BIOCHEMISTRY

Protein folding studies go global

A large-scale approach to analyze how protein sequence determines folding provides new insights into an old question.

Data-driven protein design enables the construction of novel protein sequences that fold into desired structures. Reprinted with permission from AAAS.

Proteins fold as a result of the multitudes of weak interactions between amino acids. Understanding how sequence determines folding is a question that has been around since it was discovered that every protein has a unique, three-dimensional structure, says David Baker of the University of Washington.

To address this question, researchers have mostly applied mutagenesis to simple protein domains and characterized the resulting effects on stability. Baker and his colleagues recently took a large-scale approach to tackle the question of how proteins fold from a new angle. "In any experiment, if you're collecting a small amount of data, the generality of the conclusions you can make is much smaller than if you have a large amount of data," he says.

His team, led by postdoc Gabe Rocklin, combined computational protein design, large-scale oligo library synthesis, yeast display, and a newly developed protease susceptibility assay to study protein stability in an expanded protein folding space. They first used computational modeling to design short (~40 amino acid) protein sequences intended to fold into desired topologies. To characterize the stabilities of thousands of protein designs at the same time, they borrowed a recently published, cost-effective method for massively parallel oligo library synthesis. They expressed the sequences and displayed them on the surface of yeast cells, and they used a 'protease susceptibility assay' to measure each sequence's tolerance to increasing concentrations of enzymes that break down protein

STEM CELLS

MANY PATHS FROM A XEN-LIKE STATE

An intermediate cell state of chemical reprogramming can be induced to convert directly between cell types.

A few years ago, Hongkui Deng and colleagues at Peking University in Beijing discovered a way to avoid supplying transcription factors by using chemicals to reprogram mouse somatic cells into pluripotent stem cells. As cells underwent chemical reprogramming, they passed through an intermediate state that resembles extra-embryonic endoderm (XEN). New work led by Deng and Zhen Chai at Peking University reveals that the XEN-like state is a unique multipotent condition from which cells can embark on diverse specification paths.

The original search for a chemical reprogramming cocktail was motivated by a wish to establish whether the Yamanaka factors used in traditional reprogramming were necessary, and to circumvent technical and safety concerns around transcription factor expression. "Compared to transgenic methods, small molecules are cell permeable, cost-effective, easy to synthesize, preserve and standardize, and their effects can be reversible," Deng and first author Xiang Li write in a joint e-mail. Chemical reprogramming presented the opportunity to study an alternative route to pluripotency.

In their latest work, the researchers determined that fibroblast-specific gene expression is downregulated in XEN-like cells induced from fibroblasts. "We found that the induced XEN-like cells expressed master genes governing cell fate choices toward three germ layers and cell lineages, suggesting broader lineage plasticity," write Deng and Li. By applying a version of a medium used to specify neurons from pluripotent stem cells, they found that XEN-like cells readily assume a neuronal fate.

The partially reprogrammed cells have some attractive properties. Their numbers can be expanded dramatically without losing key features. "XEN-like cells do not compromise their neuronal induction efficiency and retain genetic integrity and genome stability

bonds. Cells that displayed stable designed protein sequences were isolated using fluorescence-activated cell sorting (FACS). These enriched, stable sequences were identified using deep sequencing; and, finally, each design was assigned a stability score.

Baker's team designed tens of thousands of sequences intended to fold into one of four different topologies. They selected the best thousand designed sequences of each topology for experimental testing; each designed sequence was paired with two control scrambled sequences that were expected not to fold properly. Initially, the team was moderately successful in designing stable sequences to fold into just one of the four desired topologies. However, they used their large data set, which consisted of both positive and negative results, to iteratively inform and improve their computational design model. After four such rounds of design and experimental testing, they achieved many more successfully designed sequences for all but one of the intended folds, and they verified some of these folds by NMR-based structure determination. "The general idea that you can improve the scientific model by iterative learning is really pretty exciting," says Baker.

Altogether, the team's efforts generated 2,788 newly designed, minimal proteins that fold into desired structures. This represents an increase in at least an order of magnitude over the number of naturally occurring stable proteins of this size, says Baker. These novel proteins may also be useful in bioengineering or pharmacological applications.

Baker notes that most protein engineering to date has been done by taking a natural protein and tweaking the structure a bit to give it a new function, something he compares to how primitive humans made crude tools out of available materials such as bone. He foresees a future where, when one wants a new protein to carry out a new function, "you won't look around in nature for something to tweak, but you just build it from scratch to do what you want it to do."

Allison Doerr

RESEARCH PAPERS

Rocklin, G.J. *et al.* Global analysis of protein folding using massively parallel design, synthesis and testing. *Science* **357**, 168–175 (2017).

after long-term expansion," state Deng and Li. Neurons induced from XEN-like cells closely resemble primary neurons with respect to their gene expression, functional properties and ability to engraft in the brains of adult mice upon transplantation. Despite their proliferative potential, they do not generate tumors in animals.

The conveniences of chemical programming can be somewhat offset, however, by the narrow concentration windows and specific application durations needed for small molecules to be effective. To achieve the best results, the researchers recommend closely following the details of their protocols.

In addition to generating neurons, which are cells of ectodermal origin, the researchers converted XEN-like cells to hepatocyte-like cells, which are of endodermal origin. The developmental plasticity of the XEN-like state is still something of a mystery. During development, extra-embryonic endoderm is an inductive tissue that does not contribute cells to the embryo. Gene expression in XEN-like cells clearly differs from expression in the pluripotent state induced by the Yamanaka transcription factors. But XEN-like cells do express at least two pluripotency factors found in fully chemically reprogrammed cells. The researchers are actively looking into the mechanisms behind XEN-like plasticity. It will be interesting to understand the epigenetic state of these cells, in addition to their other properties.

The teams are working to generate other functional cell types and are interested in determining whether cells of mesodermal origin, such as cardiomyocytes and blood cells, can be differentiated from XEN-like cells. They are also pursuing the question of whether an analogous cell type can be induced in human cells. Deng and Li predict that "chemical reprogramming will become a very easy, simple, and popular approach." An expandable multipotent state that circumvents full pluripotency may have some practical advantages and shed light on the nature of pluripotency in cultured cells.

Tal Nawy

RESEARCH PAPERS

Li, X. *et al.* Direct reprogramming of fibroblasts via a chemically induced XEN-like state. *Cell Stem Cell* **21**, 264–273.e7 (2017).