REVIEW ARTICLE

An *in silico* overview on the usefulness of tags and linkers in plant molecular pharming

Abdullah Makhzoum[1] ✉, Shifa Tahir[2], Marjorie Elizabeth Osborn Locke[3], Jocelyne Trémouillaux-Guiller[4] and Kathleen Hefferon[5]

Abstract

Plant molecular pharming is a promising concept based on the large-scale production of recombinant proteins encompassing antibodies, vaccines, enzymes and peptides for human or veterinary uses and treatments. This new branch of the biopharmaceutical industry offers practical and safety advantages over other traditional production systems. In higher plants, the complex cellular machinery makes possible synthesis and posttranslational modifications of heterologous protein macromolecules. The limiting obstacle to using this system at an industrial scale is most often the low yield of the recombinant proteins produced in host plants. To improve production levels, many studies have been focusing on the choice of plant species, tissues, organs and cell suspension cultures and/or various upstream and downstream constituents in the expression cassette. New engineering technologies in plant molecular pharming have emerged that rely on the usefulness of using soybean agglutinin (SBA), hydrophobin, zein and elastin-like peptide tags which are employed to extract and purify recombinant proteins in some host systems and are under the control of different expression cassettes. Known to be very useful tools in recombinant proteins, linkers separate different domains or units of the heterologous gene and thereby retain the functionality of the protein of interest. Here, we compare computationally one tag SBA as a part of fusion with a pharmaceutical human protein ADA joint directly or by the specific flexible (GGGGS)3 linker. The *in silico* analysis focuses on the mRNA stability and fusions of tagged and tagged-linked ADA recombinant proteins.

Keywords: plant molecular pharming; adenosine deaminase (ADA); soybean agglutinin (SBA); linkers; tags; protein fusions

AUTHOR'S AFFILIATION
1. Department of Biology, The University of Western Ontario, London, ON N6A 5B7, Canada
2. Quaid-i-Azam University, National Center for Bioinformatics Pakistan
3. Department of Computer Science, The University of Western Ontario, London, ON N6A 5B7, Canada
4. Université F. Rabelais, Tours, France
5. Cornell University, Department of Cell Biology and Genetics, United States

CORRESPONDENCE
✉ Dr. Abdullah Makhzoum, e-mail: E-mail: abmakhzoum@gmail.com

Introduction

Characteristics of plant molecular pharming

Genetically engineered plants have emerged as one of most attractive platforms for producing therapeutic recombinant proteins such as antibodies, vaccines, enzymes, and other pharmaceutical entities, such as human alpha-1-antitrypsin or human serum albumin, necessary for the improvement of veterinary and human healthcare systems throughout the world (He *et al.*, 2011; Hefferon, 2012; Hefferon, 2014; Ou *et al.*, 2014; Paul, Teh, Twyman, & Ma, 2013; Zhang *et al.*, 2012). Plants offer scalability, low cost and safety, as they are devoid of mammalian pathogens. Moreover, plant cells have the capacity to correctly assemble and fold complex proteins by providing post-translational modifications indispensable to the functionality and stability of foreign multimeric proteins (Fu *et al.*, 2014; Ma *et al.*, 2013; Stoger, Fischer, Moloney, & Ma, 2014). One drawback to the production of pharmaceutical proteins in plants is the relatively low yield of protein products that are generated. However, maize, wheat, barley and rice are the most important cereals capable of producing recombinant proteins at high levels ranging from 7 to 10% of total soluble protein in seeds (Yang, Wakasa, & Takaiwa, 2008).

He and collaborators indicated that the expression levels of human serum albumin (HSA) in transgenic rice seeds is high enough for downstream and commercialized processing (He *et al.*, 2011; Kuo *et al.*, 2013). The rice endosperm is capable of accumulating pharmaceutical proteins such as human alpha-1-antitrypsin, and thus appears to be a promising production platform for high-value plant-made pharmaceuticals (Zhang *et al.*, 2012).

Plants also offer many host species and culture modes, the choice of which, as well as the subcellular compartmentalization of the expected protein, is crucial, as no single system will be optimal for the expression of all recombinant proteins (Paul *et al.*, 2013). Several different systems have been investigated, including whole plants, seeds, hairy roots, cell suspension cultures and microalgae, or still specific cellular organelles such as chloroplasts, in order to improve recombinant protein levels (Del *et al.*, 2012; Makhzoum, Benyammi, Moustafa, & Tremouillaux-Guiller, 2013; Makhzoum, Sharma, Bernards, & Trémouillaux-Guiller, 2013; Ou *et al.*, 2014; Sabalza, Vamvaka, Christou, & Capell, 2013; Skarjinskaia *et al.*, 2013; Specht & Mayfield, 2014).

Other approaches used to improve yield production

To increase the level of targeted proteins in the host plant, scientists have been employing other physiological, biochemical and molecular approaches (Matic *et al.*, 2012). At the gene-transfer-system level, some interesting assays have focused on using bacterial plasmids containing expression cassettes, while others exploit viral vector systems to express heterologous genes and their proteins (Hefferon, 2014; Hefferon, 2012). Foreign genes can be expressed transiently using specific vectors and/or vacuum infiltration-centrifugation based on the stable integration of the gene into the chromatin structure (Kingsbury & McDonald, 2014; Shah, Almaghrabi, & Bohlmann, 2013). For example, in a study conducted by Chen and his co-workers on the factors controlling biosynthesis and accumulation of the wild and mutated human immune-regulatory interleukin-10 transformed in *Arabidopsis*, mutagenized transcripts (at 2762 and at 3262) demonstrated similar nascent mRNA synthesis but improved stability. Likewise, the mutated transcripts from plant lines with a more stable and efficient translational machinery than the wild IL-10 lines (Chen *et al.*, 2013).

Innovative genomic and epigenetic approaches in plant molecular pharming have recently emerged. Some genome re-engineering and epigenetic approaches can have tremendous potential through the suppression or decrease of protease activities. For example, RNA interference is capable of shifting the metabolism of a given host plant and can repress the silencers that inhibit the expression of transiently transformed genes (Arzola, Chen, Rattanaporn, Maclean, & McDonald, 2011; Hakkinen *et al.*, 2013; Rigano,

De Guzman, Walmsley, Frusciante, & Barone, 2013; Tremblay, Diao, Huner, Jevnikar, & Ma, 2011).

Re-engineering the expression cassette

The yield of plant-made pharmaceuticals (PMPs) is in general low. To overcome this main limiting factor, many studies have been conducted to improve the production of therapeutic proteins by reengineering the expressed genes and proteins at upstream and downstream processing levels (Makhzoum *et al.*, 2013; Twyman, Schillberg, & Fischer, 2013). Many promoters, codon-optimized ORFs, 5'UTR, 3'UTR and peptides have been tested to enhance the transcription and translation levels and therefore the yield production of recombinant proteins (Buyel, Kaever, Buyel, & Fischer, 2013; Gallie, 2002; Kanoria & Burma, 2012; Laguia-Becher *et al.*, 2010). Promoter studies are very useful for increasing the expression and accumulation of the heterologous pharmaceutical recombinant proteins, such as the strong and ubiquitous 35S CaMV promoter, or for targeting spatiotemporal expression features by using promoters specific to certain cells, tissues or organs (Abdullah, Rahmah, Sinskey, & Rha, 2008; Makhzoum *et al.*, 2013; Makhzoum, Petit-Paly, St Pierre, & Bernards, 2011).

Gene expression can also be enhanced through the 5' UTR regions. In addition to well known translational enhancers, including the viral 5' UTRs Ω and AMV for transgene expression in plant systems (especially in dicots such as tobacco) (Gallie, 2002), 5' UTR sequences from the photosystem I (PHOTO) gene and the geranyl geranyl reductase (GGR) gene in the model plant *A. thaliana*, have shown a high level of GUS expression in comparison to 5' UTR of Ω and AMV. In contrast, the alcohol dehydrogenase gene (NtADH) 5' UTR revealed similar levels of GUS expression, nonetheless, this was 30 to 100-fold level greater than the control (Agarwal *et al.*, 2014; Satoh, Kato, & Shinmyo, 2004). As another example, a new synthetic (20 nt) 5' UTR enabled GUS and GFP reporter gene expression to reach levels 10 to 50-fold higher than controls in transgenic tobacco and cotton plants. Moreover, its promoting role was confirmed either under the control of the strong 35S promoter or the weak nos promoter (Kanoria & Burma, 2012). In addition to improving promoter strength, codon adaptation, 5'UTR and 3'UTR choice in expression cassettes, and other constituents such as signal peptide tags have been utilized to target accumulation to specific organelles in plant cells. For example, inclusion of the KDEL endoplasmic reticulum retention signal will cause the foreign protein of interest to accumulate in the endoplasmic reticulum, where it is protected against protease activity present in the cytosol (Bundo *et al.*, 2014).

Targeting recombinant proteins to specific cellular compartments is also an important parameter that affects plant-made pharmaceutical expression levels. This

targeting influences post-translational modifications playing a key role on the biological activity and accumulation of recombinant proteins in plants (Warzecha, 2008). For example the endosperm of rice provides an ideal site offering correct folding and adequate protection for the foreign protein against proteolysis, without any detectable loss of activity (Rademacher *et al.*, 2008).

Fusion tags in plant molecular pharming

Short peptide tags can be fused to constructs harbouring foreign genes in order to facilitate recovery during protein purification. Some short affinity tags (His tag, Stripe tag, Argo tag, and FLAG tag) have been successfully used in bacterial and yeast biosynthesis and purification systems (Agarwal *et al.*, 2014; Zhao, Li, & Liang, 2013). However, the small tags in plant molecular pharming are characterized by their ineffectiveness, lack of scalability and high cost (Joensuu *et al.*, 2010; Waugh, 2005). Some fusion tags have increasingly been gaining more attention as a new, effective way of facilitating the expression, extraction and purification of pharmaceutical proteins, such as hydrophobin (Joensuu *et al.*, 2010; Joensuu, Conley, Linder, & Menassa, 2012), zein domain (Alvarez, Topal, Martin, & Cardineau, 2010), elastin-like polypeptides (Conley, Joensuu, Richman, & Menassa, 2011; Floss *et al.*, 2009) and soybean agglutinin (SBA). The SBA affinity tag is an effective system which can be used for quick and efficient purification of recombinant proteins (Tremblay, *et al.* 2011). Elastin-like polypeptides (ELP) were employed by many researchers in recent studies to increase the stability and purification yield of human immunodeficiency virus (HIV)-neutralizing antibodies 2G12 (light or heavy chain), transformed and expressed in transgenic tobacco plants (with 1% total soluble proteins (TSP) from leaves and seeds) (Conley, Joensuu, Jevnikar, Menassa, & Brandle, 2009; Floss *et al.*, 2009). Another interesting tag is the hydrophobin HFBI gene from *Trichoderma reesei* (hydrophobicity changer) which was used as part of a fusion with green fluorescent protein (GFP) (GFP- HFBI) and accumulated into protein bodies in transiently agro-infiltrated tobacco, with a high yield of 51% of TSP and a 91% recovery of GFP-HFBI fusion protein (Joensuu *et al.*, 2010; Joensuu *et al.*, 2012). The Zera® domain, (N-terminal proline-rich domain: γZein ER-accumulating domain) of maize Zein gene is able to form PB bodies. This domain, in addition to other constituents, led to a threefold increase in the *Yersinia pestis* F1-V antigen fusion protein accumulation when using the fusion construct, in comparison to the construct lacking it, in three separate transient or stable transformation systems in *Nicotiana benthamiana*, *Medicago sativa* (alfalfa) and *Nicotiana tabacum* NT1 cells (Alvarez *et al.*, 2010). To facilitate the removal of these tags after purification, the recognition sequence of the

protease TEV (the sequence Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser)(ENLYFQ(G/S)) can be added between the protein of interest and the tag (Kapust & Waugh, 2000; Tropea, Cherry, & Waugh, 2009; Waugh, 2011).

Linkers in plant molecular pharming

Linkers are short peptide sequences that occur between protein domains. An important consideration in recombinant protein production is the use of linker(s) between two fusion genes instead of direct fusion, and is showing promise in maintaining protein functionality and activity while avoiding interference from other connected proteins. On the same expression cassette, linkers are very attractive tools for expressing some proteins, ORFs or peptides together without perturbing the structure of each unit of the fusion or having a negative effect on their activity and stability. Linkers are often composed of flexible residues such as glycine and serine so that the adjacent protein domains are free to move relative to one another. Longer linkers are used when it is necessary to ensure that two adjacent domains do not sterically interfere with one another. Some scientists have agreed that short linkers can have a negative effect on SCFV antibody folding due to spatial occupancy. Conversely, long linkers can influence the functionality and enhance the antigenicity of ScFv antibodies (Guo *et al.*, 2006; Le Gall, Reusch, Little, & Kipriyanov, 2004; J. Zhang, Yun, Shang, Zhang, & Pan, 2009).

Several short peptides have been used toward this goal. For example, the usefulness of the linker EAAAK to design a synthetic gene encoding a combination of the carboxy-terminal fragment of intimin, the middle region of Tir and the carboxy-terminal part of EspA of Enterohemorrhagic *Escherichia coli* O157:H7 (EHEC) has been demonstrated, reaching 0.2% to 0.33% TSP in transgenic canola seeds, and 0.1% to 0.3% TSP in transgenic tobacco lines (Amani, Mousavi, Rafati, & Salmanian, 2009, 2011). Another example is the linker AG, a simple alanine-glycine linker of six amino acids with three AG repeats (AGAGAG). When added between LT-B and GFP (green fluorescent protein),using oligonucleotide extensions in the PCR (polymerase chain reaction) primers as part of the expression cassette in the vectors pLM03, pLM08, pLM09 with the percentage of TSP of LT-B::GFP being line dependent, and the accumulation went from undetectable to 0.059 extracted in maize kernels (Moeller, Gan, & Wang, 2009). Modified linkers can improve the level of protein accumulation in the case of the FV single chain linked to the CH3 of a human anti-rat transfer in receptor IgG3 heavy chain by the flexible and optimized linker (GGGGS)3, and transformed stably into mammalian cells (Trinh, Gurbaxani, Morrison, & Seyfzadeh, 2004). Linkers have also been previously designed using a genetic algorithm, taking into consideration their lengths, flexibilities and compositions (Zhang *et al.*, 2009). For

more examples about using linkers as part of the expression cassette, Soria *et al.* mentioned several used to express chimeric proteins which are comprised of subunit vaccines produced in transgenic plants (Soria-Guerra, Moreno-Fierros, & Rosales-Mendoza, 2011).

In silico and computational analysis and optimization

In molecular pharming as well as many other fields in applied molecular biology, gene design and optimization is a very useful strategy to enhance protein production yield. In order to estimate the quantity, quality and function of recombinant proteins, it is necessary to predict gene structure optimization, mRNA and protein structure and function. For these purposes, gene and protein design plays an important role in the quality and activity of production yield. Indeed, by using computational tools we can study gene structure, mRNA and the activities of proteins, as well as predict the yield production of the proteins.

Software such as Visual Gene Developer may help to optimize the study of pharmaceutical recombinant genes and proteins by analyzing synthetic genes and testing new algorithms in bioinformatics (Jung & McDonald, 2011). The software can be found and downloaded from http://www.visualgenedeveloper.net.

In silico analysis is a very good tool for the design of new drugs and for studying existing vaccines and antibodies, as

Table 1. Summarize some linkers have used in recombinant proteins and have a good potential to be used in plant molecular pharming. The linkers below are generated from the International Genetically Engineered Machine (iGEM) Foundation website.

| Name	Description	Sequence	Length (bp)	Reference/designer
BBa_J176131	PLrigid	(gaagctgctgcaagagaagctgcagctaggg Aggctgcagctagggaggctgctgcaaga)	60	(Merutka, Shalongo, & Stellwagen, 1991; Sommese, Sivaramakrishnan, Baldwin, & Spudich, 2010; Yan, Imanishi, Futaki, & Sugiura, 2007)/ Designed by: Karmella Haynes Group:Haynes_Lab (2011-10-28)
BBa_J18920	2aa GS linker	(ggcagc)	6	(Grunberg, Ferrar, van der Sloot, Constante, & Serrano, 2010)
BBa_J18921	6aa [GS]x linker	(ggtagcggcagcggtagc)	18	(Grunberg *et al.*, 2010)
BBa_J18922	10aa [GS]x linker	(ggtagcggcagcggtagcggtagcggcagc)	30	(Grunberg *et al.*, 2010)
BBa_K105012	10 aa flexible protein domain linker	(ggtgaaaatttgtattttcaatctggtggt)	30	Manuel Gersbacher, Katja Karstens Group: iGEM08_ESBS-Strasbourg (2008-10-19)
BBa_K133132	8 aa protein domain linker	(tccgcttgttactgtgagctttcc)	24	Jan Lonzaric Group: iGEM08_Slovenia (2008-10-24)
BBa_K157009	Split fluorophore linker; Freiburg standard	(cgaccagcctgtaagattccaaatgacctgaa gcagaaagttatgaatcac)	51	(Hu, Chinenov, & Kerppola, 2002; Hu & Kerppola, 2003; Kerppola, 2006)/ iGEM08_Freiburg (2008-10-26)
BBa_K157013	15 aa flexible glycine-serine protein domain linker; Freiburg standard	(ggtggaggaggttctggaggcggtggaagtgg tggcggaggtagc)	45	Kristian Muller Group: iGEM08_Freiburg (2008-10-26)
BBa_K243004	Short Linker (Gly-Gly-Ser-Gly)	(ggtggttctggt)	12	iGEM09_Freiburg_bioware (2009-10-12)
BBa_K243005	Middle Linker (Gly-Gly-Ser-Gly)x2	(ggtggttctggtggtggttctggt)	24	iGEM09_Freiburg_bioware (2009-10-14)
BBa_K243006	Long Linker (Gly-Gly-Ser-Gly)x3	(ggtggttctggtggtggttctggtggtggttc tggt)	36	iGEM09_Freiburg_bioware (2009-10-14)
BBa_K243029	GSAT Linker	(ggtggttctgccggtggctccggttctggctc cagcggtggcagctctggtgcgtccggcacggg tactgcgggtggcactgccagcggttccggtac tg gctctggc)	108	iGEM09_Freiburg_bioware (2009-10-20)
BBa_K243030	SEG	(ggtggttctggcggcggttctggtgaaggtggcgg ctccgaaggcggcagcggcgagggcggtggtag cgaaggtggtggctccgaaggtggcggttccgg	108	iGEM09_Freiburg_bioware (2009-10-20)

BBa_K404300	SEG-Linker	(ggtggttctggcggcggttctgaaggtggcgg ctccgaaggcggcggcagcgagggcggtggtag cgaaggtggtggctccgagggtggcggttccgg cg gcggtagc)	108	iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K404301	GSAT-Linker	(ggtggttctgccggtggctccggttctggctc cagcggtggcagctctggtgcgtccggcacggg tactgcgggtggcactggcagcggttccggtac tg gctctggc)	108	iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K404303	Z-EGFR-1907_Short-Linker	(gtggataacaaatttaacaaagaaatgtgggc ggcgtgggaagaaattcgtaacctgccgaacct gaacggctggcagatgaccgcgtttattgcgag cc Tggtggatgatccgagccagagcgcgaacctgc tggcggaagcgaaaaaactgaacgatgcgcagg cgccgaaaaccggcggtggttctggt)	192	iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K404304	Z-EGFR-1907_Middle-Linker	(gtggataacaaatttaacaaagaaatgtgggc ggcgtgggaagaaattcgtaacctgccgaacct gaacggctggcagatgaccgcgtttattgcgag cc tggtggatgatccgagccagagcgcgaacctgc tggcggaagcgaaaaaactgaacgatgcgcagg cgccgaaaaccggcggtggttctggtggtggt c tggt)	204	(Friedman *et al.*, 2008; Gostring *et al.*, 2010)/ iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K404305	Z-EGFR-1907_Long-Linker	(gtggataacaaatttaacaaagaaatgtgggc ggcgtgggaagaaattcgtaacctgccgaacct gaacggctggcagatgaccgcgtttattgcgag cc tggtggatgatccgagccagagcgcgaacctgc tggcggaagcgaaaaaactgaacgatgcgcagg cgccgaaaaccggcggtggttctggtggtggt c tggtggtggttctggt)	216	(Colquhoun & Mellon, 2002; Hirsch *et al.*, 2003)/ iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K404306	Z-EGFR-1907_SEG-Linker	(gtggataacaaatttaacaaagaaatgtgggc ggcgtgggaagaaattcgtaacctgccgaacct gaacggctggcagatgaccgcgtttattgcgag cc tggtggatgatccgagccagagcgcgaacctgc tggcggaagcgaaaaaactgaacgatgcgcagg cgccgaaaaccggcggtggttctggcggcggtt c tgaaggtggcggctccgaaggcggcggcagcga gggcggtggtagcgaaggtggtggctccgaggg tggcggttccggcggcggtagc)	288	(Orlova *et al.*, 2007; Walker & Dearing, 1999)/ iGEM10_Freiburg_Bioware (2010-08-28)
BBa_K416001	(Gly4Ser)3 Flexible Peptide Linker	(ggtggaggaggctctggtggaggcggtagcgg aggcggagggtcg)	45	iGEM10_NYU (2010-05-29)
BBa_K648005	Short Fusion Protein Linker: GGSG with standard 25 prefix/suffix	(ggtggttctggt)	12	Jim Rose, Alex Bina, Ben Alouidor, Brian Avison Group: iGEM11_Penn_State (2011-07-03)
BBa_K648006	Long 10AA Fusion Protein Linker with Standard 25 Prefix/Suffix	(ggtgaaaatttgtattttcaatctggtggt)	30	Jim Rose, Alex Bina, Ben Alouidor, Brian Avison Group: iGEM11_Penn_State (2011-07-03)
BBa_K648007	Medium 6AA Fusion Protein Linker: GGSGGS with Standard 25 Prefix/Suffix	(ggaggttcaggaggcagc)	18	Jim Rose Group: iGEM11_Penn_State (2011-07-03)

it is necessary to optimize drug design by the simulation of spatial structures and molecules (Jiang & Zhou, 2005; Kamphausen *et al.*, 2002)

The design of suitable linkers to separate domains of biofunctional proteins is an indispensable tool in protein engineering (Arai, Ueda, Kitayama, Kamiya, & Nagamune, 2001). A few scientists have suggested and created a web server and software for this purpose, such as "Linker", a program to generate linkers for protein fusions (Crasto & Feng, 2000). Additionally, a more recent web server, also called LINKER, serves to generate peptide sequences with extended conformation. Table 1 summarizes some

potentially useful linkers employed in recombinant protein studies by many researchers. These linkers can be studied for their positive roles in pharmaceutical plant recombinant protein assays and for *in silico* optimization of pharmaceutical protein fusions.

Based on fusions with or without using linkers, it will be interesting to study and show that modified gene and protein structures are as stable and functional as their natural homologs. As an example, we demonstrate this feature using an interesting human enzyme of pharmaceutical interest such as human adenosine deaminase (ADA). Deficiency in ADA (41 kDa), a crucial
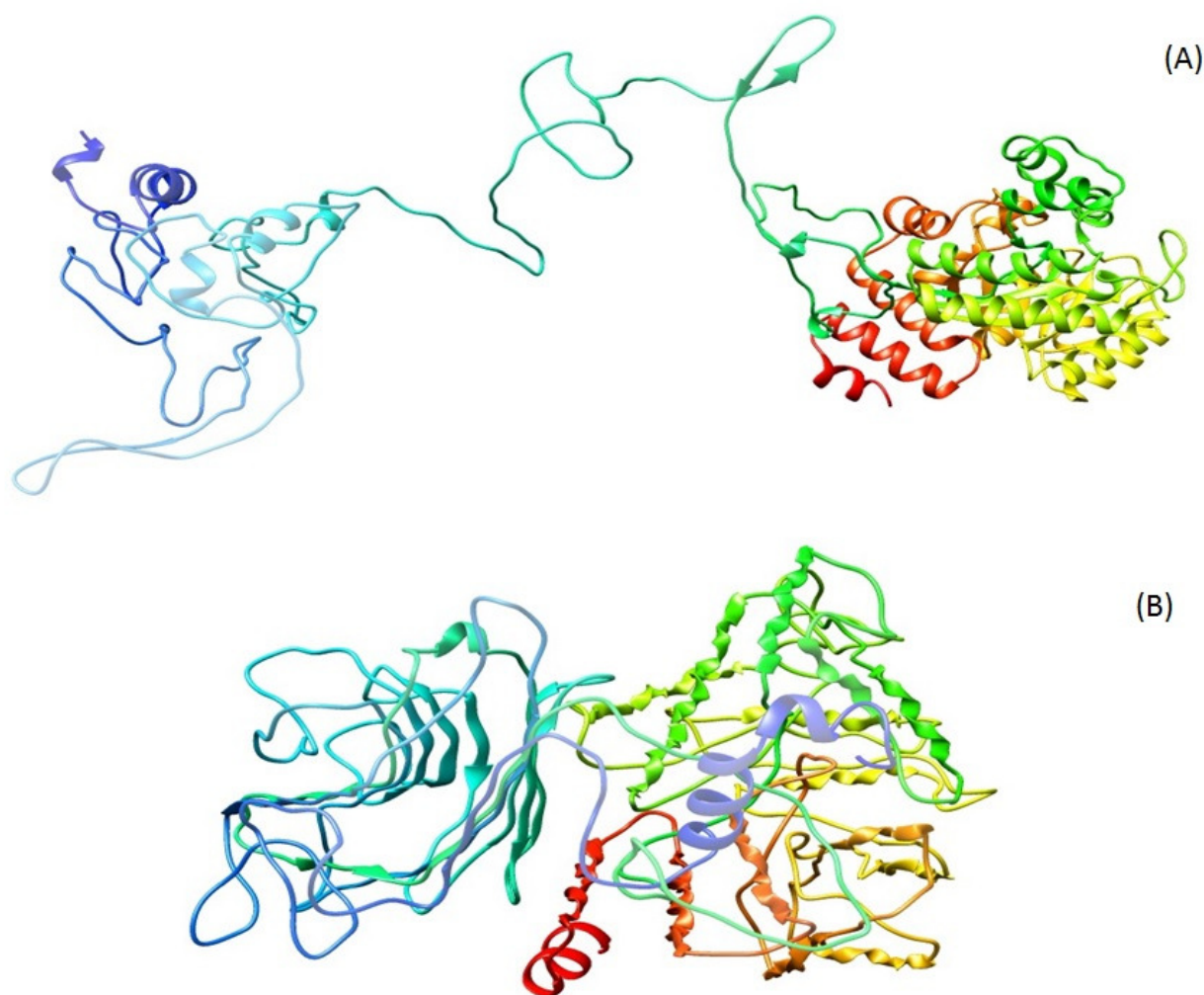
Fig. 1. 3D structures of Proteins. (A) SBA-linker-ADA (B) SBA-ADA

enzyme of the purine salvage pathway, causes a genetically inherited disorder of the immune system that can be treated by producing functional ADA treatment. This gene was empirically expressed and produced in some assays such as transgenic tobacco plants and transgenic tobacco BY-2 cell suspensions with specific activities of (0.001 and 0.003 units per mg (TSP) of leaves) and approximately 16 mg/L in transgenic tobacco BY-2 cell suspensions (Singhabahu, George, & Bringloe, 2013, 2014).

Therefore, as an interesting case study, we attempt to compare *in silico* modeling and design of ADA recombinant protein in plant molecular pharming with an SBA tag as a fusion with or without a specific linker. The goal of this paper is to show the useful role of linkers and tags in plant molecular pharming of recombinant proteins and gene fusions based on gene structure, mRNA and protein in comparison to proteins which lack tags and linkers. We compare computationally and predict some features leading to pharmaceutical protein yield production such as transcription (gene codon and structure optimization, mRNA frequency and stability) and translation (protein structure and function, stability and instability and post-translation).

Materials and Methods

Sequence retrieval:

DNA sequences of linker, SBA and ADA genes were obtained from NCBI website (http://www.ncbi.nlm.nih.gov/gene). To obtain fused proteins, SBA and ADA sequences were joined with and without linker. The resulting merged sequences were submitted to EXPAY Translate tool (http://web.expasy.org/translate/) to achieve protein sequences with all possible open reading frames.

Table **2**: Functional Characterization of SBA-ADA and SBA-linker-ADA proteins

Features	SBA-ADA	SBA-linker-ADA
Number of Amino Acids	648	664
Molecular weight	61674	72707.3
Theoretical pI	5.64	5.64
Total number of negatively charged residues (ASP+GLU)	77	77
Total number of positively charged residues (ARG+LYS)	62	62
Total number of atoms	10086	10214
Instability Index	35.63	36.68
Aliphatic Index	90	87.83
Grand Average Hydropathicity (GRAVY)	-0.221	-0.228
Ext. Coefficient	86540	86540

Table 3. PROCHECK values for SBA-ADA and SBA-linker-ADA proteins

Regions	SBA-ADA	SBA-linker-ADA
Residues in most favoured regions [A,B,L]	339 59.3%	407 70.7%
Residues in additional allowed regions [a,b,l,p]	170 29.7%	128 22.2%
Residues in generously allowed regions[~a,~b,~l,~p]	40 7.0%	28 4.9%
Residues in disallowed regions	23 4.0%	13 2.3%
Number of non-glycine and non-proline residues	572 100%	576 100%
Number of end-residues (excl, Gly and Pro)	2	2
Number of glycine residues (shown as triangles)	37	49
Number of proline residues	37	37
Total Number of Residues	648	664

Functional characterization:

Physico-chemical properties like molecular weight, theoretical IEP (isoelectric point), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) were calculated for both proteins using EXAPASY ProtParam server (http://web.expasy.org/protparam/).
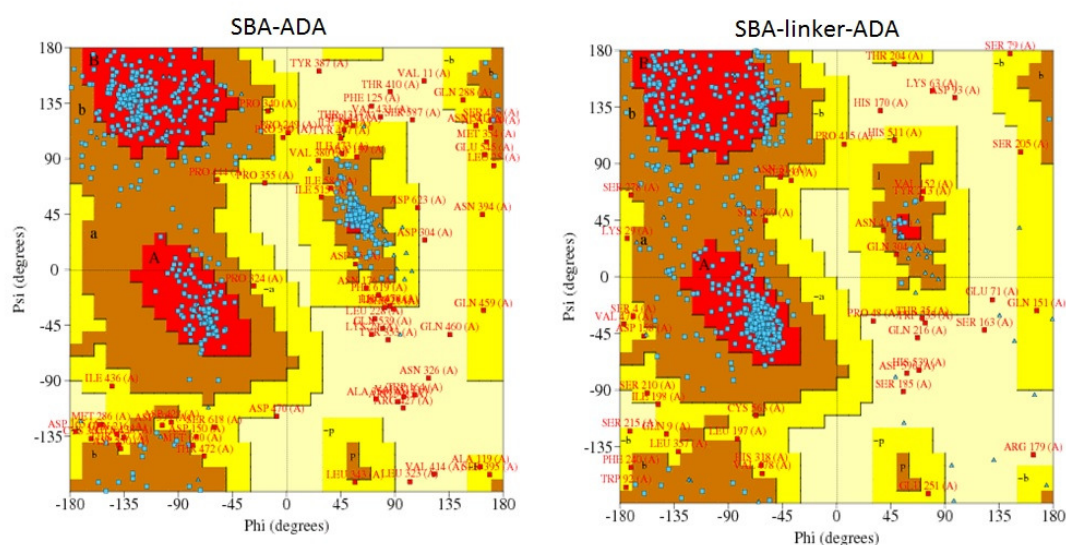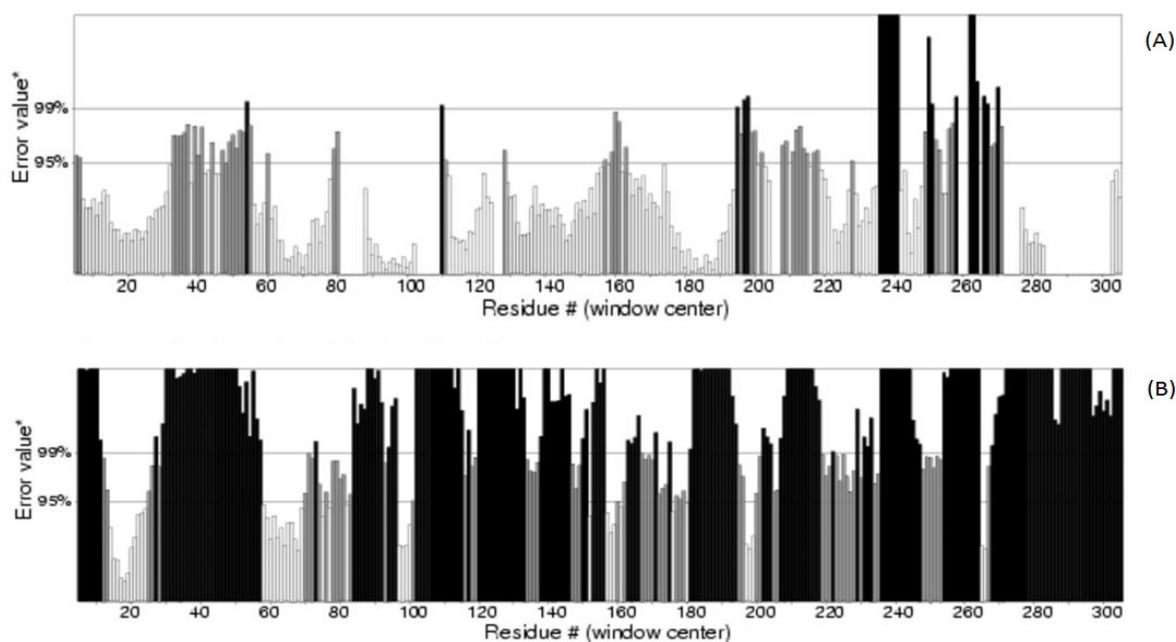
Fig. 2. Ramachandran Plots for SBA-ADA and SBA-linker-ADA

Fig. 3. ERRAT Graphs for Proteins. (A) SBA-linker-ADA (B) SBA-ADA

Structure prediction:

The mRNA structure was predicted using (http://www.bioinfo.rpi.edu/applications/mfold). The secondary structures for both the proteins were calculated using SOPMA (self-optimized prediction method) (Geourjon & Deleage, 1995) whereas 3D structures were obtained from I-TASSER (Zhang, 2008). Solvent accessibility of different residues was evaluated by DSSP and other online programs (VADAR) (http://redpoll.pharmacy.ualberta.ca/vadar/).

Structure validation:

The accurate structure assessment of model's quality is based on biological knowledge and results of analytical tools (Vaseeharan & Valli, 2011). For this purpose highly prominent tool PROCHECK (Laskowski, Watson, & Thornton, 2005) and ERRAT (Colovos & Yeates, 1993) were used.

Results and discussion

The proteins obtained after translation with maximum

amino acid sequences were selected. The functional characterization of both proteins is shown in Table 2. The hydrophilic nature of protein is associated with a low GRAVY value. The GRAVY value for a peptide or protein can be calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. The proteins which have large negative values indicated that these proteins had relatively more hydropathicity as compared to proteins those possess less negative values (Roy, Maheshwari, Chauhan, Sen, & Sharma, 2011). The instability index of both proteins was calculated less than 40 which classified them as stable molecules (Roy *et al.,* 2011) but the protein without a linker is slightly more stable as compared to the protein with a linker. Similarly, pI values of both proteins were determined to be less than 7 indicating an acidic nature. The aliphatic index for SBA-ADA is slightly higher than SBA-linker-ADA, which suggested that the SBA-ADA has a higher thermal stability (Gasteiger *et al.,* 2005; Wilkins et al., 1999). The secondary structure analysis of SBA-ADA and SBA-linker-ADA describes almost identical results. Both proteins have a high number of alpha helices, which is as expected because alpha helices are the most common structural motifs. The mRNA structure of SBA-linker-ADA protein shows a lower energy value of -564.10 kcal/mol, which is an attribute of structural stability, whereas the value for SBA-ADA is -536.10 kcal/mol.

The best predicted 3D structures for SBA-ADA and SBA-linker-ADA proteins are shown in Fig. 1. The structural stability for these proteins is validated through different servers. The PROCHECK analysis revealed that SBA-linker-ADA protein has a more stable structure than SBA-ADA (Fig. 2). The stereochemical properties, including Ramachandran plots, G-factor, and the number of bad contacts are illustrated in Table 3. The protein with the linker has more residues in allowed regions and fewer residues in disallowed regions, and similarly has a better G-factor value. The statistical analysis of non-bonded interactions between different atom types was performed using ERRAT 2.0, in a graphical form in Fig. 3. The ERRAT score for SBA-linker-ADA is 86.304 which indicates a better structure. The solvent accessibility for residues of both proteins indicates that SBA-linker-ADA has more accessible residues to solvent (Fig. 4).

Conclusions and perspectives

The *in silico* studies on SBA-ADA and SBA-linker-ADA proteins suggested that in terms of sequence analysis, SBA-ADA is slightly more stable but the overall combined sequence (Physico-chemical properties) and structural (mRNA and 3D structure) analysis of these proteins revealed notable differences

This *in silico* analysis illustrates the potential role of tags and linkers to increase the yield and the efficacy of

pharmaceutical recombinant proteins in plant molecular pharming. It will be very interesting to extend these analyses by employing more tags and linkers based on comparative modeling and design to optimize expression cassettes and predict the putative effects of these constituents on the level of foreign protein production. Other tags like hydrophobin, zein and elastin-like peptide tags and linkers (Table1 and other linkers) can be studied for their suitability and enhancing effects on extracting and purifying recombinant proteins in plant host systems.

Other computational measurements like molecular dynamics simulations may also be employed to understand time dependent changes in protein linker association. Such simulations provide deep understanding of protein structure and function at the atomic level.

References

Abdullah, M. A., Rahmah, A. U., Sinskey, A. J., & Rha, C. K. (2008). Cell engineering and molecular pharming for biopharmaceuticals. *The Open Medicinal Chemistry Journal, 2,* 49-61. http://dx.doi.org/10.2174/1874104500802010049

Agarwal, P., Garg, V., Gautam, T., Pillai, B., Kanoria, S., & Burma, P. K. (2014). A study on the influence of different promoter and 5'UTR (URM) cassettes from *Arabidopsis thaliana* on the expression level of the reporter gene beta glucuronidase in tobacco and cotton. *Transgenic Research, 23*(2), 351-363. http://dx.doi.org/10.1007/s11248-013-9757-9

Alvarez, M. L., Topal, E., Martin, F., & Cardineau, G. A. (2010). Higher accumulation of F1-V fusion recombinant protein in plants after induction of protein body formation. *Plant Molecular Biology, 72*(1-2), 75-89. http://dx.doi.org/10.1007/s11103-009-9552-4

Amani, J., Mousavi, S. L., Rafati, S., & Salmanian, A. H. (2009). In silico analysis of chimeric espA, eae and tir fragments of Escherichia coli O157:H7 for oral immunogenic applications. *Theoretical Biology and Medical Modelling, 6,* 28. http://dx.doi.org/10.1186/1742-4682-6-28

Amani, J., Mousavi, S. L., Rafati, S., & Salmanian, A. H. (2011). Immunogenicity of a plant-derived edible chimeric EspA, Intimin and Tir of *Escherichia coli* O157:H7 in mice. *Plant Science, 180*(4), 620-627. http://dx.doi.org/10.1016/j.plantsci.2011.01.004

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering, 14*(8), 529-532. http://dx.doi.org/10.1093/protein/14.8.529

Arzola, L., Chen, J., Rattanaporn, K., Maclean, J. M., & McDonald, K. A. (2011). Transient co-expression of post-transcriptional gene silencing suppressors for increased in planta expression of a recombinant anthrax receptor fusion protein. *International Journal of Molecular Sciences, 12*(8), 4975-4990. http://dx.doi.org/10.3390/ijms12084975

Bundo, M., Montesinos, L., Izquierdo, E., Campo, S., Mieulet, D., Guiderdoni, E., ... Coca, M. (2014). Production of cecropin A antimicrobial peptide in rice seed endosperm. *BMC Plant Biology,* 14(1), 102. http://dx.doi.org/10.1186/1471-2229-14-102

Buyel, J. F., Kaever, T., Buyel, J. J., & Fischer, R. (2013). Predictive models for the accumulation of a fluorescent marker protein in tobacco leaves according to the promoter/5'UTR combination. *Biotechnol Bioengineering, 110*(2), 471-482. http://dx.doi.org/10.1002/bit.24715

Chen, L., Dempsey, B. R., Gyenis, L., Menassa, R., Brandle, J. E., & Dhaubhadel, S. (2013). Identification of the factors that control synthesis and accumulation of a therapeutic protein, human immune-regulatory interleukin-10, in *Arabidopsis thaliana*. *Plant Biotechnology Journal, 11*(5), 546-554. http://dx.doi.org/10.1111/pbi.12042

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science, 2*(9), 1511-1519. http://dx.doi.org/10.1002/pro.5560020916

Colquhoun, A. J., & Mellon, J. K. (2002). Epidermal growth factor receptor and bladder cancer. *Postgraduate Medical Journal, 78*(924), 584-589. http://dx.doi.org/10.1136/pmj.78.924.584

Conley, A. J., Joensuu, J. J., Jevnikar, A. M., Menassa, R., & Brandle, J. E. (2009). Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants. *Biotechnology and Bioengineering, 103*(3), 562-573. http://dx.doi.org/10.1002/bit.22278

Conley, A. J., Joensuu, J. J., Richman, A., & Menassa, R. (2011). Protein body-inducing fusions for high-level production and purification of recombinant proteins in plants. *Plant Biotechnol Journal, 9*(4), 419-433. http://dx.doi.org/10.1111/j.1467-7652.2011.00596.x

Crasto, C. J., & Feng, J. A. (2000). LINKER: a program to generate linker sequences for fusion proteins. *Protein Engineering, 13*(5), 309-312. http://dx.doi.org/10.1093/protein/13.5.309

Del, L. Y. M., Farran, I., Becher, M. L., Sander, V., Sanchez, V. R., Martin, V., ... Clemente, M. (2012). A chloroplast-derived *Toxoplasma gondii* GRA4 antigen used as an oral vaccine protects against toxoplasmosis in mice. *Plant Biotechnol Journal, 10*(9), 1136-1144. http://dx.doi.org/10.1111/pbi.12001

Floss, D. M., Sack, M., Arcalis, E., Stadlmann, J., Quendler, H., Rademacher, T., ... Conrad, U. (2009). Influence of elastin-like peptide fusions on the quantity and quality of a tobacco-derived human immunodeficiency virus-neutralizing antibody. *Plant Biotechnol Journal, 7*(9), 899-913. http://dx.doi.org/10.1111/j.1467-7652.2009.00452.x

Friedman, M., Orlova, A., Johansson, E., Eriksson, T. L., Hoiden-Guthenberg, I., Tolmachev, V., ... Stahl, S. (2008). Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. *Journal of Molecular Biology, 376*(5), 1388-1402. http://dx.doi.org/10.1016/j.jmb.2007.12.060

Fu, K., Cheng, Q., Liu, Z., Chen, Z., Wang, Y., Ruan, H., ... Yang, D. (2014). Immunotoxicity assessment of rice-derived recombinant human serum albumin using human peripheral blood mononuclear cells. *PLoS One, 9*(8), e104426. http://dx.doi.org/10.1371/journal.pone.0104426

Gallie, D. R. (2002). The 5'-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F. *Nucleic Acids Research, 30*(15), 3401-3411. http://dx.doi.org/10.1093/nar/gkf457

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana Press. http://dx.doi.org/10.1385/1-59259-890-0:571

Geourjon, C., & Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computational Applied Bioscience, 11*(6), 681-684.

Gostring, L., Chew, M. T., Orlova, A., Hoiden-Guthenberg, I., Wennborg, A., Carlsson, J., & Frejd, F. Y. (2010). Quantification of internalization of EGFR-binding Affibody molecules: Methodological aspects. *International Journal of Oncology, 36*(4), 757-763. http://dx.doi.org/10.3892/ijo_00000551

Grunberg, R., Ferrar, T. S., van der Sloot, A. M., Constante, M., & Serrano, L. (2010). Building blocks for protein interaction devices. *Nucleic Acids Research, 38*(8), 2645-2662. http://dx.doi.org/10.1093/nar/gkq152

Guo, J. Q., Li, Q. M., Zhou, J. Y., Zhang, G. P., Yang, Y. Y., Xing, G. X., ... Zhang, C. Y. (2006). Efficient recovery of the functional IP10-scFv fusion protein from inclusion bodies with an on-column refolding system. *Protein Expression and Purification, 45*(1), 168-174. http://dx.doi.org/10.1016/j.pep.2005.05.016

Hakkinen, S. T., Raven, N., Henquet, M., Laukkanen, M. L., Anderlei, T., Pitkanen, J. P., ... Ritala, A. (2013). Molecular farming in tobacco hairy roots by triggering the secretion of a pharmaceutical antibody. *Biotechnology and Bioengineering, 111*(2), 336–346. http://dx.doi.org/10.1002/bit.25113

He, Y., Ning, T., Xie, T., Qiu, Q., Zhang, L., Sun, Y., ... Yang, D. (2011). Large-scale production of functional human serum albumin from transgenic rice seeds. *Proceedings of the National Academy of Sciences USA, 108*(47), 19078-19083. http://dx.doi.org/10.1073/pnas.1109736108

Hefferon, K. (2014). Plant virus expression vector development: new perspectives. *Biomedical Research International, 2014*, 785382. http://dx.doi.org/10.1155/2014/785382

Hefferon, K. L. (2012). Plant virus expression vectors set the stage as production platforms for biopharmaceutical proteins. *Virology, 433*(1), 1-6. http://dx.doi.org/10.1016/j.virol.2012.06.012

Hefferon, K. L. (2014). Broadly neutralizing antibodies and the promise of universal vaccines: where are we now? *Immunotherapy, 6*(1), 51-57. http://dx.doi.org/10.2217/imt.13.150

Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Jr., Di Maria, M. V., Veve, R., Bremmes, R. M., ... Franklin, W. A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *Journal of Clinical Oncology, 21*(20), 3798-3807. http://dx.doi.org/10.1200/jco.2003.11.069

Hefferon, K. L. (2012). Plant virus expression vectors set the stage as production platforms for biopharmaceutical proteins. *Virology, 433*(1), 1-6. http://dx.doi.org/10.1016/j.virol.2012.06.012

Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Jr., Di Maria, M. V., Veve, R., Bremmes, R. M., ... Franklin, W. A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *Journal of Clinical Oncology, 21*(20), 3798-3807. http://dx.doi.org/10.1200/jco.2003.11.069

Hu, C. D., Chinenov, Y., & Kerppola, T. K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell, 9*(4), 789-798. http://dx.doi.org/10.1016/S1097-2765(02)00496-3

Hu, C. D., & Kerppola, T. K. (2003). Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nature Biotechnology, 21*(5), 539-545. http://dx.doi.org/10.1038/nbt816

Jiang, Z., & Zhou, Y. (2005). Using bioinformatics for drug target identification from the genome. *American Journal of Pharmacogenomics, 5*(6), 387-396. http://dx.doi.org/10.2165/00129785-200505060-00005

Joensuu, J. J., Conley, A. J., Lienemann, M., Brandle, J. E., Linder, M. B., & Menassa, R. (2010). Hydrophobin fusions for high-level transient protein expression and purification in *Nicotiana benthamiana*. *Plant Physiology, 152*(2), 622-633. http://dx.doi.org/10.1104/pp.109.149021

Joensuu, J. J., Conley, A. J., Linder, M. B., & Menassa, R. (2012). Bioseparation of recombinant proteins from plant extract with hydrophobin fusion technology. Methods in Molecular Biology, 824, 527-534. http://dx.doi.org/10.1007/978-1-61779-433-9_28

Jung, S. K., & McDonald, K. (2011). Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. *BMC Bioinformatics, 12*, 340. http://dx.doi.org/10.1186/1471-2105-12-340

Kamphausen, S., Holtge, N., Wirsching, F., Morys-Wortmann, C., Riester, D., Goetz, R., ... Schwienhorst, A. (2002). Genetic algorithm for the design of molecules with desired properties. *Journal of Computer-Aided Molecular Design, 16*(8-9), 551-567. http://dx.doi.org/10.1023/A:1021928016359

Kanoria, S., & Burma, P. K. (2012). A 28 nt long synthetic 5'UTR (synJ) as an enhancer of transgene expression in dicotyledonous plants. *BMC Biotechnology, 12*, 85. http://dx.doi.org/10.1186/1472-6750-12-85

Kapust, R. B., & Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expression and Purification, 19*(2), 312-318. http://dx.doi.org/10.1006/prep.2000.1251

Kerppola, T. K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nature Protocols, 1*(3), 1278-1286. http://dx.doi.org/10.1038/nprot.2006.201

Kingsbury, N. J., & McDonald, K. A. (2014). Quantitative Evaluation of E1 Endoglucanase Recovery from Tobacco Leaves Using the Vacuum Infiltration-Centrifugation Method. *BioMedical Research International, 2014*, 10. http://dx.doi.org/10.1155/2014/483596

Kuo, Y. C., Tan, C. C., Ku, J. T., Hsu, W. C., Su, S. C., Lu, C. A., & Huang, L. F. (2013). Improving Pharmaceutical Protein Production in *Oryza sativa*. *International Journal of Molecular Science, 14*(5), 8719-8739. http://dx.doi.org/10.3390/ijms14058719

Laguia-Becher, M., Martin, V., Kraemer, M., Corigliano, M., Yacono, M. L., Goldman, A., & Clemente, M. (2010). Effect of codon optimization and subcellular targeting on *Toxoplasma gondii* antigen SAG1 expression in tobacco leaves to use in subcutaneous and oral immunization in mice. *BMC Biotechnol,ogy 10*, 52. http://dx.doi.org/10.1186/1472-6750-10-52

Laskowski, R. A., Watson, J. D., & Thornton, J. M. (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Research, 33*(Web Server issue), W89-93. http://dx.doi.org/10.1093/nar/gki414

Le Gall, F., Reusch, U., Little, M., & Kipriyanov, S. M. (2004). Effect of linker sequences between the antibody variable domains on the formation, stability and biological activity of a bispecific tandem diabody. *Protein Engineering, Design and Selection, 17*(4), 357-366. http://dx.doi.org/10.1093/protein/gzh039

Ma, J. K., Christou, P., Chikwamba, R., Haydon, H., Paul, M., Ferrer, M. P., ... Thangaraj, H. (2013). Realising the value of plant molecular pharming to benefit the poor in developing countries and emerging economies. *Plant Biotechnology Journal, 11*(9), 1029-1033. http://dx.doi.org/10.1111/pbi.12127

Makhzoum, A., Benyammi, R., Moustafa, K., & Tremouillaux-Guiller, J. (2013). Recent advances on host plants and expression cassettes' structure and function in plant molecular pharming. *BioDrugs, 28*, 145-159 http://dx.doi.org/10.1007/s40259-013-0062-1

Makhzoum, A., Petit-Paly, G., St Pierre, B., & Bernards, M. A. (2011). Functional analysis of the DAT gene promoter using transient *Catharanthus roseus* and stable *Nicotiana tabacum* transformation systems. *Plant Cell Reporter, 30*(7), 1173-1182. http://dx.doi.org/10.1007/s00299-011-1025-y

Makhzoum, A. B., Sharma, P., Bernards, M. A., & Trémouillaux-Guiller, J. (2013). Hairy Roots: An Ideal Platform for Transgenic Plant Production and Other Promising Applications. *Recent Advances in Phytochemistry,* Volume 42, Phytochemicals, Plant Growth, and the Environment, pp. 95-142.

Matic, S., Masenga, V., Poli, A., Rinaldi, R., Milne, R. G., Vecchiati, M., & Noris, E. (2012). Comparative analysis of recombinant Human Papillomavirus 8 L1 production in plants by a variety of expression systems and purification methods. *Plant Biotechnology Journal, 10*(4), 410-421. http://dx.doi.org/10.1111/j.1467-7652.2011.00671.x

Merutka, G., Shalongo, W., & Stellwagen, E. (1991). A model peptide with enhanced helicity. *Biochemistry, 30*(17), 4245-4248. http://dx.doi.org/10.1021/bi00231a020

Moeller, L., Gan, Q., & Wang, K. (2009). A bacterial signal peptide is functional in plants and directs proteins to the secretory pathway. *Journal of Experimental Botany, 60*(12), 3337-3352. http://dx.doi.org/10.1093/jxb/erp167

Orlova, A., Tolmachev, V., Pehrson, R., Lindborg, M., Tran, T., Sandstrom, M., ... Feldwisch, J. (2007). Synthetic affibody molecules: a novel class of affinity ligands for molecular imaging of HER2-expressing malignant tumors. *Cancer Research, 67*(5), 2178-2186. http://dx.doi.org/10.1158/0008-5472.can-06-2887

Ou, J., Guo, Z., Shi, J., Wang, X., Liu, J., Shi, B., ... Yang, D. (2014). Transgenic rice endosperm as a bioreactor for molecular pharming. *Plant Cell Reporter, 33*(4), 585-594. http://dx.doi.org/10.1007/s00299-013-1559-2

Paul, M. J., Teh, A. Y., Twyman, R. M., & Ma, J. K. (2013). Target product selection - where can Molecular Pharming make the difference? *Current Pharmaceutical Design, 19*(31), 5478-5485. http://dx.doi.org/10.2174/1381612811319310003

Rademacher, T., Sack, M., Arcalis, E., Stadlmann, J., Balzer, S., Altmann, F., ... Stoger, E. (2008). Recombinant antibody 2G12 produced in maize endosperm efficiently neutralizes

HIV-1 and contains predominantly single-GlcNAc N-glycans. *Plant Biotechnology Journal, 6*(2), 189-201. http://dx.doi.org/10.1111/j.1467-7652.2007.00306.x

Rigano, M. M., De Guzman, G., Walmsley, A. M., Frusciante, L., & Barone, A. (2013). Production of pharmaceutical proteins in solanaceae food crops. *International Journal of Molecular Science, 14*(2), 2753-2773. http://dx.doi.org/10.3390/ijms14022753

Roy, S., Maheshwari, N., Chauhan, R., Sen, N. K., & Sharma, A. (2011). Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. *Bioinformation, 6*(8), 315-319. http://dx.doi.org/10.6026/97320630006315

Sabalza, M., Vamvaka, E., Christou, P., & Capell, T. (2013). Seeds as a production system for molecular pharming applications: status and prospects. *Current Pharmaceutical Design, 19*(31), 5543-5552. http://dx.doi.org/10.2174/1381612811319310009

Satoh, J., Kato, K., & Shinmyo, A. (2004). The 5'-untranslated region of the tobacco alcohol dehydrogenase gene functions as an effective translational enhancer in plant. *Journal of Bioscience and Bioengineering, 98*(1), 1-8. http://dx.doi.org/10.1016/s1389-1723(04)70234-0

Shah, K., Almaghrabi, B., & Bohlmann, H. (2013). Comparison of Expression Vectors for Transient Expression of Recombinant Proteins in Plants. *Plant Molecular Biology Reporter, 31*(6), 1529-1538. http://dx.doi.org/10.1007/s11105-013-0614-z

Singhabahu, S., George, J., & Bringloe, D. (2013). Expression of a functional human adenosine deaminase in transgenic tobacco plants. *Transgenic Research, 22*(3), 643-649. http://dx.doi.org/10.1007/s11248-012-9676-1

Singhabahu, S., George, J., & Bringloe, D. (2014). High yield production of apoplast-directed human adenosine deaminase in transgenic tobacco BY-2 cell suspensions. *Biotechnology and Applied Biochemistry*. Article first published online: 14 JUL 2014 http://dx.doi.org/10.1002/bab.1240

Skarjinskaia, M., Ruby, K., Araujo, A., Taylor, K., Gopalasamy-Raju, V., Musiychuk, K., ... Yusibov, V. (2013). Hairy roots as a vaccine production and delivery system. *Advances in Biochemical Engineering/Biotechnology, 134*, 115-134. http://dx.doi.org/10.1007/10_2013_184

Sommese, R. F., Sivaramakrishnan, S., Baldwin, R. L., & Spudich, J. A. (2010). Helicity of short E-R/K peptides. *Protein Science, 19*(10), 2001-2005. http://dx.doi.org/10.1002/pro.469

Soria-Guerra, R. E., Moreno-Fierros, L., & Rosales-Mendoza, S. (2011). Two decades of plant-based candidate vaccines: a review of the chimeric protein approaches. *Plant Cell Reporter, 30*(8), 1367-1382. http://dx.doi.org/10.1007/s00299-011-1065-3

Specht, E. A., & Mayfield, S. P. (2014). Algae-based oral recombinant vaccines. *Frontiers in Microbiology, 5*, 60. http://dx.doi.org/10.3389/fmicb.2014.00060

Stoger, E., Fischer, R., Moloney, M., & Ma, J. K. (2014). Plant molecular pharming for the treatment of chronic and infectious diseases. *Annual Review of Plant Biology, 65*, 743-768. http://dx.doi.org/10.1146/annurev-arplant-050213-035850

Tremblay, R., Diao, H., Huner, N., Jevnikar, A. M., & Ma, S. (2011). The development of a high-yield recombinant protein bioreactor through RNAi induced knockdown of ATP/ADP transporter in *Solanum tuberosum*. *Journal of Biotechnology, 156*(1), 59-66. http://dx.doi.org/10.1016/j.jbiotec.2011.08.005

Trinh, R., Gurbaxani, B., Morrison, S. L., & Seyfzadeh, M. (2004). Optimization of codon pair use within the (GGGGS)3 linker sequence results in enhanced protein expression. *Molecular Immunology, 40*(10), 717-722. http://dx.doi.org/10.1016/j.molimm.2003.08.006

Tropea, J. E., Cherry, S., & Waugh, D. S. (2009). Expression and purification of soluble His(6)-tagged TEV protease. *Methods in Molecular Biology, 498*, 297-307. http://dx.doi.org/10.1007/978-1-59745-196-3_19

Twyman, R. M., Schillberg, S., & Fischer, R. (2013). Optimizing the yield of recombinant pharmaceutical proteins in plants. *Current Pharmaceutical Design, 19*(31), 5486-5494. http://dx.doi.org/10.2174/1381612811319310004

Vaseeharan, B., & Valli, S. J. (2011). *In silico* homology modeling of prophenoloxidase activating factor Serine Proteinase gene from the haemocytes of *Fenneropenaeus indicus. Journal of Proteomics and Bioinformatics, 4,* 053-057.

Walker, R. A., & Dearing, S. J. (1999). Expression of epidermal growth factor receptor mRNA and protein in primary breast carcinomas. *Breast Cancer Research and Treatment, 53*(2), 167-176. http://dx.doi.org/10.1023/A:1006194700667

Warzecha, H. (2008). Biopharmaceuticals from plants: a multitude of options for posttranslational modifications. *Biotechnology & Genetic Engineering Reviews, 25,* 315-330. http://dx.doi.org/10.5661/bger-25-315

Waugh, D. S. (2005). Making the most of affinity tags. *Trends in Biotechnology, 23*(6), 316-320. http://dx.doi.org/10.1016/j.tibtech.2005.03.012

Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification, 80*(2), 283-293. http://dx.doi.org/10.1016/j.pep.2011.08.005

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology, 112*, 531-552.

Yan, W., Imanishi, M., Futaki, S., & Sugiura, Y. (2007). Alpha-helical linker of an artificial 6-zinc finger peptide contributes to selective DNA binding to a discontinuous recognition sequence. *Biochemistry, 46*(29), 8517-8524. http://dx.doi.org/10.1021/bi7006417

Yang, L., Wakasa, Y., & Takaiwa, F. (2008). Biopharming to increase bioactive peptides in rice seed. *Journal of AOAC International, 91*(4), 957-964.

Zhang, J., Yun, J., Shang, Z., Zhang, X., & Pan, B. (2009). Design and optimization of a linker for fusion protein construction. *Progress in Natural Science, 19*(10), 1197-1200. http://dx.doi.org/http://dx.doi.org/10.1016/j.pnsc.2008.12.007

Zhang, L., Shi, J., Jiang, D., Stupak, J., Ou, J., Qiu, Q., ... Yang, D. (2012). Expression and characterization of recombinant human alpha-antitrypsin in transgenic rice seed. *Journal of Biotechnology, 164*(2), 300-308. http://dx.doi.org/10.1016/j.jbiotec.2013.01.008

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40. http://dx.doi.org/10.1186/1471-2105-9-40

Zhao, X., Li, G., & Liang, S. (2013). Several Affinity Tags Commonly Used in Chromatographic Purification. *Journal of Analytical Methods in Chemistry*, 2013, 581093. http://dx.doi.org/10.1155/2013/581093