

Research Progress of Fusion Enzyme and the Effect of Peptide on Fusion Protein

Chunling Ma ^{a*}, Fang Chen ^b and Li Wang ^c

School of Qilu University of Technology (Shandong Academy of Sciences), Shandong, China

^amachunling20022@163.com, ^b2937201813@qq.com, ^c2267782834@qq.com

Keywords: Fusion enzyme; Peptide.

Abstract: Fusion enzyme refers to a comprehensive enzyme binding body composed of two or more independent enzymes or domains. It is mainly obtained through the connection of two or more enzymes, which largely retains the function of several domains and is an efficient protein binding body. Recently, with the development of fusion protein technology, the research on the function of fusion enzyme has become a hot spot. Scientists believe that there is a need for the natural evolution of fusion proteins with linkers. In the process of building fusion enzymes, the introduction of linkers can avoid the interference of different domains in the process of protein folding and catalysis, so as to achieve the efficient expression of fusion enzyme.

1. Introduction of fusion enzyme technology

The fusion enzyme technology is mainly to connect different kinds of enzymes by molecular means and express them on the same plasmid, so as to obtain the fusion enzyme with multiple functions at the same time.

Compared with the mixed enzyme system, the fusion enzyme has the following three advantages [1]. First of all, we can integrate a variety of different enzymes together to realize the multi-function of the enzyme. Only through a single culture induction, we can obtain a variety of enzymes, saving production costs. Secondly, the fusion enzyme can separate the two proteins and control the distance between them, so as to accelerate the catalytic process, improve the reaction rate and save the reaction time. Finally, by fusing the target protein and the cell surface protein, the protease can be displayed on the cell surface without breaking the cell, which simplifies the operation process.

2. Protein fusion technology

There are two fusion technologies of protein, including protein indirect fusion based on gene fusion and protein direct fusion. From the perspective of the spatial structure of protein fusion, there are two ways of protein fusion: head to tail fusion and insertion fusion.

2.1 Protein indirect fusion.

Protein indirect fusion is a technology system based on gene fusion, that is, first fusing each unit coding gene to obtain the fusion gene, and then selecting the appropriate host to express the corresponding fusion protein, the key of which is gene fusion technology. At present, there are two main gene fusion technologies, one is based on ligase gene fusion, the other is splicing-by-overlap extension (SOE).

Gene fusion based on ligase: there are two methods of gene fusion based on ligase. One is to obtain the fusion gene through direct blunt junction of ligase, which is also the earliest gene fusion method. The main disadvantage of this method is that it is inefficient and needs to identify the linkage direction of genes [2]. The second is the connection between restriction endonuclease sites, which can be carried out through the polyclonal sites on the vector, and the genes can be inserted into the polyclonal sites in sequence, or the same endonuclease site sequence can be embedded in the downstream of the previous gene and the upstream of the latter gene, and the fusion gene can be

obtained through enzyme cutting and connection [3-5]. The advantage is that there is no additional amino acid residue in the fusion protein, At present, there is no report on the application of gene fusion.

Gene overlap extension technology is the most widely used gene fusion technology, which was invented by Horton et al. (1988). Compared with the above ligase based fusion method, the outstanding advantage is that no additional amino acid residues will be introduced into the fusion gene, and there is also no need to identify the connection direction of genes. Through the use of primers with complementary ends, if there is linker, the linker coding series can be designed into the primers to be amplified respectively, so as to ensure that the primers of the two genes have 9-15 complementary base sequences. After PCR amplification respectively, the two gene fragments to be fused containing the overlapping base sequences are obtained, and the amplification products of the two fragments are mixed and used as templates for subsequent amplification. Through the extension of overlapping chain, the amplification fragments from different sources should be overlapped and spliced.

2.2 protein direct fusion technology.

Protein direct fusion technology refers to the fusion of two proteins directly on the basis of protein, mainly by means of protein intron (intein).

Protein introns can be used to fuse different proteins, known as expression protein linking (EPL) [6,7]. The main step is to fuse two identical integins into the downstream and upstream of the two genes respectively, and express two fusion proteins. The C-terminal and N-terminal of the first protein are both integins, Then, the two proteins are mixed together to induce the intron self shearing under certain conditions. The C-terminal of the former protein will react with the N-terminal of the latter protein to produce the target fusion protein. In fact, there are few reports about protein fusion using EPL. There are two disadvantages of EPL. One is that the length of fusion protein can't exceed 120 amino acid residues. For the fusion of large protein, multi-step EPL is needed. The second is that the segments needed to be fused need to be specially modified. For the specific problems of EPL, please refer to the relevant review [8].

2.3 Head tail connection fusion insertion fusion.

According to the spatial arrangement and combination of fusion protein units, there are two ways of protein fusion, i.e. head to tail fusion and insertion fusion.

Head to tail connection refers to the connection of one end of a protein or domain to the other end of a protein or domain. Insertion fusion refers to the insertion of a complete protein or domain into the appropriate site of another complete protein or domain [9]. Generally, the fusion protein constructed by insertion fusion has a relatively compact structure and is relatively stable, For example, the fusion protein gluxyn-1 constructed by insertion fusion has dual functional activity of β - glucanase and xylanase [10]. Insertion fusion requires detailed structural information of the inserted protein in order to select an appropriate insertion site, usually the insertion site is located in the N-terminal or C-terminal cyclopeptide region of the protein structure, otherwise more work is needed to screen the target fusion protein. Guntas & Ostermeier (2004) randomly inserted the mature peptide gene of β - lactamase into the maltose binding protein (MBP) gene, constructed an insertion fusion gene library, and then expressed and screened two isomeric β - lactamases regulated by the effectors in *E.coli*. In contrast, head to tail fusion does not require too much structural information for the protein or domain to be fused. Therefore, most fusion proteins have been fused by head to tail fusion.

2.4 Application of protein fusion technology.

In recent years, protein fusion technology has been widely used in biomedical and other fields, mainly in the following three aspects. One is to construct the recombinant fusion protein, improve the expression and solubility of the target protein, and the fusion purification tag is conducive to the rapid purification of the protein. Second, through the fusion of different proteins (or protein domains), we can obtain functional proteins, functional specific antibodies, bivalent antibodies or multivalent

antibodies and so on. The third is the development of surface display technology through the fusion expression of target protein and cell surface protein.

3. The effect of conjugated peptide on fusion enzyme

A large number of studies have shown that when the head and tail of the fusion protein are connected and fused, adding the appropriate connecting peptide between the fusion proteins can ensure that the two fusion units fold relatively independently, so that each unit can function relatively independently [11].

Connecting peptide is a polypeptide existing between two fused enzymes or domains (it is generally believed that it is not directly related to the catalytic process of the enzyme). Its length varies from several to hundreds of amino acid residues, and the connecting peptide will not have a greater impact on the folding between proteins or interact with the substrate, which is a need to promote the development of fusion enzyme technology.

3.1 composition and classification of linkers.

In the selection of amino acids suitable for connecting skin, the preferred amino acid residues are arginine, threonine, proline, glutamic acid, glutamine, glycine, phenylalanine, lysine, aspartic acid, asparagine and dexserine. At present, it is generally divided into three categories (Table 1).

Table 1. Comparison of characteristics of linkers

Linker type	Example	Characteristic	Reference
Flexible peptide	(GGGGS) _n (n≤6)	The secondary structure has irregular curl, and the isolation effect of the two connected domains is not stable	[13, 14]
Rigid peptide	(EAAAK) _n (n≤6)	The number of N determines the effect of isolation, Keeping the structure and function of enzyme stable. Resistant to proteolysis	[12]
Protease-resistant peptide		Preventing the fusion enzyme from being degraded by protease and improving the stability of the fusion protein, Obtaining the sequence of connecting peptide difficultly	[15]

The amino acids of non polarity and low charge effect, glycine (Gly) and serine (SER), are the most widely used in the flexible binding peptides. It is characterized by the smallest molecular weight of glycine in all amino acids, the best flexibility, no chiral carbon, good extendibility, simple structure, small volume, small steric hindrance, good for movement, and the smallest influence on the conformation and function of the fusion protein when it is formed. Serine is the most hydrophilic amino acid, which can increase the solubility of the fusion protein when it is formed It can play the role of fusion protein. The first proposed (GGGGS) _n (n ≤ 6) flexible linker has become the most widely used linker. Trinh (2004) et al. Connected the single FV with the reconnection CH3 of human anti mouse ferritin receptorlgG3 with (GGGGS) 3. In the initial experiment, the fusion protein was not expressed significantly. Subsequently, the internal sequence of linker sequence was adjusted, and the junction sequence was not changed. In the transient transfection assay, the expression of the recombinant fusion protein was increased by 30 times.

There are rigid peptides (EAAAK) 5 studied by Wu (2009) and (EAAAK) _n (n = 2-5) studied by Arai (2001; 2004). Moreover, Arai et al. compared the fusion protein composed of blue fluorescent protein (EBFP) and green light protein (EGFP) using the flexible junction peptide (GGGGS) _n (n = 3-4) and the rigid junction peptide (EAAAK) _n (n = 2-5) respectively. The results show that EAAAK

can effectively separate the functional regions of two fused domains, which is more conducive to the overall function of the fusion protein.

The peptide of anti-protease degradation is people's pursuit of the peptide characteristics , (GGGGS)_n was initially considered to have the ability of anti protease degradation, but later the experiment denied this conclusion. Usually The proline rich peptide sequence, proline threonine (PT) repeat sequence, has been proved to be capable of anti protease degradation,. This sequence often exists as a connecting peptide in many natural multi domain proteins [13]. In 1995, Ong E et al. fused cellulose binding domain (CBD), proline threonine peptide (PT) and factor Xa cleavage sequence (IleGluGlyArg) to human IL-2 with N-terminal junction. The fusion protein CBDAPT-IL-2 can be purified by adsorption on cellulose. The purified fusion protein hydrolyzes and then reacts with factor Xa to release the active IL-2.

3.2 Design of linker.

Up to now, the sequence fusion technology using linkers is the most widely used and the highest success rate in the fusion enzyme technology.

Linker is to connect two active proteins to construct a fusion protein. The activity of the fusion protein can not be lost. It must have their own original activity, or even higher activity, which is the original intention of researchers to construct the fusion protein. Therefore, linker design must meet the following conditions: First, the linked functional protein chains do not fold and wrap, and each other will not lose its own functional activity. Second, linker length should be appropriate. Thirdly, the expression, purification and solubility of the fusion protein should be taken into account when the functional activity of the fusion protein is normal.

3.3 effect of linker on fusion protein.

The composition and length of the amino acid residues of the linker have a significant effect on the function of the fusion protein [11].

The relatively long and flexible binding peptide can make the two domains of the fusion protein remain structurally independent, thus ensuring that the fusion protein has better dual functional properties. However, if the peptide is too long, it will increase the sensitivity of the peptide to protease, leading to instability of the fusion protein. If the peptide is too short, the parent domain of the fusion protein will be too close, thus affecting their respective functions [16]. Of course, the length of the linker mainly depends on the spatial structure of the fusion protein. The presence of glycosylation sites on the linker also contributes to the stability of the linker in fungi [17].

The secondary structure of linker sequence: α - helix and β - corner structure limit the expansion of fusion protein and affect the effective folding of functional domain. Therefore, linker folding should be considered in linker design to avoid secondary structure. Crasto (2000) and Xue (2004) developed linker software LINKER, which can determine linker sequence by X-ray crystallography and NMR according to input parameters. The only input needed is the expected linker sequence length. This program has become a favorable tool for biotechnology industry and biomedical research.

4. Conclusion and Prospect

The fusion of two or more proteins is considered to be the advanced form of natural evolution.

The study of shows that the fusion enzyme technology is an effective way to improve the catalytic efficiency of many enzymes and provides a new research direction for enzyme engineering. The fusion enzyme technology can realize the fusion of many kinds of enzymes, so that one kind of fusion enzyme has the function of many kinds of enzymes, so as to simplify the operation process, save the production cost, speed up the reaction rate and save the reaction time. At present, the fusion enzyme has been widely used in the actual production. The application of linkers is a new strategy for the construction of fusion enzyme, which can be expected to achieve fine regulation of the spatial tissue relationship between fusion enzyme domains, isolate different protein domains, and improve the stability of fusion protein. However, in the process of building the fusion enzyme, there is a lack of

sufficient understanding of the interaction between the domains of the fusion enzyme and the specific mechanism of the action of the binding peptide on the fusion enzyme. It is also necessary to systematically study the influence of the properties of the binding peptide on the spatial organization of the domain.

References

- [1] Albertsen L, Chen Y, Bach L S, et al. Diversion of Flux toward Sesquiterpene Production in *Saccharomyces cerevisiae* by Fusion of Host and Heterologous Enzymes[J]. *Applied & Environmental Microbiology*, 2011, 77(3):1033-1040.
- [2] Kirn J H, Weaver R F. Construction of a recombinant expression plasmid encoding a staphylococcal protein A-ricin A fusion protein[J]. *Gene*, 1988, 68(2):315-321.
- [3] Arai R, Ueda H, Kitayama A, et al. Design of the linkers which effectively separate domains of a bifunctional fusion protein[J]. *Protein Eng*, 2001, 14(8): 529-532.
- [4] Chang H C, Kaiser C M, Hartl F U, et al. De novo Folding of GFP Fusion Proteins: High Efficiency in Eukaryotes but Not in Bacteria[J]. *journal of molecular biology*, 2005, 353(2):0-409.
- [5] Liu Dongmei, Zhang Zhuoyuan, Wang Yan, et al. Construction of expression vector of secretory rhGM-CSF/IL-2 fusion protein[J]. *chinese journal of biologicals*, 2002, 15..
- [6] Blaschke U K, Silberstein J, Muir T W . [29] Protein engineering by expressed protein ligation[J]. *Methods in Enzymology*, 2000, 328:478-496.
- [7] Xu, R, Ayers,et al. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies[J]. *Proceedings of the National Academy of Sciences*, 1999.
- [8] Ralf, David, Michael,et a. Expressed protein ligation : Method and applications[J]. *European Journal of Biochemistry*, 2004.
- [9] Doi N, Yanagawa H. Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett*, 1999, 453(3): 305–30.
- [10] Ay J , Gotz F , Borriss R , et al. Structure and function of the *Bacillus* hybrid enzyme GluXyn-1: Native-like jellyroll fold preserved after insertion of autonomous globular domain[J]. *Proceedings of the National Academy of Sciences*, 1998, 95(12):6613-6618.
- [11] Crasto C J , Jin-An F . LINKER: a program to generate linker sequences for fusion proteins[J]. *Protn Engineering*(5):5.
- [12] Ziliang H et al. Progress in design and application of fusion enzyme. *Journal of bioengineering*, 2012.28 (04): 393-409.
- [13] Holsch, K. and D. Weuster-Botz, Enantioselective reduction of prochiral ketones by engineered bifunctional fusion proteins. *Biotechnol Appl Biochem*, 2010. 56(4): p. 131-40.
- [14] Arai, R., et al., Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. *Proteins*, 2004. 57(4): p. 829-38.
- [15] Rawlings, N.D., A.J. Barrett and A. Bateman, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*, 2012. 40(Database issue): p. D343-50.
- [16] Gustavsson M, Lehtio J Denman S, Teeri n Hult K, Martinelle M. Stable linker peptides for acellulose-binding domain-lipase fusion protein expression in *Pichia pastoris*. *Protein Engineering Design and Selection* 2001,14:711-715.
- [17] Remy I, Wilson IA, Stephen WM. Erythropoietin receptor activation by a ligand-induced conformation change. *Science* 1999, 283:990-993.