Available online at www.derpharmachemica.com

ISSN 0975-413X
CODEN (USA): PCHHAX

Der Pharma Chemica, 2016, 8(22):22-26
(http://www.derpharmachemica.com/archive.html)

Theoretically Analysis of Fusion Protein, Sequence and Folding

Mona Al-Terehi*, Ali H Al-Saadi, Haider K Zaidan, Abbas Al-Shariaf

College of Science, University of Babylon, Iraq

ABSTRACT

Fusion protein strategies was one of the important technologies in biotechnology fields, thus the present study introduced novel analysis for fusion protein sequence and folding using theoretically design GFP and RFP used for constructed fusion protein in mammalian expression vector, the results show there were three types of protein possible appearances resulted from DNA sequence translated these types were, unmolded, Fluorescent proteins, luminescent protein.

Keywords: Unmolded, Fluorescent proteins, Luminescent protein, Fusion protein

INTRODUCTION

The strategy of co-expression gene or dicistronic genes is fusion protein, the fusion protein mean that tow genes translate as one polypeptide [1], this feature has been used in expression vectors using reporter gene, the artificial plasmids that expressed in host cell used reporter gene such as GFP, RFP, YFP and BFP used fusion protein strategy to insert interested protein gene, all vectors that produced by scientific company used fusion protein gene in C-termini and N-termini, this mean that the target protein can fused in Amino termini or Carbone termini of reporter gene according to site of multiple cloning site. The purpose of fusion or manufactures chimeric protein is increasing solubility of protein, made its purification easier and different application reported in biotechnology such as immunoassay and antibody production also in bio functional enzyme.

The previous study used fusion protein strategy for constitute multiple domin vectors, also it characterized cassette vector that based on four specific restriction site. Fusion protein of target protein which used in researches implied data such as steady state distribution of protein, dynamic, kinetic and association with other protein [2].

The fusion protein design was effected on the localization and target protein thus investigator suggested tow strategies for fusion protein by study proteins behavior and its environment in the cells when functions and targeting domains was un known, the tow strategy based on the loci of interested protein with fusion protein, the first was fused target protein at the NH_2-terminus and second was at COOH-terminus which effected on protein surface after its folding rather than burned in core of protein [3]. Fusion protein strategy was developed by researchers in order to provide flexibility between FP and target protein, a small linker consist of 2-10 amino acid such as glycine interspersed with serine residue was added between two proteins, also the linker can help enhanced suitable folding and functioning of target protein, this linker can be added by PCR to cDNA [4]. Also in the early studies Kozak [5] pointed that fusion protein and target protein frame sequence must be having unambiguous methionine within kozak sequence.

In the other hand when function of protein assayed and its domain was identified the fusion protein can be inserted in optimal position which not interference with target protein or effected on protein folding, this strategy help researchers to study the target protein features, Martoglio, [6] explained that the fusion protein must be placed after the signal sequence, which important for targeting and translocation of the nascent peptide into the lumen of the endoplasmic reticulum, the functional domino of the interested protein near the COOH-terminus the FP placed toward the COOH-terminus, but it not necessarily at the absolute terminus if the functional domain of the protein was toward the NH_2-terminus.

A linker that must be added to fusion protein is importance to increased fabricate stable and bioactive fusion proteins, linkers were designed in three categories according to their structures: Flexible linkers, rigid linkers, and *in vivo* cleavable linkers contain restriction site sequence, linkers have many other advantages for the production of fusion proteins, like improving biological activity, increasing expression yield, and achieving desirable pharmacokinetic profiles (Figure 1) [7].

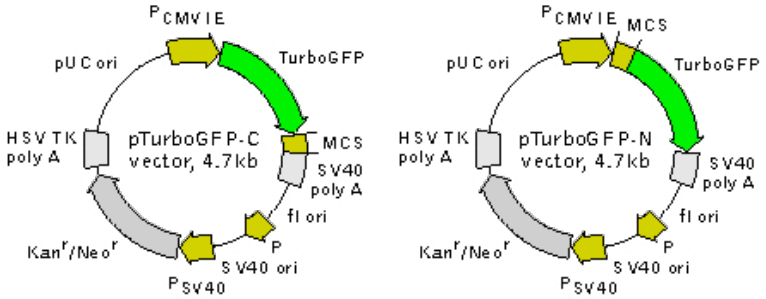

Figure 1: p-Turbo GFP C and N vector from Evrogen company

Fused two proteins by linked it performed using peptide linker the choosing of linker was suggested by studies that flexibility and hydrophobicity of linker was very important for protein function and folding [8] Arai and Ueda [9] used linker consist of helix forming peptide linkers between two florescence protein A (EAAAK)n A for control on distance between domain, the result show that the helical linker have ability to control the distance and decrease interference between domains. Engineering of linker design predict more flexible and proper function of fusion protein, the previous study suggested that the helical linker can separate fusion protein domine and the controlled of distance performed by change the repetition motif of EAAAK motif, so the linker of mltidomaine protein is very important because it effect on domain orientation and the distance between its.

The factors that effect on fusion protein biological activity were mention by Zhang et al. [10] and others the position of derived molecules which determined by active domain and the structure of domain receptors; also the size and complexity of linkers can effect on fusion protein activity in addition of others factor such as the structure.

Linker which used in antibody fusion protein manifacturing was important for correct folding and corresponding biological activity, many studies explained the different linker design effected on the fusion protein expression, they found that linker can effect on binding between fusion protein binding activity and instability *in vivo* [11].

The size of linker also effect on fusion protein, researchers improved that the longer linker is better for protection the folding and biological activity and can be cleaves by cellular protease, these features promote investigators to design artificial antibody for expression it in a large scale. Yan and others successfully to produced anti-human colectral cancer bivalent single chain with linker consist of GGGGS, they found that this design was highly expressed in *E. coli* and have good immune activity. Some of these linkers were added as a sequence at DNA level in vectors design [9].

Subjects and software's for protein sequence and folding analysis

1. Fusion protein construed theoretically by GFP and RFP which theoretically used to Construction Expression Vector Consist of Two Reporter Gene as a Fusion Protein by Al-Terehi et al. [12].
2. All open reading frame which creation in the DI was translated to protein using Addgene site https://www.addgene.org/analyze-sequence/, then it folding using protein folding software http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index.

RESULTS

According to DNA sequence there are three open reading frame in DI vector, according to Addgene analysis sequence the three open reading frames were translated and transcription to the following proteins.

1. The first open reading frame, its translated to the amino acid in Table 1 and the secondary structure in Figure 2 according to Phyren[2] software, also this protein is unmolded.

Table 1: The first open reading frame translation

AA only	With DNA	
1	MADYDQLSRS GGSRARGTVD CRIRSLSSRS ESGTSSTFTS ITSATSRVSG	50
51	HSHLVLLDVG LIVGLLDSFQ SVVHVVDAGH LEVRSGFLGS VCGLKVADQV	10
10	APAHELQGHV TCAFQAAVSG VHRLGGGLPA ECFLLHHRAV GWEVHPSNL	0
1	D	15
15	VVDEAAVLEA GVLGSGQHAP VFVCGDSLPC EALREGLLKE VGDARRVLDE	0
1	GSAAVHEAGS QDVEGEGERA ALDDLDSHGL GALVGLAFAL GCALEVVVV	20
20	H	0
1	GALHVQLHGH VLLNQLAHHG GDR*	25
25	-	0
1	-	-

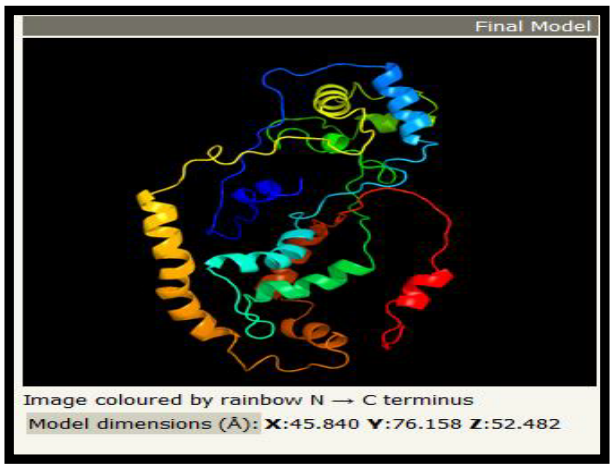

Figure 2: The structure of the first protein according to phyren², 0% of residues modeled at >90% confidence

2. The second open reading frame was translated to the amino acid in Tables 2 and 3 and the secondary structure in the Figure 3.

Table 2: The amino acid sequence of the second open reading frame ORF 16-804 (263 aa)

AA only	With DNA	
1	MVSELIKENM PMKLYMEGTV NNHHFKCTSE GEGKPYEGTQ TMRIKVVEG	50
51	G	10
10	PLPFAFDILA TSFMYGSRTF IKHPPGIPDF FKQSFPEGFT WERVTTYEDG	0
1	GVLTATQDTS LQDGCLIYNV KVRGVNFPAN GPVMQKKTLG WEASTETMY	15
15	P	0
1	ADGGLEGACD MALKLVGGGH LICNLETTYR SKKPATNLKM PGVYNVDHR	20
20	L	0
1	ERIKEADDET YVEQHEVAVA RYSTGGAGDG GKGGGGSGLR SRAQASNSAV	25
25	DGTAGPGSTG SR*	0
1	-	-

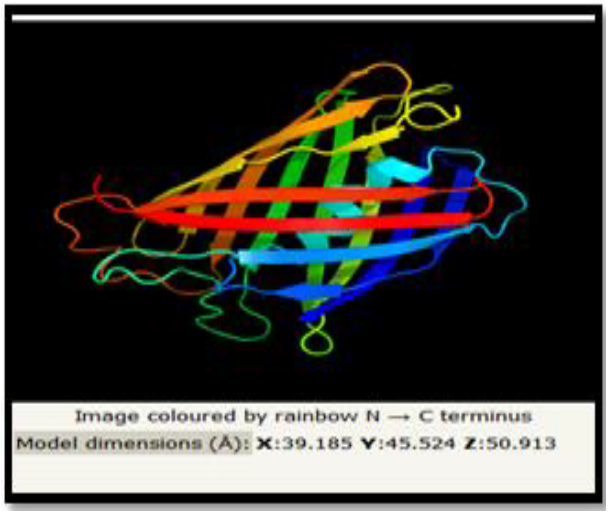

Figure 3: The secondary structure of protein translated from the second open reading frame

3. The third open reading frame was translated to the amino acid sequences in the Table 4 and the secondary structure in Figure 4.

Table 3: The second protein features according to Phyren² software

Top template information			
Fold:GFP-like Superfamily:GFP-like Family: Fluorescent proteins			
Confidence and coverage			
Confidence:	**100.0%**	**Coverage:**	**84%**
219 residues (84% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.			

Table 4: Amino acid sequence of the third open reading frame ORF 1-699 (233 aa)

1	MESDESGLPA MEIECRITGT LNGVEFELVG GGEGTPEQGR MTNKMKSTKG	50
51	ALTFSPYLLS HVMGYGFYHF GTYPSGYENP FLHAINNGGY TNTRIEKYED	100
101	GGVLHVSFSY RYEAGRVIGD FKVMGTGFPE DSVIFTDKII RSNATVEHLH	150
151	PMGDNDLDGS FTRTFSLRDG GYYSSVVDSH MHFKSAIHPS ILQNGGPMFA	200
201	FRRVEEDHSN TELGIVEYQH AFKTPDADAG EE*	

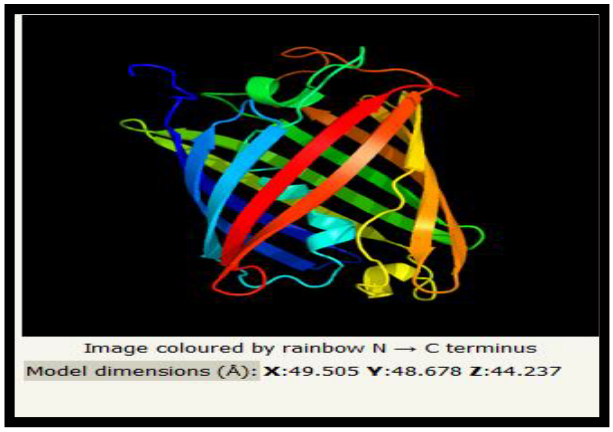

Figure 4: The secondary structure of protein translated from the third open reading frame

DISCUSSION

From the three open reading frame resulted from fusion GFP-RFP, three type of protein was detected theoretically using bioinformatics translation and its folding using online Phyre[2] software, these protein have different properties and when its alignment with database it was GFP, luminescence protein and unmolded pattern as show in Figures 1-3 and Tables 1-5, the proteomics need further analysis such as protein size (kDa) and detection it's in practical approaches, indeed, fusion protein technology used in different application and using bioinformatics design have been applied in drugs and therapeutic protein linker design also important in proteomic analysis for determinate its role in fusion protein, also this is benefit to determination natural rules of duplication hybridization, and biological activity of proteins [13].

Table 5: The third protein feature according to Phyren[2], translated from third open reading frame

PDB header: luminescent protein Chain: B: PDB Molecule: green fluorescent protein 2; PDB Title: the 2.1a crystal structure of copgfp			
Confidence and coverage			
Confidence:	100.0%	Coverage:	91%
211 residues (91% of your sequence) have been modeled with 100.0% confidence by the single highest scoring template.			

Fusion protein in present study was suggested to express two reporter genes at the same time and in the same levels [14] used new strategy to creation basic cassettes for protein targeting to cellular organelles and for purification by tagging it.

The most important applications of fusion protein are in pharmacology for produced antibody and recombinant protein therapy, some of therapeutic protein drugs were produced as fusion protein such as Etanercept (Enbrel) which used Ankylosing spondylitis, juvenile rheumatoid arthritis, plaque psoriasis, psoriatic arthritis and rheumatoid arthritis it composed of p75 TNFαR and IgG1 Fc [15], Romiplostim (Nplate) for treated Immune thrombocytopenic purpura it composed from Peptide and IgG1 Fc, Rilonacept (Arcalyst) for Cryopyrin-associated periodic syndromes Ligand-binding domains of IL-1 receptor and IL-1 receptor accessory protein (IL-1RAcP)and gG1 Fc [16].

REFERENCES

[1] S.R. Edwards, T.J. Wandless, *Yeast (Chichester, England)*, **2010**, 27, 4, 229-236.

[2] J. Lippincott-Schwartz, N. Altan-Bonnet, G.H Patterson, *Nat. Cell. Biol.,* **2003**, 5,7-14

[3] S. Hovmöller, T. Zhou, *Proteins.,* **2004,** 55, 219-222.

[4] A. Miyawaki, A. Sawano, T. Kogure, *Nat. Cell. Biol.,* **2003,** 5, S1-S7.

[5] M. Kozak M. *Annu. Rev. Cell. Biol.,* **2003,** 8, 197-225.

[6] B. Martoglio, B. Dobberstein, *Cell. Biol.,* **1998,** 8, 410-415.

[7] X. Chen, L. Jennica, W. Zaro, *Adv Drug Del Rev.,* **2013,** 10, 1357-1369.

[8] C.J Crasto, J.A Feng, *Protein. Eng.,* **2000**, 13, 309-312.

[9] R. Arai, H. Ueda, A. Kitayama, *Protein. Eng.,* **2011,** 14, 529-532.
[10] J. Zhang, J. Yun, Z. Shang, X. Zhang, B. Pan, *Progress. Nat. Sci.,* **2009,** 1197-1202.
[11] M. Fang, X. Jiang, Z. Yang, *Chin. Sci. Bull.*, **2003,** 14, 1912.
[12] M. Al-Terehi, A. Al-Saadi, H. Zaidan, A. Al-shariafe, *Australian J Basic Applied Sci.,* **2015,** 9, 7 586-595.
[13] J. Zhang, J. Yun, Z. Shang, X. Zhang, B. Pan, *Prog. Nat. Sci.,* 19, **2009,** 1197-1202.
[14] K. Truong, M. Ikura, *Curr. Opin. Struct. Biol.,* **2001,** 11, 573-578.
[15] E. Ducharme, J.M Weinberg, *Expert. Opin. Biol. Ther.,* **2008,** 8, 491-502.
[16] D.B Cines, U. Yasothan, P. Kirkpatrick, *Nat. Rev. Drug. Dis.,* **2008,** 7, 11, 887-888.