PROTEIN POLYMERS

Gene libraries open up

By combining gene cloning and amplification techniques, a new one-pot, parallel synthesis method for the generation of long, repetitive genes is realized. The method promises to open up the discovery of protein polymer biomaterials.

Sheng Ding, Xiaoxiao Wang and Annelise E. Barron*

Protein polymers — engineered repetitive polypeptides, up to 500 or more amino acids in length — are a uniquely tunable family of functional biomaterials. They can mimic aspects of natural protein structure and function, or be designed *de novo* to solve problems that nature never contemplated[1–3]. With 20 years of work, the palette of useable non-natural amino acids has grown in breadth and nuance, and the methods of production improved. Still, the number of biomolecular engineering research groups tinkering with protein polymers is relatively small. The spectrum of innovative users must broaden if protein polymers are to fulfil their considerable potential.

Design and discovery of useful protein polymers has been, until now, labour intensive and slow. Sequence domains (~30 amino acids or fewer) from proteins such as elastin are 'polymerized' via genetic engineering. Tedious, low-yielding cloning work has been unavoidable in a quest for long, synthetic genes that *Escherichia coli* expresses well.

This is about to change. Writing in *Nature Materials*, Chilkoti and colleagues report a new way to achieve the rapid, parallel synthesis of long, repetitive genes, which they developed through a clever amalgamation of prior gene cloning and amplification methods. Obtaining new genes that encode long protein polymers (>500 amino acids) is made easier, faster and more reliable. In the present study, this method is demonstrated for the discovery of thermoresponsive proteins, as well as pharmaceutically relevant glucagon-like peptide-1 (GLP-1) polymers.

Protein polymer synthesis requires the generation of repeating DNA sequences encoding proteins of different lengths. A few cloning methods were previously used to achieve this goal. Gene concatemerization, although widely used and straightforward, is inefficient by virtue of the required enzymatic digestion and ligation steps, and offers low utility for the parallel creation of large gene libraries (Fig. 1a). A more efficient method, overlap-extension polymerase chain reaction (OEPCR), replaces ligation-based gene multimerization with just one PCR step (Fig. 1b). Although amenable to high-throughput projects, OEPCR has the disadvantage of low fidelity as a result of non-specific self-priming of short gene sequences, and denies access to the long genes that are often ideal for biomaterials[4].

The highly parallel method for gene synthesis reported by Chilkoti and colleagues, called 'overlap-extension rolling circle amplification (OERCA)', combines RCA (Fig. 1c) with OEPCR, in one reaction tube[5]. Fast, high-throughput gene oligomerization can be accomplished with excellent retention of sequence fidelity. An RCA step replicates a circularized, single-stranded DNA (ssDNA) template, providing a library of repetitive nucleic acid polymers. Next, a linear gene-extension step creates complementary DNA fragments. Subsequent thermal cycling uses both forward and reverse primers. In later extension steps, ssDNA molecules with longer overlapping regions initiate overlap-extension reactions, thus generating long, repetitive DNA oligomers with high sequence fidelity. The size distribution of oligomers is tuned by varying the primer-to-template ratio or the thermal cycling protocol. Chilkoti and colleagues show that this approach generates diverse gene libraries, in a single reaction, by creating genes as long as 2,500 base pairs (bp) from 18-bp 'monomers'. Prior cloning methods used in protein polymer production yield shorter genes, often with higher error rates.

Elastin-like protein polymers (ELPs) are of interest for their emergent biomaterial properties and especially for their stimuli-responsiveness. Depending on amino acid sequence and length, an ELP is soluble in aqueous media below a certain temperature, but partitions into a water-insoluble 'coacervate' phase at higher temperatures. The ELPs are based on a pentameric consensus sequence, Val-Pro-Gly-Xaa-Gly, where Xaa is any amino acid, and have been engineered for a variety of applications. Amino acid sequence and total chain length affect the phase-transition temperature, as well as protein biodegradability[6]. The highly repetitive, guanine/cytosine-rich sequences of ELP-encoding genes

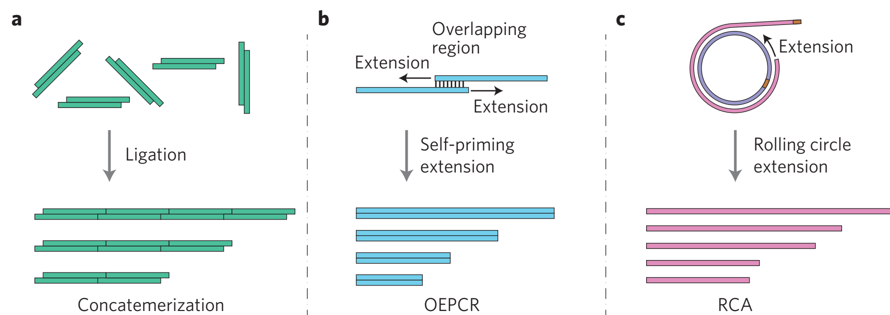

Figure 1 | A schematic of strategies to synthesize oligomer genes. **a**, Concatemerization. The double-stranded monomer genes are generated either by enzyme digestion from original plasmid or by annealing of two complementary single-stranded DNA fragments. Then, monomer fragments are ligated to form oligomers of different lengths. Type IIS restriction endonuclease can be used to eliminate sequence constraints and introduce direction control in oligomization. **b**, Overlap-extension PCR (OEPCR). Two ssDNA fragments anneal to each other through their overlapping regions, and are extended to longer sizes during chain extension steps. **c**, Rolling circle amplification (RCA). Cyclized DNA is used as the template in RCA. During the elongation cycle the polymerase rolls over the template and extends continuously to yield linear complementary single-stranded DNA fragments of various lengths.

© 2011 Macmillan Publishers Limited. All rights reserved

have made it hard to synthesize ELPs with widely ranging compositions and lengths, hampering exploration of the relationships between phase-transition behaviour, protein sequence, chain length and coacervate secondary and tertiary structure. Using OERCA, Chilkoti and colleagues show the synthesis of an ELP gene library with large breadths of sequences and lengths. Using this library, they were able to investigate how variations of the Val-Pro-Gly-Xaa-Gly sequence motif affect ELP coacervation behaviour. Alanine insertion and substitution mutants of poly(Val-Pro-Gly-Xaa-Gly) sequences yielded proteins capable of new types of reversible stimuli-responsive phase behaviour. They compare the biophysical properties of the newly discovered variant ELPs with previously studied canonical ELPs, gaining novel insights into the influence of hydrophobic interactions on ELP phase-transition behaviour, which suggest that unordered and dynamic conformations of ELPs are the key to thermal reversibility. More thorough investigation of the sequence-related properties of ELPs should enable the creation of materials and hydrogels with easily tuned, fully reversible stimuli-responses, sharper phase transitions, and controlled biophysical structures.

To demonstrate the versatility of OERCA in generating repetitive genes, Chilkoti and colleagues synthesized another gene with a markedly different motif length and composition. By engineering proteins with weak, intermediate or strong thrombin protease cleavage sequences embedded between the repeating peptide units, a library of protease-sensitive polymers based on GLP-1 variants of differing potencies was generated and studied. Molecules of this type are under study for the treatment for type-2 diabetes, and may provide better control over blood glucose through slow *in vivo* release of GLP-1 peptides as the polymers are degraded by thrombin. The OERCA method could facilitate the screening of these and other pharmaceutically relevant protein polymers, and aid in the optimization of clinical efficacy by accessing a wide diversity of sequences and lengths.

Combining OEPCR and RCA to achieve rapid, highly parallel and well-controlled synthesis of highly repetitive genes is a significant advance in the preparation of protein polymers. Although traditional step-wise cloning strategies[7,8] are useful in the cases where precise control of protein polymer block location and distribution are required to achieve desired functions, OERCA offers advantages for screening new protein families. All cloning methods display some degree of sequence-dependence, however — the true robustness of OERCA, and its advantages and drawbacks, must be further explored with the synthesis of protein polymers that are based on a broader array of sequence motifs, which, thankfully, will be a relatively easy task through the use of this new method. ❐

Sheng Ding, Xiaoxiao Wang, Annelise E. Barron are at Stanford University in the department of bioengineering, James H. Clark Center, 318 Campus Drive, Stanford, California 94305-5444, USA.
**e-mail: aebarron@stanford.edu*

References

1. Banta, S., Wheeldon, I. R. & Blenner, M. *Annu. Rev. Biomed. Eng.* **12,** 167–186 (2010).
2. Davis, N. E., Ding, S., Forster, R. E., Pinkas, D. M. & Barron, A. E. *Biomaterials* **31,** 7288–7297 (2010).
3. Sengupta, D. & Heilshorn, S. C. *Tissue Eng. B* **16,** 285–293 (2010).
4. Kurihara, H., Morita, T., Shinkai, M. & Nagamune, T. *Biotechnol. Lett.* **27,** 665–670 (2005).
5. Amirama, M., Quiroza, F. G., Callahana, D. J. & Chilkoti, A. *Nature Mater.* **10,** 141–148 (2011).
6. MacEwan, S. R. & Chilkoti, A. *Biopolymers* **94,** 60–77 (2010).
7. Meyer, D. E. & Chilkoti, A. *Biomacromolecules* **3,** 357–367 (2002).
8. Won, J. I., Meagher, R. J. & Barron, A. E. *Biomacromolecules* **5,** 618–627 (2004).

PROTEIN FOLDING

To knot or not to knot?

A knot-containing protein is found to fold reversibly at biologically relevant timescales despite not having naturally evolved for this ability.

Eugene Shakhnovich

The discovery ten years ago of deep knots in proteins[1] posed a challenge to protein-folding researchers: are knotted proteins able to fold spontaneously, or are there (yet undiscovered) dedicated chaperones that 'tie the knots'? The answer came in recent work by Sophie Jackson and co-workers[2,3], who showed spontaneous refolding of a knotted protein, YibK, a member of the family of the methyltransferases. Subsequent folding simulations confirmed the spontaneous folding scenario[4,5]. The apparent ease with which YibK and its computational counterpart actually fold, and the apparent topological difficulty in folding a knotted protein, have led to suggestions that, for such reliable folding to occur, evolution had to optimize the folding pathways of knotted proteins. Although some evidence for this was found in a bioinformatics analysis of possible knot-promoting non-native interactions[4], it has been especially important to determine whether a designed protein that has not evolved naturally to fold into a knotted conformation is capable of undergoing folding at timescales that are biologically reasonable. Writing in *Proceedings of the National Academy of Sciences of the USA*, Todd Yeates and colleagues have achieved exactly that: they engineered a knot-containing protein and its almost-identical, unknotted twin[6]. Their most striking finding is that the designed knotted protein does indeed fold, albeit 20 times slower than its unknotted twin. This result is highly significant in light of the debate on whether special selection for folding kinetics is necessary to obtain foldable proteins.

Indeed, because none of the early-discovered folds in the 1960s contained knots, it was suggested that knot-avoidance could be an evolutionary principle of selection of protein structures, presumably for fast and efficient folding. This view was also stimulated by the 'Levinthal paradox' — the observation that proteins cannot fold spontaneously into the unique native state by direct exhaustive search over the astronomical number of accessible configurations. However, this situation started to change 10 years ago, as more and more structures with knots were

© 2011 Macmillan Publishers Limited. All rights reserved