解　説

Recombinant Protein Synthesis for Nanomaterial Assembly: Technical Overview

中野　綾菜[1,2]・加藤　功一[2,3]
Ayana Nakano[1,2] and Koichi Kato[2,3]

Recombinant DNA technology has emerged as a useful method for providing protein-based building blocks for nanomaterial assembly. This article provides an overview of the technical aspects of recombinant protein synthesis, especially for nanomaterial scientists and engineers. First, the rationale for the use of protein-based nanomaterials synthesized through recombinant DNA technology is presented. This article then focuses on simple standard protocols for protein expression in *Escherichia coli*. The last section presents several methods for creating artificial proteins with complex structures.

1. Introduction

In nature, various nanoscale objects that are structurally and functionally unique are encountered; these objects have attracted the attention of nanoscientists and nanoengineers who are interested in biomimicry[1]. Inspired by nature, a growing number of research groups have proposed design principles that permit the generation of innovative nanomaterials through the self-assembly of biomolecular building blocks[2–5], among which proteins are one of the most effective elements. This is primarily because proteins exhibit structural diversity and the ability to undergo biological functions through specific interactions with partner proteins. The possibility of molecular design through genetic engineering technology adds to their usefulness as a building block for nanomaterials.

To date, naturally occurring proteins such as collagen[6], gelatin[7], and fibrin[8] have been utilized to construct nanomaterials for various medical applications. In addition, proteins such as antibodies[9] and growth factors[10] are attractive as functional components in nanomaterials because of their capability to have interactions with specific targets. In most cases, proteins are extracted from various organisms. In these cases, however, the proteins that can be obtained are limited to those that have previously been identified, and it is not easy to extract and purify trace proteins. Importantly, modification of the primary structure of naturally occurring proteins is normally impossible.

In contrast, recombinant protein synthesis, which is the focus of this article, is an attractive route for creating building blocks for assembling nanomaterials. This type of protein is prepared by genetic engineering (genetic recombination) of living cells such as *Escherichia coli* (*E. coli*), yeast, and animal cells[11]. In addition to facilitating scale-up, this method allows for the molecular design of proteins by deleting and replacing amino acids, as well as the addition of small peptides[12–14], polypeptide domains[15–17], and even other large proteins[18]. Such versatility is attractive to scientists and engineers seeking new protein-based building blocks for nanomaterial assembly.

The following sections focus on the technical aspects of recombinant protein synthesis. First, taking a simple recombinant protein as an example, one of the most standard methods will be described. Subsequently, several synthetic methods will be introduced as feasible strategies for preparing chimeric genes that encode recombinant proteins with complex structures. Most of the methods described in

[1]Department of Orthodontics, Hiroshima University Hospital (Hiroshima 734-8551, Japan)
[2]Department of Biomaterials, Graduate School of Biomedical and Health Sciences, Hiroshima University (Hiroshima 734-8553, Japan)
[3]Nanomedicine Research Division, Research Institute for Nanodevice and Bio Systems, Hiroshima University (Higashi-Hiroshima 739-8527, Japan)
Received October 28, 2021

this article will be familiar to molecular biologists, but we believe that this overview will provide new perspectives for nanoscientists and nanoengineers.

2. Process of Recombinant Protein Production

Recombinant proteins are produced in cells through a pathway similar to that of endogenous proteins (Figure 1). According to the central dogma of molecular biology, genetic information flows from DNA to RNA and is finally translated to generate proteins in living organisms. When a heterologous gene is introduced into a cell together with an appropriate promoter sequence, the gene is expressed and translated to generate the respective protein using the endogenous systems for gene expression by RNA polymerase and protein synthesis by ribosomes. The recombinant proteins derived from the heterologous genes can be extracted from the cells and purified by removing other endogenous proteins. The following section describes the details of the entire procedure for recombinant protein production.

Recombinant proteins are typically prepared via the following ten steps: (i) acquisition of the protein sequence information, (ii) design of primers for the polymerase chain reaction (PCR), (iii) amplification of the target gene by PCR using a cDNA library as a template; (iv) construction of an expression vector (plasmid); (v) cell transformation; (vi) protein expression; (vii) extraction; (viii) purification; (ix) refolding; and (x) cryopreservation. In the following subsections, these steps will be explained in detail, using a simple case for the sake of clarity: the expression of epidermal growth factor (EGF) with a hexahistidine sequence (His-tag) at the C-terminus in *E. coli* using Novagen's pET vector system (Merck KGaA, Darmstadt, Germany), specifically, the pET-22b(+) plasmid[19]. As will be described later, the His-tag facilitates the purification of recombinant proteins.

2.1. Selection of Host Cells

Various types of host cells can be used for recombinant protein expression. The most popular cells include *E. coli*, yeast, insect, and animal cells[11]. Each host has different properties; therefore, it is necessary to select suitable host cells. *E. coli* provides the simplest and easiest method for recombinant protein production, although post-translational modifications, such as glycosylation, do not occur in *E. coli*. Therefore, when a sugar chain is necessary for the function of the recombinant protein, we need to consider another host, such as animal cells.

2.2 Acquisition of Genetic Information

The first step in the process is to acquire DNA sequence information for the protein of interest. This can be achieved using the nucleotide database available online from the National Center for Biotechnology Information (NCBI, Bethesda, MD, USA; https://www.ncbi.nlm.nih.gov/). The

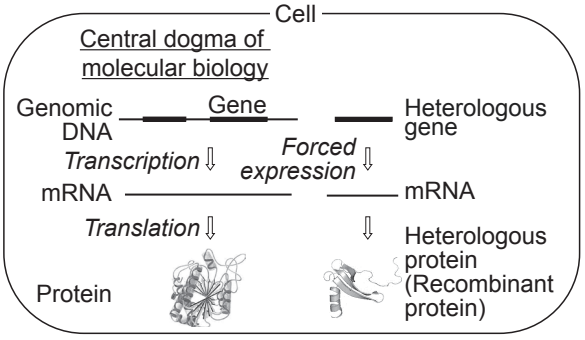

Figure 1. Production of a recombinant protein in a host cell via a pathway similar to that of endogenous proteins.

```
  1 aatagtgact ctgaatgtcc cctgtcccac
 31 gatgggtact gcctccatga tggtgtgtgc
 61 atgtatattg aagcattgga caagtatgca
 91 tgcaactgtg ttgttggcta catcggggag
121 cgatgtcagt accgagacct gaagtggtgg
151 gaactgcgc
```

Scheme 1. Coding sequence of mature human epidermal growth factor (EGF) (Accession: AY548762).

sequence information for human EGF can be found in the search result "Homo sapiens epidermal growth factor (EGF) mRNA, partial coding sequence (cds)." The term "partial" indicates that the entry corresponds to part of the entire EGF sequence, namely, 53 amino acid residues of mature EGF generated by post-translational cleavage of a large EGF precursor. The sequence of EGF cds (Scheme 1) can be found at the following link: https://www.ncbi.nlm.nih.gov/nuccore/JQ346088.1, along with other information such as the accession number, related literature, and comments on featured regions.

2.3. Acquisition of EGF-Coding DNA

The next step in the process is to obtain EGF-coding DNA via the PCR. For this purpose, forward and reverse primers must be designed (Figure 2). The primers are designed to hybridize the 5' and 3' ends of the DNA to be amplified. Approximately 20 base-pair residues from both ends are normally considered. To insert the PCR products into the specific position in the multiple cloning site of the plasmid to be used, DNA sequences which can be digested by restriction enzymes are flanked to the 5' ends of the primers (*Nde* I and *Xho* I in the specific case in Figure 2). Because restriction reactions are often attenuated when the digestion sequences are located at the end of the DNA, several nucleotides are added. Importantly, the *Nde* I digestion site contains a start codon, ATG, at which ribosomal translation begins. In contrast, a stop codon, TGA, is present in the pET-22b(+) downstream of His-tag. Various commercial services for

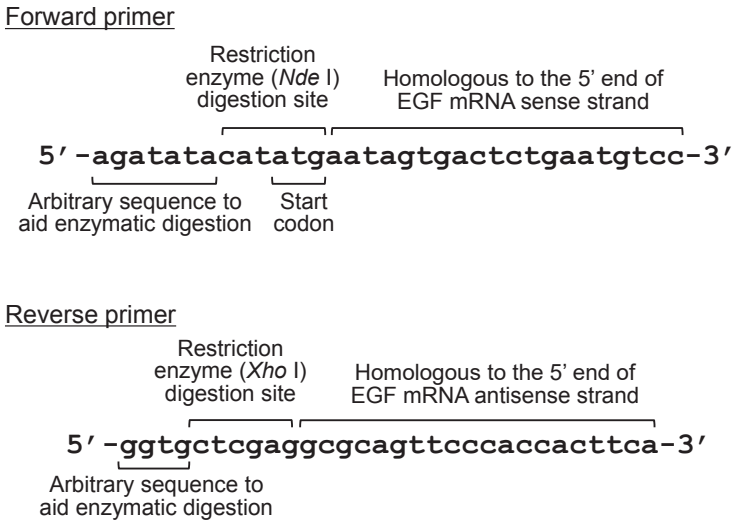

Figure 2. Primer design for PCR amplification of the EGF gene.

obtaining custom primers are available.

To amplify the EGF-coding DNA, PCR is performed using the forward and reverse primers designed above and any cDNA source as a template, such as a human cDNA library and first-strand cDNA synthesized from total RNA isolated from EGF-expressing cells (Figure 3a). In PCR, DNA polymerase with high fidelity should be used to ensure the accuracy of the sequence of the amplified DNA. The PCR products are then purified and digested with restriction enzymes as described below.

2.4. Cloning into a Plasmid

The purified PCR products are then digested with the restriction enzymes *Nde* I and *Xho* I to form sticky ends (Figure 3b). pET-22b(+) is also digested with the same combination of enzymes to generate linear DNA with sticky ends. After purification, the digested products are ligated to each other using T4 DNA ligase and then introduced into competent cells (such as DH5α *E. coli*) via the heat-shock method at 42 °C (Figure 3c). The competent cells represent *E. coli* whose cell membrane has been destabilized with, for example, calcium chloride. The

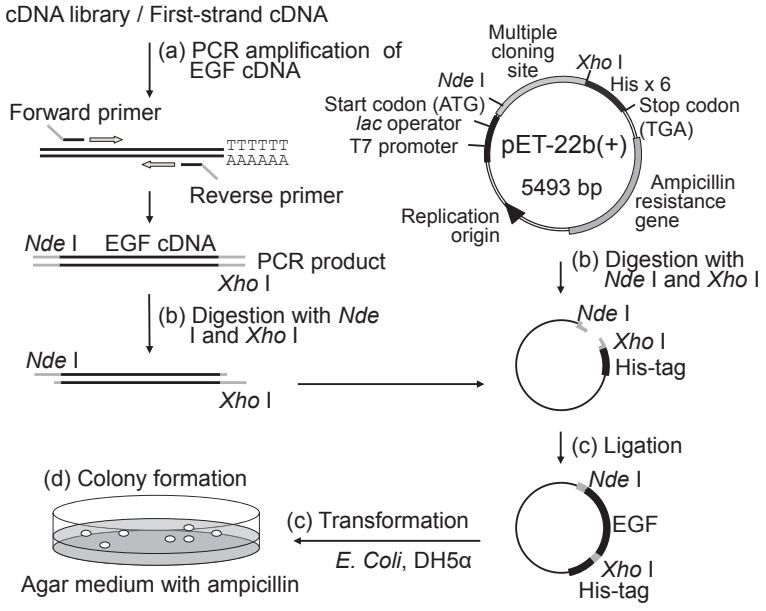

Figure 3. Entire procedure for cloning the EGF gene in the pET-22b(+) plasmid. (a) PCR amplification. (b) Digestion with restriction enzymes. (c) Ligation and transformation. (d) Colony formation.

bacteria are plated on an agar medium containing ampicillin. After being cultured overnight at 37 °C, bacteria containing the plasmid form colonies on the agar plate (Figure 3d), which can be confirmed by colony PCR. The entire procedure is called "cloning". The plasmid-containing bacteria are expanded by culturing them in a liquid LB medium at 37 °C overnight, after which the plasmid DNA (pET22-EGF) is isolated. The accuracy of the inserted DNA can be confirmed by sequencing.

2.5. Transformation and Protein Expression

The obtained plasmid pET22-EGF is then introduced into the host cells for protein expression. This procedure is called transformation. *E. coli* of the BL21 strain that has been modified with the DE3 gene is suitable for use with pET vectors with the *lac* operator system. Since the codons used in *E. coli* are slightly different from those in eukaryotic cells, it is advisable to use a strain that has been modified to efficiently express human-derived proteins. The plasmid is introduced into the BL21(DE3) competent cells via heat-shock and plated on an ampicillin-containing agar plate. The cells are cultured overnight at 37 °C, and the colonies formed on the plate are analyzed using colony PCR to confirm the presence of the plasmid and obtain transformants.

For protein expression, the *E. coli* bacteria harboring pET22-EGF are cultured, typically in 100–200 mL of LB medium containing ampicillin with vigorous shaking at 37 °C. During culturing, the turbidity of the suspension (absorbance of light at 600 nm) is measured as a function of time. When the absorbance reaches 0.6 in the logarithmic growth phase, isopropyl-β-D-thiogalactopyranoside (IPTG) is added to induce the expression of EGF-His. The culture is then continued for several hours to express the protein.

The DE3 gene is integrated into the chromosomal DNA of the BL21(DE3) strain. This is a DNA sequence that contains the *lac* promoter, *lac* operator, and T7 RNA polymerase gene and is usually suppressed by the *LacI* repressor. IPTG binds to the *lac* repressor to release the inhibition, leading to the initiation of T7 RNA polymerase synthesis. The polymerase transcribes the gene in the plasmid; however, the gene expression is again suppressed by the *lac* repressor because the plasmid also contains *lac* operator. Therefore, the addition of IPTG also initiates the transcription of the gene carried by the plasmid. This conditional expression system is rather complicated, but useful, as exogenous proteins are sometimes harmful to bacterial growth.

2.6. Extraction, Purification, and Refolding of Proteins

The bacteria containing the expressed EGF-His are collected by centrifugation. A solution containing a surfactant, lysozyme, and DNase is added to the pellet to destroy the cell membranes. Alternatively, the bacterial suspensions can be subjected to sonication for the same purpose. In many cases, the recombinant proteins are obtained as an insoluble substance, which is also known as an inclusion body. The precipitate is washed extensively with a surfactant solution. Finally, a buffer solution containing 8 M urea is added to the insoluble products to denature and solubilize the recombinant protein. It is recommended to further add 2-mercaptoethanol or dithiothreitol to cleave any spontaneously formed disulfide linkages.

Since EGF-His has a His-tag at the C-terminal, the recombinant protein can be purified using a nickel chelate column in which the Ni(II) ions carried by the adsorbents bind the histidine residues in EGF-His. In the case of insoluble recombinant proteins such as EGF-His, all purification steps are performed in the presence of 8 M urea. The protein trapped in the column is eluted by adding a high concentration of imidazole, an analog of histidine.

The final step is to remove urea and refold the denatured protein into the correct form. For this purpose, step-wise dialysis is carried out using dialysates of gradually decreasing urea concentrations. To promote the formation of disulfide bonds between cysteine residues, it is advisable to add oxidized and reduced glutathione. In fact, the correct form of EGF contains three S–S linkages. Additives such as Tween20 and arginine sometimes promote the solubility of recombinant proteins at low urea concentrations. After refolding, the protein is sterilized using 0.22 μm filters as needed and stored at –80 °C until use.

3. Preparation of Chimeric Proteins

Recombinant protein synthesis techniques have great potential in the design of protein molecules. One of the most useful designs may be chimeric proteins in which two or more polypeptides are connected to each other. Such chimeric proteins are expected to exert two or more distinct functions based on the individual polypeptides. The key step in the preparation of chimeric proteins is the preparation of the chimeric genes. Depending on the attributes of the sequences (artificial or natural) and the length of the polypeptides to be fused, an appropriate method for chimeric gene construction should be adopted. We will provide an overview of the effective methods in the following subsections.

3.1. Fusion with a Short Peptide

Proteins that have a short functional peptide at their N- or C-terminus have many potential applications. To recombinantly synthesize such chimeric proteins, chimeric genes that encode the total chimeric polypeptides must first be prepared. This is performed by simply incorporating the peptide-coding DNA sequences into forward (for N-terminus fusion) or reverse (for C-terminus fusion) primers. Figure 4 shows the reverse primer sequence that was used to link a tandem repeat decapeptide (KLKLKLKLKL; KL5) to the

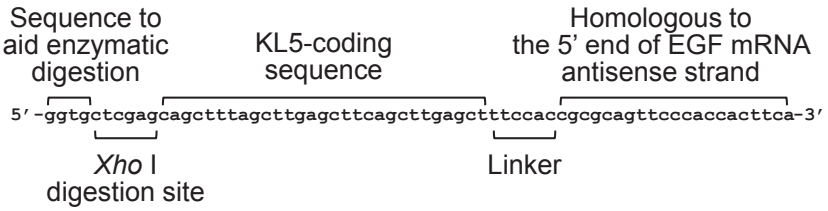

Figure 4. Design of a reverse primer used for the PCR amplification of the EGF-KL5 chimeric gene.

C-terminus of EGF[12]. Two glycine residues were inserted as flexible linkers between EGF and KL5. The relatively long primer (68 nucleotides) had no adverse effect on the PCR amplification of the chimeric gene.

3.2. Fusion of Large Polypeptide Domains

Overlap extension PCR[20] is suitable for synthesizing a chimeric gene that encodes a protein consisting of two relatively large polypeptide domains. In this method, unique primers are used in the first PCR amplification.

Consider the case in which polypeptide A is linked to the N-terminus of polypeptide B to obtain a chimeric polypeptide A–B. First, the genes for polypeptides A and B are separately amplified by PCR (Figure 5). In the reaction for gene A, a reverse primer (complementary to the 3' region of the sense sequence) is designed to have an overhang that is complementary to the 5' region of the sense sequence of gene B. Similarly, a forward primer for gene B (complementary to the 3' region of the antisense sequence) is designed to have an overhang that is complementary to the 5' region of the antisense sequence of gene A. The PCR products thus obtained are genes A and B with an extension overlapping the upstream of the B gene and the downstream of the A gene, respectively. Therefore, the sense sequence of product A can partially hybridize to the antisense sequence of product B. The unhybridized region of the partially hybridized chain can be filled in using DNA polymerase in the presence of deoxyribonucleotide triphosphates to generate completely filled double-stranded DNA. This chain is further amplified by PCR using the forward primer for gene A and the reverse primer for gene B.

This method was successfully used to prepare EGF fused with a collagen-binding domain derived from von Willebrand factor[16], the cell-adhesive G3 domain of the laminin α-chain fused with a helical polypeptide and a collagen-binding peptide derived from decorin[15], and a self-assembling chimeric protein consisting of an α-helical polypeptide and a cell-adhesive domain derived from laminin[21].

3.3. Artificial Protein with Repeated Segments

The multimerization of DNA fragments by head-to-tail ligation (concatemerization) can be used to produce a

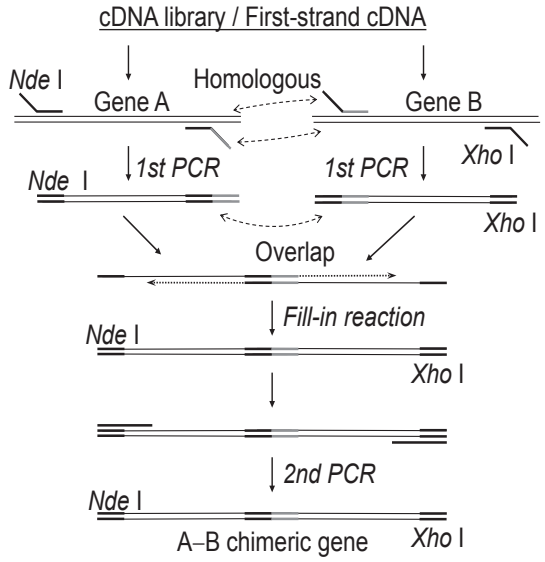

Figure 5. Preparation of a chimeric gene by overlap-extension PCR. *Nde* I and *Xho* I digestion sites are shown as examples.

synthetic gene with repeated sequences (Figure 6). This method was utilized by Cappello[22] to genetically synthesize silk-like and elastin-like homoblock polymers and by McGrath[23] to prepare polypeptides consisting of repetitive

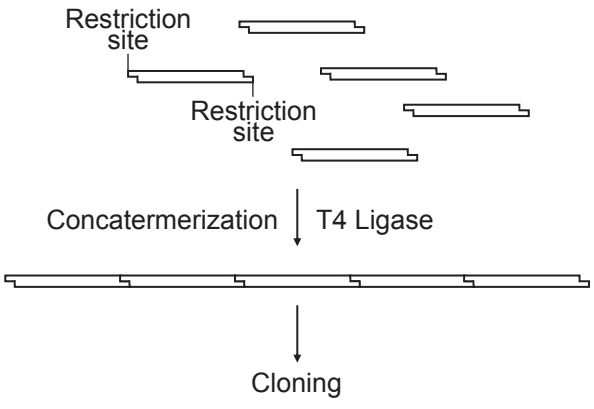

Figure 6. Concatemerization of polypeptide genes.

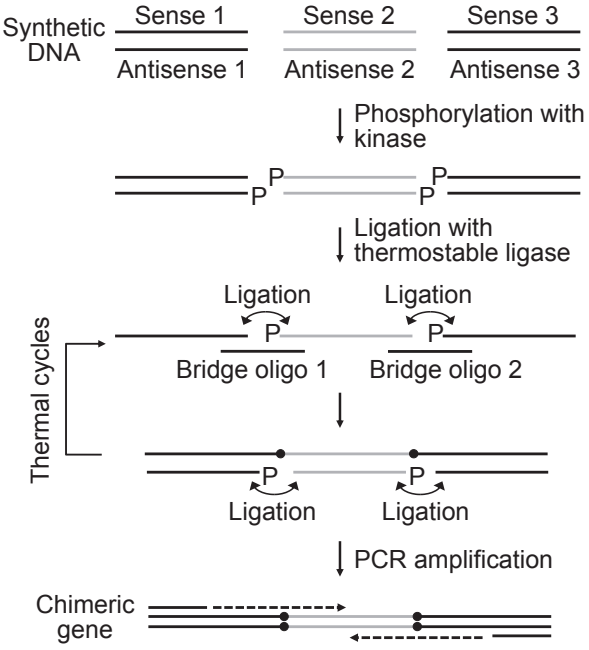

Figure 7. Preparation of a chimeric gene by chain cloning.

artificial sequences. Such repetitive protein polymers are attractive analogs of extracellular matrix proteins. However, the early methods had limitations, such as low yields of the long concatemers and the requirement of using nonpalindromic restriction enzymes for cloning. A modification of the method to generate size-controlled DNA concatemers with no requirement for amino acid sequences was reported[24].

3.4. Proteins Consisting of Artificial Sequences

The chain cloning method[25] is especially effective when the chimeric gene to be synthesized contains artificial sequences that cannot be amplified by PCR. Any length and number of DNA fragments can be ligated in the correct order and orientation. As shown in Figure 7, the 5' ends of all DNA fragments are first phosphorylated with kinase and then ligated to each other in the presence of thermostable DNA ligase and bridging oligonucleotides. The bridging oligonucleotides are designed to hybridize the head and tail of the DNA fragments to be ligated, so that the order and orientation of the ligated fragments are specified. After ligation by thermal cycling, the entire sequence is amplified using PCR. The chain cloning method was used to synthesize EGF fused with artificially designed helical polypeptides that spontaneously dimerize through coiled-coil association with a neighboring molecule[26].

3.5. Methods Applicable to Various Fusion Proteins

Stemmer *et al*. reported[27] that assembly PCR is a versatile method with a wide range of applications. The first step in this method is to synthesize a pool of oligonucleotides, 40 nucleotides in length, that encode the sense and antisense

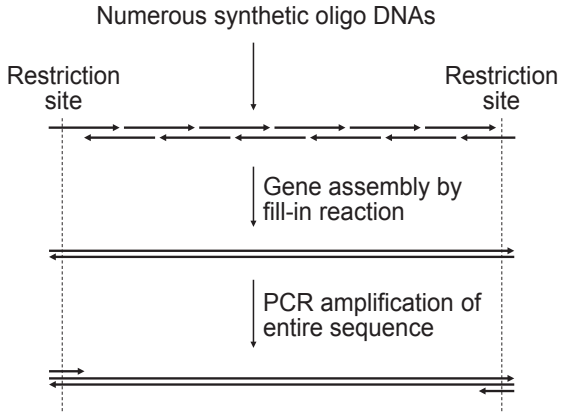

Figure 8. Preparation of a chimeric gene by assembly PCR.

strands of the entire DNA of interest. As shown in Figure 8, these fragments, which overlap with the complementary oligos by 20 nt, are assembled by PCR and further amplified, and restriction sites are added for cloning in a plasmid vector. A similar method called “recursive PCR” was also reported by Prodromou and Pearl[28]. Recently, commercial services for custom DNA synthesis based on the assembly PCR method have become available.

4. Conclusions

Recombinant protein synthesis is a useful and attractive way to produce bioinspired building blocks for nanomaterials. Currently, various techniques and resources are available for the synthesis of recombinant proteins. Most importantly, unlike natural proteins, protein molecules obtained through recombinant DNA technology can be rationally designed. This advantage will serve to significantly expand the possibilities of using nanomaterials.

Acknowledgements

This work was supported by Grant-in-Aid for Research Activity Start-up (No. 21 K20l4) and Fund for the Promotion of Joint Intemational Research (No. 19KK0278), JSPS.

References

1) Qian, Z.-G., Pan, F. and Xia, X.-X.: *Curr. Opin. Biotechnol*. **65**, 197 (2020).
2) Hamley, I. W.: *Biomacromolecules* **20**, 1829 (2019).
3) Zhang, X., Liu, Y., Zheng, B., Zang, J., Lv, C., Zhang, T., Wang, H. and Zhao, G.: *Nat. Commun*. **12**, 4849 (2021).
4) Wargacki, A. J., Wörner, T. P., van de Waterbeemd, M., Ellis, D., Heck, A. J. R. and King, N. P.: *Nat. Commun*. **12**, 883 (2021).
5) Glover, D. J., Giger, L. and Kim, S. S., Naik, R. R. and

Clark, D. S.: *Nat. Commun.* **7**, 11771 (2016).
6) Lo, S. and Fauzi, M. B.: *Pharmaceutics* **13**, 316 (2021).
7) Rehana, Y., Mohsin, S., Ahmad, K. S. and Roshan, A.: *Nanotechnol. Rev.* **6**, 191 (2017).
8) Muhamed, I., Sproul, E. P., Ligler, F. S. and Brown, A. C.: *ACS Appl. Mater. Interfaces* **11**, 3771 (2019).
9) Farahavar, G., Abolmaali, S. S., Gholijani, N. and Nejatollahi, F.: *Biomater. Sci.* **7**, 4000 (2019).
10) Chu, C., Deng, J., Liu, L., Cao, Y., Wei, X., Li, J. and Man, Y.: *RSC Adv.* **6**, 90856 (2016).
11) Tripathi, N. K. and Shrivastava, A.: *Front. Bioeng. Biotechnol.* **7**, 420 (2019).
12) Nakano, A., Hirata, I., Pham, B. V., Shakya, A., Tanimoto, K. and Kato, K.: *J. Biomater. Sci. Polym. Ed.* **32**, 76 (2021).
13) Yamauchi, Y., Hirata, I., Tanimoto, K. and Kato, K.: *Biotechnol. Bioeng.* **117**, 2741 (2020).
14) Shakya, A., Imado, E., Nguyen, P. K., Matsuyama, T., Horimoto, K., Hirata, I. and Kato, K.: *Sci. Rep.* **10**, 8762 (2020).
15) Nakaji-Hirabayashi, T., Kato, K. and Iwata, H.: *Bioconjugate Chem.* **24**, 1798 (2013).
16) Egawa, E. Y., Kato, K., Hiraoka, M., Nakaji-Hirabayashi, T. and Iwata, H.: *Biomaterials* **32**, 4737 (2011).
17) Nakaji-Hirabayashi, T., Kato, K. and Iwata, H.: *Biomaterials* **30**, 4581 (2009).
18) Nakaji-Hirabayashi, T., Kato, K., Arima, Y. and Iwata, H.: *Bioconjugate Chem.* **19**, 516 (2008).
19) Kato, K. Sato, H. and Iwata. H.: *Langmuir* **21**, 7071 (2005).
20) Higuchi, R., Krummel, B. and Saiki, R. K.: *Nucleic Acids Res.* **16**, 7351 (1988).
21) Nakaji-Hirabayashi, T., Kato, K. and Iwata, H.: *Biomacromolecules* **9**, 1411 (2008).
22) Cappello, J., Crissman, J., Dorman, M., Mikolajczak, M., Textor, G., Marquet, M. and Ferrari, F.: *Biotechnol. Prog.* **6**, 198 (1990).
23) McGrath, K. P., Fournier, M. J., Mason, T. L. and Tirrell, D. A.: *J. Am. Chem. Soc.* **114**, 727 (1992).
24) Won, J.-I. and Barron, A. E.: *Macromolecules* **35**, 8281 (2002).
25) Pachuk, C. J., Samuel, M., Zurawski, J. A., Snyder, L., Phillips, P. and Satishchandran, C.: *Gene* **243**, 19 (2000).
26) Nakaji-Hirabayashi, T., Kato, K. and Iwata, H.: *Bioconjugate Chem.* **20**, 102 (2009).
27) Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M. and Heyneker, H. L.: *Gene* **164**, 49 (1995).
28) Prodromou, C. and Pearl, L. H.: *Protein Eng.* **5**, 827 (1992).