The Traveling Salesman Enables the Rapid Synthesis of Repetitive Polypeptides

A codon-scrambling algorithm enables the PCR-based synthesis of repetitive proteins by finding the least-repetitive synonymous gene sequence

Table 1 | Computational results for the optimization of repetitive proteins.

	Sequence	Objective value	Time (s)
Designed ankyrin repeat protein (Darpin E_01; ref. 21)	DLGKKLLEAARAGQDDEVRILMANGADVNADDTWGWTPLHLAAYQGHLEIVEVLLKNGADVNYDYIGWTPLHLAADGHLEIVEVLLKNGADVNASDYIGDTPLHLAAHNGHLEIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQ	1.65×10^3	684
Glucagon-like peptide (GLP-1; ref. 16)	[GAHGEGTFTSDVSSYLEEQAAKEFIAWLVKGR]$_{6}$	7.09×10^3	10.9
Adenovirus-like construct (Ad I; ref. 29)	[LSVQTSAPLTVSDGK]$_{14}$	1.20×10^4	72.2
Mussel adhesive protein (MAP; ref. 9)	[AKPSYPPTYK]$_{16}$	3.41×10^4	1.33
Repeats in toxin (BRT17; ref. 20)	[GGAGNDTLY]$_{17}$	5.04×10^4	31.3
Wheat gliadin (SPR16; ref. 32)	[PQQPY]$_{16}$	9.23×10^4	0.460
Cell adhesive substrate (poly-RGD; ref. 22)	[GSGSGSGRGDS]$_{20}$	1.05×10^5	79.5
β-sheet-forming polypeptide (poly(alanylglycine); ref. 24)	[AGAGAGPEG]$_{10}$	1.22×10^5	5.06
Resilin-like polypeptide (Dros16; ref. 10)	[GAPGGGNGGRPSDTY]$_{16}$	1.74×10^5	603
Abductin (AB12; ref. 11)	[FGGMGGGNAGFGGMGGGKAGFGGMGGGNAG]$_{4}$	2.00×10^5	24.6
Transglutaminase substrate peptide (BQ$_6$; ref. 25)	[[GQQQLGGAGTGSA]$_2$[GAGQGEA]$_3$]$_6$	3.44×10^5	285
Silk-elastin-like polypeptide (SELPOK; ref. 12)	[[GAGAGS]$_2$[GVGVP]$_4$GKGVP[GVGVP]$_3$]$_6$[GAGAGS]$_2$	4.72×10^5	913
Alanine-rich polypeptides (35-H-6; ref. 18)	[AAAQAAQAQAAAEAAAQAAQAQ]$_{6}$	8.11×10^5	67.4
Elastin-like polypeptide (ELP[$V_5A_2G_3$-60]; ref. 36)	[[GVGVP]$_2$GGGVPGAGVP[GVGVP]$_3$GGGVPGAGVPGGGVP]$_6$	9.77×10^5	2,020
β-sheet-forming polypeptide (poly-EAK9; ref. 31)	[AEAEAKAK]$_{18}$	1.16×10^6	2.45
Collagen-like protein (CLP3.7; ref. 13)	[GAPGTPGPQGLPGSP]$_{24}$	1.22×10^6	18.0
Elastin-like polypeptide (ELP[AV-60]; ref. 19)	[GAGVPGVGVP]$_{30}$	1.35×10^6	14.1
Elastin-like polypeptide (ELP[V-60]; ref. 36)	[GVGVP]$_{60}$	1.48×10^6	4.02

The synthesis of genes encoding highly repetitive polypeptides is one of the unsolved problems in synthetic biology. While fast, scalable, high-throughput methods for the synthesis of non-repetitive genes are readily available, these methods rely on piecing together oligonucleotides, or gene fragments. For highly repetitive proteins, these methods fail because the gene fragments are too similar to yield precise results. However, because synthetic biologists can get the same amino acid from multiple DNA codons, they can avoid troublesome DNA repeats by swapping in different codons that achieve the same effect. The challenge is finding the least repetitive genetic code that still yields the desired polypeptide or protein. In their publication in *Nature Materials,* Research Triangle MRSEC professor **Ashutosh Chilkoti** and graduate fellow **Nicholas Tang** from Duke University have removed this hurdle by developing a freely available computer program based on the "traveling salesman" mathematics problem. Using this program, they successfully synthesized 19 different repetitive proteins using commercial biotechnology services. Synthetic biologists can now find the least-repetitive genetic code to build the molecule they want to study. The researchers say their program will allow those with limited resources or expertise to easily explore synthetic biomaterials that were once available to only a small fraction of the field. "This advance really democratizes the field of synthetic biology and levels the playing field," said Tang. "Before, you had to have a lot of expertise and patience to work with repetitive sequences, but now anyone can just order them online. We think this could really break open the bottleneck that has held the field back and hopefully recruit more people into the field."

"Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins." Nicholas C. Tang and Ashutosh Chilkoti. Nature Materials, December, 2015. DOI: 10.1038/NMAT4521