Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 1197–1200
Progress in Natural Science
www.elsevier.com/locate/pnsc

Review

Design and optimization of a linker for fusion protein construction

Jianhua Zhang [a,*], Jun Yun [b], Zhigang Shang [a], Xiaohui Zhang [c], Borong Pan [b]

[a] Department of Biomedical Engineering, Zhengzhou University, Zhengzhou 450001, China
[b] Department of General Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an 710033, China
[c] College of Medicine, University of South Florida, FL 33647, USA

Received 23 November 2008; received in revised form 9 December 2008; accepted 14 December 2008

Abstract

Bivalent, bispecific single-chain antibody fusion protein construction appears to be a promising tool for tumor therapy. One of its drawbacks is that the function and activity of the fusion proteins have varied affinity and/or anti-tumor activity compared with the original molecules from which they are derived. A more optimized linker for fusion proteins would confer more favorable biological activities upon bispecific single-chain antibodies. Thus, it is critical to design and optimize an inter-peptide linker. With different functional domains and optimized linkers, fusion proteins provide a solid base for targeted immune therapy for malignancies. In this paper, we review the inter-peptide linker studies and the design of an optimized linker using genetic algorithms. The spatial structure of the fusion protein can be predicted by using genetic and bioinformatics research. Based on current research, the future focus will be on different correlation models to perform simulations of spatial structures and drug molecule design.

Keywords: Homology modeling; Improved genetic algorithm; Single-chain bispecific antibody; Design of inter-peptide linker

1. Introduction

With the development of molecular biology, tumor therapy has entered an era of molecule and gene therapy [1–3].

The combination of gene therapy with conventional therapeutic approaches enables a more favorable efficacy to be achieved, thus representing a promising treatment option for the future.

There have been quite a few achievements in tumor gene therapy; for example, several genes are under study in clinical trials [4,5]. However, most of them are experimental and many key techniques and theories still need more exploration, such as the identification of reliable target genes, safe and efficient vectors, targeted regulation of gene therapy, and the combination of multiple gene therapies or the combination of gene therapy with conventional therapies. These problems are currently being faced in tumor gene therapy and are the focus of future research.

Single-chain Fv antibodies (for example, ScFv) are ideal tools for construction of single-chain bispecific antibody fusion proteins [1,6,7]. Bivalent antibodies derived from ScFv using genetic engineering have promising future applications in a clinical setting, since ScFv itself is therapeutic and since it can be used as a vector to combine with a toxin [8]. One of the problems with this approach is that the fusion proteins have varied affinity and anti-tumor activity compared with the original molecules from which they are derived. Determining the optimal inter-peptide linker in fusion proteins is one of the key issues to be solved. Thus, the design and analysis of linker proteins will be a future research focus.

2. Current research on inter-peptide linkers

The reason that fusion proteins have varied affinity and anti-tumor activity compared with the original molecules is

* Corresponding author. Tel./fax: +86 29 84775271.
E-mail address: petermails@zzu.edu.cn (J. Zhang).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.12.007

the structural alterations of the fusion proteins [1,9–12]. The factors determining the biological activities of the fusion proteins include (1) the spatial position of the derived molecules, which is determined by its active domains; (2) the structure of individual receptors for the derived molecules and their correlation; (3) the size and complexity of the linker; and (4) other factors that could affect the spatial structures of the fusion proteins.

In the construction of antibody fusion proteins, researchers pointed out that optimized linkers contribute not only to the expression efficiency of the fusion protein, but also to the correct folding and corresponding biological activities of the different domains of the fusion proteins. Fang et al. [13] designed and constructed single-chain bispecific antibodies with different inter-peptide linkers, using an altered anti-human CD3 ScFv antibody and an anti-human ovarian-associated antigen OC183B2 ScFv antibody. They studied the effects of different linkers on the expression of the single-chain bispecific antibody, binding activities to the antigen, and its *in vivo* stability. Their results showed that the inter-peptide linkers can affect the antigen-binding activities and the *in vivo* stability of the bispecific antibody. Another study by Liu et al. [14] showed that the longer the inter-peptide linkers were, the better the preservation of the independent folding and biological activities of the two molecules. However, longer linkers can also be easily cleaved by the proteases of the host cells, because the structures of the linkers and the adjacent regions are more loosely connected. Therefore, recombinant proteins, especially those in the eukaryotic cells with complex proteases, often have weak linking regions, which are subject to digestion. Yan et al. [15] successfully constructed an anti-human colorectal cancer bivalent single-chain ND-a SC(Fv)2 with a linker of GGGGS. It was highly expressed in *Escherichia coli* and the fusion protein has good immune activities.

Other researchers [16,17] manipulated peptide linking at the DNA level. These studies showed that the length of the linkers, composition of amino acids, glycosylation status, and flexibility of the linkers with the half-molecules are the factors that affect the function and stability of fusion proteins.

Some researchers [18] believe that sufficient length and certain sequence characteristics are the key factors that provide the two half-molecules with sufficient free space to fulfill their functions, and avoiding the formation of the α-helix and β-sheet is important for stability. The size of the antibody and different degrees of affinity can all affect the hemodynamic features of the single-chain bispecific antibodies [19]. Increasing the molecular weight of an antibody slowed the clearance rate, but the permeation into the tumors was decreased. Besides the molecular weight, other factors like conformation of the antibody, and charge features and binding characteristics can also affect the metabolism of the fusion protein. In studies by Gustavsson [16], linkers with a length of 4–44 amino acid residues were proposed to be optimized, whereas in the studies of Le Gall [9] 6–27 residues were considered to be good. Both studies agreed that the length and composition of the linkers affect the functionality and stability of the ScFv. Linkers that are too short negatively affect protein folding by spatial occupancy, and those that are too long enhance the antigenicity of the ScFv antibody and also affect the functionality and activity of ScFv antibodies.

To summarize, the factors that affect the functionality and activity of the ScFv antibody include its composition, length, flexibility, and also the conformation, charge and binding features of the original molecules. Thus, an optimized inter-peptide linker is a critical factor for the ScFv antibody's biological activity, although further mechanisms remain to be explored. Further research on inter-peptide linkers will be of great significance in the screening and construction of single-chain bispecific antibodies.

3. Bioinformatics in the design of inter-peptide linkers

With the development of proteomics and protein composition, structure and function are more frequently studied. In studies on the relationship between the structure and activity of fusion proteins, bioinformatics has proven to be a powerful tool. An important area of bioinformatics is drug design based on structural simulation of biological molecules. The simulation of spatial structure and molecular drug design, including studies on fusion proteins with various functional domains and inter-peptide linkers, is a focus of current research [20–22].

3.1. Analyses of the structures of the derived factors using homology modeling

It is well known that a protein's function is determined by its structure, and its structure is determined by the amino acid sequence. Bioinformatics can be used to predict the function of unknown proteins based on the current knowledge of protein sequences, structure and function [23–26]. This approach is cost-effective, time saving and reliable.

Protein structures are more conserved than protein sequences. If two proteins have 50% homologs of amino acid residues, 90% of the α-carbon atoms are positioned similarly, with a variation of 3 Å. This guarantees the success of homology modeling in protein structure prediction. Homology modeling relies on the identification of one or more known protein structures, known as "templates", likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. This can be done by searching databases of protein sequences and structures such as the Protein Data Bank (PDB). The sequence alignment and template structure are then used to produce a structural model of the target. Detectable levels of sequence similarity usually imply significant structural similarity. The homology modeling study is the main successful method to predict a protein's

three dimensional structure [27]. An example is structural studies of G-protein-coupled receptors [28]. G-protein-coupled receptors are a family of receptors on the cell membrane. They play important roles in signal transduction pathways, and, thus, are important targets for drug development. Because there is no experimental method to determine their native protein structures, researchers used Insight II software developed by Accelrys, in which molecular simulation and homology modeling are utilized, generated 3D structures, and then studied the protein functions and designed drugs based on the structures [28].

Homology modeling has been successfully applied to the correlation of protein sequences, structures and functions. Particularly, orthologous proteins often have conserved residues, and these residues can be interpreted by protein 3D modeling. Using such a structure model, multiple sequences of orthologous proteins can be compared and interpreted according to the restrictions of natural selection, requirements of protein folding, stability, dynamics, and function. Indeed, homology modeling can help us to determine which functional groups a protein possesses based on analyses of critically conserved residues at the active sites and binding sites. Thus, homology modeling plays an important role in computer-aided drug design [22,29].

3.2. Design of inter-peptide linkers with improved genetic algorithms

Since Holland developed the first genetic algorithm (GA) in 1975 [30], GAs have widely been used in the issue of optimization in many fields, because GA has characteristics of independence of model of problems, global optimization, implicit parallelism, high efficiency, and robustness in solving non-linear problems. GAs were introduced into computer-aided molecular drug design in the 1990s and have become among the most widely used non-numerical optimized algorithms.

In recent years, there have been many advances in GAs [31]. GAs have been improved to include hierarchical distributed GAs, adaptive GAs, and GAs based on migration and artificial selection (GAMAS). In addition, researchers have also combined GAs and other optimization methods such as the GA/simulated annealing hybrid algorithm, the GA/fuzzy hybrid algorithm, and the GA/Tabu search hybrid algorithm.

There have also been advances in the application of GAs in the prediction and design of protein structures [23,32,33]. GAs have a lot of unique advantages in protein analyses. For example, a GA might be used to simulate natural biological evolution and to follow the natural rules of duplication, hybridization, and mutation. However, standard GAs also have disadvantages such as slow convergence rate and are frequently premature. Thus, GAs have been improved for fitness sharing, crossover, mutation operator, and selections of control parameters. These algorithms have improved searching capability, searching efficiency, and convergence. These improved algorithms and hybrid algorithms are of great significance for molecular drug design.

4. Prospective view

At present, fusion proteins have varied affinity and antitumor activity compared with the original molecules; this is due to structural alterations of the fusion proteins [1,9–12], so we can optimize inter-peptide linkers with the computer-aided design. Based on homology modeling of derivatives, the future design of inter-peptide linkers can be regarded as solving an equation. Not only are the structures and characteristics of target molecules taken into consideration, but the composition, length, and flexibility of the inter-peptide linker must also be considered. Improved GAs can be applied to linker design and optimization. The selection of the fitness function may be one of the key factors in the improvement of these algorithms to optimize linker design.

Using a model of molecular quantization, different correlation models can be established to perform the simulation of spatial structures and molecular drug design. Based on databases of natural protein structures and functions, the spatial structures and functions of fusion proteins can be predicted using homology modeling. Combined with bioinformatics and genetic research, this approach can be used to search for new agents to use in gene therapy for malignant neoplasms.

References

[1] Geng SS, Feng JN, Li Y, et al. Binding activity difference of anti-CD20 scFv-Fc fusion protein derived from variable domain exchange. Cell Mol Immunol 2006;3:439–43.
[2] Park K, Kim WJ, Cho YH, et al. Cancer gene therapy using adeno-associated virus vectors. Front Biosci 2008;13:2653–9.
[3] Favaro E, Indraccolo S. Gene therapy of cancer in the clinic: good news in sight from Asia? Curr Opin Mol Ther 2007;9:477–82.
[4] Ramnaraine M, Pan W, Goblirsch M, et al. Direct and bystander killing of sarcomas by novel cytosine deaminase fusion gene. Cancer Res 2003;63:6847–54.
[5] Benjamin R, Helman L, Meyers P, et al. A phase I/II dose escalation and activity study of intravenous injections of OCaP1 for subjects with refractory osteosarcoma metastatic to lung. Hum Gene Ther 2001;12:1591–3.
[6] Muller D, Karle A, Meissburger B, et al. Improved pharmacokinetics of recombinant bispecific antibody molecules by fusion to human serum albumin. J Biol Chem 2007;282:12650–60.
[7] Stone E, Hirama T, Tanha J, et al. The assembly of single domain antibodies into bispecific decavalent molecules. J Immunol Methods 2007;318:88–94.
[8] Guo JQ, Li QM, Zhou JY, et al. Efficient recovery of the functional IP10-scFv fusion protein from inclusion bodies with an on-column refolding system. Protein Expr Purif 2006;45:168–74.
[9] Le Gall F, Reusch U, Little M, et al. Effect of linker sequences between the antibody variable domains on the formation, stability and biological activity of a bispecific tandem diabody. Protein Eng Des Sel 2004;17:357–66.
[10] Tranchant I, Herve AC, Carlisle S, et al. Design and synthesis of ferrocene probe molecules for detection by electrochemical methods. Bioconjug Chem 2006;17:1256–64.

[11] Bello-Rivero I, Torrez-Ruiz Y, Blanco-Garces E, et al. Construction, purification, and characterization of a chimeric TH1 antagonist. BMC Biotechnol 2006;6:25.
[12] Shibata K, Maruyama-Takahashi K, Yamasaki M, et al. G-CSF receptor-binding cyclic peptides designed with artificial amino-acid linkers. Biochem Biophys Res Commun 2006;341:483–8.
[13] Fang M, Jiang X, Yang Z, et al. Effects of inter-peptide linkers to the biological activities of bispecific antibodies. Chin Sci Bull 2003;48:1912–8.
[14] Liu ZG, Lin JB, Du W, et al. Anti-proteolysis study of recombinant IIn-UK fusion protein in CHO cell. Prog Biochem Biophys 2005;32:544–50.
[15] Yan DD, Yang FH, Fang J. Construction and expression of bivalent single-chain antibody against human colorectal carcinoma. World J Gastroenterol 2006;14:2395–400.
[16] Gustavsson M, Lehtio J, Denman S, et al. Stable linker peptides for a cellulose-binding domain-lipase fusion protein expressed in *Pichia pastoris*. Protein Eng 2001;14:711–5.
[17] Arai R, Ueda H, Kitayama A, et al. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 2001;14:529–32.
[18] Xue F, Gu Z, Feng JA. LINKER: a web server to generate peptide sequences with extended conformation. Nucleic Acids Res 2004;32:W562–5, [Web Server issue].
[19] Cao Y, Lam L. Bispecific antibody conjugates in therapeutics. Adv Drug Deliv Rev 2003;55:171–97.
[20] Kamphausen S, Holtge N, Wirsching F, et al. Genetic algorithm for the design of molecules with desired properties. J Comput Aided Mol Des 2002;16:551–67.
[21] Hajduk PJ, Huth JR, Tse C. Predicting protein drugability. Drug Discov Today 2005;10:1675–82.
[22] Jiang Z, Zhou Y. Using bioinformatics for drug target identification from the genome. Am J Pharmacogenomics 2005;5:387–96.
[23] Dietmann S, Aguilar D, Mader M, et al. Resources and tools for investigating biomolecular networks in mammals. Curr Pharm Des 2006;12:3723–34.
[24] Roshan U, Livesay DR. Livesay, Probalign: multiple sequence alignment using partition function posterior probabilities. Bioinformatics 2006;22:2715–21.
[25] Tosatto SC, Toppo S. Large-scale prediction of protein structure and function from sequence. Curr Pharm Des 2006;12:2067–86.
[26] Petrova NV, Wu CH. Prediction of catalytic residues using support vector machine with selected protein sequence and structural properties. BMC Bioinform 2006;7:312.
[27] Paxton JW, Kestell P, Chiang D, et al. Inhibition of human CYP1A2 oxidation of 5,6-dimethyl-xanthenone-4-acetic acid by acridines: a molecular modelling study. Clin Exp Pharmacol Physiol 2005;32:633–9.
[28] Park SJ. A study of fragment-based protein structure prediction: biased fragment replacement for searching low-energy conformation. Genome Inform 2005;16:104–13.
[29] Szantai-Kis C, Kovesdi I, Eros D, et al. Prediction oriented QSAR modelling of EGFR inhibition. Curr Med Chem 2006;13:277–87.
[30] Sumida BH, Houston AI, McNamara JM, et al. Genetic algorithms and evolution. J Theor Biol 1990;147:59–84.
[31] Wahab HA, Ahmad Khairudin NB, Samian MR, et al. Sequence analysis and structure prediction of type II *Pseudomonas* sp. USM 4-55 PHA synthase and an insight into its catalytic mechanism. BMC Struct Biol 2006;6:23.
[32] Douguet D, Munier-Lehmann H, Labesse G, et al. LEA3D: a computer-aided ligand design for structure-based drug design. J Med Chem 2005;48:2457–68.
[33] Arunachalam J, Kanagasabai V, Gautham N. Protein structure prediction using mutually orthogonal Latin squares and a genetic algorithm. Biochem Biophys Res Commun 2006;342:424–33.