

chemistry

May 2017

in Australia

Using biology
to make
chemistry

chemaust.raci.org.au



- Trump's Order and the threat to scientific exchange
- New developments in helium microscopy
- Turning yoghurt waste into new products



Directed evolution

DNA polymerase, the enzyme that synthesises DNA molecules.

iStockphoto/somersault18:24

Following nature's lead

Dave Sammut speaks to Caltech professor Frances Arnold, who is 'using biology to do chemistry'.

If DNA is a language of life, we have so far only learned the alphabet, and to read and write the simplest sentences. Designing and synthesising known sequences of DNA is one thing; predicting the function of the resulting proteins is quite another. Thus, tailoring biochemical agents for desired functions is a significant challenge. Evolution has the answer.

In February, I spoke with Frances Arnold, the Dick and Barbara Dickinson Professor of Chemical Engineering, Bioengineering and Biochemistry at the California Institute of Technology, US. She will be a plenary speaker at the RACI Centenary Congress later this year.

'How a protein sequence encodes what the protein does remains an open question'

Arnold has been awarded the 2016 Millennium Technology Prize (€1 million) for her work in a different approach to the design–function problem. 'I don't want to insult computational [scientists], but we cannot design enzymes in any reliable fashion, especially when it comes to the functions that are really unique to chemistry, to how enzymes make new molecules. We cannot predict what amino acid sequence would give rise to a desired function. But we would not be talking to one another unless there were a reliable design process for

making them – that process is evolution.'

Arnold uses the science of 'directed evolution' to create improved proteins in the laboratory, with specific outcomes in mind. In a 2015 lecture at Oak Ridge National Laboratory (bit.ly/2lrXLHI), Arnold stated that: 'DNA code was solved even before I was born – how DNA encodes a protein sequence, but how a protein sequence encodes the folded protein structure remains an open question. And that's not even the question that matters. It's "How does the sequence encode function?" Those are two big questions that I don't have time to wait for people to solve'. Arnold instead elected to use nature's own processes to solve real human problems.

'To me, the greatest engineer of all time is nature'

Arnold's point is that nature has already evolved immensely powerful tools to drive efficient, selective and renewable chemistry. 'To me, the greatest engineer of all time is nature', she says. Arnold's vision is to use nature's existing materials and inventive processes to efficiently produce chemicals in a sustainable fashion: 'Life is poised to do chemistry that we never dreamed of. We are using biology to do chemistry that human chemists thought only they could do, and more'.

Arnold's vision is to use nature's existing materials and inventive processes to efficiently produce chemicals in a sustainable fashion.



Frances Arnold

Of course, humans have been utilising nature's processes for some time to build new biological variants. The chihuahua is a long process of selective breeding away from the wolf. After generations of selection, staple crops and grains have yields far beyond their early agricultural ancestors, and with modern genetic techniques we are experimenting with other properties such as pest protection.

Directed evolution's different approach is to use the empirical 'better or worse' selection from evolutionary processes, rather than more theoretical methods of protein design. In Arnold's words: 'Chemistry is hard to design, but easy to evolve'.

'Nothing comes from nothing'

Arnold emphasises that 'nothing comes from nothing'. The starting point to approach any new chemical function is to find some biological agent – a protein – that does the desired job to some extent, no matter how infinitesimally small.

Take Arnold's recent work on creating enzymes that catalyse the formation of a carbon-silicon bond, a bond nature is not known to make (Kan et al., *Science* 2016, vol. 354, pp. 1048–51). Human chemists have never made such organosilicon compounds using iron-based catalysts. Arnold's co-workers started by screening haem proteins to find ones that functioned even at extremely low levels. From these, they knew that the chemistry

was at least possible. 'Planning is important. Choosing where to start is a rational, knowledge-based process.'

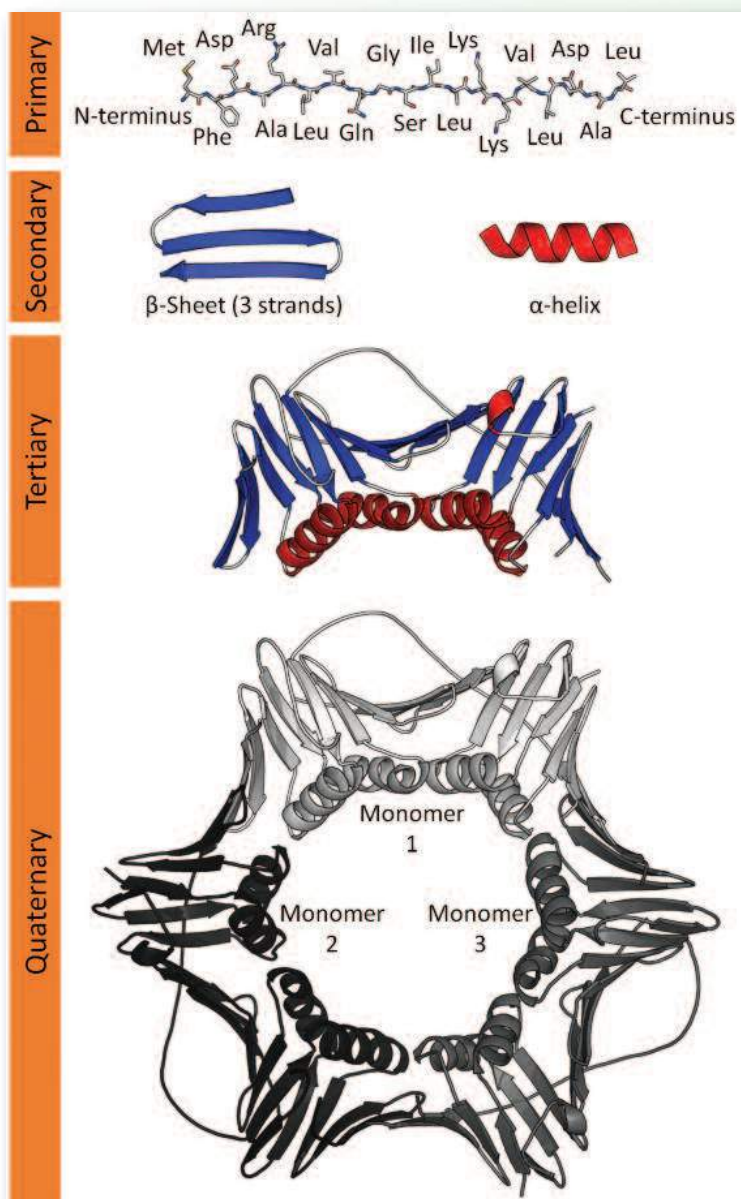
Next, Arnold applied processes for directed and random mutation of the proteins (for example, controlled mutation using the polymerase chain reaction), creating hundreds of physical samples that they screened by well-established analytical methods for improved activity in the new reaction. The genes encoding the improved proteins were recombined or mutated again to enable beneficial

mutations to accumulate. The evolved enzyme they reported formed organosilicon compounds inside bacterial cells, essentially 'bringing silicon to life' for the first time.

By starting with functional proteins that are homologous or evolutionarily related, Arnold significantly increases the odds of success. With 20 amino acids and at least 400 peptides in a protein chain, that makes for an astronomical number of combinations – at least 20^{400} . Most of these combinations won't be

functional. With more than just two or three random mutations, the probability of creating a functional protein is very low. Arnold's approach allows her to accumulate dozens or sometimes hundreds of mutations while still achieving a reasonable chance of successful variants.

'When we do this, we create so-called "chimeric" structures. What we've learned is that recombination is conservative. When we do this, we can create large numbers of [DNA] sequences. We can synthesise these



Proteins 101

Amino acids: These organic molecules are used by living organisms to make proteins. There are about 20 types of amino acids, which we source by breaking down the proteins in the food we eat.

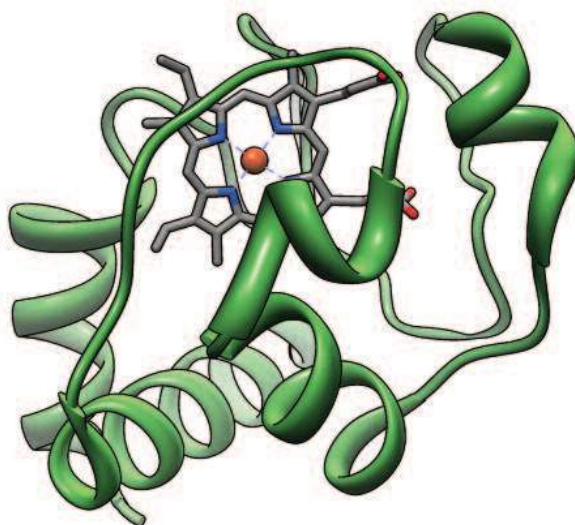
Proteins: These are long chains of amino acids. They are the most diverse group of biologically active substances, and are effectively the central component to life. There are thousands of different proteins just in the human body, with many different types, including structural (such as collagen and keratin), defensive (such as antibodies), transport (such as haemoglobin) and catalysts (such as enzymes).

Making a protein: Via 'transcription', a copy of the DNA (deoxyribonucleic acid) in a cell is created, called RNA (ribonucleic acid). This is something like the 'printer instructions'. Via 'translation', the RNA is transported to a ribosome, something like the 'protein printer'. The ribosome finds a 'beginner' sequence on the RNA called a 'codon', then moves sequentially down the strand of RNA, where every three 'letter' combination represents another amino acid 'peptide' molecule. The ribosome builds a string of peptides, until it finds the 'stop' point on the RNA, and the protein synthesis is complete.

Protein structure: The particular three-dimensional shape of each protein defines its highly specific function. If this shape is altered, the protein becomes denatured and does not function as expected. Four levels of structure are used to describe the protein:

- primary structure – the linear sequence of amino acids
- secondary structure – where hydrogen bonding between amino acids creates three-dimensional geometry, such as an alpha helix or a pleated sheet
- tertiary structure – the overall shape of the protein: globular (such as enzymes) or fibrous (such as muscle tissue)
- quaternary structure – the appearance when a protein is composed of two or more polypeptide chains.

By starting with functional proteins that are homologous or evolutionarily related, Arnold significantly increases the odds of success.



Three-dimensional structure of cytochrome c (green) with a haem molecule coordinating a central iron atom (orange). Frances Arnold's group have discovered 'that haem proteins catalyse the formation of organosilicon compounds under physiological conditions via carbene insertion into silicon-hydrogen bonds' (doi: 10.1126/science.aah6219).

Vossman/CCA-BYSA-3.0

and put them into bacteria and see who folds and functions.'

'Still, we do it safely. We have a computational algorithm so you can take more disparate sequences. We can choose the recombination cross-over points so as to maximise the number of these that will be folded and functional.'

'A simple process of random uphill walk'

A key point is that, rather than seeking the optimum solution, the process seeks one or more *incrementally better* solutions. According to Arnold, it's 'a simple process of optimisation by a random uphill walk – making mutations over multiple generations and collecting the beneficial effects'.

'The science of molecular breeding is choosing how big a barrier you have to jump in each level of the optimisation. If you choose that properly, you can continually walk uphill to adapt the protein to its new function.'

The surprising outcome is that with just a few generations – not tens or

hundreds – considerable improvements in yield, selectivity and/or other targeted properties can be achieved. 'We can do what nature takes millions of years to do in a matter of weeks. Proteins are very adaptable'.

For this, Arnold explains, the selection of the screening method is absolutely critical. 'You get what you screen for.' Arnold gives the example of a manufacturer who approached her years ago to develop cold-temperature enzymes for laundry detergents. She explained to the manufacturer that while she could definitely develop enzymes that would function well over a wide range of temperatures, the enzymes could easily lose their primary useful function – of taking stains off of clothes – unless she screened them in washing machines.

'Science is a small fraction of commercial success'

Having come to this area from an engineering background, Arnold describes herself as very much an applied scientist. While her laboratory

does a great deal of pure research, she has been involved in 13 start-ups to commercialise technologies arising from her work and others', six of which have gone public. She describes these as 'all great fun', but is particularly enthusiastic about her most recent start-up (Provivi), which is seeking to develop processes for commercial production of insect pheromones that can be safely sprayed onto crops to disrupt pest breeding cycles.

On the topic of commercialising new technologies, Arnold is passionate in the expression of her views: 'Science is a small fraction of commercial success. It takes business skills, timing and a great story'. She says that too many scientists develop a technology, then try to find an application for it. 'First', says Arnold, 'find a problem worth solving'.

Arnold is an engaging speaker, and we'll expect her to be addressing a packed house in July.

Dave Sammut FRACI CChem is principal of DCS Technical, a boutique scientific consultancy, providing services to the Australian and international minerals, waste recycling and general scientific industries.