SOFT COMPUTING TECHNIQUES FOR FUSION PROTEIN STRUCTURE AND LINKER PREDICTION

U.Subhashini[1], P.Bhargavi[2], S.Jyothi[3]
[1]Research Scholar, [2]Assistant professor, [3]Professor
[1,2,3]Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam Tirupati

Abstract

A fusion protein is one whose polypeptide sequence is composed by joining two or more different protein sequences with a suitable linker. Fusion protein exhibits multifunctional properties derived from each of their parent proteins. Fusion proteins can be used in a wide variety of applications and playing an important role in structural biology and biotechnology. The major challenges in computational biology for the design of the novel bifunctional fusion protein are the prediction of the structure and linker. This review highlights the applications of the fusion protein, linkers and soft computing techniques for the prediction of the fusion protein structure.

KEYWORDS: Computational Biology, Fusion Protein, Linkers, Protein, Soft computing techniques.

1. Introduction

Proteins are large, complex molecules in our cells and are the essence of life processes. They are the fundamental constituents of all protoplasm, involved in the structure of the living cell and in its function. Proteins are organic compounds made up of smaller units called amino acids, which are covalently linked to each other by peptide bonds. There are 20 standard amino acids that can be combined to make a protein [1]. Proteins have widespread applications in pharmaceuticals and medical diagnosis [2]. In the beginning, natural proteins were extracted from animal, human sources, or from plants. Then, Recombinant DNA technology started a new area of research and applied aspects of biology [3]. Since then, a significant increase has been seen in reproducing natural proteins by Recombinant DNA technology. It also focused toward developing de novo proteins that do not exist in nature and are called as fusion proteins. Fusion proteins are a class of proteins, constructed by joining two or more different domain proteins [4]. Fusion proteins are also called as chimeric proteins or hybrid proteins. Fusion protein obtains many functional properties derived from each of the original proteins, including biological activity [5]. Many of the research studies revealed that some fusion proteins have greater stability and effectiveness over naturally occurring proteins [6]-[8]. Over the years, researchers have been using the recombinant DNA technology for the construction of the fusion proteins due to the wide variety of its applications such as tissue engineering, improving enzyme activity, drug development, half-life extension[7][9], biomaterial design[10] and analyses of protein-protein interactions [11][12]. Successful construction of the fusion protein initially requires the desired proteins and its compatibility. If component domains are not compatible then it leads to misfolding [13]. The simplest method of fusing selected domains is an end to end genetic fusion. In some cases, direct fusing is simple and works best where N or C-terminal regions of the component proteins act as a “bridge” to provide enough space between protein domains for correct folding[14][15]. However, this strategy fails when the N or C-terminal is not flexible or long enough to avoid steric hindrance, which reduces the degrees of freedom in protein bioactivity and may give rise to undesirable outcomes such as lack of proper protein folding, low yield in protein production and decreased bioactivity[16]-[18]. For this reason, several protein bioactivity studies

yielded fusing of the selected proteins without a linker results in decreased biological activity [19][20]. In this context, most commonly used method for construction of the fusion protein is linker mediated tandem fusion method. In this method, fusion protein is achieved by fusing two proteins with a suitable linker [4].

2. Application of the Fusion Proteins

2.1 Fusion Proteins in the field of Tissue Engineering

Hayashi M et al. [21] introduced EGF-Collagen fusion protein and showed various purposes of this fusion protein like in the field of tissue engineering biocompatible, biodegradable and adhesive fibrous mitogen. Huang J et al. [22] constructed a novel fusion protein with the combination of spider silk and HA nucleating domain of DMPI and they specified this new fusion protein has potential applications in bone tissue engineering.

A new spider silk-bone sialoprotein fusion protein was designed by Gomes S et al. [23]. They fused spider silk protein with the human bone sialoprotein (BSP) and result showed that fusion protein retained same functionalities of each of the domains. They also showed that this fusion protein can be used as a biomaterial in tissue engineering field (especially for the construction of grafts for bone regeneration). A novel fusion protein FNIII7–10/CDH 11 EC 1–2 was derived by Zhang Y et al. [24] for the construction of tissue engineering bones and cartilage products and the manufacture of cell-loaded plastic materials. Asakura et al. [25] developed a silk-like fusion protein film, which showed high cell growth activities of kidney VERO cells.

2.2 Fusion Protein for Half-life Extension and Drug Delivery

One of the useful applications of the fusion proteins is to prolong the in vivo half-life of pharmaceutical proteins because many of the proteins have short half-life profiles. For example, pharmacokinetic studies predicted that half-life of soluble CD4 is short in human. So that Capon et al. [27] fused CD4 derivative with the Fc portion of immunoglobulinG(IgG) and showed rapidly cleared protein can be stabilized by fusion. Generally GLP-1(Glucagon-like peptide-1) is considered as a therapeutic agent for type-2 diabetes. Unfortunately this GLP-1 has short half-life in vivo. To prolong the life of the GLP-1 Gao Z et al. [28] designed a novel fusion protein of KGLP-1/HAS. Here, the N-terminal of GLP-1 is fused with HAS (human serum albumin). Like KGLP-1/HAS, Glaesner et al. [29] developed a GLP-1 immunoglobulin G(IgG4)Fc fusion protein(LY21892625) and proved that LY21892625 retains the activity of GLP-1 with increased half-life. Xin Yan et al. [30] explained vascular endothelial growth factor (VEGF) protein which is an important active protein that stimulates angiogenesis and actively participates in the tissue regeneration of diabetic wounds. However, there are some problems with VEGF: low doses of VEGF may lead to retarded angiogenesis, half-life of VEGF is too short etc. Due to these reasons Zhang J et al. [77] designed CBD-VEGF fusion protein by fusing vascular endothelial growth factor (VEGF) protein with a collagen-binding domain (CBD). VEGF-CBD fusion protein showed better results in diabetic wound healing with extended half-life. Wen ZL et al. [26] fused a gene encoding diphetheria toxin with a gene of α-melanocyte stimulating hormone and also demonstrated that the resulted fusion protein acts as a potential drug for the treatment of malignant melanoma.

In the recent years, several classes of fusion proteins such as structure-based recombinant fusion protein, elastin-and silk-like polymers [31] [32] and cadherin-based fusion protein has been used in different applications.

3. Role of Linkers in the Design of Fusion Proteins

Linkers are short peptide sequences and are widely used in the fusion protein design. Linkers can be considered as independent units in the design of fusion proteins. Linkers do not affect the function of the individual proteins to which they are attached [33] [34]. Linkers are classified into two types: Natural linkers, Empirical linkers.

3.1 Natural Linkers

Naturally occurring multi-domain proteins are common in a variety of cellular processes and their domains are interconnected through linkers. Linkers perform a key role in cooperative inter-domain interactions, function regulation, regulate different folded domains, and domain-domain orientation [35]. In recombinant technology, studying the natural linkers can help in the rational design of

empirical linkers [33]. In the recent years researchers have been identified various natural linkers in several multi-domain proteins [36]. Natural linkers mostly consists of Pro, Arg, Phe, Thr, Glu and Gln amino acid residues and many of linker studies revealed Proline-rich sequences form rigid structures to prevent unfavourable interactions between the domains. Flexible Gly-rich regions have been treated as natural linkers in multi-domain proteins [37]. For example, Crystal structure analysis of PAX6 complex revealed that Pax6 Gly-rich linker is much better ordered and makes many more contacts with the DNA [38]. Wilson1 et al, [39] showed Gly-rich linkers plays an important functional roles in transmembrane glycoproteins(TMs) of retroviruses and Cho S. et al. [40] showed RNA-binding domain (RBD) of SRSF1, consists of two RRMs connected through a glycine-rich linker.

3.2 Empirical Linkers

In addition to the natural linkers, researchers have designed many empirical linkers for the construction of fusion proteins [41]. Empirical linkers are classified into three types: flexible, rigid and cleavable linkers.

3.2.1 Flexible Linkers

Flexible linkers provide an important consequence of the flexibility between the linked domains. Flexible linkers maintain structural stability by forming hydrogen bonds with the water molecules. Flexible linkers are used whenever fused domains require the ability to move to and from spatial proximity. Argos [42] analysed flexible linkers are generally consists of small, polar or non-polar amino acids such as Gly, Ser and Thr. Generally flexible linkers have the $(Gly\text{-}Ser)_n$ amino acid sequence, where "n" represents number of repeats of the motif and by adjusting the number "n" the length of the flexible linker (Gly-Ser) can be fine-tuned to achieve necessary interactions between fused domains. Argos [42] also suggested that flexible linkers designed with Gly and Ser are also contains additional residues like Thr and Ala in order to reduce unfavourable interactions between linker and the fused domains. Yun Bai et al. [43] used a flexible linker of $(GGGGS)_3$ and showed inserted flexible linker has improved both in vitro and in vivo myelopoietic activity over non-linker (G-CSF-Tf) fusion protein. In another study, Hu W et al. [44] inserted $(GGGGS)_3$ linker between two copies of HBsAg preS1(21-47) to produce a fusion protein and showed immunoactivity of the fusion protein is much stronger than the fusion protein without a linker. Michelle S et al. [45] used an eight amino acid flexible linker $(Gly)_8$ in the construction of Myc-Est2p fusion protein and analysed that inserted flexible linker improved in vivo ability of the telemerase sub units(Est2p and Est1p) to maintain telomere length. In the construction of the HAS-ANF fusion protein Sheffield WP et al. [46] fused HAS and ANF with a flexible linker i.e $(Gly)_6$ in order to increase biological activity of ANF, Zhao HL et al. [47] showed that direct fusion of human serum albumin(HSA) with interferon-alpha2b(IFN-alpha2b) resulted as unstable. To get stability of HSA-IFN-alpha2b they inserted flexible linker in between HSA and IFN-alpha2b.

3.2.2 Rigid Linkers

Flexible linkers have several advantages for constructing a wide variety of fusion proteins, even though they have limitations such as ineffectiveness in separating functional domains due to its high flexibility. Maeda et al. [17] described immunoglobulin binding activity of the protein G domain in a G-Vargula luciferase fusion protein is not retained with the inserted flexible GGGGS linker. Some other linker studies revealed use of the flexible linkers resulted in poor expression or loss of biological activity [19]. In this context, rigid linkers have been successfully applied to keep a fixed distance between domains. Rigid linkers also prevent unfavourable interactions between the protein domains [33]. Aria et al. [18] were first introduced a rigid linker with the peptide sequence of $A(EAAAK)_nA$ (n=2-5) for the construction of bioactive fusion proteins. George and Heringa [48] suggested stiff alpha-helical linkers to act as rigid spacers between protein domains. Rigid linkers often applied in cases where the spatial separation of the individual domains is crucial to maintain the stability and biological activity of the entire fusion protein. Amet N et al. [49] used a helical linker in the design of hGH-Tf fusion protein and this helical linker was inserted as a rigid spacer between the fused domains in order to enhance the biological activity. Rigid linkers exhibit relatively stiff structures by adopting α-helical structures or by containing multiple Pro residues. Lu et al. [50] used α-helical peptides linker $(EAAAK)_n$ (n=1-

3) to construct bifunctional β-glucanase-xylanase fusion enzyme. The results showed that the inserted α-helical linker improved the catalytic effect of both moieties.

George and Heringa [48] also specified another type of rigid linkers which are composed of a proline-rich such as $(XP)_n$. Here, X indicates any amino acid, with preference of Lys, Ala or Glu. Proline is a very unique amino acid and it structurally prevents the unfavourable interaction between protein domains. For instance, recombinant fusion proteins comprised of interferon-y (IFN-y) and gp120 of the human immunodeficiency virus were constructed using $(Ala\text{-}Pro)_n$ linkers of different lengths[51].

3.2.3 Cleavable Linkers

Flexible and rigid likers generally consist of stable peptide sequences and provide many facilities like structure flexibility and increase biological activity. However, these linkers do not allow for the separation of the two fusion protein domains *in vivo*. Flexible and rigid linkers have several drawbacks including decreased bioactivity, steric hindrance between functional domains, and altered bio-distribution [33]. To overcome some of these potential pitfalls, Chen et al. [52] introduced an in vivo cleavable linker using the reversible nature of the disulfide bond. Cleavable linkers are large size category of linkers used to construct recombinant fusion proteins [33]. These linkers are often designed with the aim of release free functional domains in vivo, to reduce steric hindrance, improve bioactivity, or achieve independent actions/metabolism of individual domains of recombinant fusion proteins after linker cleave. Xiaoying ch et al. [52] designed a cleavable linker for *in vivo* release of protein domains from a fusion protein. They inserted *in vivo* cleavable linker between transferrin (Tf) and granulocyte colony-stimulating (G-CSF) and achieved an improved therapeutic effect, desired pharmacokinetic profile and bio-distribution.

3.3 Linker Prediction using Soft Computing Methods

Understanding of linkers and their biochemical properties is crucial in designing linkers for construction of various recombinant fusion proteins. Linkers are short peptide sequences that occur between protein domains. Crystallographers usually identify domains and linkers while determining the structure of multi-domain proteins. However, in many cases, the domain linkers are not identified explicitly and there is a need to employ computational methods to determine the linkers based on sequence or structure of the protein. In this context, soft computing provides several possibilities by generating low-cost good solutions. Recent soft computing approaches used for domain-linker prediction are Artificial Neural Networks (ANN) and variants of Support Vector Machines (SVM).In 2002 Miyazaki S et al. [53] described prediction of domain linkers using neural network. Here the neural network was trained to differentiate between domain linker sequences and non-linker sequences using a SCOP-defined domain library. The neural network tested by a jack-knife test validation technique and predicted linker regions with 58% specificity and 36% sensitivity of domain linkers. Again in 2006, Miyazaki,S. et al. [54] used neural network to identify putative domain linker regions in the SWISSPROT database consisting data set of 74 multi-domain proteins [55]. Sim et al. [56] introduced a neural network method, PPRODO (Prediction of PROtein DOmain boundaries) for the prediction of protein domain boundaries. Paul D et al. [57] used an enhanced general regression neural network(EGRN) for the prediction of linkers from the multi-domain proteins and this model achieved 71% accuracy for domain boundary identification in multi-domains proteins.

Ebina et al. [58] constructed a protein linker predictor called DROP (Domain linker PRediction using OPtimal features) which utilizes a SVM with a Radial Basis Function (RBF) kernel. SVM classifier is trained using 25 optimal features. The optimal features were selected from a set of 3000 features using a Random Forest (RF) algorithm. The selected features were related to secondary structures and PSSM (position specific scoring matrix) elements of hydrophilic residues. The accuracy of DROP was evaluated by two domain-linker datasets (DS-All and CASP8 FM). DS-All contains 169 protein sequences, with a maximum sequence identity of 28.6%, and 201 linkers. In protein linker predictions DROP showed a sensitivity and precision of 41.3% and 49.4%, respectively. Ebina T et al. [59] used a variant of SVM called Loop-Length-Dependent SVM was used to predict the domain linkers.

4. Role of Soft Computing in Fusion Protein Design

Several biochemical experiments have shown that the protein function is determined by its structure [60]-[62]. Thus, explicating a fusion protein structure is a key to understanding its function, and is essential for any related biological, medical, or biomaterial applications. In spite of progress in Recombinant DNA technology, experimental determination of a protein structure using the techniques like X-ray, crystallography [63][64] are still expensive, labor intensive, time consuming, and not always possible. Therefore, computational methods to predict structures have been rigorously explored. On this scenario, Soft Computing plays an important role as it provides techniques that are especially well suited to resolve imprecision and uncertainty that the bioinformatics problems have [65]. Various bioinformatics problems like DNA, RNA, protein structure prediction and gene finding etc. are often affected by uncertainty [66]. For all these problems, soft computing techniques are the promising approaches to achieve efficient and reliable heuristic solutions. This review describes the application of these soft computing techniques in the area of fusion protein structure prediction.

4.1 Fusion Protein Structure Prediction using Soft Computing Methods

Once the fusion protein has been fused with the linker then prediction of its functionality is important because individual proteins have a defined shape and structure. Observing the resulted fusion protein structure in detail could reveal what it acquired from its parental domains. Now a day's Structural Bioinformatics mainly focused on prediction of 3D structure of proteins through different experimental methods such as X-ray diffraction, electron microscopy or NMR (nuclear magnetic resonance). In this context, Fusion proteins are complex proteins as they consist of multiple structural domains. Determination of fusion protein structure using experimental methods is expensive due to the costs associated to crystallography, electron microscopy or NMR and time consuming [67].

In order to know the 3D structure of the designed fusion protein (TGFαL3-SEB) Abbas ali et al. [68] used I-TASSER server. For accurate prediction of the fusion protein structure I-TASSER used a simple neural network algorithm. Golshani et al. [69] also used neural network based I-TASSEER for the structure prediction of the L7/L12-TOmp31 fusion protein. Initially Qian et al. [70] used a fully connected feed forward neural network for prediction of protein secondary structure. Here, input layer used a sliding window of 13 consecutive amino acids and the output layer predicts secondary structure classes of the amino acids (helex,beta,coil). Another method to predict the structure of the protein is PSIPRED that was designed by McGuffin LJ et al. [71] which incorporates two feed forward neural networks. The sequences of amino acids are given as input to the first neural network. The network has one hidden layer and produces an initial prediction. Second neural network is used to filter the predicted structures of the first neural network. Vapnik et al. [72] proposed support vector machine method for prediction of protein structure. Also several researchers used support vector machines (SVM's) for the prediction of protein secondary structure [73].

The genetic algorithm is used by Dandekar et al. [74] to predict the protein structure by using tetrahedral lattice and fitness function for the protein structure prediction. Later, Day Ro et al. [75] used multi-objective genetic algorithm to increase the effectiveness of the algorithm. In the combination of genetic algorithm and tabu search algorithm, Mansour RF et al. [76] proposed a hybrid algorithm in off-lattice AB model. Thang N. Bui and Gnanasekaran Sundarraj [81] used genetic algorithm for the prediction of protein tertiary structure in the 2D HP model.

5. Conclusion

This review highlighted various applications of fusion proteins, different types of linkers and various soft computing techniques used in the design of the fusion protein. Linkers offer many advantages in the construction of the fusion proteins such as improving stability and bioactivity and achieving required pharmacokinetic profiles. Construction of the fusion protein using an empirical linker will increase the proximity between protein domain partners and preserve the natural interaction. Till now only ANN and SVM soft computing techniques were used for fusion protein structure and linker prediction. However, computational methods for accurate fusion protein structure prediction are still in the stage of development.

References

[1] JR. Hoffman , and MJ. Falvo, "Protein–which is best?." Journal of sports science & medicine, vol. 3(3), pp:118-130, Sep 2004.

[2] MP. Ratnaparkhi, SP. Chaudhari, and VA. Pandya, "Peptides and proteins in pharmaceuticals", International Journal of Current Pharmaceutical Research, vol.3(2), pp:1-9, 2011.

[3] S. Khan, et al., "Role of Recombinant DNA Technology to Improve Life" , International journal of genomics, vol.2016, 2016.

[4] JS. Klein, S.Jiang, RP. Galimidi, JR. Keeffe, and PJ. Bjorkman, "Design and characterization of structured protein linkers with differing flexibilities", Protein Engineering, Design & Selection, vol. 27(10), pp: 325–330, Oct 2014.

[5] M. Uhlén, G.Forsberg, T. Moks, M. Hartmanis, and B. Nilsson, "Fusion proteins in biotechnology", Current opinion in biotechnology, vol. 3(4), pp: 363-369, August 1992.

[6] M. Nagaoka, HL. Jiang, T.Hoshiba, T. Akaike, and C. Cho, "Application of recombinant fusion proteins for tissue engineering", Annals of biomedical engineering, vol. 38(3), pp: 683-693, Mar 2010.

[7] WR. Strohl, "Fusion proteins for half-life extension of biologics as a strategy to make biobetters", BioDrugs, vol. 29(4), pp: 215-239, Aug 2015.

[8] B. Nilsson, G. Forsberg, T. Moks, M. Hartmanis, and M. Uhlén, "Fusion proteins in biotechnology and structural biology", Current Opinion in Structural Biology, vol. 2(4), pp: 569-75, Aug 1992.

[9] GN. Cox, DJ. Smith, SJ. Carlson, AM. Bendele, EA. Chlipala, and DH. Doherty, "Enhanced circulating half-life and hematopoietic properties of a human granulocyte colony-stimulating factor/immunoglobulin fusion protein" , Experimental hematology, vol. 32(5), pp:441-9, May 2004.

[10] R. Plowright, N. Dinjaski, S. Zhou, DJ. Belton, DL. Kaplan, and CC. Perry, "Influence of silk–silica fusion protein design on silica condensation in vitro and cellular calcification", RSC advances, vol. 6(26), pp:21776-88, Feb 2016.

[11] UH. Weidle, B. Schneider, G. Georges, and U. Brinkmann, "Genetically engineered fusion proteins for treatment of cancer", Cancer Genomics-Proteomics, vol. 9(6), pp:357-72, Nov 2012.

[12] A. Beck, and JM. Reichert, "Therapeutic Fc-fusion proteins and peptides as successful alternatives to antibodies", In.MAbs, vol. 3, No. 5, pp: 415-41, Sep 2011.

[13] D. Sachdev, and JM. Chirgwin, "Order of Fusions between Bacterial and Mammalian Proteins Can Determine Solubility inEscherichia coli", Biochemical and biophysical research communications, vol. 244(3), pp:933-7, Mar 1998.

[14] M. Rizk, G. Antranikian, and S. Elleuche, "End-to-end gene fusions and their impact on the production of multifunctional biomass degrading enzymes", Biochemical and biophysical research communications, vol. 428(1), pp:1-5, Nov 2012.

[15] T. Nagamune, "Biomolecular engineering for nanobio/bionanotechnology", Nano Convergence, vol. 4(9), Apr 2017.

[16] K. Yu, C. Liu, BG. Kim, and DY. Lee, "Synthetic fusion protein design and applications", Biotechnology advances, vol. 33(1), pp:155-64, Feb 2015.

[17] Y. Maeda, H. Ueda, J. Kazami, G. Kawano, E. Suzuki, and T. Nagamune, "Engineering of functional chimeric protein G–VargulaLuciferase", Analytical biochemistry, vol.249(2), pp:147-52, Jul 1997.

[18] R . Arai, H. Ueda, A. Kitayama, N. Kamiya, and T. Nagamune, "Design of the linkers which effectively separate domains of a bifunctional fusion protein", Protein engineering, vol.14(8), pp:529-32, Aug 2001.

[19] N. Amet, HF. Lee, and WC. Shen, "Insertion of the designed helical linker led to increased expression of tf-based fusion proteins", Pharmaceutical research, vol. 26(3), pp: 523, Mar 2009.

[20] VP. Reddy Chichili, V. Kumar, and J. Sivaraman, "Linkers in the structural

biology of protein–protein interactions", Protein Science, vol. 22(2), pp: 153-67, Feb 2013.

[21] M. Hayashi, M. Tomita, and K. Yoshizato, "Production of EGF–collagen chimeric protein which shows the mitogenic activity", Biochimica et Biophysica Acta (BBA)-General Subjects, vol. 1528(2), pp:187-95, Oct 2001.

[22] J. Huang, C. Wong, A. George, DL. Kaplan, "The effect of genetically engineered spider silk-dentin matrix protein 1 chimeric protein on hydroxyapatite nucleation", Biomaterials, vol. 28(14), pp: 2358-67, May 2007.

[23] S. Gomes, IB. Leonor, JF. Mano, RL. Reis, and DL. Kaplan, "Spider silk-bone sialoprotein fusion proteins for bone tissue engineering", Soft Matter, vol. 7(10), pp:4964-73, 2011.

[24] Y. Zhang, Y. Zhou, J. Zhu, S. Dong, C. Li, and Q. Xiang, " Effect of a novel recombinant protein of fibronectinIII7-10/cadherin 11 EC1-2 on osteoblastic adhesion and differentiation", Bioscience, biotechnology, and biochemistry, vol.73(9), pp:1999-2006, Sep 2009.

[25] T. Asakura, C. Tanaka, M. Yang, J. Yao, and M. Kurokawa, "Production and characterization of a silk-like hybrid protein, based on the polyalanine region of Samia cynthia ricini silk fibroin and a cell adhesive region derived from fibronectin", Biomaterials, vol. 25(4), pp:617-24, Feb 2004.

[26] ZL. Wen, X. Tao, F. Lakkis, T. Kiyokawa, and JR. Murphy, "Diphtheria toxin-related alpha-melanocyte-stimulating hormone fusion toxin. Internal in-frame deletion from Thr387 to His485 results in the formation of a highly potent fusion toxin which is resistant to proteolytic degradation", Journal of Biological Chemistry, vol. 266(19), pp:12289-93, Jul 1991.

[27] DJ. Capon et al., "Designing CD4 immunoadhesins for AIDS therapy", Nature, vol. 337(6207), pp:525-3, Feb 1989.

[28] Gao Z, Bai G, Chen J, Q. Zhang, P. Pan, F. Bai, P. Geng, "Development, characterization, and evaluation of a fusion protein of a novel glucagon-like peptide-1 (GLP-1) analog and human serum albumin in Pichia pastoris", Bioscience, biotechnology, and biochemistry, vol.73(3), pp:688-94, Mar 2009.

[29] W. Glaesner et al., "Engineering and characterization of the long-acting glucagon-like peptide-1 analogue LY2189265, an Fc fusion protein", Diabetes/metabolism research and reviews, vol. 26(4) ,pp:287-96, May 2010.

[30] X. Yan, B. Chen, Y. Lin, Y. Li, Z. Xiao, X. Hou, Q. Tan, and J. Dai, "Acceleration of diabetic wound healing by collagen-binding vascular endothelial growth factor in diabetic rat model",Diabetes Res Clin Pract, vol. 90, pp:66–72, 2010.

[31] W. Teng, J. Cappello, and X. Wu, "Recombinant silk-elastinlike protein polymer displays elasticity comparable to elastin", Biomacromolecules, vol. 10(11), pp:3028-36, Sep 2009.

[32] W. Huang, A. Rollett, and DL. Kaplan, "Silk-elastin-like protein biomaterials for the controlled delivery of therapeutics", Expert opinion on drug delivery, vol. 12(5), pp: 779-91, May 2015.

[33] X. Chen, JL. Zaro, and WC. Shen, "Fusion protein linkers: property, design and functionality", Adv. Drug Deliv. Rev, vol.65 (10), PP: 1357–1369, 2013.

[34] K. Maximilian, LC. Cheng, B. Loo Liat Wen, A. Choo Boon Hwa, D. Ow Siak Wei, and LD. Yup, "Effect of linker flexibility and length on the functionality of a cytotoxic engineered antibody fragment, Journal of Biotechnology, vol.199,pp:90-97, 2015.

[35] RS. Gokhale, and C. Khosla, "Role of linkers in communication between protein modules", Current opinion in chemical biology, vol. 4(1), pp:22-7, Feb 2000.

[36] B. Kobe, T. Ve, and SJ. Williams, "Fusion-protein-assisted protein crystallization", Acta Crystallographica Section F: Structural Biology Communications, vol.71(7), pp:861-9, Jul 2015.

[37] PM. Steinert, JW. Mack, BP. Korge, SQ. GAN, SR. Haynes, and AC. Steven,

"Glycine loops in proteins: their occurrence in certain intermediate occurrence in certain intermedi[4 filament chains, loricrins and single-stranded RNA binding proteins", International journal of biological macromolecules, vol. 13(3), pp:130-9, 1991.

[38] HE. Xu et al., "Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding", Genes & development, vol. 13(10), pp:1263-75, May 1999.

[39] KA. Wilson, S. Bär, AL. Maerz, M. Alizon, and P. Poumbourios, "The conserved glycine-rich segment linking the N-terminal fusion peptide to the coiled coil of human T-cell leukemia virus type 1 transmembrane glycoprotein gp21 is a determinant of membrane fusion function", Journal of virology, vol.79(7), pp:4533-9, Apr 2005.

[40] S. Cho et al., "The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs", Nucleic acids research, vol. 39(21), pp: 9413-21, Aug 2011.

[41] C. Liu, JX. Chin, and DY. Lee, "SynLinker: an integrated system for designing linkers and synthetic fusion proteins", Bioinformatics, vol. 31(22), pp:3700-2, Jul 2015.

[42] P. Argos, "An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion", Journal of molecular biology, vol. 211(4), pp: 943-58, Feb 1990.

[43] Y. Bai, and WC. Shen, "Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization", Pharmaceutical research, vol. 23(9), pp:2116-21, Sep 2006.

[44] W. Hu et al., "A flexible peptide linker enhances the immunoreactivity of two copies HBsAg preS1 (21-47) fusion protein", Journal of Biotechnology, vol.107, 83–90, 2004.

[45] M. Sabourin, CT. Tuzon, TS. Fisher, and VA. Zakian, " A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae", Yeast, vol.24(1), pp:39-45, Jan 2007.

[46] MK. de Bold et al., "Characterization of a long-acting recombinant human serum albumin-atrial natriuretic factor (ANF) expressed in Pichia pastoris", Regulatory peptides, vol.175(1), pp:7-10, Apr 2012.

[47] HL. Zhao et al., "Increasing the homogeneity, stability and activity of human serum albumin and interferon-α2b fusion protein by linker engineering", Protein expression and purification, vol. 61(1), pp: 73-7, Sep 2008.

[48] RA. George, and J. Heringa, "An analysis of protein domain linkers: their classification and role in protein folding", Protein Eng, vol. 15, pp: 871–879, 2002.

[49] N. Amet, W. Wang, and WC. Shen, " Human growth hormone–transferrin fusion protein for oral delivery in hypophysectomized rats", Journal of Controlled Release, vol.141(2), pp:177-82, Jan 2010.

[50] P. Lu, and MG. Feng, "Bifunctional enhancement of a β-glucanase-xylanase fusion enzyme by optimization of peptide linkers, "Applied microbiology and biotechnology", vol. 79(4), pp: 579-87, Jun 2008.

[51] AL. McCormick, MS. Thomas, and AW. Heath, "Immunization with an interferon-gamma-gp120 fusion protein induces enhanced immune responses to human immunodeficiency virus gp120J", Infect Dis, vol.184, pp: 1423–1430, 2001.

[52] X. Chen, Y. Bai, J. Zaro, and W.C. Shen, "Design of an in vivo cleavable disulfide linker in recombinant fusion proteins", Biotechniques, vol. 49, pp: 513–518, 2010.

[53] S. Miyazaki, Y.Kuroda, and Y. Shigeyuki, "Characterization and prediction of linker sequences of multidomain proteins by a neural network", Journal of Structural and Functional Genomics, vol.2, pp: 37–51, 2002.

[54] S. Miyazaki, Y.Kuroda, and Y. Shigeyuki, "Identification of putative domain linkers by a neural network -

application to a large sequence database", BMC Bioinformatics, vol.**7**:323, 2006.

[55] A. Bairoch, R. Apweiler, " The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000", Nucleic Acids Res, vol.28(1), pp:45-8, Jan 2000.

[56] J. Sim, SY. Kim, J. Le, " PPRODO: prediction of protein domain boundaries using neural networks", Proteins, vol.59(3),pp:627-32, May 2005.

[57] PD. Yoo, AR. Sikder, J. Taheri, BB. Zhou, and AY. Zomaya, "DomNet: protein domain boundary prediction using enhanced general regression network and new profiles", IEEE transactions on nanobioscience, vol. 7(2), pp:172-81, Jun 2008.

[58] T. Ebina, H. Toh, and Y. Kuroda, "DROP: an SVM domain linker predictor trained with optimal features selected by random forest", Bioinformatics, vol. 27(4), pp: 487-94, Dec 2010.

[59] T. Ebina, H. Toh, and Y. Kuroda, "Loop-length-dependent SVM prediction of domain linkers for high-throughput structural proteomics", Peptide Science, vol. 92(1), pp:1-8, Jan 2009.

[60] JM. Thornton, AE. Todd, D. Milburn, N. Borkakoti, and CA. Orengo, "From structure to function: approaches and limitations", Nature Structural & Molecular Biology, vol.7, pp: 991-994, Nov 2000.

[61] CL. Mills, PJ. Beuning, and MJ. Ondrechen, "Biochemical functional predictions for protein structures of unknown or uncertain function", Computational and structural biotechnology journal, vol.13, pp: 182-91, Dec 2015.

[62] OC. Redfern, B. Dessailly B, and CA. Orengo, "Exploring the structure and function paradigm", Current opinion in structural biology, vol.18(3), pp:394-402, Jun 2008.

[63] WL. Bragg, DC. Phillips, and H. Lipson, "Development of X-ray analysis", G. Bell, 1975.

[64] K. Wüthrich, "NMR with Proteins and Nucleic Acids", Europhysics News, vol. 17(1), pp:11-3, 1986.

[65] JR. González, DA. Pelta, and JL. Verdegay, "Solving Bioinformatics Problems by Soft Computing Techniques: Protein Structure Comparison as Example", Intel. Sys. and Tech., SCI, vol.217, pp:123-36, Jun 2009.

[66] A. Gaurav, V. Kumar, and D. Nigam, "New Applications of Soft Computing in Bioinformatics: A Review", Journal of Pure & Applied Science & Technology. vol.2(2), pp:12-22, Jul 2012.

[67] L. Lee, JL. Leopold, and RL. Frank, "Protein secondary structure prediction using BLAST and relaxed threshold rule induction from coverings", Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE Symposium on. IEEE, Jul 2011.

[68] J. Yang, R.Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER Suite: protein structure and function prediction", Nature methods, vol. 12(1), pp:7-8, Jan 2015.

[69] M. Golshani, "In silico design, cloning and high level expression of L7/L12-TOmp31 fusion protein of Brucella antigens", Research in pharmaceutical sciences, vol.10(5), pp:436-445, Sep 2015.

[70] N. Qian, TJ. Sejnowski, "Predicting the secondary structure of globular proteins using neural network models", journal of Molecular Biology, vol. 202(4), pp:865–884, Aug 1988.

[71] LJ. McGuffin, K. Bryson, DT. Jones, "The PSIPRED protein structure prediction server", Bioinformatics, vol.16, pp: 404–405, 2000.

[72] C. Cortes, and V. Vapnik, "Support vector networks", Machine Learning, vol.20, pp: 273- 293, 1995.

[73] S. Hua, and Z. Sun, "A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach", Journal of molecular biology, vol. 308(2), pp: 397-407, Apr 2001.

[74] T. Dandekar, and P. Argos, "Potential of genetic algorithms in protein folding and protein engineering simulations", Protein Engineering, Design and Selection vol.5(7), pp:637-45, Oct 1992.

[75] RO. Day, JB. Zydallis, GB. Lamont, and R. Pachter, "Solving the protein structure prediction problem through a

multiobjective genetic algorithm", Nanotechnology. Vol.2, pp:32-5, Apr 2002.

[76] N. Thang, Bui, and Gnanasekaran Sundarraj, "An Efficient Genetic Algorithm for Predicting Protein Tertiary Structures in the 2D HP Model", GECCO '05 Proceedings of the 7th annual conference on Genetic and Evolutionary computation, pp. 385-392, 2005.

[77] J. Zhang, "Collagen-targeting vascular endothelial growth factor improves cardiac performance after myocardial infarction", Circulation, vol. 119(13), pp:1776-84, Apr 2009.