}(z_s, M_{1500}, EW)$ randomly with 100 000 trials using the python module numpy.random.binomial for each $N_{1500}(z_p, M_{1500})$ and the true LAE fraction ($X_{\mathrm{LAE}}^{\mathrm{true}}$). Then we generate $N_{\mathrm{LAE}}^{\mathrm{det}}(z_s, M_{1500}, EW)$ randomly for each $N_{\mathrm{LAE}}^{\mathrm{true}}(z_s, M_{1500}, EW)$ again using numpy.random.binomial for a given completeness correction value as a probability of detec-
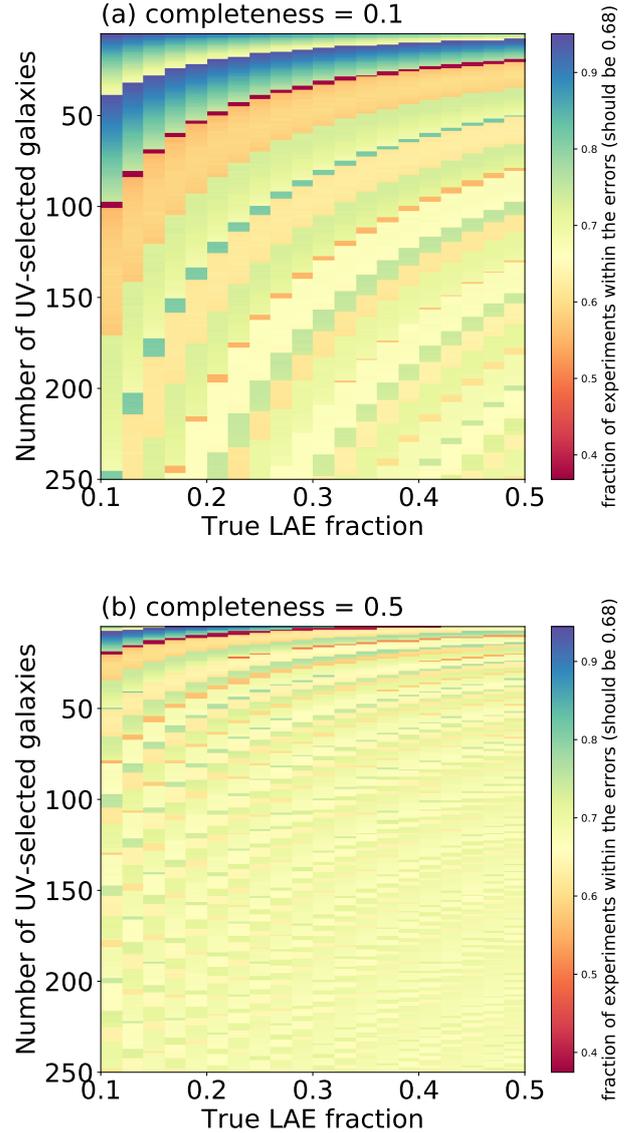

Fig. E.1. Test of the accuracy of our uncertainty estimation of X_{LAE}. We generate mock X_{LAE} distribution numerically for each $N_{1500}(z_p, M_{1500})$ and $X_{\mathrm{LAE}}^{\mathrm{true}}$ with a given completeness of 0.1 in *panel a* and 0.5 in *panel b*. We then derive the fraction of experiments within our upper and lower limits calculated with BPCI. The colors encode the fraction of the experiment for each $N_{1500}(z_p, M_{1500})$ and $X_{\mathrm{LAE}}^{\mathrm{true}}$ in the x and y axes, respectively.

tion. We can obtain the probability distribution of X_{LAE} with a given completeness correction value, $N_{1500}(z_p, M_{1500})$, and $X_{\mathrm{LAE}}^{\mathrm{true}}$. We compare the 1σ upper and lower uncertainties from the probability distribution function of the mock X_{LAE} with those derived from two methods of BPCI (binom_conf_interval) with input parameters of the number of LAEs and $N_{1500}(z_p, M_{1500})$ if the incompleteness of the number of LAEs are corrected ($N_{\mathrm{LAE}}^{\mathrm{true}}(z_s, M_{1500}, EW)$) or not ($N_{\mathrm{LAE}}^{\mathrm{det}}(z_s, M_{1500}, EW)$). We confirm that it is better to input $N_{\mathrm{LAE}}^{\mathrm{det}}(z_s, M_{1500}, EW)$ and to multiply the obtained uncertainties by the correction value (see also Sect. 2.3).

Second, we check the accuracy of the method above in the plane of $N_{1500}(z_p, M_{1500})$ and $X_{\mathrm{LAE}}^{\mathrm{true}}$ for a given completeness correction value. We calculate the 1σ upper and lower limits of X_{LAE} with this method. Again we generate 100 000 mock values

of $N_{\mathrm{LAE}}^{\mathrm{det}}(z_{\mathrm{s}}, M_{1500}, EW)$ and then X_{LAE} numerically. The fractions of the mock X_{LAE} within the range of the 1σ upper and lower limits among all the experiments are calculated. If the method is accurate, this fraction would be $\approx 68\%$. We check the fraction of experiments for $N_{1500}(z_{\mathrm{p}}, M_{1500})$ from 0 to 250 with a step of 1 and for $X_{\mathrm{LAE}}^{\mathrm{true}}$ from 0.1 to 0.5 with a step of 0.02. In Fig. E.1, the example with low completeness correction values, 0.1 and 0.5, is shown. In the panel a for completeness correction = 0.1, most of the plane is colored with yellow green corresponding to a fraction close to 0.68, while it is colored with blue or red corresponding to overestimation and underestimation of the errors, respectively, for cases with poor statistics (i.e., low $N_{1500}(z_{\mathrm{p}}, M_{1500})$ and $X_{\mathrm{LAE}}^{\mathrm{true}}$). In panel b for completeness correction = 0.5, the method is found to produce errors accurately except for the very poor statistics cases (at the upper left region in the panel). We can estimate the error more accurately with this method for a higher completeness case. Even with a low completeness and smallest numbers, the uncertainties are overestimated. Therefore, we adopt this method conservatively.

Appendix F: The best-fit linear relations of X_{LAE} as a function of z and M_{1500}

We show the best-fit linear relations of X_{LAE} as a function of z and M_{1500} in Fig. F.1. The equations of relations are shown in Sects. 3.1 and 3.2.
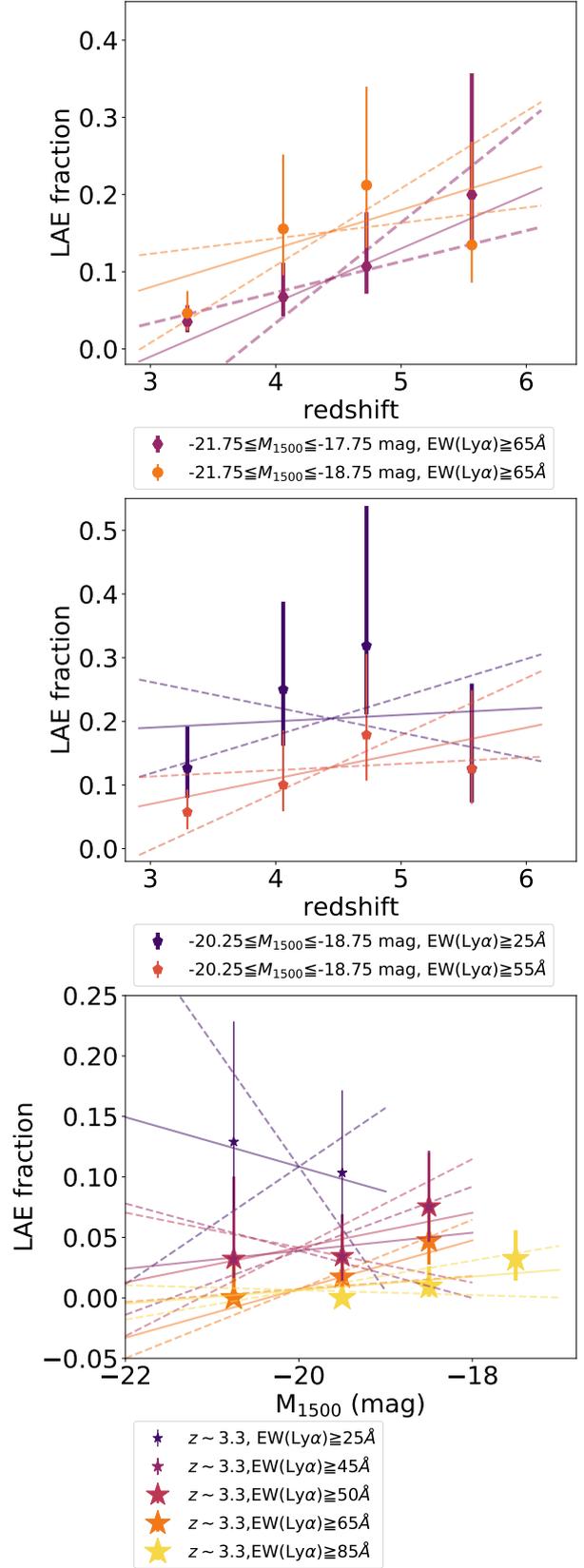

Fig. F.1. Slopes of the best-fit linear relations of of X_{LAE} as a function of z (*top and middle panels*) and M_{1500} (*bottom panel*). Symbols are the same as those in Figs. 7–6. The best-fit linear relations and the $\pm 1\sigma$ slopes are shown by the solid and dashed lines, respectively, with lighter colors of those for the symbols.