A review of chimeric proteins/enzymes

Jiakang Zhu, Song Han and Le Gao *

School of Life Science and Technology, Changchun, Jilin, 130013, China

Abstract. The low homologous protein chimeric recombination technique by rational design of mutant library proteins in the construction of chimeric esterases by structural domain recombination often produces a large number of inclusion bodies due to misfolding of nascent proteins caused by clashes between residue side-chains, leading to a decrease in soluble protein content. The formation of inclusion bodies is evidence of erroneous interactions of intermediates and the lack of essential cofactors in the cell. Heterotrimeric proteins obtained during chimerisation of low homologous parental chimeric esterases tend to produce misfolding, resulting in inactive chimeric esterases constructed or low soluble protein content. Chimeric enzymes obtained by using structural domain recombination to alter the nature and function of the enzyme play an important role in medicine and industry, among others. Although low homology parental chimerism is more likely to produce chimeric enzymes with new functions, the lower the homology of the parents the more likely it is to produce conflicts between residues, causing the resulting heterotrimeric proteins to misfold and form a large number of inclusion bodies. With the development of bioinformatics technology and improved understanding of protein properties, chimerisation of some low homology protein molecules is now possible. In this article we review several methods and important strategies for directed protein evolution and discuss recent research advances in reducing inclusion bodies to increase soluble protein content.

1 Preface

Protein engineering is an important part of modern biological technology. By directed evolution and rational design of protein molecules, target proteins are modified and modified to obtain biological macromolecules that meet human expectations. In 1983, the special topic on protein engineering published in SCIENCE was the earliest elaboration of protein engineering. Modern protein engineering can be divided into rational design, directed evolution and semi-rational design.

Directed evolution technology is a kind of simulated evolution at molecular level, which simulates the evolution process of nature in vitro, artificially creates a large number of random mutations, and gives specific selection pressure according to the desired purpose to screen mutants with desired characteristics. Directed evolution is widely used in protein engineering, enzyme engineering and metabolic engineering. Rational protein design is based on the understanding of protein structure-function relationship, through site-directed mutation, domain recombination and other techniques to modify the properties and functions of proteins, in order to obtain biological macromolecules with expected values. The evolution of the second gene of β-galactosidase in Escherichia coli in 1973 is the first successful example of this technique. Semi-rational design is a combination of rational design and directed evolution, creating mutations at specific sites, reducing the number of mutant pools and reducing the difficulty of screening.

The rational design of protein is the main method of protein modification, which can be divided into site-directed mutation and domain recombination. Site-directed mutation refers to the introduction of specific base insertion and deletion into the target fragment by PCR technology to improve protein function. In 1992, site-directed mutation was introduced into two residues of lipase Lip4, and a mutant with hydrolysis activity much higher than that of wild type was obtained. Site-directed mutagenesis has been widely used in protein modification, but it only focuses on the modification of protein molecules in a small range, and can not achieve functional transition. Domain recombination can change the function of enzymes in a large range through domain exchange, and construct proteins with new functions. This paper reviews several main methods of protein recombination, and puts forward solutions to two main problems encountered in the process of domain recombination. Chimeric enzymes constructed by low homologous parents often obtain completely new properties and functions, which is a key research project in protein engineering. However, the heterotrimeric proteins obtained when chimeric esterases are chimerised with low homologous parents often produce misfolding, resulting in inactive chimeric esterases constructed with low soluble protein content. Therefore, it is valuable to successfully construct chimeric esterases with stable activity and high soluble fractions after expression.

2 Text

2.1 Chimeric recombination between highly homologous parents

* Corresponding author: 2499638254@qq.com

2.1.1 domain recombination

Traditional domain recombination involves linking domains of different parents by flexible linkers, and many protein properties have been successfully modified by this method, such as new activities not present in the parent protein by recombination of domains of GST protein from human and mouse. and Carriere et al. altered the substrate specificity of human pancreatic lipase (HPL) by linking the lid domain of porcine pancreatic lipase-related protein to the catalytic domain of human pancreatic lipase. After constructing a chimeric esterase by linking the recombination sites in the flexible linkage region of the structural domain, the denovo chimeric esterase was used as a template for optimisation by staggered extension PCR, targeted mutagenesis, etc., and finally a viable chimeric esterase was obtained. This experimental group successfully constructed the chimeric esterases PAR and AAM7 of the hyperthermophilic mineralogical archaeon ap APH and the thermophilic archaeon thermophilic esterase AFEST, which are chimeric esterases, by this method. The enzymes have only 15% primary sequence identity and both inherited the thermophilicity and thermal stability of their parents in nature.

2.1.2 Evolution-based recombination

The structure and function of protein must be understood deeply through rational design of protein, so this method is only suitable for proteins whose three-dimensional structure has been correctly identified and structure-function relationship has been understood. In order to break the limitation, a series of evolution-based recombination methods have been developed through protein directed evolution technology in protein engineering combined with high-throughput screening.

2.1.2.1 DNA shuffling

Originally created in 1994 by Stemmer.A novel in vitro directed molecular evolution technique based on PCR technology is proposed, which is characterized in that a target gene is cut into several fragments by endonuclease, then the fragments are randomly recombined by PCR, and the recombined fragments are used as templates for extension until a brand-new mutant gene is obtained. Crameri et al. selected four 1.6 kb genes from different bacterial genera, which belong to the same gene encoding cephalosporin C enzyme and have 58~ 82 % sequence identity, and used them as starting sequences for DNA shuffling. The results showed that the enzyme activity increased by 270~540 times.

2.1.2.2 Staggered extension recombination

The earliest was Zhao = 1998 The main method proposed is to create different point mutations on multiple templates, mix them and repeat the process of denaturation, annealing and extension by PCR. In each cycle, different templates are paired with each other by using extended fragments to synthesize mutant genes containing different template information. Compared with conventional DNA recombination technology, this method combines annealing and extension into one step, which greatly saves reaction time.

Both common domain recombination and evolution-based recombination methods have been widely used in chimeric recombination between highly homologous parents, making great contributions to industry and medicine, but there are still great difficulties in chimeric recombination between proteins with low parental sequence identity.

2.2 Chimeric recombination between low homologous parents

It is found that the lower the sequence identity of parents, the farther the genetic relationship, the higher the possibility of obtaining new functional proteins, but at the same time, it will also lead to residue side chain conflicts between domains. When the homology of parents is too low, traditional domain recombination often leads to protein misfolding, that is, cross destruction of low homology protein domain recombination. At present, chimeric recombination between proteins with less than 70 % sequence identity of parents is extremely difficult.To overcome these constraints, novel approaches have been developed to construct proteins with novel functions.

2.2.1 Chimeric enzyme obtained by recombination of traditional domains and optimization of domain interface

Traditional domain recombination methods, which construct chimeric enzymes through recombination sites linking flexible linker regions of domains, can not obtain functional chimeric proteins due to the destruction of domain interfaces caused by the interaction of residues in the case of low parental homology. Optimization of domain interfaces after recombination (directed evolution method) can reduce the impact of destruction.

The experimental group successfully constructed the chimeric enzymes PAR and AAM 7 of thermophilic esterase ap APH of hyperthermophilic mineral archaea and thermophilic esterase AFEST of thermophilic archaea by the method, the primary sequence identity of the two chimeric enzymes was only 15%, They inherit thermophilicity and thermostability from their parents.

2.2.2 Key motif-directed domain recombination

Because they evolved from the same ancestral protein, members of the same protein superfamily or even different protein families may have the same sequence and structural motifs. Sequences with conserved structure and sequence in these sequences can be defined as key motifs.1 and LAF, a chimeric enzyme of thermophilic ester AFEST, inherited the thermophilicity and thermostability of AFEST and the activity of lipase Lip 1 decomposing long chain fatty acids, respectively.

2.3 Problems to be solved during restructuring

2.3.1 Selection of recombination sites

With the development of bioinformatics technology, many methods have been developed to select recombination sites by computer-aided design.

The original SCHEMA algorithm was to calculate the parent protein to be chimeric, analyze the sequence motifs and structural motifs, identify the interactions between residues, score the number of damage caused by residue interactions at each recombination site, and plot an S curve by scores. The lower the S value of the residue corresponding to the site, the less damage caused by recombination. Three β-lactamases were calculated and recombined by SCHEMA program, and their sequence identity was less than 34 %. Of the 553 chimeric enzymes constructed, 111 remained active.

LPC/CSU is an on-line server for analyzing residues in protein molecules developed by Tel Aviv University, Israel. It can analyze the interaction, distance and contact area between residues, and select residues with less interaction as recombination sites, which can greatly reduce the damage caused by residue interaction during recombination.

Fam Clash., OPTCOMB, SCAA || of them can be used to construct chimeric enzymes by computer aided design.

2.3.2 Increasing soluble expression of chimeric enzymes

When constructing chimeric enzymes by domain recombination, the low homology protein chimeric recombination techniques mentioned above often produce a large number of inclusion bodies due to misfolding of the newly formed proteins caused by the collision between residues, resulting in the decrease of soluble protein content.

Reducing inclusion body and increasing soluble protein content is one of the key research problems in chimeric enzyme engineering. At present, there are many methods to reduce inclusion body content, such as controlling temperature and pHprotein renaturationand co-expression with chaperones.

Adding protein fusion tags (such as maltose binding protein) and co-expression of archaeal chaperonin is one of the main methods to increase the soluble protein content of chimeras. Co-expression of the target protein with another protein can bypass the solubility problem and help purification. However, since fusion tags are large proteins, fusion will affect protein crystals and lead to structural and functional disorders of the target protein. Therefore, this method also has problems such as removal of subsequent tags and difficulty in maintaining natural activity of synthetic proteins.[1] In order to avoid the influence of fusion tags on the target protein, protease cleavage or chemical cleavage methods should be adopted in the later stage of transformation. However, these methods often have problems such as low yield, precipitation of target protein and failure to recover active protein.

Peptide fusion protein tag is a functional peptide that enhances the specific interaction between a substrate binding protein and a specific subunit. Compared with common protein fusion tags, the tag peptide is relatively short, does not need to be removed, and has little effect on the structure and function of the target protein. By computer aided design of tag peptide with the highest affinity to the target protein, and adding molecular chaperone and target protein co-expression, the soluble protein content can be significantly increased. This experimental group has successfully increased the soluble protein content after commercial heparinase HepA expression by constructing tag peptide cpks and molecular chaperone co-expression system. To make it reach 40% of the total expression.

Table 1. Main methods for increasing soluble expression of chimeric enzymes

main methods	specific measures
Selection of suitable promoters	Promoter addition can significantly affect the transcription intensity and persistence of target gene, and then affect the yield of target protein
optimization of culture conditions	Production temperature, extracellular pH and carbon source all affect the formation of inclusion bodies. Strict control of culture conditions can increase the content of soluble protein.
protein renaturation	Denaturing agents and sulfhydryl reagents for protein renaturation were obtained from inclusion bodies by ion exchange, size exclusion or metal affinity methods, and protein folding was allowed.
high-density fermentation	Higher volume yield of recombinant protein was obtained by batch feeding, constant pH feeding and constant dissolved oxygen feeding, thus increasing soluble protein content.
Co-expression with other proteins	1. Co-expression with nuclease: remove high molecular weight nucleic acid of E. coli by adding nuclease, avoid difficulties in filtration due to excessive viscosity of lysate, and reduce loss of target protein 2. Co-expression with chaperonin: adding appropriate protein fusion tags and co-expression of chaperonin to assist in correct folding of nascent polypeptide chains, reducing inclusion bodies and increasing soluble protein content

2.3.2.1 Current status of research

Protein is the material basis of life, as a part of metabolism, is the main bearer of all life activities. With the rapid development of genetic engineering and protein engineering in recent years, natural and recombinant proteins have been commercially produced, benefiting major sectors such as biopharmaceutical industry, enzyme industry and agriculture.1976 Itakura, K.et al. successfully synthesized a polypeptide containing amino acid sequence corresponding to somatostatin gene by chemical method for the first time, which proved that gene can be fused to plasmid element by chemical synthesis of DNA and

recombinant DNA method, and expressed in Escherichia coli or other bacteria, becoming a new model expression system

When we select an appropriate expression system for recombinant protein production, we must first consider the quality, functionality, production rate and yield of the protein. Generally speaking, proteins larger than 100 kD are expressed in eukaryotic systems, while proteins smaller than 30 kD are expressed in prokaryotic systems. For proteins requiring glycosylation, mammalian cells, fungi or baculovirus systems are selected.

Escherichia coli is one of the earliest and most widely used hosts for heterologous protein production. With its advantages of relatively simple operation, low cost and high success rate, Escherichia coli can be widely used in biotechnology to promote protein expression, especially for functional expression of non-glycosylated proteins.E. coli genome can be modified rapidly and accurately, promoter control is not difficult, plasmid copy number is easy to change, and it has better controllability.However, many heterologous proteins and proteins lacking interaction partners may encounter obstacles when expressed in E. coli cytoplasm, such as poor host growth, inclusion body formation, protein inactivity, or even no protein product. To solve these problems, a variety of culture strategies have been derived for high level production of heterologous proteins., such as using different promoters to regulate expression; using different host strains; co-expression of chaperones or folding enzymes; secreting proteins into periplasmic space or culture medium; reducing the rate of protein synthesis; adding fusion partners; and (X) in vitro denaturation and renaturation of proteins, etc.

2.3.2.2 Formation and avoidance of inclusion bodies

Expression of heterologous proteins in bacteria usually results in the formation of insoluble protein clusters, i.e., inclusion bodies (IBS), which are aggregates of hydrophobic plaques exposed to misfolded or partially folded proteins, produced in the cytoplasmic or periplasmic space, possibly related to an imbalance in the ratio of overexpressed proteins to cellular component partners and folding assistants dedicated to correct folding.[2] Inclusion body formation in recombinant expression systems is the result of protein aggregation and solubilization imbalance in vivo. Aggregation in recombinant systems can be minimized by controlling temperature, expression rate, host metabolism, target protein engineering including solubility tagging technology, and co-expression of plasmid encoded chaperone proteins. In 1989, ki, A. et al. found that inclusion body formation represents the interaction of folding intermediates due to partial degeneration or lack of necessary cofactors in heterologous cytoplasm. Carrio, M et al. Bacterial inclusion bodies are organized aggregates with important structural and biological characteristics associated with amyloid protein. FT-IR spectra data indicate that highly ordered β-sheet structures exist among molecules, and there is a residual natural-like secondary structure of recombinant protein. In addition, it is also believed that the existence of natural-like structures allows for solubilization based renaturation schemes at low denaturant concentrations, thus significantly improving protein recovery. Wang, L. et al. found that inclusion bodies are structured, amyloid formation is a ubiquitous process in eukaryotes and prokaryotes, and amino acid sequence evolution is to avoid amyloid conformation.

Protein aggregation is closely related to protein folding and stability, and is also closely related to molecular chaperones in cells. Protein aggregation can be divided into the following categories: in vivo and in vitro, ordered and disordered. Amyloid fibers (in vivo + in vitro) are examples of ordered aggregates, while inclusion bodies are examples of disordered aggregates in vivo. Several factors that may lead to the formation of inclusion bodies have been proposed, including: Three possible mechanisms of inclusion body formation were proposed: aggregation of natural proteins with limited solubility, aggregation of unfolded proteins, and improper interaction with partners and other cofactors. The exact nature of intermolecular interactions is unclear and may vary among different aggregates. What is clear, however, is that aggregation generally results in an increased number of secondary structures. Thus, both ordered and disordered aggregates exhibit increased beta structures relative to the native conformation, and aggregates of a given protein may have similar structures regardless of how they are formed (in vivo or in vitro), Aggregation is likely to involve specific intermolecular interactions; after amino acid sequence, the next most critical factor determining aggregation is protein concentration. Factors that increase the concentration of partially folded intermediates also increase aggregation propensity. At the molecular level, solubility has been considered to be the main macroscopic signal for successful conformation and functional quality of proteins. More and more evidences show that inclusion body formation is a highly specific process, which is formed by the combination of intermediates with obvious secondary structure.

Inclusion bodies can be produced or avoided by selecting appropriate culture conditions, such as production temperature, extracellular pH and carbon source, and induction temperature is also important to optimize product yield, demonstrating the importance of strict control of culture conditions for recombinant protein production to control inclusion body formation. and his colleagues showed that inclusion bodies were formed only in the first few hours after induction of the fusion protein of Staphylococcus aureus protein A and Escherichia coli I-galactosidase at 42 ° C, and inclusion body formation increased with decreasing pH. Inclusion body formation initially increased with increasing temperature, and at 42 °C-39 °C, the tendency of aggregation gradually decreased with time. This suggests that heat shock proteins may be involved in preventing misfolded proteins or intermediates from forming inclusion bodies. The formation of existing soluble proteins observed at pH drop also suggests that inclusion bodies are not always formed from folding intermediates.

Interestingly, functional and correctly folded proteins are embedded in amyloid proteins of the same polypeptide,

and these amyloid peptides tend to be organized into fibers, forming a rigid skeleton.This particular structure leads to a high mechanical stability and allows the particles to retain a large degree of biological activity, making IBS an unexpected functional material. However, the inherent compositional heterogeneity of IBS or the possible presence of endotoxins thus weakens its medical applicability.

Protein renaturation is a rapid method of obtaining renatured proteins from inclusion bodies using minimal, complex and easily automated procedures. After dissolution, the purpose of the renaturation process is to efficiently remove denaturants and sulfhydryl reagents and allow protein folding. Therefore, in order to minimize aggregation reactions, several issues are important: The final concentration of protein to be renatured, composition of renaturation buffer, and method of renaturation, where neutral pH buffer is suitable for solubilization. When solubilizing inclusion bodies, sufficient incubation time should be included to allow complete development. Sometimes, inclusion bodies may be difficult to dissolve, so some form of agitation and elevated temperature may be required. Some researchers also choose to further purify inclusion bodies by combination column chromatography, such as ion exchange, size exclusion or metal affinity. Again, this is thought to enrich the proportion of monodisperse proteins, thereby increasing the yield of renatured target proteins.

2.3.2.3 Protein fusion tag

Soluble expression of heterologous proteins in E. coli has always been an important problem in protein production, and previous experience has suggested that expression levels can be improved by changing expression conditions, but the best tool so far is fusion tags that enhance the solubility of the expressed protein. As the name suggests, the protein of interest is "fused" with another protein to bypass the solubility problem and facilitate purification. The fusion protein greatly improves the solubility of heterologous protein, which is helpful for affinity chromatography purification of heterologous protein. Unfortunately, the addition of fusion tags also raises a new set of questions, including how these tags are ultimately removed and whether proteins made in this way retain their natural structure and activity.

One of the earliest protein fusion tags used was the 26 kDa glutathione S-transferase (GST) encoded by the Schistosoma japonicum parasite, which was refolded from inclusion bodies using E. coli as host. However, this approach was flawed in that refolding was time-consuming and the protein might not recover its full function. The fusion protein can be dissolved in aqueous solution in most cases and purified by affinity chromatography with immobilized glutathione from crude bacterial lysate under non-denaturing conditions. Multiple fusion proteins can be purified in parallel in less than 2 h by intermittent washing. The modified vector can be cleaved from the fusion protein by site specific proteases such as thrombin or coagulation factor. The vector and any uncleaved fusion protein are then removed by glutathione agarose gel adsorption.

The other is a universal E. coli expression system using E. coli thioredoxin (TrxA) as a gene fusion partner. TrxA fusion protein can improve solubility and promote the expression of various mammalian cytokines and growth factors. It has a wide range of uses. It has thermal stability and provides conditions for the correct folding of fusion protein.

Maltose binding protein (MBP), a homologous E. coli protein as a solubility fusion tag, is one of the most commonly used protein expression tags because of its excellent performance in improving the solubility of the target protein and has a molecular weight of 42 kDa. It has recently been shown to have little effect on the crystallization process, which is an advantage because of the general impression that the "peptide tag" needs to be removed to aid crystallography. MBP, the periplasmic part of ATP binding maltose/maltodextrin transporter, exists in many bacteria. Although the solubilization mechanism of MBP is not fully understood, its large size, inherent solubility, abundant flexible loops and high proportion of exposed hydrophilic residues may be responsible for its solubilization. Recent studies have shown that MBP does not have the chaperone activity that promotes fusion protein folding as originally thought. Part of the reason why MBP is used as a crystallization partner may be due to its high solubility and expandable hydrophilic surface providing a variety of crystal filling interfaces [3].

A large number of solubility labels have been reported in previous studies. These labels can be divided into several categories, including purification labels, stability labels and functional labels. For example, maltose binding protein (MBP), nitrogen utilization substance A(NUSA), thioredoxin (TRX) and glutathione-S-transferase (GST) were mentioned above. GST and MBP were selected as fusion partners because of their advantages in combination of high expression and affinity purification, while thioredoxin was selected as fusion partner due to its high level of soluble protein expression.Both MBP and GST have an additional benefit because they can act as affinity tags; MBP binds tightly to amylose resin, while GST binds to glutathione resin. However, overwhelming evidence suggests that GST is at best a poor solubility enhancer in E. coli, making MBP one of the most well studied solubility factors.

Notably, we found that MBP is a more efficient solubilizer than the other two fusion partners. In some cases, fusion with MBP can promote correct folding of the attachment protein into a biologically active conformation. Thus, MBP seems to be able to function as a common molecular chaperone in the context of fusion proteins. Moreover, there is ample evidence that N-terminal MBP fusions often produce soluble proteins when the unfused partner is insoluble, making MBP one of the most well-studied factors, and there is ample evidence that MBP fusions often produce soluble proteins when the unfused partner is insoluble. Testing a different set of crystallization partners is beneficial for different protein targets because different fusion tags can provide different packaging interfaces.However, such fusion tags do not work for all proteins.

2.3.2.4 Small peptide tag

Small peptide tags called SET tags, characterized by highly acidic amino acid sequences, have also been shown to stimulate the solubility of a small number of paired proteins. If SET tags prove to have a more general effect, they may be highly valued because their small size (<30 amino acids) may lead to less folding interference and make them easier to study structurally without removing the tag. A potential recent development is sumo tags, This is a ubiquitin related protein that has been reported to increase solubility and in some cases appears to be as effective as MBP. Polycationic amino acid tags using arginine and lysine have been developed to facilitate purification of proteins by ion exchange columns. Using this strategy, molecular chaperones and folding enzymes are fused to their C-terminal 10-arginine tags to simultaneously purify and renature insoluble target proteins on column platforms.

Ubiquitin is a highly conserved protein of 76 amino acids in eukaryotic organisms. It plays a chaperone role in ribosomal biogenesis and is also a marker for protein targeting to proteasome. Numerous studies have reported that fusion of ubiquitin with the N terminal of the target protein gene improves the quality and quantity of fusion protein expression in recombinant E. coli. Ubiquitin fusion enhances the production of cloned gene products in E. coli. Despite the existence of efficient transcriptional and translational signals, some heterologous gene products are not fully expressed after introduction into prokaryotes and eukaryotes. In order to enhance the production of cloned gene products, an expression system was established in E. coli, in which ubiquitin was fused with yeast metallothionein or adenylate cyclase-stimulated GTP α subunit, The increase in yield of cloned gene products is likely due to improved fusion protein stability and/or more efficient translation.

2.3.2.5 New Fusion Tags

New fusion partners continue to emerge and complement traditional solutions, for example, the recently studied FH8 fusion tag was named one of the best solubility enhancer partners FH8 fusion tags act as potent solubility enhancers, and their low molecular weight potentially gives them an advantage over larger solubility tags by providing a more reliable assessment of the solubility of target proteins when expressed as fusion proteins.TRX, GST, MBP and NUSA fusion partners are commonly used as solubility enhancing vectors, but these fusion partners must often be removed when producing recombinant proteins for structural and functional applications. Removal of fusion partners is usually performed by specific protease sites between the fusion tag and the target protein. However, after the soluble fusion partner is cleaved, the target protein may precipitate. Here, a major bottleneck arises, Because the solubility of the target protein changes dramatically in the presence and absence of fusion partners.

However, fusion tags are not always effective and therefore cannot be considered as omnipotent solubility enhancers. Among protein attributes considered to be related to protein solubility, protein acidity is of particular concern. The more acidic the protein fusion partner, the greater the ability to enhance the solubility of the target protein.msyB, a highly acidic E. coli derived protein, greatly improves the solubility of the target protein when used as a fusion partner. In order to test the hypothesis that highly or ultraacidic fusion tags are effective means to improve the solubility of recombinant proteins, the ORFs of E. coli were screened for ultraacidic proteins. The results showed that these ultraacidic proteins were superior to other commonly used fusion tags, and could even be successfully used for target proteins that are easy to aggregate, which cannot be overcome by conventional solubility enhancement methods.

2.3.2.6 Chaperone protein

Molecular chaperones such as trigger factors, DnaK, GroEL, Hsp 70 and Hsp 60 family members (hsHsp proteins), etc., individually or cooperatively contribute to the solubility of target proteins In vitro, chaperones can cooperate as part of a functional network, where "holder" chaperones prevent misfolded proteins from aggregating, while "folding" chaperones actively assist in folding. Hsp70 chaperone DnaK participates in the folding of a subset of newly synthesized proteins, prevents protein aggregation under denaturing conditions, and refolds misfolded proteins. Molecular chaperones are essential for the correct folding of a significant portion of proteins in cells under physiological and stress conditions, Hsp70 chaperone DnaK is a central element of the chaperone network that assists in de novo protein folding, protein translocation, protein translocation, protein folding, protein refolding and protein solubilization. Oligomer dissociation and prevention of stress-induced protein aggregation. DnaK mediates protein folding by ATP-dependent interaction with short-stranded peptides exposed to unfolded proteins.

The hallmark of a true substrate protein is that the protein coprecipitated with the chaperone complex undergoes ATP-dependent chaperone release.It was found that CpkB, a subunit of T. kodakarensis chaperone protein, protects positively charged labeled proteins more effectively than unlabeled targets, and attachment of positively charged tags may enhance target recognition by chaperones carrying negatively charged C-terminal regions.[5]Archaeal class II chaperones have a negatively charged C-terminal region, and their specificity is enhanced by attaching a positively charged tag to their target protein, thereby increasing electrostatic interaction between the target protein and the negatively charged carboxy-terminal region of CpkB, protecting the protein from thermal denaturation and correct folding.

In cells, molecular chaperones bind to proteins that tend to aggregate and impede aggregation. The ability of proteins to resist aggregation and remain soluble in aqueous solutions is related to their physical properties. Numerous physical studies have shown that charged atoms facilitate solubility. Many molecular chaperones have considerable negative charges, which can be said to be a common feature of molecular chaperones. HSP 90 is

one such negatively charged chaperone. The charges on Hsp90 are concentrated mainly in two highly acidic regions. In order to study the relationship between partner charges and protein solubility, Wayne, N et al deleted these charge-rich regions and analyzed the anti-aggregation activity of the resulting Hsp90 structure. Finally, it was found that The deletion of these two charge-rich regions significantly weakens the anti-aggregation activity of Hsp90. The anti-aggregation effect of the deleted charge-rich regions may be due to net charge or sequence specific characteristics, indicating that net charge of Hsp90 is involved in its anti-aggregation activity.

3 Summary

Construction of highly efficient and stable chimeric enzymes by domain recombination is an important research project in protein engineering. In this paper, some strategies such as DNA shuffling, staggered extension recombination, etc. are summarized by referring to previous research experience. There are still bottlenecks in chimerism between parents with low sequence identity and low soluble expression of hybrid protein after chimerism, so computer assisted selection of suitable recombination sites and addition of molecular chaperones to improve soluble expression of chimeric protein may be more appropriate methods. So that it can play a bigger role in biological modification.

References

1. Vargas Cortez, T; Morones Ramirez, J.R; Balderas Centeria, I; Zarate, X. Expression and purification of composite proteins in Escherichia coli tagged with a small metal binding protein from Nitrosomonas Europaea Protein Expr Purif , **118**, 49-54 (2016).
2. Ami, D; Natalello, A; Gatti Lafranconi, P; Lotti, M; D Trends Biochem Sci, 42726-737 (2017).
3. Jin, T; Chuenchor, W; Jiang, J; Cheng, J; Li, Y; Fang, K; Huang, M; Smith, P; Xiao, T.S. Design of an expression system to enhance MBP mediated crystallization Sci Rep., **7**, 40991 (2017).
4. Han, Y; Guo, W; Su, B; Guo, Y; Wang, J; Chu, B; Yang, G. High level expression of double recombinant proteins in Escherichia coli using an HE maltotriose binding protein fusion tag Protein Expr Purif , **142**, 25-31 (2018).
5. Gao, L; Hidese, R; Fujiwara, S. Function of a thermodynamic archeal chapter is enhanced by electrostatics Interactions with its targets J. Biosci Bioeng, **124**, 283-288 (2017).