

To be, or not to be: Balmer breaks in high- z galaxies with *JWST*

Anton Vikaeus¹,^{*} Erik Zackrisson^{1,2}, Stephen Wilkins^{3,4}, Armin Nabizadeh¹, Vasily Kokorev⁵, Abdurro'uf^{6,7}, Larry D. Bradley⁷, Dan Coe^{6,7,8}, Pratika Dayal⁵ and Massimo Ricotti⁹

¹*Observational Astrophysics, Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden*

²*Institute of Space Sciences and Astronomy, University of Malta, Msida MSD 2080, Malta*

³*Kapteyn Astronomical Institute, University of Groningen, 9700 AV Groningen, The Netherlands*

⁴*Center for Astrophysical Sciences, Department of Physics and Astronomy, The Johns Hopkins University, 3400 N Charles St. Baltimore, MD 21218, USA*

⁵*Space Telescope Science Institute (STScI), 3700 San Martin Drive, Baltimore, MD 21218, USA*

⁶*Association of Universities for Research in Astronomy (AURA), Inc. for the European Space Agency (ESA), STScI, Baltimore, MD 21218, USA*

⁷*Department of Astronomy, University of Maryland, College Park, 20742, USA*

⁸*Swedish Collegium for Advanced Study, Linneanum, Thunbergsvägen 2 SE-752 38 Uppsala, Sweden*

⁹*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK*

Accepted 2024 January 10. Received 2024 January 10; in original form 2023 September 12

ABSTRACT

Standard models of structure formation allow us to predict the cosmic timescales relevant for the onset of star formation and the assembly history of galaxies at high redshifts ($z > 10$). The strength of the Balmer break represents a well-known diagnostic of the age and star formation history of galaxies, which enables us to compare observations with contemporary simulations – thus shedding light on the predictive power of our current models of star formation in the early Universe. Here, we measure the Balmer break strength for 23 spectroscopically confirmed galaxies at redshifts $6 \lesssim z \lesssim 12$ using public *JWST* NIRSpec data from the cycle 1 GO 1433 and GO 2282 programmes (PI Coe), as well as public spectroscopic data from the *JWST* Deep Extragalactic Survey (JADES). We find that the range of observed Balmer break strengths agree well with that of current simulations given our measurement uncertainties. No cases of anomalously strong Balmer breaks are detected, and therefore no severe departures from the predictions of contemporary models of star formation. However, there are indications of a number of outliers in the observed distribution which have weaker Balmer breaks than predicted by simulations.

Key words: techniques: spectroscopic – galaxies: formation – galaxies: high-redshift – galaxies: star formation – infrared: general.

1 INTRODUCTION

When estimating the star-forming age of an evolved galaxy, one of the more powerful proxies is the strength of the so-called Balmer break. This apparent discontinuity in the spectra of galaxies appears as a result of the complete ionization of hydrogen atoms occupying the second excited atomic state – producing a break in the restframe continuum emission around a wavelength of $\sim 3600 \text{ \AA}$ (see Fig. 1). The strength of this continuum break is primarily governed by stellar physics which identifies stars of spectral type A as the dominant stellar component responsible for strong Balmer breaks in galaxy spectra. As such, the Balmer break will evolve with time due to stellar evolution and the galaxy's star-forming history. The break grows strongest at around 0.3–1 Gyr for a simple, single-age stellar population, while leveling out as the galaxy continues to age (e.g. Kriek et al. 2006). Furthermore, in galaxies with ongoing star formation, the continuous replenishing of young stars enhances the integrated flux at shorter wavelengths, resulting in a bluer spectrum with a less pronounced Balmer break. In normal circumstances, one

therefore generally requires a mature stellar population with an age of $\gtrsim 0.3$ Gyr that dominates the integrated flux in order to expect a significant Balmer break (e.g. Steinhardt et al. 2023).

The Balmer break is also sensitive to nebular reprocessing of the stellar radiation which becomes evident in cases of ongoing star formation and young stellar populations. The nebular continuum emission strongly enhances the flux at wavelengths blueward of the break – resulting in a smaller Balmer break ratio. Furthermore, dust attenuation has the opposite effect such that a significant dust component enhances the Balmer break due to reddening (e.g. Wilkins et al. 2023).

The Balmer break, in conjunction with detections of strong Balmer emission lines, such as $H\alpha$ or $H\beta$, makes it possible to constrain the star formation history (SFH) and the overall age of the stellar population making up a galaxy. The utility of the Balmer break as an age indicator is commonly used when analyzing spectra of galaxies at lower redshifts, where an evolved stellar population is likely a more significant component in the integrated galaxy spectra. Reaching towards the very high redshifts ($z \gtrsim 10$) now probed by *The James Webb Space Telescope (JWST)* suggests a theoretically declining strength in the Balmer break due to the predominantly young stellar populations that formed close to the onset of galaxy formation – thus

* E-mail: anton.vikaeus@physics.uu.se

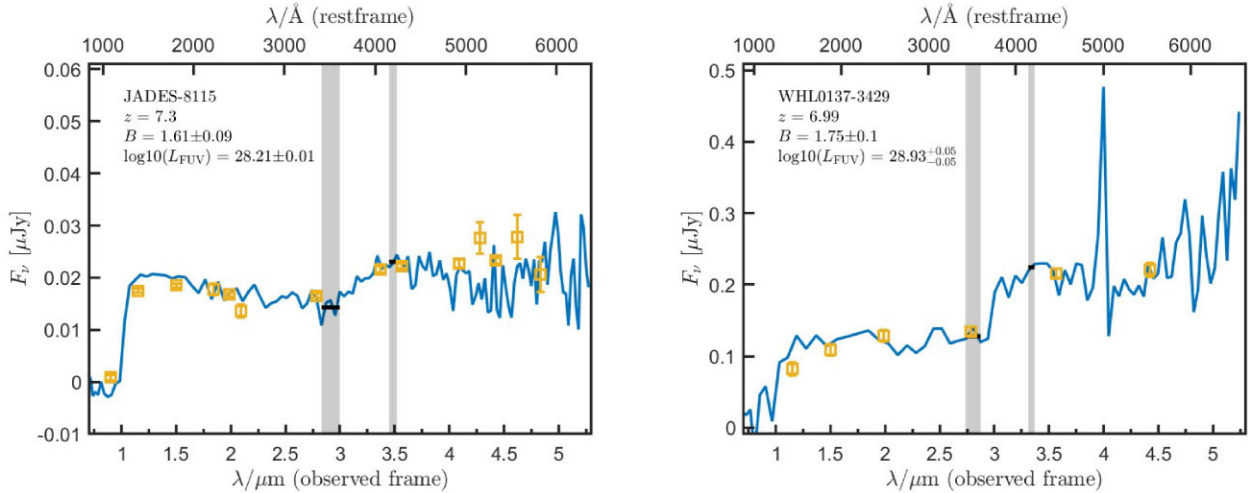


Figure 1. *JWST* NIRSpec/PRISM spectra and NIRCам photometry (yellow boxes) for two of the high-redshift galaxies (JADES-8115 and WHL0137-3429) from our sample with notable Balmer breaks. The grey shaded area marks the two wavelength ranges used to calculate the Balmer break. Inside the shaded regions, two solid black lines indicate the continuum level fitted and used for the Balmer break calculations.

encouraging our endeavour for observational confirmation of this very assumption.

Recent discoveries (Roberts-Borsani, Ellis & Laporte 2020; Laporte et al. 2021) have complicated this picture due to the identification of $z \approx 7$ –9 galaxies that, if based solely on the Balmer break, argue for an onset of star formation very early in the Universe through short (~ 10 –100 Myr) bursts of star formation followed by quiescent phases or very low star formation rates. Hashimoto et al. (2018) provided additional evidence for such a $z \approx 9$ galaxy (MACS1149-JD1), showing a strong Balmer break that suggests the galaxy likely formed the bulk of its stars around 250 million years after the Big Bang. Whether such bursts and following quiescent phases are typical for the assembly of the first galaxies, is currently being investigated by *JWST* which enables us to study their SFHs back to the very first 100 Myrs of cosmic time. This was just recently subject to investigation by new *JWST* data, showing evidence for a much weaker Balmer break in MACS1149-JD1, and thus a less extreme SFH (Bradac et al. 2023; Stiavelli et al. 2023).

As of now, a few observations by *JWST* consistent with Balmer breaks at high redshift have been presented. Photometrically, Labbé et al. (2023b) provides evidence of a Balmer break in six galaxies with $7.4 \leq z \leq 9.1$. However, at the same time, objects with similar photometric signatures have been linked with active galactic nuclei (Kokorev et al. 2023; Labbe et al. 2023a), revealing the challenge of photometrically identifying clear and unambiguous Balmer breaks. So far, spectroscopic confirmation of a high-redshift galaxy with a distinct Balmer break has only been reported by Looser et al. (2023a) at $z = 7.3$. More recently, Curtis-Lake et al. (2023) spectroscopically confirmed multiple galaxies at extreme redshifts $z \approx 10$ –13, but ruled out the existence of any prominent Balmer breaks in the data presented. Trussler et al. (2023) presents photometric data for a large number of galaxies at $7 < z < 12$ where they search for an excess in the F444W and F356W NIRCам filters of *JWST* as a means to identify candidate high- z Balmer break galaxies. Similarly (Atek et al. 2023) reports three galaxies at $z \sim 9.5$ –10.2 with a strong F444W excess.

A key aspect in determining the star-forming age of a galaxy from the Balmer break is the assumptions made regarding the SFH and chemical enrichment. Recent high-redshift simulations such as

Feedback In Realistic Environments (FIRE) by Ma et al. (2020), First Light And Reionization Epoch Simulations (FLARES) by Vijayan et al. (2020) and Lovell et al. (2020), the SPHINX simulations (Rosdahl et al. 2018; Katz et al. 2021) as well as the simulations in Garcia et al. (2023), provide a picture different than the perhaps oversimplified models assuming a constant, or slowly varying star formation rate. Feedback processes, chemical enrichment, magnetic fields, supernova explosions, etc. all play significant roles in the way star formation proceeds after onset. Oversimplifications would evidently render our predictions less accurate and could therefore call for some revision of our contemporary models of star formation – advocating a more detailed treatment which at this point in time is only amenable to numerical simulations such as those mentioned above.

In this work, we present spectroscopically-determined Balmer break strengths for a sample of 23 galaxies (see Table 1) studied with *JWST* and explore the agreement of these measurements with Balmer breaks derived for simulated galaxies (presented in Figs 2–4) from the FLARES (Lovell et al. 2020; Vijayan et al. 2020; Wilkins et al. 2023), the DELPHI (Dayal et al. 2014, 2022; Mauerhofer & Dayal 2023) simulation suites and the high-resolution zoom-in simulations of an individual high-redshift galaxy (Fig. 5) presented in Garcia et al. (2023).

The paper is structured as follows: In Section 2, we outline our methodology for measuring the Balmer break strength and introduce the theoretical and observational data used in this study. In Section 3 we present our observational findings regarding the Balmer break of high-redshift galaxies while in Section 4 we discuss and summarize our conclusions. A flat Lambda cold dark matter (Λ CDM) cosmological model with $H_0 = 67.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_\Lambda = 0.685$, $\Omega_m = 0.315$, and $\Omega_b = 0.0487$ (Planck Collaboration 2014) is adopted throughout the paper.

2 METHODS

We define the strength of the Balmer break following the approach explained in Binggeli et al. (2019), where the continuum (in units of F_ν) is fitted both at wavelengths longward (at 4200 Å) and shortward

Table 1. The observed Balmer break strength B measured for 23 galaxies.

| ID | RA | Dec | $B = \frac{F_{\nu}(4200 \text{ \AA})}{F_{\nu}(3500 \text{ \AA})}$ | z_{spec} | μ | $\log_{10} M_{\star}$ | $\log_{10} L_{\text{FUV}}$ | F200W |
|-----------------|------------|-----------|---|-------------------|---------------------|------------------------|---|-----------------|
| v4 (v7) | deg | deg | | | | (M_{\odot}) | ($\text{erg s}^{-1} \text{ Hz}^{-1}$) | nJy |
| MACS0647 | | | | | | | | |
| 3568 (4922) | 101.903980 | 70.242978 | 0.87 ± 0.25 | 9.25 | $2.5^{+0.1}_{-0.1}$ | $8.99^{+0.66}_{-0.65}$ | $28.30^{+0.03}_{-0.03}$ | 28.5 ± 2.2 |
| 3754 (5191) | 101.920529 | 70.244899 | 0.73 ± 0.13 | 7.47 | $6.3^{+0.5}_{-0.4}$ | $8.77^{+0.54}_{-0.49}$ | $28.45^{+0.04}_{-0.04}$ | 216 ± 5.1 |
| 1715 (2121) | 101.983171 | 70.209787 | 1.05 ± 0.15 | 6.14 | $1.1^{+0.1}_{-0.1}$ | $9.37^{+0.59}_{-0.62}$ | $28.23^{+0.05}_{-0.05}$ | 37.6 ± 4.0 |
| 3989 (5591) | 101.940754 | 70.249101 | 1.12 ± 0.15 | 6.14 | $1.7^{+0.1}_{-0.1}$ | $9.40^{+0.58}_{-0.67}$ | $27.51^{+0.04}_{-0.04}$ | 7.6 ± 2.6 |
| 3308 (4533) | 101.952449 | 70.239294 | 1.13 ± 0.05 | 6.13 | $2.5^{+0.2}_{-0.1}$ | $8.84^{+0.60}_{-0.50}$ | $27.94^{+0.03}_{-0.04}$ | 68.4 ± 3.0 |
| 3208 (4411) | 101.918791 | 70.237614 | 1.41 ± 0.13 | 6.12 | $1.9^{+0.1}_{-0.1}$ | $9.37^{+0.61}_{-0.47}$ | $28.59^{+0.04}_{-0.04}$ | 144 ± 4.2 |
| WHL0137 | | | | | | | | |
| 1968 (5609) | 24.355077 | -8.447929 | 0.96 ± 0.18 | 8.22 | 8.6^{+14}_{-6} | $7.97^{+1.10}_{-0.92}$ | $27.92^{+0.57}_{-0.44}$ | 72.3 ± 7.4 |
| 3429 (7424) | 24.324384 | -8.418989 | 1.75 ± 0.10 | 6.99 | $1.1^{+0.1}_{-0.1}$ | $9.88^{+0.57}_{-0.52}$ | $28.93^{+0.50}_{-0.50}$ | 128.4 ± 7.8 |
| JADES | | | | | | | | |
| 10014220 | 53.16477 | -27.77463 | 1.25 ± 0.44 | 11.58 | - | - | 28.39 ± 0.03 | 14.6 ± 0.7 |
| 10014177 | 53.15884 | -27.77349 | 0.86 ± 0.34 | 10.38 | - | - | 28.09 ± 0.02 | 7.7 ± 0.6 |
| 6438 | 53.16735 | -27.80750 | 0.67 ± 0.15 | 9.70 | - | - | 28.36 ± 0.01 | 19.1 ± 0.6 |
| 10058975 | 53.11243 | -27.77461 | 0.74 ± 0.06 | 9.43 | - | - | 28.80 ± 0.01 | - |
| 8013 | 53.16446 | -27.80218 | 0.65 ± 0.35 | 8.47 | - | - | 28.03 ± 0.04 | 12.1 ± 0.8 |
| 21842 | 53.15683 | -27.76716 | 0.73 ± 0.13 | 7.98 | - | - | 28.13 ± 0.01 | 11.5 ± 0.6 |
| 8115 | 53.15508 | -27.80177 | 1.61 ± 0.09 | 7.30 | - | - | 28.21 ± 0.01 | 16.7 ± 0.5 |
| 8079 | 53.15283 | -27.80194 | 1.28 ± 0.29 | 7.26 | - | - | 27.85 ± 0.02 | 8.6 ± 0.5 |
| 10013905 | 53.11833 | -27.76901 | 0.71 ± 0.16 | 7.20 | - | - | 28.17 ± 0.01 | - |
| 4297 | 53.15579 | -27.81520 | 0.80 ± 0.22 | 6.71 | - | - | 28.03 ± 0.02 | 10.8 ± 0.6 |
| 3334 | 53.15138 | -27.81916 | 1.30 ± 0.23 | 6.71 | - | - | 27.82 ± 0.01 | 6.1 ± 0.8 |
| 16625 | 53.16904 | -27.77884 | 0.80 ± 0.07 | 6.63 | - | - | 28.10 ± 0.01 | 13 ± 0.5 |
| 18179 | 53.17582 | -27.77446 | 1.41 ± 0.14 | 6.34 | - | - | 28.26 ± 0.02 | 24 ± 0.7 |
| 18846 | 53.13492 | -27.77271 | 0.90 ± 0.03 | 6.33 | - | - | 28.59 ± 0.01 | 43.5 ± 0.6 |
| 18976 | 53.16660 | -27.77240 | 1.07 ± 0.12 | 6.33 | - | - | 28.05 ± 0.01 | 11.6 ± 0.6 |

MACS0647 and WHL0137 IDs are from the v4^a catalogue (v7^b IDs in parentheses). The redshifts for the MACS0647 and WHL0137 galaxy cluster fields are fitted using MSAEXP, while the spectra and estimated redshifts for the galaxies in the JADES survey are retrieved from Bunker et al. (2023). The magnification from gravitational lensing μ is estimated from the lens models discussed in Section 2.2. The total stellar mass (M_{\star}) is fitted using piXedfit (Abdurro'uf et al. 2021), while the far ultraviolet luminosity (L_{FUV}) is calculated from the spectra using the restframe continuum at 1500 Å. The NIRCam F200W filter flux is uncorrected for lensing. ^a<https://s3.amazonaws.com/grizli-v2/JwstMosaics/v4/index.html> ^b<https://dawn-cph.github.io/dja/imaging/v7/>

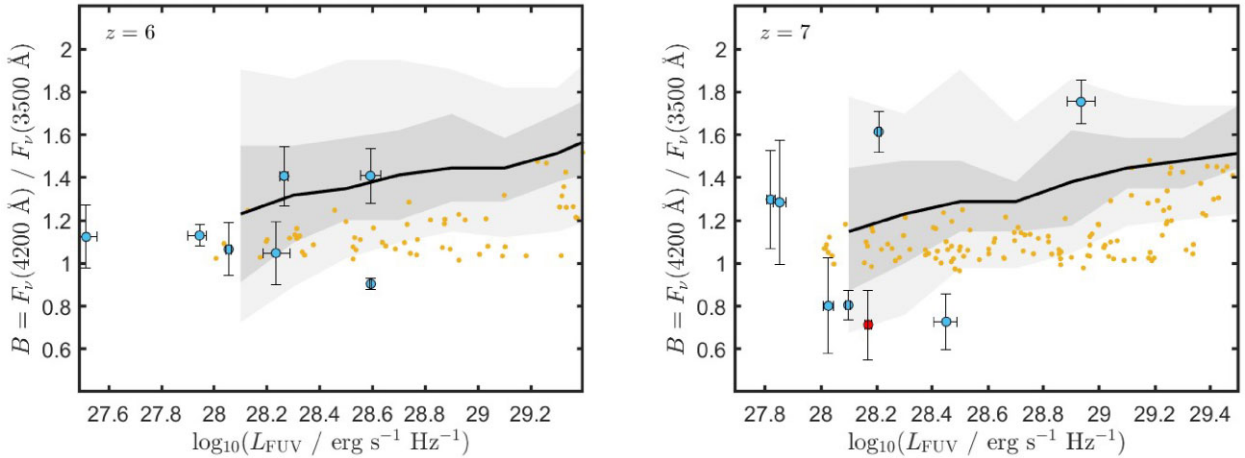


Figure 2. The Balmer break strength $B = F_{\nu}(4200 \text{ \AA})/F_{\nu}(3500 \text{ \AA})$ as a function of far ultraviolet luminosity (L_{FUV}) at redshifts $z = 6$ and $z = 7$. The light blue dots with accompanying error bars represent the measured Balmer break strength derived from the observed galaxies with spectroscopic redshifts within $z = 6 \pm 0.5$ (left) and $z = 7 \pm 0.5$ (right). Red markers indicate galaxies with either missing NIRCam photometry, or photometric data that cannot be trivially reconciled with spectroscopy. The large errors in L_{FUV} for some of the galaxies are due to gravitational lensing. The black solid line is the median Balmer break strength from the FLARES simulations, while the shaded regions correspond to the 2.2–97.8th (lighter shade) and 15.8–84.2th (darker shade) percentiles of the distribution. Orange dots correspond to simulated Balmer break strengths from the DELPHI models.

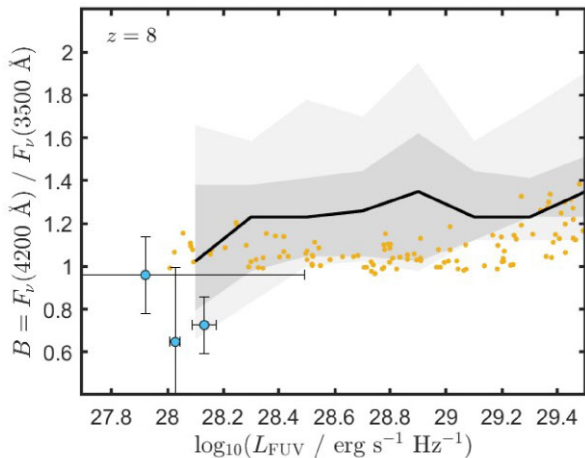


Figure 3. Same as in Fig. 2, here for $z = 8$. Strong gravitational lensing ($\mu \sim 7.9^{+1.2}_{-0.6}$) is responsible for the wide error bars in L_{FUV} for one of the objects (identified as WHL0137-1968).

(at 3500 Å) of the break (grey shaded regions in e.g. Fig. 1). The ratio of the two continuum levels forms an index quantifying the strength of the Balmer break, i.e. $B = F_{\nu}(4200 \text{ \AA})/F_{\nu}(3500 \text{ \AA})$.

Entangled within the Balmer break region is also another feature known as the 4000 Å break. This feature has a similar effect on the spectrum and is therefore usually merged with the Balmer break into one single break. However, the mechanisms behind this break are different, with a notably stronger dependence on metallicity. For older stellar populations the closely spaced absorption features from ionized metals reduce the flux at redder wavelengths therefore typically weakening the break feature (Bruzual 1983; Kriek et al. 2011). For the purposes of high redshift observations with telescopes like *JWST*, the dim nature of distant and faint objects generally does not provide sufficient resolution to resolve the two break features. Therefore, in this paper, we do not distinguish the two and instead quantify the Balmer/4000 Å break through wavelength ranges covering either sides of the break region as a whole (Binggeli et al. 2019; Wilkins et al. 2023).

Utilizing the above-mentioned definition, we determine the Balmer break from spectroscopic continuum measurements, coupled to spectroscopic redshifts, obtained from *JWST/NIRSpec*. This ensures that we deal with well-established redshift values for the galaxies under consideration, which allows us to constrain the Balmer break while minimizing potential contamination from strong emission lines.

The necessity for spectroscopic studies of the Balmer break becomes evident due to the prevalent presence of strong emission lines such as $[\text{O III}]_{\lambda 4959, 5007}$ and $\text{H}\beta$, which can complicate the analysis of photometric data and mimic the presence of a strong Balmer break as they enter into the filter depending on redshift (e.g. Stefanon et al. 2023).

2.1 Theoretical predictions of the Balmer break from simulations

2.1.1 FLARES

We present the simulated Balmer break strengths derived from the FLARES simulations within the redshift range of $z = 6\text{--}10$. Wilkins et al. (2023) provide a thorough analysis of the physical mechanisms affecting the Balmer break in high- z galaxies. They

showed the various factors contributing to the observed strength of the break, including aspects such as the SFH, metallicity, dust, escape of Lyman-continuum radiation, stellar initial mass function (IMF), and also the employed stellar population synthesis (SPS) models. These simulations demonstrate that a galaxy with a given far ultraviolet luminosity (or total stellar mass) exhibit Balmer breaks that may deviate by $\sim 20\text{--}60$ per cent from the median value given the 2.2–97.8th percentiles (the variation are largest at the lower luminosities). Considering the variations seen in the simulations, we expect that observations should frequently fall within the 2.2–97.8th percentile if our contemporary models of galaxy evolution are accurate.

2.1.2 DELPHI

We also compare our observational results to those from the DELPHI semi-analytical model (Dayal et al. 2014, 2022; Mauerhofer & Dayal 2023). This model includes all the key processes of mergers and accretion history in assembling the dark matter halo mass and gas mass up to $z \sim 40$, starting at $z \sim 4.5$ with a time resolution of 30 Myr and a halo mass resolution of $10^8 M_{\odot}$. The available gas mass in any halo can form stars with an ‘effective efficiency’ of f_{*}^{eff} , which is the minimum between the efficiency that produces enough Type II Supernova (SNII) energy to eject the remainder of the gas (f_{*}^{ej}) and an upper maximum (mass- and redshift-independent) threshold (f_{*}^{u}). The models include the key processes of production, astration, destruction (of dust into metals), ejection and dust grain growth in the interstellar medium (that leads to a corresponding decrease in the metal mass) to calculate the total dust and metal masses for each galaxy. Crucially, this model contains only two mass- and redshift-independent free parameters to match observations. These are the maximum (instantaneous) star formation efficiency of $f_{*} = 8$ per cent and the fraction f_w (~ 7.5 per cent) of the SNII explosion energy that is available to drive an outflow. These parameters have been tuned to simultaneously reproduce the observed stellar mass function and the UV luminosity function at $z \sim 5\text{--}12$. The integrated spectrum for each galaxy is obtained by summing the spectrum from each burst of star formation, accounting for its metallicity, and using a Salpeter IMF between 0.1 and $100 M_{\odot}$ in the STARBURST99 (Leitherer et al. 1999) SPS model. For the DELPHI model we consider the fiducial case of $f_{\text{esc}} = 0$ resulting in a maximal contribution from both continuum nebular emission and nebular emission lines. For nebular emission lines we use the metallicity-dependent results tabulated in Anders & Fritze-v. Alvensleben (2003).

2.1.3 Garcia + 23

As a complement to the statistical distributions available from the simulations on larger scales discussed above, we also present results based on the Garcia et al. (2023) zoom-in simulations of a low-mass galaxy at $z > 8$. Here, the formation and evolution of compact star clusters is resolved with high accuracy thanks to a spatial resolution of 0.1 pc and stellar particle masses of $10 M_{\odot}$. The simulations are ran with the code RAMSES-RT and they include, in addition to supernova feedback, non-equilibrium chemistry and radiation transfer of hydrogen and helium ionizing radiation and molecular hydrogen dissociating radiation [we refer to Garcia et al. (2023) for details on the methods]. The morphology of these simulated galaxies also resembles closely the clumpy structures observed in strongly lensed high-redshift galaxies where no prominent disc or bulge is present.

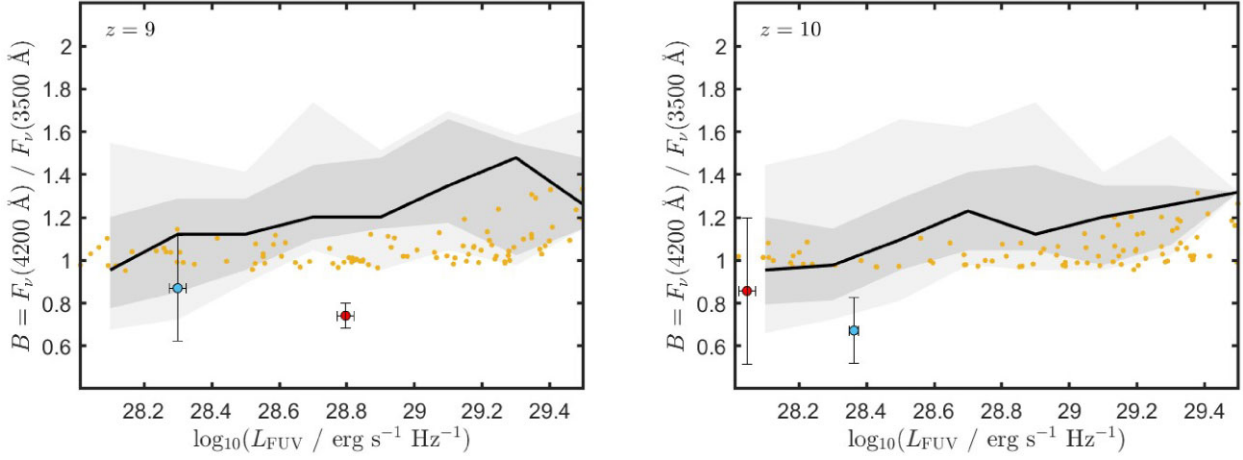


Figure 4. Same as in Fig. 2, here for $z = 9$ and $z = 10$. Spectra for the outliers at $z = 9$ can be seen in Fig. 7, where the clear outlier is identified as JADES-10058975. Similarly, spectra for the $z = 10$ galaxies can be seen in Fig. 8, where the weaker break is identified with JADES-6438. The red markers indicate galaxies with either missing NIRCcam photometry, or, photometric data that cannot be trivially reconciled with spectroscopy.

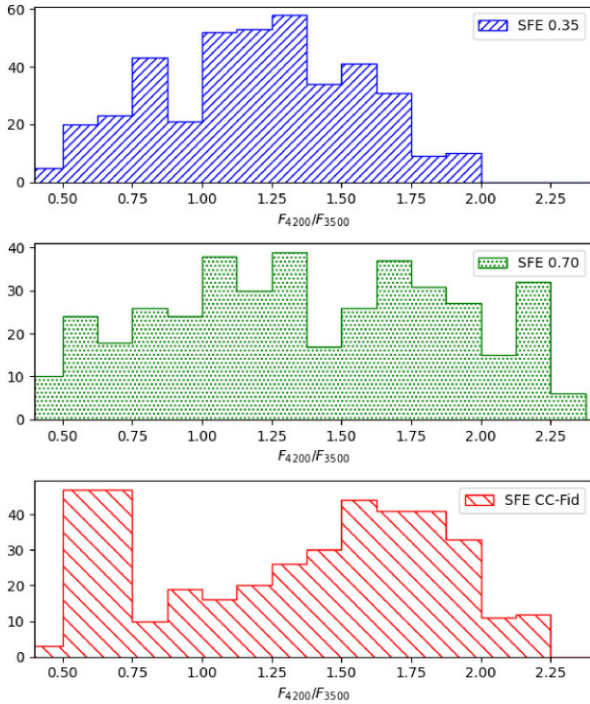


Figure 5. Probability distribution function (PDF) of $B = F_v(4200 \text{ \AA}) / F_v(3500 \text{ \AA})$ for the low-star formation efficiency (low-SFE) run (top panel) and the high sub-grid SFE run (middle panel) in Garcia et al. (2023) and a run assuming ‘realistic’ sub-grid SFE derived in He, Ricotti & Geen (2019) using au-scale simulations of star formation in molecular clouds (bottom panel). The run with high sub-grid SFE and ‘realistic’ sub-grid SFE show the most bursty/stochastic star formation, which is reflected in the wider range of B values. The run with ‘realistic’ sub-grid SFE shows a bimodal PDF with peaks at $B \sim 0.7$ and $B \sim 1.75$. The PDFs are obtained by extracting spectra for the simulated galaxies at random times uniformly sampled in the redshift range $7.5 < z < 12$, regardless of their rest-frame FUV luminosity, i.e. no selection bias is applied to derive these PDFs.

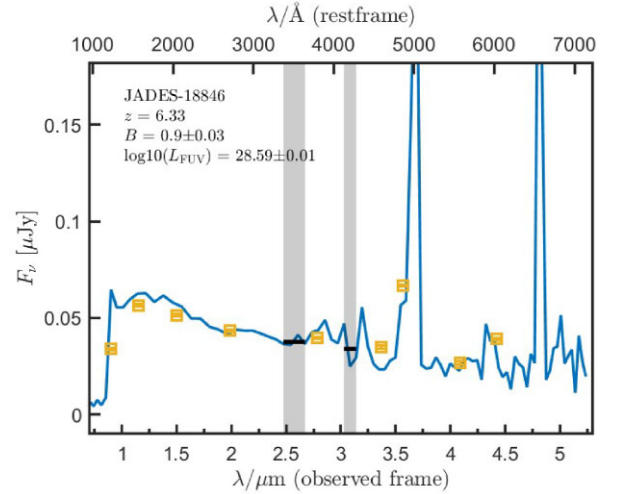


Figure 6. JWST NIRSpec/PRISM spectra and NIRCcam photometry for the outlier at $z = 6.33$, seen in the left panel of Fig. 2. This galaxy has a quite small Balmer break strength with low error margin, placing it outside the 2.2–97.8th percentile even if considering the error. The spectrum shows that the break is extracted in a region where a notable fluctuation in the continuum shape is present. This has an impact on the calculated break strength and could therefore possibly explain the low value.

One defining result of these simulations is the formation of compact star clusters, most of which remain gravitationally bound, producing a remarkably stochastic SFH with short burst followed by periods of quiescence that in some simulations can last for 100–150 Myr at $z > 8$. We therefore expect that, since the simulated galaxy can be extracted either during a burst or during a long quiescent phase, the Balmer break strength would exhibit a large scatter. This is interesting because, as discussed later, the observed galaxy sample presented in this paper shows a rather large range of Balmer break strengths in their continuum spectra, where the large scale simulations tend to underestimate outliers with weak Balmer breaks in the distribution.

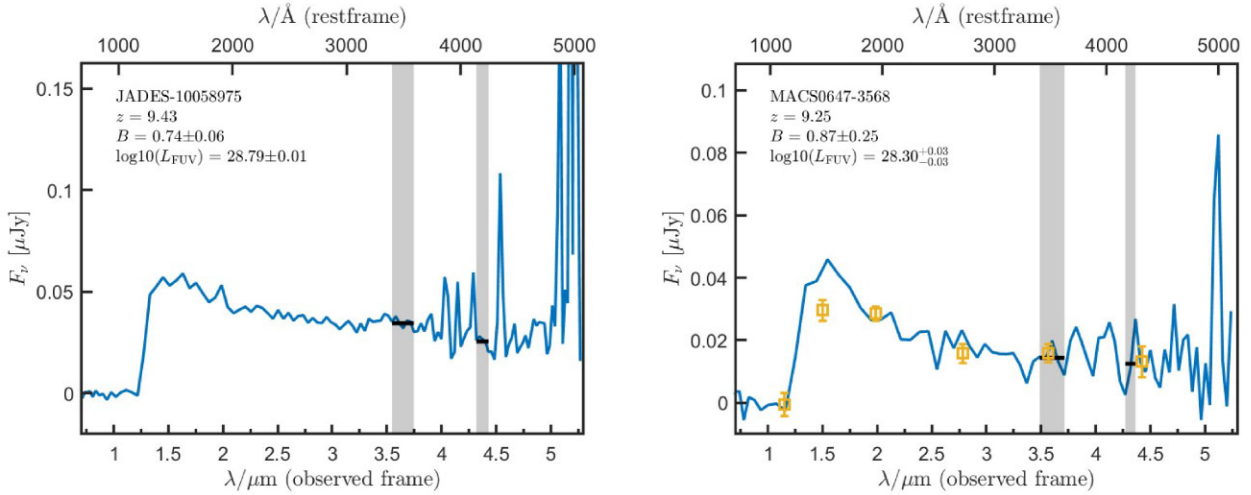


Figure 7. *JWST* NIRSpec/PRISM spectra and NIRCам photometry for the two $z \approx 9$ outliers in the left panel of Fig. 4. The JADES-10058975 galaxy lacks photometric data, which raises some concern due to its significantly weak Balmer break strength ratio with a suspiciously small error margin. MACS0647-3568 lies outside of the 2.2–97.8th percentile of the FLARES simulations but has error margins that places it within the distribution and shows an adequate agreement with photometry/spectroscopy in the break region.

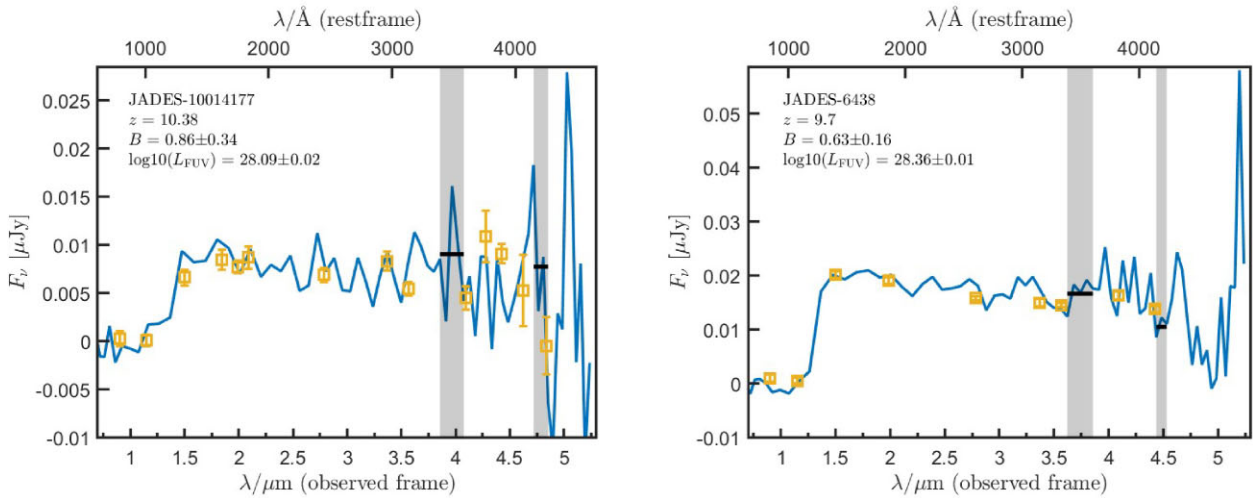


Figure 8. *JWST* NIRSpec/PRISM spectra and NIRCам photometry for the $z = 10.38$ (left) and $z = 9.7$ (right) galaxies seen in Fig. 4. JADES-10014177 shows a spectrum with significant noise/variation in the Balmer break region where spectra and photometry are hard to reconcile.

Fig. 5 shows the probability distribution function (PDF) of the Balmer break $B = F_v(4200 \text{ \AA})/F_v(3500 \text{ \AA})$ calculated from three simulation suites of the same galaxy with different assumptions for the *sub-grid* star formation efficiency (SFE).¹ We consider:

- (i) A constant sub-grid SFE = 35 percent (this is the least bursty simulation that produces fewer compact star clusters);
- (ii) A constant sub-grid SFE = 70 percent;
- (iii) A ‘realistic’ sub-grid SFE that depends on the density, mass and metallicity of the molecular cloud, derived from au-scale magnetohydrodynamic

¹The data for the constant SFE runs are the high- and low-SFE runs in Garcia et al. (2023), while the data for the run with ‘realistic’ SFE is work in preparation not yet published. The continuum spectrum for each galaxy is obtained by summing the spectrum (including nebular emission) from each burst of star formation using a 0.1 and 100 M_{\odot} Kroupa IMF between and constant metallicity $Z = 0.001$ in the STARBURST99 (Leitherer et al. 1999) SPS model.

simulations of star formation presented in He, Ricotti & Geen (2019, 2020).

Given that for these simulations we do not have a statistical sample of galaxies, the PDFs are calculated by extracting spectra for the simulated galaxies at random epochs (uniformly sampled in time) within the redshift range $7.5 < z < 12$ without binning with respect to the rest-frame far ultraviolet luminosity of the galaxy. However, we should point out that quiescent galaxies with a stronger Balmer break tend to be dimmer in the rest-frame far ultraviolet, hence they may escape detection more often, and vice versa – no such observational selection bias is applied to the PDFs. Fig. 5 shows that the run with high sub-grid SFE and ‘realistic’ sub-grid SFE show the most bursty/stochastic star formation, which is reflected in the wider range of B values. The run with ‘realistic’ sub-grid SFE shows a bimodal PDF with peaks at $B \sim 0.7$ and $B \sim 1.75$. We will discuss the implications of these results later in Section 4.

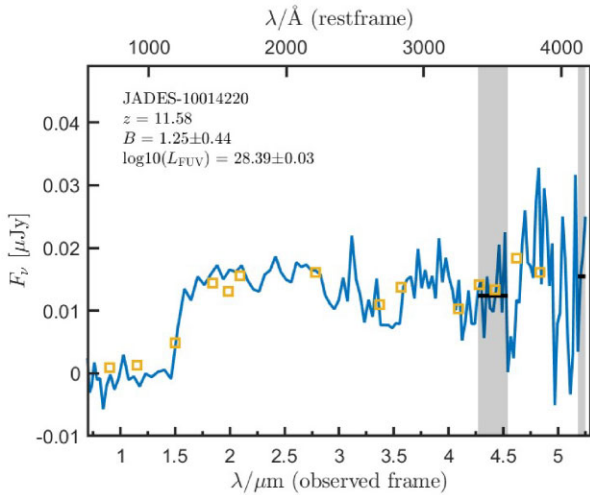


Figure 9. *JWST* NIRSpec/PRISM spectra and NIRCam photometry for the $z = 11.58$ galaxy in Table 1. The high redshift of this galaxy makes the extraction of the Balmer break strength challenging due to severe noise at the wavelengths where the break is measured.

2.2 NIRSpec/NIRCam observations of high- z candidates

In order to spectroscopically constrain the Balmer break in high redshift galaxies, we utilized prism data from the NIRSpec instrument on *JWST*, which covers the near-infrared wavelength range (0.6–5.3 μm) and allows continuum measurements beyond the rest-wavelength Balmer break for redshifts up to $z \lesssim 11.5$.

The data analyzed here partly comes from the *JWST* cycle-1 GO 1433 programme (PI Coe), which observed the cluster MACS J0647.7 + 7015 (hereafter MACS0647), and GO 2282 (PI Coe), which observed the cluster WHL0137-08 (hereafter WHL0137). These data were retrieved from MAST, which goes through STScI *JWST* pipeline² version 1.9.2. Spectral lines and features are fitted using MSAEXP³ version 0.6.0 to determine the redshift. Photometric data is processed through GRIZLI (Brammer et al. 2022).⁴

MACS0647 and WHL0137 are known cluster lenses which implies that gravitational lensing of the observed flux is to be expected for several of the galaxies. The galaxy WHL0137-1968 is gravitationally lensed into an arc and estimated to have a magnification of $\mu \sim 7.9_{-6}^{+12}$ (Bradley et al. 2022) based on the photometric redshift estimate of $z_{\text{ph}} = 9$. With a spectroscopic redshift of $z = 8.22$ we find a similar magnification estimate of $\mu \sim 8.6_{-6}^{+14}$ using the same lensing models as in Bradley et al. (2022). These models include Lenstool (Jullo & Kneib 2009), WSLAP (Diego et al. 2005, 2007), glafic (Oguri 2010), and Light-traces-mass (Broadhurst et al. 2005; Zitrin et al. 2009, 2015). The galaxy WHL0137-3429 is not located near the cluster core which places it in a region with minor magnifications. Hence, we here use a conservative estimate of $\mu \sim 1.1 \pm 0.1$. The galaxy MACS0647-3754 is also strongly lensed as indicated in imaging. Here, we estimate a magnification $\mu \sim 6.3_{-0.4}^{+0.5}$ using the glafic lens model. The remaining galaxies from the MACS0647 field were estimated to have smaller magnifications, found to be in the range $1.1 \lesssim \mu \lesssim 2.5$ with errors in magnification of around ± 0.1 . This introduces wider error bars in the calculated restframe

²<https://github.com/spacetelescope/jwst>

³<https://github.com/gbrammer/msaexp>

⁴Repositories at: <https://dawn-cph.github.io/dja>

far ultraviolet luminosity (L_{FUV}), but will not be large enough (for the weaker lensing estimates) to impact our conclusions regarding the agreement between observations and the simulated distributions, shown in Figs 2–4.

Additional galaxies from the JADES programme are retrieved from fully reduced public data (Bunker et al. 2023) with spectroscopic redshift estimates (see spectra in Fig. 1 and Figs. 6–9).

In order to reduce the impact of noise and unwanted features the spectrum is rebinned, resulting in a smoother continuum from which we can calculate the Balmer break strength. This slightly affects the width (i.e. wavelength range) of the regions we picked for the calculation. However, care has been taken to make sure that parts of the spectrum that the feature strong emission lines have been left out.

The far ultraviolet luminosity is calculated from the spectra using the restframe continuum at 1500 \AA after scaling the observed spectra to the photometric measurements. Scaling the spectrum also enables us to check the consistency between photometry and spectroscopy in the region where the Balmer break is measured. Discrepancies between the two would make the inferred Balmer break strength less robust and such a case was noted with JADES-10014177 (see Fig. 8). We discuss this further in Section 4.

The total stellar masses for the objects in MACS0647 and WHL0137 are estimated using piXedfit (Abdurro’uf et al. 2021, 2023) and corrected for gravitational lensing. Note that we include no corresponding plots for the Balmer break strength as a function of total stellar mass in this paper (see Wilkins et al. 2023, for simulations). We furthermore include no total stellar mass estimates for the galaxies taken from the JADES programme.

Table 1 is not an exhaustive list of all spectroscopically confirmed galaxies with possible Balmer breaks, but this data set should serve as a quantitative sample for the purposes of seeing whether present simulations are in line with observations and to rule out frequent departures from model predictions.

Presently there are several galaxies in the literature showing significantly strong Balmer breaks. Hashimoto et al. (2018) estimates a Balmer break strength $\sim 2.3 \pm 0.4$ (recently challenged by *JWST* observations; Bradac et al. 2023) using only slightly different wavelength ranges to define the break. Wilkins et al. (2023) also presents a Balmer break strength estimate from Carnall et al. (2023) which suggests a break strength of ~ 2.5 , although in a galaxy at lower redshift ($z \approx 4.7$). The galaxy studied by Looser et al. (2023a) is also represented in our sample (JADES-8115) where we find a clear Balmer break with a strength of ~ 1.6 (see Fig. 1), which is still within the distribution of the simulations by Wilkins et al. (2023).

We moreover include one galaxy with $z = 11.58$ (see Fig. 9) for which the rest-frame 4200 \AA continuum data point longward of the Balmer break falls at the very edge of the NIRSpec range. The shifting of the Balmer break to these very red wavelengths combined with the low brightness introduces severe noise, which showcases that the extraction of a reliable Balmer break measurements at such high redshifts is very challenging and thus limits the redshift range where the Balmer break is applicable as a proxy for age. No comparison to simulations is presented for this object, since we currently lack simulated data at $z \gtrsim 11$. However, extrapolation from the $z = 10$ data suggests it would clearly be consistent with simulations.

3 RESULTS

We assemble spectra from 23 galaxies with spectroscopically confirmed redshifts and calculated their Balmer break strength (see Table

1). Based on the simulations included in this paper we find that the majority (19 out of 23) of our observed Balmer breaks (including the $z \sim 11$ galaxy, JADES-10014220) can be accounted for considering the overall distribution of simulated galaxies as the calculated values fall within the shaded areas surrounding the median in Figs 2–4. Considering the uncertainties in the calculated Balmer break strength we find that 20 out of 23 galaxies are consistent with the simulations. Some of the galaxies have far ultraviolet luminosities below the resolution limit of the simulations (or outside the simulated redshift range in the case of JADES-10014220) but can be extrapolated to fall within the predictions. However, we find three clear candidates (JADES-18846 and 10058975, MACS0647-3754) which show Balmer breaks that deviate significantly from the simulated data such that they are consistently lower than those predicted by the simulations, even when considering the 2.2–97.8th percentile variations from the median in FLARES and the measurement uncertainties of the break strength. Since MACS0647-3754 is gravitationally lensed with resulting large errors in L_{FUV} , this outlier can potentially be alleviated, such that it falls into the distribution if sufficiently lensed. Using one lens model (glafic), we estimate that MACS0647-3754 is lensed by $\mu \sim 6.3 \pm 0.5$, which is just slightly too low to place it within the simulated distribution. The discrepancy between observation and simulation is more considerable for the higher redshifts ($z \gtrsim 8$) where we albeit have fewer observed galaxies with a well-constrained Balmer break.

Our analysis shows that the galaxies included in the JADES programme reveal an acceptable level of statistical agreement between the observed and simulated data sets. The majority of our JADES observations exhibit consistency with the simulated results for redshifts $z \gtrsim 6$. However, an exception arises in the case of JADES-10058975 (see left panel Fig. 7 for spectra), showing a weak Balmer break strength of 0.74 ± 0.06 at $z \sim 9$. Considering the relatively minor uncertainty associated with this measurement, it deviates significantly from the corresponding simulations by more than 3σ from the 97.8 per cent significance interval.

The galaxies in the MACS0647 and WHL0137 clusters also show good consistency with the simulations.

There are no objects in our sample with extreme Balmer break strengths ($\gtrsim 2$) that suggest any major tensions with standard models of galaxy formation.

4 DISCUSSION AND CONCLUSION

In this paper, we evaluated the magnitude of the Balmer break in 23 spectroscopically confirmed galaxies spanning redshifts from 6.1 to 11.6. For this, we utilized *JWST*/NIRSpec data obtained from the GO 1433 and GO 2282 programmes together with the publicly available JADES observations. Despite our analysis revealing a reasonable agreement between the observed strength of the Balmer break and contemporary simulations, given the uncertainties inherent in our measurements, some outliers were identified (see Section 3). These Balmer breaks fall below the predictions by about 60 per cent from the median with respect to FLARES, while slightly less (by ~ 10 per cent) than that when comparing to the median of the DELPHI models.

Currently, other observations of galaxies at high redshifts suggest a significantly more bursty SFH than normally expected in standard models of star formation (Endsley et al. 2023a, b; Looser et al. 2023b; Sun et al. 2023a, b). Such bursty star formation implies a sporadically replenished young stellar component in the galaxy, resulting in a highly varying Balmer break strength which cannot grow particularly strong if burst reoccurs on timescales shorter than

~ 300 Myr. Overall, this would push the simulated Balmer breaks down and increase the scatter in Figs 2–4, and in doing so somewhat improving the agreement between models and observations.

Our considerations of the Garcia et al. (2023) model highlights how the high spatial and mass resolution of the simulation, as well as the sub-grid star formation recipe assumed on scales of molecular clouds, can affect the stochasticity of star formation and therefore introduce a scatter in the Balmer break strength. Fig. 5 shows how the Balmer break strength can vary over ranges $B \sim 0.5\text{--}2.5$ given realistic sub-grid star formation efficiencies. The predictions in Fig. 5 are for simulations of a single galaxy where the spectrum is extracted in the range $7.5 < z < 12$, with no specified far ultraviolet luminosity dependence or cut. The large range in redshift and the unspecified far ultraviolet luminosity will clearly imprint itself in the Balmer break strength and therefore explains some of the scatter observed in the Garcia et al. (2023) PDFs with respect to the other simulations presented in this paper. Wilkins et al. (2023) shows that in FLARES, a typical galaxy can have a $B \sim 0.8$ scatter in the median Balmer break strength due to the combined effects of such a large redshift span and their simulated range of far ultraviolet luminosities. The bimodal PDF found for the realistic sub-grid SFE model in Garcia et al. (2023) show a somewhat larger scatter, and could potentially provide a mechanism that explains some of the outliers found in our observed sample.

In this paper we find no cases of extreme Balmer breaks ($B \gtrsim 2$) at high redshifts. Taking our observations at face value suggests that such extreme Balmer breaks are indeed rare. However, one must acknowledge that the 23 galaxies presented in this paper do not provide a sufficiently large sample in order to draw statistically relevant conclusions regarding the observed and simulated distributions for all included redshifts. We therefore risk missing some of the important details of the true overall distribution of Balmer breaks, especially for high-redshift galaxies ($z \gtrsim 10$) where the inherent challenge of attaining high-quality spectra of significant numbers of galaxies is evident.

We conclude that FLARES (Wilkins et al. 2023) make predictions that agree fairly well with observations as ~ 87 per cent of the galaxies included in this paper agree with the predictions given that we allow reasonable uncertainties and variations from the simulated median. We do however find several accounts of particularly low Balmer break strengths, falling below the simulated median with as much as ~ 60 per cent in the more extreme cases. Looking at the spectra of one of these galaxies (JADES-10058975, left panel Fig. 7) reveals a very blue slope and strong spectral features in the break region, resulting in the weak break strength. The underlying reason for this is speculative but could be explained by a significantly young stellar population with strong nebular emission from highly ionizing stars (see e.g. Cameron et al. 2023).

While we find reasonably good agreement between simulations and our observed galaxy population, we observe several galaxies with Balmer breaks quite far from the median predictions of FLARES. If this is the actual case, one could argue that there is some missing ingredient from the simulations related to feedback of e.g. early growth of black holes, stochastic SFHs etc.

The predictions on the Balmer break strength from the DELPHI model have a slightly lower median and a smaller distribution than those seen in FLARES. The discrepancy between model and observation for the outliers with weaker Balmer breaks is therefore lower here. However, the objects with particularly weak Balmer break strengths are still to be considered as outliers in the distribution. Considering the lower median and more narrow distribution of the DELPHI models, we find that the galaxies with high observed Balmer

break strengths (see e.g. $z = 7$ in Fig. 2) are harder to reconcile with the models.

Several galaxies show spectral features and noise that make the measurement of the Balmer break strength challenging. This can be seen clearly in, e.g. JADES-10014177 (left panel, Fig. 8). The spectrum shows a lot of noise which is reflected in the relatively large error in the measured Balmer break strength. This also reveals an example where photometry and spectroscopy are hard to reconcile in order to help constrain the Balmer break. JADES-18846 (Fig. 6) also reveals a scenario where a spectral feature (potentially instrumental) complicates the break measurement. The estimated errors are very small due to a relatively clean spectrum in the break region, but the spectral feature pushes the break strength down slightly. The true Balmer break strength could potentially be closer to ~ 1 , which would bring this galaxy right to the very edge of the 97.8th percentile of the simulated FLARES distribution.

While one of the aims of this paper was to utilize spectroscopy as a powerful tool to alleviate some of the obstacles related to photometric measurements of the Balmer break strength, we see that even with spectroscopic data our analysis is rarely straightforward. Due to noise and spectral features in the observed wavelength ranges corresponding to the Balmer break for high-redshift galaxies, we find few clear examples of a Balmer break (such as in Fig. 1), but often rather measure the slope of the continuum. This, on the other hand, is still revealing of the underlying physical processes of star formation that is at play and therefore serves an important purpose when measured.

Larger observational data sets with more calculated Balmer breaks will improve the robustness of our findings. We conclude however that given our current data set, we can find no significant deviations from predictions based on standard models of structure and star formation.

So the conspicuous title: To be or not to be; Balmer breaks in high-*z* galaxies with *JWST*, cannot be unambiguously answered by this paper alone. While our findings indicate that the simulated predictions agree fairly well with observations we find no cases similar to the extreme Balmer breaks presented in the literature – suggesting such cases are indeed rare.

ACKNOWLEDGEMENTS

AV and EZ acknowledge funding from the Swedish National Space Agency. AN and EZ acknowledge funding from Olle Engkvists Stiftelse. EZ also acknowledges grant 2022-03804 from the Swedish Research Council. EZ has benefited from a sabbatical at the Swedish Collegium for Advanced Study. PD acknowledges support from the Dutch Research Council (NWO) through the award of the VIDI Grant 016.VIDI.189.162 (‘ODIN’) and the European Commission’s and University of Groningen’s CO-FUND Rosalind Franklin program.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Abdurro’uf, Lin Y.-T., Wu P.-F., Akiyama M., 2021, *ApJS*, 254, 15
 Abdurro’uf et al., 2023, *ApJ*, 945, 117
 Anders P., Fritze-v. Alvensleben U., 2003, *A&A*, 401, 1063

- Atek H. et al., 2023, *MNRAS*, 524, 5486
 Binggeli C. et al., 2019, *MNRAS*, 489, 3827
 Bradac M. et al., 2024, *ApJL*, 961, 21
 Bradley L. D. et al., 2022, *ApJ*, 955, 13
 Brammer G., Strait V., Matharu J., Momcheva I., 2022, Zenodo, grizli.
 Broadhurst T. et al., 2005, *ApJ*, 621, 53
 Bruzual A. G., 1983, *ApJ*, 273, 105
 Bunker A. J. et al., 2023, preprint (arXiv:2306.02467)
 Cameron A. J., Katz H., Witten C., Saxena A., Laporte N., Bunker A. J., 2023, preprint (arXiv:2311.02051)
 Carnall A. C. et al., 2023, *Nature*, 619, 716
 Curtis-Lake E. et al., 2023, *Nature Astron.*, 7, 622–632.
 Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, *MNRAS*, 445, 2545
 Dayal P. et al., 2022, *MNRAS*, 512, 989
 Diego J. M., Protopapas P., Sandvik H. B., Tegmark M., 2005, *MNRAS*, 360, 477
 Diego J. M., Tegmark M., Protopapas P., Sandvik H. B., 2007, *MNRAS*, 375, 958
 Endsley R. et al., 2023a, preprint (arXiv:2306.05295)
 Endsley R., Stark D. P., Whitler L., Topping M. W., Chen Z., Plat A., Chisholm J., Charlot S., 2023b, *MNRAS*, 524, 2312
 Garcia F. A. B., Ricotti M., Sugimura K., Park J., 2023, *MNRAS*, 522, 2495
 Hashimoto T. et al., 2018, *Nature*, 557, 392
 He C.-C., Ricotti M., Geen S., 2019, *MNRAS*, 489, 1880
 He C.-C., Ricotti M., Geen S., 2020, *MNRAS*, 492, 4858
 Jullo E., Kneib J. P., 2009, *MNRAS*, 395, 1319
 Katz H. et al., 2021, *MNRAS*, 507, 1254
 Kokorev V. et al., 2023, *ApJL*, 957, 7
 Kriek M. et al., 2006, *ApJ*, 645, 44
 Kriek M., van Dokkum P. G., Whitaker K. E., Labbé I., Franx M., Brammer G. B., 2011, *ApJ*, 743, 168
 Labbé I. et al., 2023a, preprint (arXiv:2306.07320)
 Labbé I. et al., 2023b, *Nature*, 616, 266
 Laporte N., Meyer R. A., Ellis R. S., Robertson B. E., Chisholm J., Roberts-Borsani G. W., 2021, *MNRAS*, 505, 3336
 Leitherer C. et al., 1999, *ApJS*, 123, 3
 Looser T. J. et al., 2023a, preprint (arXiv:2302.14155)
 Looser T. J. et al., 2023b, preprint (arXiv:2306.02470)
 Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2020, *MNRAS*, 500, 2127
 Ma X. et al., 2020, *MNRAS*, 493, 4315
 Mauerhofer V., Dayal P., 2023, *MNRAS*, 526, 2196
 Oguri M., 2010, *PASJ*, 62, 1017
 Planck Collaboration, 2014, *A&A*, 571, A16
 Roberts-Borsani G. W., Ellis R. S., Laporte N., 2020, *MNRAS*, 497, 3440
 Rosdahl J. et al., 2018, *MNRAS*, 479, 994
 Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2023, *ApJ*, 943, 81
 Steinhardt C. L., Sneppen A., Clausen T., Katz H., Rey M. P., Stahlschmidt J., 2023, preprint (arXiv:2305.15459)
 Stiavelli M., Morishita T., Chiaberge M., Grillo C., Rosati P., Schuldt S., Trenti M., Treu T., 2023, *MNRAS*, 957, 18
 Sun G., Faucher-Giguère C.-A., Hayward C. C., Shen X., 2023a, *MNRAS*, 526, 2665
 Sun G., Faucher-Giguère C.-A., Hayward C. C., Shen X., Wetzel A., Cochrane R. K., 2023b, *ApJL*, 955, 35
 Trussler J. A. A. et al., 2023, *MNRAS*, 527, 11627
 Vijayan A. P., Lovell C. C., Wilkins S. M., Thomas P. A., Barnes D. J., Irodotou D., Kuusisto J., Roper W. J., 2020, *MNRAS*, 501, 3289
 Wilkins S. M. et al., 2023, *MNRAS*, 527, 7965
 Zitrin A. et al., 2009, *MNRAS*, 396, 1985
 Zitrin A. et al., 2015, *ApJ*, 801, 44

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.