1 Introduction

It is now well established that the three-dimensional structure of the genome is important for cellular function (Lieberman-Aiden et al., 2009) and with the increasing amount of high resolution and throughput chromosome conformation capture (3C) data becoming available, such as Hi-C (Lieberman-Aiden et al., 2009), Promoter Capture Hi-C (Schoenfelder et al., 2018), Capture-C (Hughes et al., 2014) and Tri-C (Davies et al., 2017), there is a need to understand chromatin structure beyond visualizing data on a 2D genome browser and using heatmaps. The advent of sophisticated microscope imaging of chromatin to observe these structures using super resolution microscopy (Prakash, 2017) and electron microscopy (Ou et al., 2017) offers the ultimate means of visualising and understanding 3D genome architecture but these methods are laborious and expensive. Computational modelling offers a way to gain a better understanding of the complexity of chromatin in the nucleus and how the differences in structure cause enhancer/promoter/gene interactions in different cell types and disease. There are a number of methods for modelling chromatin 3D structures from 3C data but still a lack of easy to use tools (Oluwadare et al., 2019) so C data may be better understood by bench scientists.

Here, we present CSynth, an easy to use web-based portal that allows uploading of multiple 3C datasets, PDB models, annotations and quantitative data to generate 3D models of chromatin structure. The models and their parameters are interactive and may be manipulated in real time and compared in a high-quality fully rendered 3D genome browser that can be shared online and used in publications.

2 Materials and methods

2.1 Using CSynth

CSynth was developed to lower the barriers to the interrogation of complex multi-genomics data in the 3D rather than 2D genome. While the generation of genome-scale 3C data, such as Hi-C, is becoming commonplace, the computational barriers to generate 3D models from such data remain high. Even more limiting are the options to interact with such models in a dynamic nature and in concert with other classes of genomics data, such as ChIP-seq, ATACseq and RNA-seq. CSynth provides a flexible platform for the generation of restraint-based models as well as a fully featured environment to interact with these or externally generated models in a...
3 Results

3.1 Comparison to other 3D visualization tools

There are currently several 3D genome browser implementations suitable for looking at 3D chromatin structure. See Supplementary Table S1 for an overview of the 3D genome viewing tools available. A key problem CSynth is addressing is to make a high-quality 3D modelling accessible that is easy and fast to run so that any person generating 3C data will use it routinely to gain further insight from their experiments. Another key factor is the results should be high quality so they can be used in publications.

Genome3d (Asbury et al., 2010) is a downloadable C++ application, which requires a computer running the Windows OS and the installation of software which makes it more limited for general use. GMOL (Nowotny et al., 2016) does not handle Hi-C data, but more recently the author has released GenomeFlow (Trieu et al., 2019) which offers a full Hi-C analysis pipeline. However, using Java requires the user to install the relevant Java version as opposed to using the desktop browser, again causing a barrier to entry to anyone who wants to rapidly and easily visualize their analysis (data). Tadkit (Serra et al., 2017) is web based and shows a 3D chromatin view in the context of a 2D browser based on IGV (Nicol et al., 2009) but there is no possibility provided to show different states (e.g. in different tissues).

3.2 Examples of CSynth modelling

In Figure 2, we show data generated from Capture-C data (Oudelaar et al., 2018) at the alpha globin region in mouse captured in erythroid cells at 4 kb resolution. Clearly visible is the chromatin looping of the x-globin (mm9, chr11:32 000 000–32 300 000) self-interacting domain. The coloured sections of the model represent genes loaded as Browser Extensible Data (BED) format and the ChIP-Seq data uploaded as Wig format. A video showing uploading and general features of CSynth can be seen in the ‘Media section’ on the CSynth website at csynth.org or directly on YouTube (see https://youtu.be/SMgw_cFeH6Q and https://youtu.be/O6W10Y 1o04). More details on the modelling may be found in Supplementary Section S1.

In Figure 3, we show an example of loading a large Hi-C dataset at 2 kb resolution from Schizosaccharomyces pombe Chromosome I, comparing the difference between mitosis and interphase states (Kakui et al., 2017) using CSynth’s dynamic GPU modelling. To find the parameters for modelling, we used the distance between certain chromosomal locations (Petrova et al., 2013). In interphase (Fig. 3a), chromatin fibre forms a characteristic structure and its telomeres are located in the vicinity as expected from Rabl orientation within the interphase nucleus in S.pombe (Funabiki, 1993). This is where the centromeric region (the centre of which is visible between the green and red arms in Fig. 3) and telomeres attach to the nuclear lamina which causes the overall structure to bend at this point. Here, CSynth shows it has several interesting folding patterns and shows looping that is not obvious in the heatmap view. In mitosis (Fig. 3b), one can see the structure is more compact, folding into the characteristic structure and each arm becomes individualized.

3.3 Modelling methods

There are a large number of 3D genome modelling methods available that can be represented by polymer, spheres or point-based models (Oluwadare et al., 2019). Benchmarking all available models is beyond the scope of this article but we compare CSynth’s modelling to ones that apply a similar point-based approach used by Chromosome3D (Adhikari et al., 2016) and LorDG (Trieu and Cheng, 2017). A key point is CSynth’s modelling, which is done quickly, in real time and is interactive which encourages the user to explore and gain an intuitive feel for the model by varying parameters. CSynth’s model is constantly being recalculated so transitions between states are animated and there is direct feedback when the user adjusts parameters on the model. An overview of the modelling process is shown in Figure 4.
on very long distances. This allows simpler (to compute) dynamics. Our dynamics are inspired by Poing (Jefferys, Kelley and Sternberg, 2010) largely based on spring-like forces. Some of the forces used in our dynamics may be related to real physical forces but the relationship is usually indirect; our dynamics are better thought of as an emulation rather than a simulation or modelling. The various forces we have built in CSynth are detailed in the Supplementary Section 1. Our dynamics work directly from IF or distance map inputs, which are held in sampler buffers on the GPU. The modelling system is based on particles (also referred to as beads), which are represented using the size of the fragment from the capture experiment. The particles generally match the Hi-C bands one to one, but we permit the use of multiple particles per cell for more refined modelling. The particles are assumed to be joined in a backbone chain (or chains). The modelling operates in conventional Newtonian dynamics steps, where in each step an overall force is computed on each particle; the force is then applied to the velocity which is used to compute a new position. The yeast model shown in figure 3 (2798 particles) are produced in less than 30 seconds on a 3.4GHz i7 machine with 16GB RAM and a GTX 1080 graphics card whereas most modelling packages take many minutes to several hours. The number of particles is limited by GPU texture constraint which is typically, as of writing, 16,000 particles. In tests, we have resolved models of 6284 particles (3 chromosomes of yeast at 2k resolution, see Supplementary Section 2) in a few seconds on an NVIDIA GTX 1080.

3.4 Model and data comparison
The main purpose of CSynth is for interactive 3D modelling of IF data, and comparison of states from multiple IF sources. It can also be used to visualize and compare data from other sources and can import and display static data in xyz or pdb format. CSynth does this by creating distance-based spring models from the distance data implicit in xyz data. These models permit inbetweening of different datasets to visualize their differences and similarities. Such model-based inbetweening is smoother and more informative than simple linear inbetweening of xyz coordinates; and also, naturally aligns the visual output. Furthermore, CSynth can move smoothly between its own models of IF data and imported distance-based models. For example, with the mouse example (Fig. 2), we have both IF data for erythroid and non-erythroid states (embryonic stem cell), and also a xyz data from an independently derived polymer model (Chiariello et al., 2020). Loading all four datasets (2 IF, 2 xyz imported) into CSynth allows the visualization of the differences between the states for each of the models, and between the CSynth and external models for each of the states. The differences can be visualized by transitions between the states or by history trace view similar to that shown in Figure 3.

3.5 Comparison with other modellers using simulated data
The principal feature of CSynth is the ability to visualize and interact with the modelling, better to understand both the data and the modelling. We do not make strong claims for the CSynth modelling, but illustrate here that it is comparable with other recent restraint-based modellers. We carried out tests to compare with LorDG and Chromosome3D, using the LorDG simulated datasets to permit some statistical verification. We added code to the modelling framework based on the LorDG Lorentzian function to allow these comparisons to be conveniently done in CSynth itself. We used chr20 from chainDres25 from the MissouriBox dataset, and loaded the IF data plus the resulting 10 pdb results files, 5 from LorDG modelling and 5 from Chromosome3D modelling which we show graphically using the history trace view shown in Figure 5. As in the section above, we could perform visual comparisons between the different models, and different runs from the imported models. Visual comparison immediately showed that 3 of the LorDG results were almost identical (apart from orientation), and brought out the differences with the rest. We were able to vary the parameters of both the LorDG model (such as c and alpha) and CSynth, and see
their impact, and that visual differences between our model and the LorDG model with corresponding parameters were very small.

We also applied statistics using multiple runs and comparing results with their ‘definitive’ simulated data. The statistics of this single experiment indicates that CSynth modelling gives marginally better results than either LorDG or Chromosome3D (Table 1). The differences are very small, and the statistical methods, scale of the experiment and the use of simulated data limit what conclusions we can safely draw.

### 3.6 Availability

CSynth can run directly in Chrome and Firefox and has been tested on all major operating systems (including tablets). It is available from csynth.org where there are several example models and instructions for use. For larger models (more than 500 contact points), it is advised to use a discrete graphics card. The absolute limit typically is 16 000, but depends on the maximum texture size depending on the browser’s WebGL implementation. Data can either be uploaded to the CSynth portal (https://csynth.molbiol.ox.ac.uk) for later use and for sharing, or can be directly drag-dropped from the local file system for quick viewing. Code is open source and available at https://github.com/csynth/csynth.

### 4 Discussion

#### 4.1 Potential enhancements

The range of features CSynth supports adds complexity to the user interface. We aim to provide simplified interfaces for common applications based on user feedback. We are extending CSynth documentation of several existing features: scripting and API (JavaScript or Python via websockets). We plan to extend CSynth VR HTC Vive support to other eXtended Reality platforms.

### 4.2 Summary

CSynth provides a high quality, interactive, user friendly and powerful way of visualizing chromatin interaction data, by combining model, heatmap and genome annotations in one display in a standard web browser. These features are critical when trying to understand how structure and biological activity are interconnected in genome function. A key improvement in CSynth, in comparison to other currently available tools, is that modelling is done on the GPU dynamically. This allows the user to load chromosome capture matrices quickly and vary model parameter values for a better understanding of their effect on the modelling process. Another unique feature of CSynth is the facility to view and compare models between any number of different samples (e.g., tissues or cell types) or even other modelling systems. Finally, we use VR to view and interact with these complex 3D structures which helps get a better intuition for the 3D modelling and is also useful for teaching and public engagement. We foresee that CSynth has the potential to be an invaluable tool to understand the structure and dynamics of more complex systems, such as data generated from different samples from existing and new 3C-based techniques such as single-cell Hi-C (Stevens et al., 2017).

### Acknowledgements

The authors acknowledge the use of the MRC WIMM Centre of Computational Biology computational infrastructure which hosts this work.

### Funding

This research was funded in part by Medical Research Council [MC\_UU\_12025] and Wellcome Trust [106130/Z/14/B].

### Conflict of Interest

None declared.

### References


Stevens, T. J. et al. (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature, 544, 59–64.

