

**Figure 1.2** Illustration of the 3D shape of a protein called T1044/6VR4. The green structure shows the ground truth as determined by X-ray crystallography, whereas the superimposed blue structure shows the prediction obtained by a deep learning model called AlphaFold. [From Jumper *et al.* (2021) with permission.]



consist of a set of proteins for which the amino acid sequence and the 3D structure are both known. Protein structure prediction is therefore another example of supervised learning. Once the system is trained it can take a new amino acid sequence as input and can predict the associated 3D structure (Jumper *et al.*, 2021). [Figure 1.2](#) compares the predicted 3D structure of a protein and the ground truth obtained by X-ray crystallography.

### 1.1.3 Image synthesis

In the two applications discussed so far, a neural network learned to transform an input (a skin image or an amino acid sequence) into an output (a lesion classification or a 3D protein structure, respectively). We turn now to an example where the training data consist simply of a set of sample images and the goal of the trained network is to create new images of the same kind. This is an example of *unsupervised learning* because the images are unlabelled, in contrast to the lesion classification and protein structure examples. [Figure 1.3](#) shows examples of synthetic images generated by a deep neural network trained on a set of images of human faces taken in a studio against a plain background. Such synthetic images are of exceptionally high quality and it can be difficult to tell them apart from photographs of real people.

This is an example of a *generative model* because it can generate new output examples that differ from those used to train the model but which share the same statistical properties. A variant of this approach allows images to be generated that depend on an input text string known, as a *prompt*, so that the image content reflects the semantics of the text input. The term *generative AI* is used to describe deep learning models that generate outputs in the form of images, video, audio, text, candidate drug molecules, or other modalities.