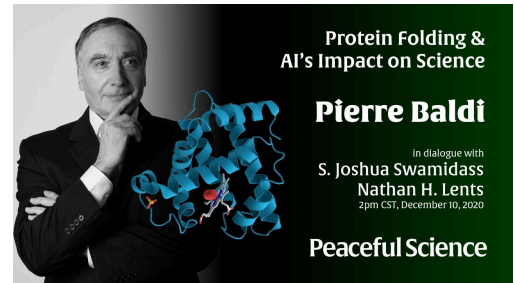




Pierre Baldi: Protein Folding and AI's Impact on Science

S. Joshua Swamidass

<https://peacefulscience.org/articles/baldi-ai-science/>



Google just announced that its Artificial Intelligence algorithm, AlphaFold, made a major advance in protein folding. To help make sense of this, I am interviewing Pierre Baldi (h-index: 106), one of the leading AI scientists in the world, and also my PhD advisor back in the day. Pierre is a Distinguished Professor at UC Irvine, and has been using AI to solve scientific problems for decades. He literally wrote the textbook I used in graduate school, and his *Deep Learning in Science* book is due to be published March, 2021.

What is Deep Learning?

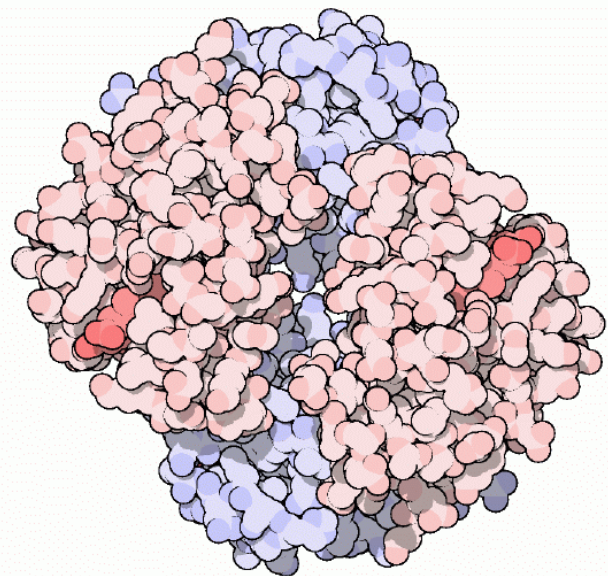
AlphaFold is a type of Deep Learning algorithm, a neural network that learns from data. That “learning” ability is one reason it is called Artificial Intelligence. Deep Learning can be used in many ways to address many tasks. In science, one task we care about is protein folding.

What is Protein Folding?

So what is “protein folding”? We want to know the three-dimensional structure of proteins from the amino acid sequence. Proteins are defined by a sequence of letters, where each letter corresponds to a different building block, like this:

```
VLSPADKTNVKAAWGKVGGAHAGEYGAELERMFL
SFPTTKTYFPHFDSLHGSQAQVKGHGKVKVADALNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHC
LLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
```

The task of protein folding is to translate this sequence of building blocks into a three dimensional structure, like this:



The protein structure of hemoglobin, in the style of David Goodsell. This is the structure encoded in the amino acid sequence above.

That structure becomes important. It can give important insight into how the protein functions. In this case, we are looking at hemoglobin, the oxygen carrying protein in our blood. Shifts in the structure when it binds oxygen are important to its function.

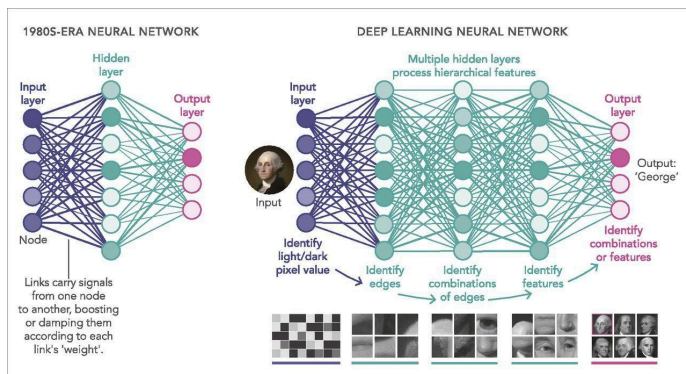
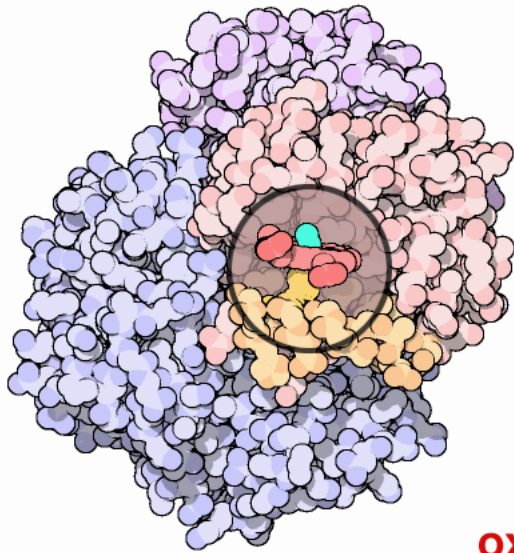


Figure from <https://www.pnas.org/content/116/4/1074> at PNAS. Deep Learning networks are often depicted in this way, with information flowing through nodes in a network to make a computation..



oxy

The shift in structure is key to how hemoglobin carries oxygen in our blood.

There are several experimental techniques to determine protein structure: X-Ray crystallography, NMR, and cryo-EM. These

References

<https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>
<https://www.youtube.com/watch?v=0Y2kQf54tpM>
<http://www.igb.uci.edu/~pfbaldi>
<https://scholar.google.com/citations?user=RhFhIIgAAAAJ&hl=en>
<https://peacefulscience.org/books/bioinformatics/>
<https://amazon.com/dp/B097HRKGDZ/>
<https://www.pnas.org/content/116/4/1074>

techniques have been used for decades, and produced thousands of protein structures that are stored in the Protein Data Bank. However, they are expensive and slow, and they often do not work.

So, could we build a computer program that could look at all the structures solved so far, and learn patterns from it, so that it could now predict the structure of a new protein? That is the basic challenge and opportunity of protein folding. As simple as that sounds, it turns out to be a very hard problem, for a whole host of reasons. But AI algorithms are proving to be particularly effective at this question

With AlphaFold's announcement, is protein folding a solved problem now? This tweet from a scientist nails the answer. "Protein folding is not a solved problem." Still, this is exciting. What exactly does it mean?

How is Deep Learning Impacting Science?

This is just one more example, on a growing list, of where Deep Learning is making a large impact on how scientists conduct their work. Pierre Baldi has been applying deep learning to solve scientific problems for decades. We will be discussing how Deep Learning is coming of age, and how it is shaping science in many ways.

https://en.wikipedia.org/wiki/Amino_acid#Table_of_standard_amino_acid_abbreviations_and_properties
<https://pdb101.rcsb.org/motm/41>
https://en.wikipedia.org/wiki/X-ray_crystallography
https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance_spectroscopy_of_proteins
https://en.wikipedia.org/wiki/Cryogenic_electron_microscopy
<https://www.rcsb.org>
<https://twitter.com/lpachter/status/1333702159850893312?s=20>