Plants as a production platform for high-value proteins

Lotte B. Westerhof

Plants as a production platform for high-value proteins

Lotte B. Westerhof

Plants as a production platform for high-value proteins

Lotte B. Westerhof

Thesis committee

Promotor

Prof. Dr J. Bakker

Professor of Nematology

Wageningen University

Co-promotor

Dr A. Schots

Associate professor, Laboratory of Nematology

Wageningen University

Other members

Prof. Dr M.M. van Oers, Wageningen University

Prof. Dr J.K-C. Ma, St George's University of London, United Kingdom

Prof. Dr S. Schillberg, Fraunhofer IME, Aachen, Germany

Dr D. Nolan, Synthon, Nijmegen

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Plants as a production platform for high-value proteins

Lotte B. Westerhof

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the Rector Magnificus
Prof. Dr M. J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 3 December 2014
at 4 p.m. in the Aula.

Lotte B. Westerhof
Plants as a production platform for high-value proteins, 192 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)
With references, with summaries in Dutch and English

ISBN 978-94-6257-146-4

Halverwege.
Halfway.

Table of contents

Chapter 1

General introduction

Lotte B. Westerhof

Since the biomedical revolution there is a growing demand for pharmaceuticals proteins, so-called biopharmaceuticals. These proteins are isolated from a natural resource or, more often, made using cell-based fermentation systems (heterologous expression systems), as the isolation from humans or other mammals does often not suffice in terms of quantity or quality. Plants provide a good alternative production platform that has unique advantages compared to the currently used heterologous expression systems.

Genetic manipulation of plants

Today, many plant species can be easily manipulated genetically. In 1982 the first plant was transformed with exogenous genes [1]. The gene transfer was accomplished by exploiting the soil-born bacterium *Agrobacterium tumefaciens*. *A. tumefaciens* can infect many plant species causing crown-gall disease. Upon infection the bacteria induces the formation of tumours that provide the bacteria with nutrients. To induce these tumours *A. tumefaciens* carries genes that lead to synthesis of plant hormones and metabolites in the plant on a plasmid called the tumour inducing (Ti) plasmid. The part of the plasmid that carries these genes, called the T-DNA, is actively integrated in the plant genome. Subsequent expression of the integrated genes induces tumour formation. The genes encoding the enzymes to synthesise hormones and other metabolites can be easily replaced by (an)other gene(s) and does not affect the transformation potency of *A. tumefaciens*. When in 1985 a simple protocol for *Agrobacterium*-mediated transformation and regeneration of plants from leaf disks was described [2], this quickly became a widely used technique. Later a very efficient procedure to transform the model plant *Arabidopsis thaliana* involved only the simple dipping of the plant's flowers in *Agrobacterium* suspension where after transformation results in transgenic seeds [3].

Unfortunately *A. tumefaciens* has a specific host range. Although plant-bacterium incompatibility can sometimes be overcome using specialised tissue culture techniques, *A. tumefaciens* cannot be used for transformation of all plant species such as many monocots. Therefore direct gene transfer methods were also developed based on the introduction of naked DNA involving the 'shooting' of DNA coated gold or tungsten particles into plant cells [4] or chemical transformation or electroporation of protoplasts, where after integration into the plant genome occurs by chance. But, despite these alternatives being available *Agrobacterium*-mediated transformation has always been greatly preferred, because it has a great advantage. The use of *A. tumefaciens* ensures that the gene of interest is more often integrated at a location that allows transcription, as it makes use of the plant's own DNA repair system for integration and DNA repair is most often needed in transcriptionally active sites.

After transformation and plant regeneration, repeated selfing can create a stable transgenic plant as the transformed genes are transferred to the plant's offspring in a Mendelian

fashion. However, the creation of stable transgenic plants is a lengthy and laborious process. Alternatively, plant transformation can be performed transiently. Expression of introduced genes upon transformation does not depend on integration into the genome [5-7]. Because the introduced genes that are not integrated are eventually broken down, the transformation is referred to as transient. Transient expression is often used as a pre-screen to evaluate the effect of expression of the gene or to obtain higher expression of the transformed gene, because in transient transformation there is no limit on expression imposed by the site of integration.

A currently widely used technique based on transient expression is the *Agrobacterium*-mediated transient transformation assay (ATTA). This technique includes the infiltration of an *Agrobacterium* suspension through the stomata into the spongy parenchyma of plant leaves using either a needle-less syringe (Figure 1) or by applying vacuum. Commonly used plant species for this technique include tobacco, tomato and potato. Gene expression can be detected from one day post infiltration (dpi) onwards. The peak of recombinant protein accumulation usually lies between 3 and 20 dpi dependent on the size and stability of the protein, whether or not a viral vector is used and whether or not gene silencing is triggered. Gene silencing is the natural mechanism of plants to combat viruses. As a response some plant viruses encode a silencing inhibitor in their genetic material. Co-expression of the silencing inhibitor p19 of tomato bushy stunt virus is often used to avoid gene silencing during transient expression. Due to the high levels of expression that can be obtained using transient transformation the ATTA technique may become the standard for recombinant protein production.

Figure 1 ◆ *Agrobacterium* transient transformation assay.

Plants as a production platform for biopharmaceuticals

Of the biopharmaceuticals that are produced in heterologous expression systems and are approved for medical use in humans in 2010, 17 were produced in the bacterium *Escherichia coli*, 4 in the yeast *Saccharomyces cerevisiae* and 32 in mammalian cell lines (mainly in Chinese hamster ovarian (CHO) and murine myeloma (SP2/0) cells) [8]. The use of *E. coli* is most cost effective, because relatively simple media and culture conditions are sufficient to support its growth. However, *E. coli* is a prokaryote and of all expression platforms is the most distantly related to humans and cannot produce all human proteins with the same quality as the original. Yeasts are single-cell eukaryotic organisms and are often able to properly fold proteins and carry out post-translational modifications, but have a tendency to hyper-glycosylate [9]. Mammalian cells produce the highest quality of proteins with regard to similarity to human proteins, but are the most expensive to maintain.

Plants have been demonstrated to be a good alternative production platform [10-13]. Plants are higher eukaryotes and have the ability to correctly fold complex proteins, assemble heteromultimeric protein complexes [14] and carry out post-translational modifications. While they have the same capabilities as mammalian cells, they require much less specialized care and are therefore much cheaper to maintain. The production of insulin in safflower demonstrated that plants could even be cheaper than the use of *E. coli* [15]. Insulin production could be easily scaled up in safflower, with an estimated production of 2,5 kilo of insulin per hectare. Thus, for the worldwide demand only 65 square kilometre of land, i.e. 3 commercial farms in North America, is required [16].

The speed of protein production in plants using transient expression is exploited in a personalized medicine approach used to treat Non-Hodgkin lymphoma (NHL). NHL is characterized by the uninhibited clonal proliferation of a B-cell. Because each B-cell is unique in the membrane-bound antibodies it carries, an anti-idiotype immune response against this antibody could help clear the malignant B-cells from the body. Thereto the variable domain of the antibody should be sequenced to allow synthetic gene synthesis followed by production of the antibody for immunization of the patient. Thus, each patient requires a personalized anti-idiotype 'vaccine'. Grams of antibody can be produced in plants in a matter of weeks using the ATTA technique. Because no expression platform can compete with plants in the terms of speed [17], it is the production platform of choice for production of the NHL-vaccine. The company Icon Genetics has recently announced the successful completion of the Phase I clinical study with plant-produced personalized NHL-vaccines whereby 73% of the patients were found to have mounted a tumour-specific immune response [18].

The ATTA technique is also used for large-scale production of biopharmaceuticals [19]. The production strategy of the company Kentucky Bioprocessing involves the large-scale growth of *Nicotiana benthamiana* plants in green houses and automated vacuum-infiltration of

Agrobacterium and plant harvest. Kentucky Bioprocessing focuses mainly on the production of monoclonal antibodies for use as microbiocides and mucosal vaccines to prevent infection with pathogens such as *Clostridium difficile*, Papillomavirus, HIV, Parvovirus, Rabies virus and Ebola virus. They have also demonstrated the feasibility to rapidly deliver large quantities of purified recombinant protein needed as vaccine in case of a pandemic threat. An important fact, as for instance the production of influenza-viruses in current egg based vaccine production systems has limited capacity and speed to respond to pandemic threats [20].

The company Protalix Biotherapeutics produces the first approved biopharmaceutical for clinical use in humans. This company exploits the fact that plants can produce the enzyme glucocerebrosidase with N-glycans terminating in mannoses, which is sold as Taliglucerase alfa. Glucocerebrosidase is a lysosomal enzyme that catalyses the hydrolysis of glycosylceramide. In patients suffering from Gaucher disease a genetic defect renders this enzyme non-functional leading to accumulation of glycosylceramide in lysosomes. Associated symptoms include bruising, fatigue, anaemia, low blood platelets, and enlargement of the liver and spleen. Gaucher disease is treated with life-long enzyme-replacement therapy. It has been shown that uptake of the enzyme by macrophages relieves the symptoms associated with Gaucher disease to a major extent. To ensure uptake of glucocerebrosidase by macrophages the enzyme needs to carry N-glycans with terminal mannose residues to enable endocytosis by the cell through the mannose receptor. Because glucocerebrosidase produced in mammalian cells does not carry N-glycans with terminal mannoses, it needs to undergo several deglycosylation steps *in vitro*. In plants N-glycans with terminal mannoses can be achieved *in vivo*. Thereto, Protalix expressed glucocerebrosidase in carrot cells with a vacuolar targeting signal [21]. This plant-based glucocerebrosidase was demonstrated to perform better than its mammalian-cell produced counterpart as *in vitro* deglycosylation leads to a heterogeneous population of N-glycans often without the required terminal mannoses. Therefore, Taliglucerase alfa is considered a 'bio-better' as opposed to a 'bio-similar'. The approval for Taliglucerase alfa was accelerated because Genzyme, the only company that produced glucocerebrosidase in CHO cells (trade name Cerezyme) at the time, had manufacturing difficulty due to contamination with *Vesivirus* 2117 [22]. While this virus is not harmful for humans, it hampers the growth of CHO cells. This illustrates another important advantage of plants as a production platform for biopharmaceuticals, namely their low risk of contamination with mammalian pathogens, especially human viruses as these cannot replicate in plant cells.

Engineering of the N-glycosylation pathway in plants

The chance of other plant-based biopharmaceuticals reaching the market is greatly enhanced by the capability of plants, upon engineering of the glycosylation pathway, to produce proteins with homogeneous human-like N-glycan profiles. The biological role of N-glycans is to aid in protein folding, provide protein stability and enable (sub)cellular protein trafficking and receptor mediated signalling. In all eukaryotes N-glycosylation occurs co-translational when the oligosaccharyl transferase complex encounters a N-glycosylation signal. The signal for N-glycosylation is the amino acid sequence asparagine-X-serine/threonine, whereby X can be any amino acid except proline. A common N-glycan precursor is than attached to the asparagine in the endoplasmic reticulum. In the Golgi several sugar residues are removed from the precursor by glycosylhydrolases whereas other sugar residues are attached by glycosyltransferases, thereby creating a specific N-glycan type dependent on species, tissue, environment and the protein itself.

The plant repertoire of glycan modifying enzymes is limited, because plants are less complex organisms and use N-glycosylation for protein folding and stability and not or to a minor extent for cellular protein trafficking and receptor mediated signalling. Unsurprisingly, plants do not produce N-glycans that are identical to human/mammalian N-glycans (Figure 2). Differences in N-glycan type can lead to a reduced protein half-life, loss of function and increased immunogenicity [23]. Increased immunogenicity is one of the main concerns of regulatory bodies when dealing with biopharmaceuticals. An immune response against the therapeutic protein may lead to allergic symptoms or even septic shock. Therefore, the N-glycosylation pathway of plants has been engineered to produce proteins carrying human N-glycan types [24].

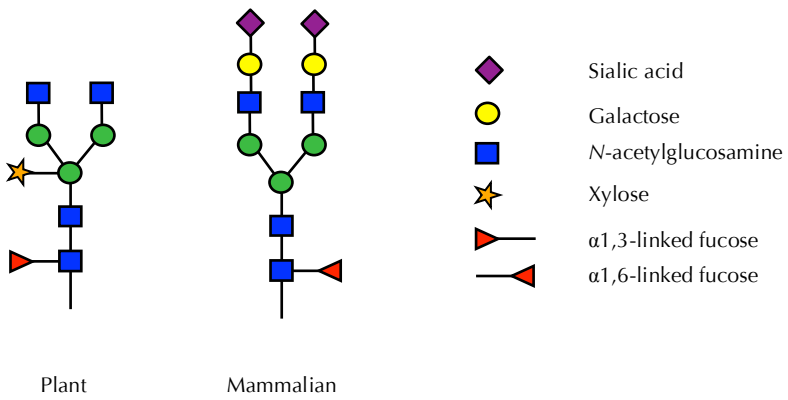


Figure 2 ◆ Schematic representation of typical plant and mammalian N-glycans.

The majority of secreted plant proteins carry a diantennary N-glycan structure with terminal *N*-acetylglucosamine or paucimannosidic oligosaccharides (Figure 2), both with β1,2-xylose and core α1,3-fucose, abbreviated as $GnGnXF^3$ and $MMXF^3$, respectively [25-28]. Only the β1,2-xylose and the core α1,3-fucose do not occur on human N-glycans. Because these sugar residues are potentially immunogenic, the glycosyltransferases responsible for the addition of these sugar residues were either silenced or mutated successfully in several plant species [26-31]. Human N-glycans often carry a core α1,6-fucose. This sugar residue was added to the N-glycans on a plant-produced antibody by co-expression of α1,6-fucosyltransferase 8 [32, 33]. Next to that, the antennae of human N-glycans are often extended with terminal β1,4-galactose and/or sialic acid. *In planta* β1,4-galactosylation and sialylation has also been achieved [25, 34-37]. Furthermore, while plants only produce diantennary N-glycans, human N-glycans can be bisected or branched (Figure 2). Plant expression of *N*-acetylglucosaminyltransferases III, IV and V resulted in the synthesis of different bisected and branched complex N-glycans [38, 39].

Engineering of human-like N-glycans in plants is in part successful due to the fact that plants rarely show aberrant phenotypes upon engineering of the N-glycosylation pathway. Furthermore, antibodies produced in glyco-engineered plants were even shown to be more homogeneously glycosylated compared to CHO cell produced antibodies, which improved antibody functioning [28, 36]. A more homogenous glycosylation pattern is most likely due to the fact that the plant repertoire of glycan modifying enzymes is limited. The advances of glyco-engineering cover over a decade of research and now lead to possibilities to engineer virtually any N-glycan. Potentially interesting N-glycan types with regards to immunogenicity and cellular targeting may be identified and exploited for biopharmaceutical and vaccination purposes [40].

Biopharmaceuticals for the treatment of inflammatory disorders

Currently, biopharmaceuticals are commonly used in the treatment of autoimmune diseases such as diabetes, rheumatoid arthritis, multiple sclerosis, psoriasis and other chronic inflammatory disorders such as inflammatory bowel diseases. In autoimmune diseases and inflammatory disorders the immune system reacts to, respectively, self-antigens or harmless non-self molecules causing chronic inflammation and tissue damage. High serum concentrations of signalling molecules used by the immune system characterize chronic inflammation. The largest group of signalling molecules of the immune system is the cytokines. During the onset of an infection, cells of the innate immune system (such as dendritic cells and macrophages) recognize the pathogen and start secreting pro-inflammatory cytokines such as tumour necrosis factor (TNF)-α, interleukin (IL)-1β, IL-6, IL-12 and IL-23. Depending on the

type of pathogen, specific cytokines are secreted by antigen presenting cells that drive the proliferation of a specific T helper (Th) cell subset of the adaptive immune system. Generally, intracellular pathogens such as viruses and intracellular microbes trigger a Th1 response, while extracellular parasites trigger Th2 cell proliferation. More recently, the Th17 subset was characterized and shown to proliferate upon infection with extracellular bacteria and fungi, including mucosal bacteria upon breaching epithelial mucosae. Dependent on the Th subset specific immune cells of both the innate and adaptive immune system are activated and B cells are induced to secrete specific immunoglobulin isotypes that are best equipped to eliminate the invading pathogen. These effects are also mediated by cytokines. Th1 cells predominantly secrete interferon (IFN)-γ and lymphotoxin (LT), while Th2 cells secrete IL-4, IL-5, and IL-13. Th17 cells secrete IL-17, IL-21 and IL-22. After the pathogen is cleared the immune system returns to homeostasis through the secretion of cytokines such as IL-10 and transforming growth factor (TGF)-β, which are mainly secreted by innate immune cells and T regulatory cells. Cytokines are pleiotropic, meaning that they can exert many different functions. Dependent on the environment and stimuli given, some cytokines, like TGF-β and IL-22, exert either pro- or anti-inflammatory responses.

Autoimmune diseases are often associated with the activation and proliferation of Th1 and Th17 cells, whereas Th2 cells dominate allergic responses. Current therapies to treat chronic inflammation often include the use of antibodies that can neutralize pro-inflammatory cytokines or immune receptors that support inflammation. The largest share of the global biopharmaceutical market ($99 billion in 2009) belong to antibody-based biopharmaceuticals, mainly antibodies that neutralize TNF-α ($18 billion in 2009) [8].

Outline of this thesis

This thesis describes the use of *Nicotiana benthamiana* as a production platform for biopharmaceuticals that are currently or may be used in the future to treat autoimmune diseases and inflammatory disorders. In **Chapter 2** we describe the expression of the cytokine IL-10. This cytokine was suggested to have great anti-inflammatory potential as it can reduce the activity of macrophages, inhibit antigen presentation by dendritic cells and inhibit the production of pro-inflammatory cytokines by innate immune cells and adaptive immune cells, such as T and B cells [41-44]. The yield of cytokines expressed in heterologous hosts is often poor. The suggested reason for this is that cytokines must be inherently unstable as they often have a short half-life *in vivo*. We demonstrated that for IL-10 the major yield-limiting factor is not intrinsic instability, but its ability to swap domains intended to result in dimers. Upon expression *in planta* this so-called '3D domain swapping' lead to large multimers, which were visible as intracellular granules of organelle size. By protein engineering we were able to

prevent multimerisation of IL-10 and increase its yield 30-fold.

In **Chapter 3** we compared the plant-based expression of IgA with IgG. The use of the IgG antibody isotype is the standard in the clinic today. But, other isotypes such as IgE and IgA may offer, yet unexplored, advantages. IgA naturally blocks pro-inflammatory responses mediated by IgG and IgE. This would be an interesting 'bonus' trait for an antibody used to reduce inflammation. Although ample experience with expression of IgG exists, reports on the production of IgA are scarce. We expressed IgA variants of the three antibodies Infliximab, Adalimumab and Ustekinumab, which are used as immunosuppressors in the clinic today, and compared it with the expression of the original IgG variants. Infliximab (trade name Remicade) and Adalimumab (trade name Humira) both neutralize TNF-α. Ustekinumab (trade name Stelara) is directed against the p40 subunit shared by IL-12 and IL-23 and able to neutralize these pro-inflammatory cytokines. We demonstrate that IgA can be produced equally well in plants in terms of yield and binding capacity, but receives a different N-glycan. The type of N-glycan a protein receives upon expression in *Nicotiana benthamiana* can be determined by protein intrinsic properties, in this case the antibody isotype.

In **Chapter 4** we continue our research on IgA and transiently express secretory (s)IgA, a heteromultimeric protein complex. sIgA consists of two IgA molecules that are kept together by the joining chain and this dimeric IgA is enveloped by a part of the polymeric-Ig-receptor called the secretory component. The challenge in transient expression of a heteromultimeric protein is to achieve expression of all necessary genes simultaneously in as many cells as possible and in the most efficient ratio that reflects the stoichiometry of the protein complex. Use of a multi-gene expression vector increased sIgA yield compared to co-infiltration of separate vectors. However, the transformation method turned out not to be the production bottleneck. We demonstrated that not joining chain expression, but its incorporation in the protein complex is the limiting factor for sIgA yield. We hypothesize that the partial N-glycosylation of the tailpiece of the alpha heavy chain and/or the joining chain are the cause of this inefficient incorporation.

In **Chapter 5** we study the plant-based expression of a helminth protein. Parasitic helminths have often evolved such that they employ immune modulatory proteins to evade the host's immune system to sustain infection. Their ability to control the immune system has even been exploited in treatment of inflammatory bowel disease and allergies [45]. In this chapter we focused on omega-1, a *Schistosoma mansoni* egg antigen. Omega-1 is a T2 RNase that modifies the activity of dendritic cells and upon antigen presentation polarizes T helper cells towards a Th2-phenotype. This effect is not only facilitated by its enzymatic activity, but also by its N-glycosylation that ensures internalisation by the dendritic cell. Upon expression in *Nicotiana benthamiana* the majority of omega-1 was secreted to the apoplast. Isolation of apoplast proteins was a good starting point for purification, as little 'contaminating' proteins reside in this extracellular compartment. Although plant-produced omega-1 was shown to predominantly carry the typical plant *N*-glycan $MMXF^3$, omega-1 from *S. mansoni*

predominantly carries N-glycan types with terminal Lewis-X motifs. Lewis-X motifs consist of a β1,4-galactose and α1,3-fucose on the terminal *N*-acetylglucosamine. To be able to investigate the influence of these Lewis-X motifs on the function of omega-1 we co-expressed omega-1 with α1,3-fucosyltransferase 9a from *Tetraodon nigriviridus* and β1,4-galactosyltransferase 1 from *Danio rerio*. Controlled expression of these glycosyltransferases resulted in the production of omega-1 with a significant proportion carrying N-glycans with one or two terminal Lewis-X motifs. This demonstrated the potential of plants to produce, often immune-modulatory, helminth proteins to elucidate their biological function. Research on these proteins may reveal new therapeutic candidates.

Finally, when expressing genes heterologously, matching of the codon use of the gene of interest to the codon use of the production host can boost protein yield tremendously. Although this is a widely known fact, a general codon optimisation strategy has never been described and the results with the various published strategies are highly variable. We developed a strategy based on a general expression-linked codon bias found across plant species and turned out to be robustly successful. This encouraged the research described in **Chapter 6** wherein we demonstrate that codon optimisation of heterologously expressed genes can improve protein yield by increasing mRNA stability as well as translatability. To support our experimental data we employed a large database study on the codon use of several species. We demonstrate that this general codon bias found in plants extends to other kingdoms of life. In our investigations to explain this general codon bias we found that structural mRNA characteristics are correlated with expression. The mRNA molecules of highly expressed genes are generally more stable due to an increased amount of nucleotide bonds, but also display a more equal distribution of these bonds over the molecule, making the mRNA structure more 'airy' to allow efficient translation.

References

1. Barton, K.A., et al., *Regeneration of Intact Tobacco Plants Containing Full Length Copies of Genetically Engineerd T-DNA and Transmission of T-DNA to R1 Progeny.* Cell, 1983. 32: p. 1033-1043.
2. Horsch, R.B., et al., *A simple and General Method for Transferring Genes into Plants.* Science, 1985. 227: p. 1229-1231.
3. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.* Plant Journal, 1998. 16(6): p. 735-43.
4. Klein, T.M., et al., *High-velocity micorprojectiles for delivering nucleic acids into living cells.* Nature, 1987. 327: p. 70-73.
5. Tzfira, T., et al., *Site-specific integration of Agrobacterium tumefaciens T-DNA via double-stranded intermediates.* Plant Physiology, 2003. 133(3): p. 1011-23.
6. Janssen, B.J. and R.C. Gardner, *Localized transient expression of GUS in leaf discs following cocultivation with Agrobacterium.* Plant Molecular Biology, 1990. 14(1): p. 61-72.
7. Narasimhulu, S.B., et al., *Early transcription of Agrobacterium T-DNA genes in tobacco and maize.* Plant Cell, 1996. 8(5): p. 873-86.
8. Walsh, G., *Biopharmaceutical benchmarks 2010.* Nature Biotechnology, 2010. 28(9): p. 917-24.
9. Gemmill, T.R. and R.B. Trimble, *Overview of N- and O-linked oligosaccharide structures found in various yeast species.* Biochimica et Biophysica Acta, 1999. 1426: p. 227-237.
10. Daniell, H., S.J. Streatfield, and K. Wycoff, *Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants.* Trends in Plant Sciences, 2001. 6(5): p. 219-26.
11. Giddings, G., et al., *Transgenic plants as factories for biopharmaceuticals.* Nature Biotechnology, 2000. 18(11): p. 1151-5.
12. Goldstein, D.A. and J.A. Thomas, *Biopharmaceuticals derived from genetically modified plants.* QJM, 2004. 97(11): p. 705-16.
13. Warzecha, H., *Biopharmaceuticals from plants: a multitude of options for posttranslational modifications.* Biotechnology & Genetic Engineering Reviews, 2008. 25: p. 315-30.
14. Ma, J.K., et al., *Generation and assembly of secretory antibodies in plants.* Science, 1995. 268(5211): p. 716-9.
15. Boothe, J., et al., *Seed-based expression systems for plant molecular farming.* Plant Biotechnol Journal, 2010. 8(5): p. 588-606.
16. CTV-news, http://www.ctvnews.ca/new-source-of-insulin-blossoming-on-the-prairies-1.479043. Visited on 23 July 2014, 2010.
17. Hiatt, A. and M. Pauly, *Monoclonal antibodies from plants: A new speed record.* PNAS, 2006. 103(40): p. 14645-14646.
18. Icon-Genetics, http://www.icongenetics.com/html/5975.htm. Visited on 23 July 2014, 2014.
19. Stoger, E., et al., *Plant molecular pharming for the treatment of chronic and infectious diseases.* Annual Review of Plant Biology, 2014. 65: p. 743-68.
20. Partridge, J., M.P. Kieny, and WHO, *Global production of seasonal and pandemic (H1N1) influenza vaccines in 2009-2010 and comparison with previous estimates and global action plan targets.* Vaccine, 2010. 28(30): p. 4709-4712.

21. Shaaltiel, Y., et al., *Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system.* Plant Biotechnology Journal, 2007. 5(5): p. 579-590.
22. Bethencourt, V., *Virus stalls Genzyme plant.* Nature Biotechnology, 2009. 27(8): p. 681.
23. Lingg, N., et al., *The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis.* Biotechnology Journal, 2012. 7(12): p. 1462-1472.
24. Bosch, D., et al., *N-Glycosylation of plant-produced recombinant proteins.* Curr Pharm Des, 2013.
25. Bakker, H., et al., *Galactose-extended glycans of antibodies produced by transgenic plants.* PNAS, 2001. 98(5): p. 2899-2904.
26. Strasser, R., et al., *Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure.* Plant Biotechnol Journal, 2008. 6(4): p. 392-402.
27. Sourrouille, C., et al., *Down-regulated expression of plant-specific glycoepitopes in alfalfa.* Plant Biotechnology Journal, 2008. 6(7): p. 702-21.
28. Cox, K.M., et al., *Glycan optimization of a human monoclonal antibody in the aquatic plant Lemna minor.* Nature Biotechnology, 2006. 24(12): p. 1591-7.
29. Strasser, R., et al., *Generation of Arabidopsis thaliana plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose.* FEBS Letters, 2004. 561(1-3): p. 132-6.
30. Koprivova, A., et al., *Targeted knockouts of Physcomitrella lacking plant-specific immunogenic N-glycans.* Plant Biotechnol Journal, 2004. 2(6): p. 517-23.
31. Shin, Y.J., et al., *Production of recombinant human granulocyte macrophage-colony stimulating factor in rice cell suspension culture with a human-like N-glycan structure.* Plant Biotechnology Journal, 2011. 9(9): p. 1109-19.
32. Forthal, D.N., et al., *Fc-Glycosylation Influences Fc gamma Receptor Binding and Cell-Mediated Anti-HIV Activity of Monoclonal Antibody 2G12.* Journal of Immunology, 2010. 185(11): p. 6876-6882.
33. Castilho, A., et al., *Rapid High Yield Production of Different Glycoforms of Ebola Virus Monoclonal Antibody.* Plos One, 2011. 6(10).
34. Palacpac, N.Q., et al., *Stable expression of human beta1,4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns.* PNAS, 1999. 96(8): p. 4692-7.
35. Bakker, H., et al., *An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes.* PNAS, 2006. 103(20): p. 7577-82.
36. Strasser, R., et al., *Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous beta1,4-galactosylated N-glycan profile.* Journal of Biological Chemistry, 2009. 284(31): p. 20479-85.
37. Castilho, A., et al., *In Planta Protein Sialylation through Overexpression of the Respective Mammalian Pathway.* Journal of Biological Chemistry , 2010. 285: p. 15923-15930.
38. Castilho, A., et al., *N-glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex N-glycans.* Glycobiology, 2011b. 21(6): p. 813-23.
39. Nagels, B., et al., *Production of complex multiantennary N-glycans in Nicotiana benthamiana plants.* Plant Physiology, 2011. 155(3): p. 1103-12.
40. Bosch, D. and A. Schots, *Plant glycans: friend or foe in vaccine development?* Expert Review of Vaccines, 2010. 9(8): p. 835-842.

41. Moore, K.W., et al., *Interleukin-10 and the interleukin-10 receptor.* Annual Review of Immunology, 2001. 19: p. 683-765.
42. Moore, K.W., et al., *Interleukin-10.* Annual Review of Immunology, 1993. 11(1): p. 165-190.
43. Mocellin, S., et al., *The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle.* Cytokine & Growth Factor Reviews, 2004. 15(1): p. 61-76.
44. Saraiva, M. and A. O'Garra, *The regulation of IL-10 production by immune cells.* Nature Reviews Immunology, 2010. 10(3): p. 170-181.
45. Elliott, D.E., R.W. Summers, and J.V. Weinstock, *Helminths as governors of immune-mediated inflammation.* International Journal for Parasitology 2007. 37: p. 457-464.

Chapter 2

3D Domain swapping causes extensive multimerisation of human interleukin-10 when expressed *in planta*

Lotte B. Westerhof,☯ Ruud H.P. Wilbers,☯ Jan Roosien, Jan van de Velde, Aska Goverse, Jaap Bakker and Arjen Schots

☯Equal contribution

This work has been published in:
PlosOne 7, 2012; doi:10.1371/journal.pone.0046460

Abstract

Heterologous expression platforms of biopharmaceutical proteins have been significantly improved over the last decade. Further improvement can be established by examining the intrinsic properties of proteins. Interleukin-10 (IL-10) is an anti-inflammatory cytokine with a short half-life that plays an important role in re-establishing immune homeostasis. This homodimeric protein of 36 kDa has significant therapeutic potential to treat inflammatory and autoimmune diseases. In this study we show that the major production bottleneck of human IL-10 is not protein instability as previously suggested, but extensive multimerisation due to its intrinsic 3D domain swapping characteristic. Extensive multimerisation of human IL-10 could be visualised as granules *in planta*. On the other hand, mouse IL-10 hardly multimerised, which could be largely attributed to its glycosylation. By introducing a short glycine-serine-linker between the fourth and fifth alpha helix of human IL-10 a stable monomeric form of IL-10 (hIL-10^{mono}) was created that no longer multimerised and increased yield 30-fold. However, hIL-10^{mono} no longer had the ability to reduce pro-inflammatory cytokine secretion from lipopolysaccharide-stimulated macrophages. Forcing dimerisation restored biological activity. This was achieved by fusing human IL-10^{mono} to the C-terminal end of constant domains 2 and 3 of human immunoglobulin A (Fcα), a natural dimer. Stable dimeric forms of IL-10, like Fcα-IL-10, may not only be a better format for improved production, but also a more suitable format for medical applications.

Introduction

Recombinant DNA technology has revolutionized the production and application of pharmaceutical proteins. Current heterologous production hosts include bacteria, yeasts, insect and mammalian cells, and, more recently, plants. Most expression systems have been rapidly improved in terms of yield and cost efficiency in the last decades. However, these advances have mainly been achieved with antibodies and hormones, which are relatively stable proteins. Recently more attention in optimizing the production of biopharmaceutical proteins is directed to their intrinsic properties that result of post-translational modifications and folding processes. Improved insight in these processes may increase the yield of still poorly expressed proteins.

Many cytokines have a promising therapeutic potential. However, several cytokine families show a short half-life *in vivo* and are poorly expressed in heterologous hosts. Human interleukin-10 (IL-10) is such a cytokine that may be used for treatment of many inflammatory and autoimmune diseases due to its immunosuppressive properties [1, 2]. Generally, IL-10 facilitates the return of the immune system to homeostasis after clearance of antigen and plays an important role in conferring oral tolerance. It exerts its function through reduction of the activity of macrophages, inhibition of antigen presentation by dendritic cells and inhibition of the production of pro-inflammatory cytokines by antigen presenting cells and T lymphocytes [3-6]. The human IL-10 gene encodes a 178 amino acid protein including a N-terminal signal peptide for secretion. An IL-10 monomer consists of six alpha helices (A-F) with two internal disulphide bridges (Cys30-Cys126 and Cys80-Cys132). Two monomers are stabilized into a biologically active dimer by exchanging their C-terminal domains composed of the helices E and F, a process called 3D domain swapping [7-9].

Human interleukin-10 has previously been produced in bacterial systems for medical purposes [10, 11], and in insect and mammalian cells for research purposes. The use of plants as a production platform for IL-10 provides a cheap alternative compared to bacterial, insect and mammalian expression systems. As plants are eukaryotes they can correctly fold and assemble proteins, and are able to perform complex post-translational processes, such as glycosylation. Plants as production hosts for IL-10 offer an extra advantage as they have a low risk of contamination with human pathogens, especially relevant when producing immunosuppressive molecules for medical application. Human IL-10 was produced for the first time *in planta* by stable transformation of a low-alkaloid *Nicotiana tabacum* variety [12]. High transcript levels were contrasted by low protein levels with a maximum of 0.000069% of total soluble protein (TSP). Biological activity of plant-derived human IL-10 was shown *in vitro* and *in vivo* without the need for purification [12, 13]. Yield could be increased to 0.55% of TSP by transient expression of human IL-10 fused to an elastin-like polypeptide combined with retention in the endoplasmic reticulum (ER), but biological activity was not confirmed [14].

From these experiments it was concluded that protein instability is a major bottleneck for human IL-10 production.

We show that 3D domain swapping is an important bottleneck for human IL-10 production in *Nicotiana benthamiana*. Human IL-10 multimerises extensively as was visualised *in situ* using GFP fusions. Domain swapping could be prevented by engineering a stable monomer [15] that regained biological activity through fusion to the Fc portion of IgA, a natural dimer. Identification of this expression bottleneck enabled us to increase yield considerably to levels that approach the economic threshold.

Results

Yield of mouse IL-10 is significantly higher compared to human IL-10

To determine the maximum yield of human (h) and mouse (m) IL-10 in a transient expression system, leaves of 5-6 weeks-old *Nicotiana benthamiana* plants were agro-infiltrated. The expression vector contained the complete native coding sequence of the human or mouse IL-10 gene with or without a 3′ tag coding for a thrombin cleavage site, a 6xHis-tag and the ER retention sequence KDEL (thk) (Figure 1a). Adequate transcription of the constructs was confirmed two and three days post infiltration (dpi) through determination of mRNA levels by means of quantitative PCR. Similar relative transcription levels (around 1000 transcripts of IL-10 per β-actin transcript) were found for all four constructs on both days (Figure 1b). Human and mouse IL-10 yield was determined on 1-6 dpi using a sandwich ELISA (Figure 1c and 1d). Maximum human and mouse IL-10 yield was obtained between 2-4 dpi. Although maximum yields were similar for native and ER-retained IL-10, expression levels of ER-retained IL-10 remained higher over time. Considering average yields from 2 to 5 dpi showed that ER-retained IL-10 resulted in 11 and 10-fold more protein for human and mouse IL-10 respectively. More striking was the observation that even though mRNA transcript levels were comparable, mouse IL-10 yield was significantly higher when compared to human IL-10 (P=0.043), regardless of ER-retention (P=0.003). The average yield of mIL-10 was 19 and 17-fold higher for secreted and ER-retained IL-10 respectively. Retention of a protein in the ER can lead to increased yield due to the presence of chaperones that assist protein folding and the absence of many proteases explaining reduced degradation over time [16]. However, as the difference in expression level between human and mouse IL-10 was significant despite ER-retention another factor than protein degradation supposedly influences yield.

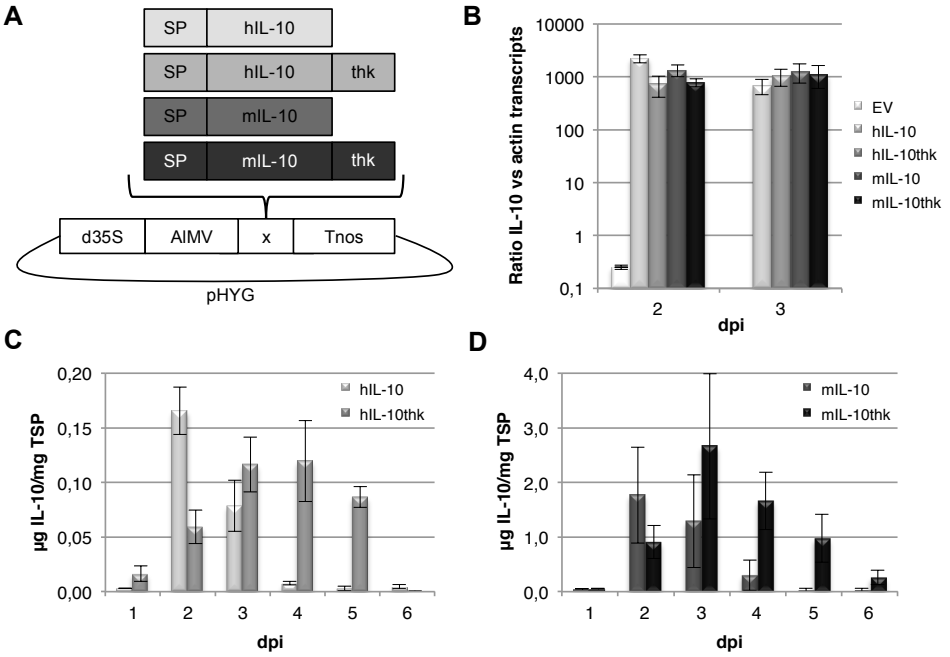


Figure 1 ◆ Expression data of human and mouse IL-10 in transiently transformed *Nicotiana benthamiana* leaves. Use of the thk-tag gives an increasing boost in yield for both human and mouse IL-10 from 2 days post infiltration (dpi). Strikingly, mouse IL-10 yield was significantly higher compared to human IL-10, regardless of ER-retention. Differences in yield could not be explained by differences in mRNA transcript levels. **A** Schematic representation of expression cassettes and vector used. Expressed genes include the native coding sequence of the human (h) or mouse (m) IL-10 gene including signal peptide for secretion (SP) with or without a 3′ tag coding for a thrombin cleavage site, a 6xHis-tag and the ER retention sequence KDEL (thk). All expression cassettes include the 35S promoter of the Cauliflower mosaic virus with duplicated enhancer (d35S), 5′ leader sequence of the Alfalfa mosaic virus RNA 4 (AlMV) and *Agrobacterium tumefaciens* nopaline synthase transcription terminator (Tnos). **B** Relative transcript levels of IL-10 versus actin as determined by Q-PCR on 2 and 3 dpi (*n*=3, error bars indicate standard error). **C/D** Human and mouse IL-10 yield in crude extracts (1 to 6 dpi) in µg per mg total soluble protein (TSP) as determined by ELISA (*n*=3, error bars indicate standard error).

Human IL-10 accumulates in granules

To investigate the cellular fate of human and mouse IL-10 *in planta*, GFP was fused C- and N-terminally to both proteins and expression was monitored by confocal microscopy. Plants expressing hIL-10-GFP showed fluorescent globular structures up to 5 m in size (Figure 2b). These granules resembled Golgi-bodies and were, regardless of their size, highly mobile as they travelled along ER strands. In contrast, mIL-10-GFP showed no or negligible signs of this phenomenon (Figure 2c). For both proteins, fluorescence was observed in the nuclear envelope, the ER and, putatively, the apoplast as expected for a secretory protein (Figure 2a). The results of C-terminal GFP fusions harbouring the native IL-10 signal peptides, like the

unfused variants, are shown. N-terminal GFP fusions gave similar results (data not shown). Formation of granules may be caused by aggregation of folding intermediates or extensive multimerisation of human IL-10. Since mouse IL-10 did not show granules as observed for human IL-10, we assumed that formation of granules hinders yield of biologically active IL-10.

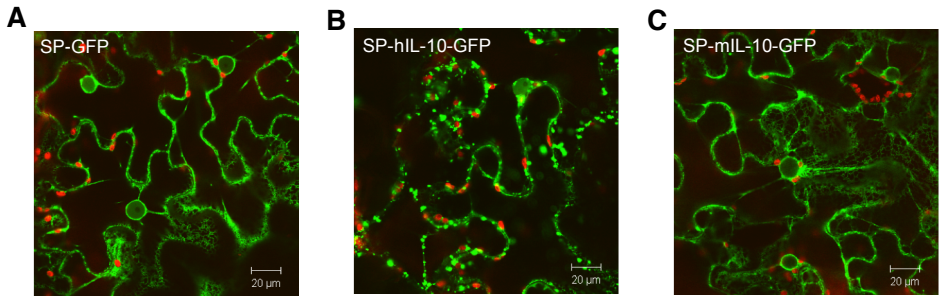


Figure 2 ◆ Whole mount confocal microscopy output of leaves expressing human or mouse IL-10 fused to GFP. Highly mobile globular granules of up to 5µm in size were observed traveling along cytoplasmic and/or ER strands for SP-hIL-10-GFP only. **A** GFP preceded by the *Arabidopsis thaliana* chitinase signal peptide for secretion (SP-GFP). **B/C** The native open reading frame of human (h) and mouse (m) IL-10 including the native signal peptide (SP) with GFP fused C-terminally.

Glycosylation of IL-10 hinders granulation

Human and mouse IL-10 have a homology of 73% at amino acid level. The most apparent difference between human and mouse IL-10 during post-translational processing is the glycosylation of mouse IL-10. Both proteins have one potential N-glycosylation site (Asn134) that is not glycosylated, while mouse IL-10 has yet another site that is glycosylated (Asn29). As glycosylation can stabilize a protein and influence protein folding by mediating interaction with chaperones, it may explain the difference in protein processing between human and mouse IL-10. To investigate the influence of glycosylation of IL-10 on granulation, the glycosylation site of mouse IL-10 was introduced in human IL-10 and removed from mouse IL-10. In both cases this was done by a single nucleotide mutation causing the serine of human IL-10 on position 29 to change into an asparagine and vice versa (hIL-10^{S29N} and mIL-10^{N29S}). Figures 3a and 3d show a micrograph of GFP fused C-terminally to hIL-10^{S29N} and mIL-10^{N29S}, respectively. Granulation of hIL-10^{S29N}-GFP was greatly reduced, however, never completely absent. For mIL-10^{N29S}-GFP granulation was induced, however, never reached the same extent as seen for hIL-10-GFP.

To confirm the presence of a N-glycan on hIL-10^{S29N} and the absence on mIL-10^{N29S}, all non-GFP fused variants were analysed by Western blot (Fig 3b and 3e). *E. coli* produced human and mouse IL-10, hence non-glycosylated, had the same molecular mass of 18 kDa

when compared to plant-produced hIL-10 and mIL-10^{N29S}. The molecular weight of mIL-10 and hIL-10^{S29N} was, as expected upon single N-glycosylation, approximately 1.5 kDa higher. Both mIL-10 and hIL-10^{S29N} samples also showed a small proportion of non-glycosylated IL-10.

The effect of glycosylation on human and mouse IL-10 yield was compared. ELISA indicated that removal of the glycosylation of mouse IL-10 decreased yield on average 4-fold, however, this was found not to be significant (Figure 3c). Introducing a glycosylation site in human IL-10 did not lead to a yield increase (Figure 3f), however, for hIL-10 the band intensities on Western blot (Figure 3b) showed that an increased amount of glycosylated hIL-10 was extracted. This discrepancy might be explained by interference of the glycan on Asn-29 of hIL-10^{S29N} with the binding of the monoclonal antibodies used in ELISA.

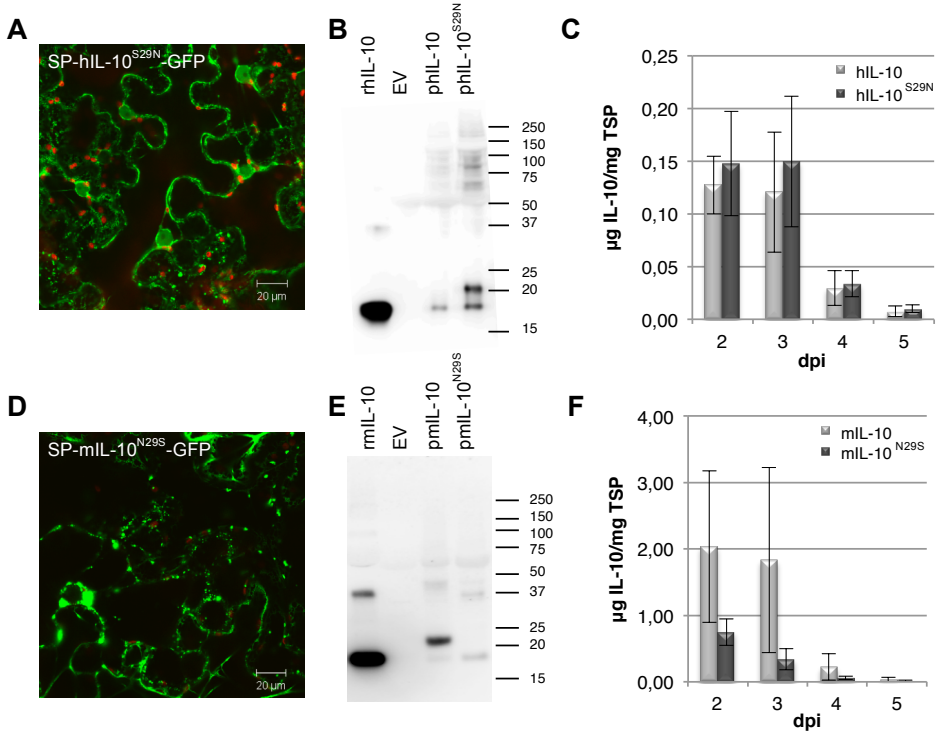


Figure 3 ◆ Analysis of the effect of N-glycosylation at Asn29 on granulation. Glycosylation of IL-10 plays a role in preventing granulation. **A/D** Whole mount confocal microscopy output of leaves expressing GFP fused C-terminally to human (h) and mouse (m) IL-10 including native signal peptide (SP) and with introduced (S29N) or removed (N29S) glycosylation site, respectively. **B/E** Western blot analysis under reducing conditions of plant produced (p) hIL-10 and mIL-10 with and without glycosylation site. As controls, empty vector (EV) and 50 ng recombinant (r) *E. coli* produced hL-10 and mIL-10 were used. A molecular weight marker is indicated in kDa. **C/F** Yield of hIL-10 and mIL-10 with and without glycosylation site in crude extracts 2 to 5 days post infiltration (dpi) as determined by ELISA (*n*=4, error bars indicate standard error).

Glycosylation does not influence biological activity

The possible effects of glycosylation on the biological activity of plant-produced human and mouse IL-10 was next assessed. The capacity of the different IL-10 variants to reduce Tumor Necrosis Factor-alpha (TNF-α) expression by human (THP-1) and mouse (RAW264.7) macrophages upon stimulation by lipopolysaccharide from *E. coli* was determined. Figure 4 shows the percentage inhibition of TNF-α secretion by macrophages when compared to the empty vector control. As expected, glycosylated as well as non-glycosylated mouse IL-10 suppressed the secretion of pro-inflammatory TNF-α from mouse macrophages, confirming that glycosylation is not necessary for mouse IL-10 activity [3]. Strikingly, our data show that glycosylated human IL-10 is as active as its native non-glycosylated form on both human and mouse cells. Apparently, the structure of human IL-10 is not negatively influenced by the glycosylation event. Neither non-glycosylated nor glycosylated mouse IL-10 was active on human cells.

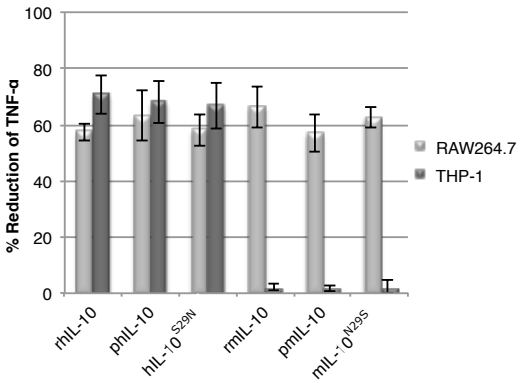


Figure 4 ◆ Biological activity of human and mouse IL-10 variants on human and mouse macrophages. Plant produced (p) and recombinant (r) *E. coli* produced human (h) or mouse (m) IL-10 were calibrated to contain the same amount of IL-10 as well as total soluble protein by using the empty vector control. Human (THP-1) and mouse (RAW264.7) macrophages were then pretreated with 10 ng/ml hIL-10 or mIL-10 for 20 min and subsequently stimulated with 1 µg/ml *E. coli* lipopolysaccharide. Tumor Necrosis Factor-alpha (TNF-α) expression was determined by ELISA and IL-10 activity is indicated as the percentage of inhibition of TNF-α expression as compared to the empty vector control (n=3, error bars indicate standard error).

Granulation of human IL-10 is caused by extensive multimerisation

Literature describes IL-10 to be a 3D domain swapping protein, a term used to describe a process wherein two or more protein chains exchange identical structural elements or "domains" [9, 17]. Two IL-10 monomers dimerise by exchanging their E-F helices. However, when the concentration of a potentially 3D domain swapping protein is high, swapping of

domains does not have to be limited to two partners [18]. For example, IL-10 can provide its A-D helices to one partner, while giving its E-F helices to another partner. This can create a chain reaction that can, in theory, continue until a stable form is reached. To determine if human IL-10 granulation was triggered by extensive multimerisation, a flexible linker was introduced between α-helices D and E of human IL-10. This allows helices E-F of one human IL-10 molecule to fold into its own A-D α-helices, creating a stable monomer (hIL-10^{mono}) as designed by Josephson and co-workers [15]. Figure 5a shows three cartoons of the expected human IL-10 (I) dimer, (II) monomer and (III) stable monomer conformations. When the monomeric form of hIL-10 was fused to GFP no signs of granulation was observed (Figure 5b). hIL-10^{mono}-GFP was detected in the ER and, putatively, the apoplast. Unlike hIL-10-GFP and mIL-10-GFP, fluorescence was also observed in the cytoplasm and nucleoplasm. This may be indicative of partial failure of the protein to be taken up into the secretory pathway. However, on Western blot a band of 21 kDa, the expected size for hIL-10^{mono} with signal peptide, was never observed. Therefore, hIL-10^{mono}-GFP must be expelled from the secretory pathway and enter the nucleoplasm by either diffusion or active transport. Active transport into the nucleus after expulsion from the ER was demonstrated with a secretory GFP fusion with the P-domain of the chaperone calreticulin [19].

Western blot analysis demonstrated that hIL-10^{mono} was indeed present as a monomer, whereas plant and *E. coli* produced hIL-10 also revealed bands corresponding to dimeric (36 kDa) and multimeric hIL-10 (Figure 5c). In repetitive experiments a small proportion of dimeric hIL-10^{mono} was observed occasionally, but never larger multimers.

Analysing average yields of hIL-10 and hIL-10^{mono} by ELISA revealed that the hIL-10^{mono} protein level was significantly higher (P=0.048) resulting in 16-fold more protein. Maximum yield obtained with hIL-10^{mono} was 3,2 µg IL-10/mg (0.32%) TSP at 3 dpi (Figure 5d). However, when hIL-10^{mono} was tested for biological activity by assessing its ability to suppress TNF-α expression by LPS stimulated macrophages, it appeared not to be functional (Figure 6a), which was expected as IL-10 receptor binding studies suggest that the dimeric form of IL-10 confers biological activity [20]. This experiment was repeated several times and on some occasions slight activity of the hIL-10^{mono} was observed, however, always less than *E. coli* produced hIL-10. It appeared that the biological activity coincided with the proportion of dimeric hIL-10^{mono} seen on Western blot. It is likely that the extent of dimerisation and multimerisation of IL-10 is dependent on sample treatment such as freeze-thaw cycles, temperature and pH of the solution after extraction.

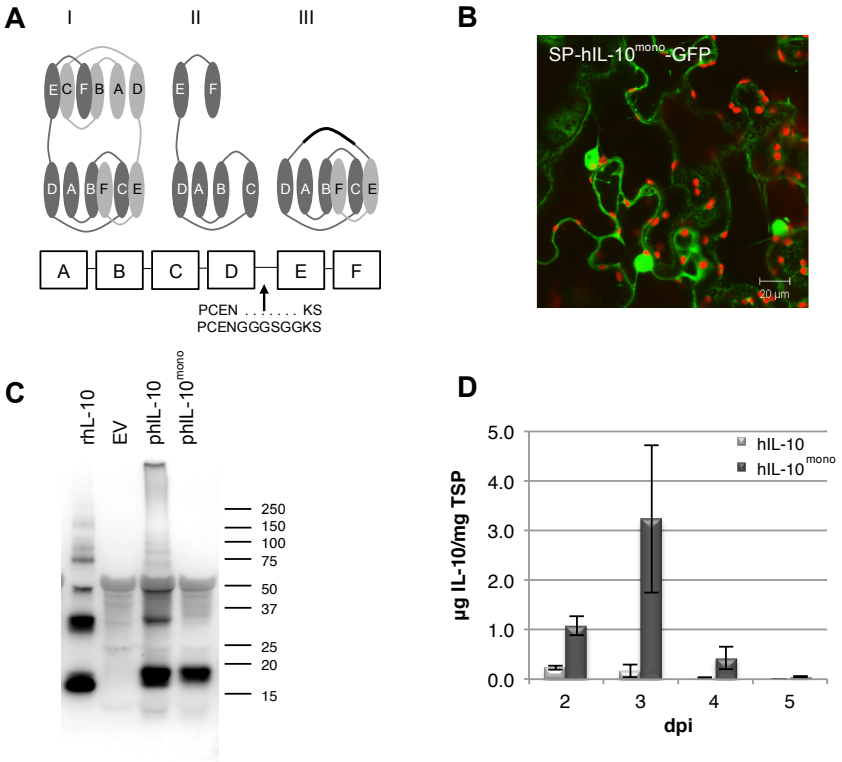


Figure 5 ◆ Analysis of expression of a stable monomeric form of human IL-10. A stable monomeric form of human IL-10 (hIL-10^{mono}) does not granulate and yield increases 30-fold. **A** Three cartoons illustrating the human IL-10 (I) dimer, (II) monomer and (III) stable monomer structure, as well as a schematic representation of the human (h) IL-10 alpha helices A-F. Helices are represented by ovals, whereby a fragment of the amino acid sequence and the location of insertion of the small GS-linker is indicated. **B** Whole mount confocal microscopy output of GFP fused C-terminally to hIL-10^{mono} including native signal peptide (SP). **C** Western blot analysis under non-reducing conditions of plant produced hIL-10 and hIL-10^{mono}. As controls, empty vector (EV) and 50 ng recombinant (r) *E. coli* produced hL-10 were used. A molecular weight marker is indicated in kDa. **D** Yield of hIL-10 and hIL-10^{mono} in crude extracts 2 to 5 days post infiltration as determined by ELISA (n=3, error bars indicate standard error). Average yield of hIL-10^{mono} was significantly higher compared to hIL-10.

Biological activity of hIL-10^{mono} was restored by fusion to Fcα

To re-establish biological activity, dimerisation of hIL-10^{mono} was forced by fusion to the C-terminus of constant domains 2 and 3 of human immunoglobulin A2m1 (Fcα). As a reference, a construct with unmodified hIL-10 fused to Fcα was also made. Figure 6a shows the percentage inhibition of TNF-α secretion by macrophages when compared to the empty vector control. As mentioned before, hIL-10^{mono} was found to have reduced biologically activity, but

by fusion to Fcα we were able to restore its biological activity (*P*=0.0230). Also, fusion of native hIL-10 to Fcα slightly increased biological activity, but not significant. While the activity of the monomeric IL-10 Fcα fusion had no significant difference with *E. coli* or plant produced native hIL-10, a significant difference was found with the native IL-10 Fcα fusion (*P*=0.0375).

Both fusion proteins showed similar accumulation patterns as observed for the unfused variants of IL-10. However, yield of Fcα-hIL-10^{mono} was slightly lower compared to the unfused variant on dpi 3, but was found not to be significant (Figure 6b). This indicates that fusion with a stable fusion partner does not ensure increase in yield of IL-10, as was the case with fusion to elastin-like polypeptides [14, 21].

Western blot analysis using antibodies raised against hIL-10 showed that the largest proportion had the expected size of dimeric Fcα-hIL-10 (95 kDa) for both hIL-10 and hIL-10^{mono} fusions, but also bands at 200 kDa and 300 kDa were observed, which could again indicate the presence of multimers of these fusion proteins (Figure 6c). Using antibodies raised against IgA revealed, in addition to all bands corresponding with the IL-10 blot, an extra band around 50 kDa (Figure 6d). Because this band does not appear on the IL-10 blot, it is probably a cleavage product whereby IL-10 is removed. Specifically, a dimeric Fcα protein without hIL-10 would have an expected size of 50 kDa. Yet, no separated IL-10, expected at 18 or 36 kDa for monomeric and dimeric IL-10 respectively, was seen and must therefore be completely degraded.

Discussion

Here we show that extensive multimerisation of human IL-10 is the most important factor limiting yield and not protein instability. Human IL-10 has previously been expressed in plants [12, 14, 21-23]. The highest yield of 0.55% of TSP was obtained by transient expression of human IL-10 fused to an elastin-like polypeptide while retained in the ER. It was suggested that retention in the protein friendly environment of the ER protects the protein against degradation and that elastin-like polypeptides might prevent aspecific aggregation by association with chaperones [12, 14]. These data suggested a role for protein instability or inefficient post-translational processing as a limiting factor for production of human IL-10. Protein instability as a bottleneck for production of IL-10 would not be surprising, as cytokines are known for their short half-life *in vivo*.

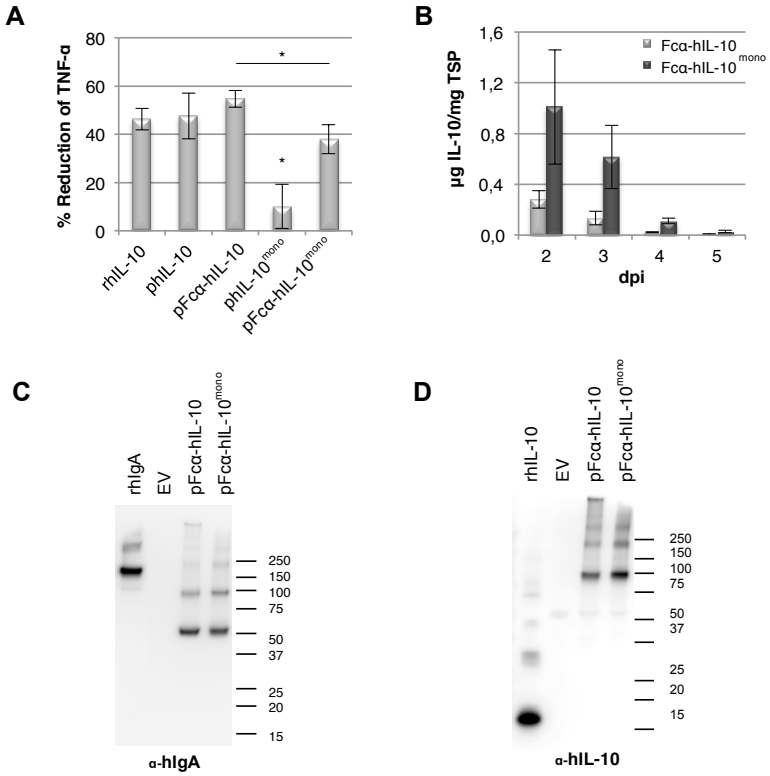


Figure 6 ◆ Analysis of biological activity and expression of human IL-10 and human IL-10^{mono} fused to Fcα. Forced dimerization of human IL-10^{mono} restores biological activity. **A** Bioactivity assay of hIL-10^{mono} and Fcα-hIL-10 fusion proteins on mouse macrophages (RAW267.4). Plant produced (p) and recombinant (r) *E. coli* produced hIL-10 were calibrated to contain the same amount of IL-10 as well as total soluble protein by using the empty vector control. Cells were then pretreated with 50 ng/ml hIL-10 for 20 min and subsequently stimulated with 1 µg/ml *E. coli* lipopolysaccharide. Tumor Necrosis Factor-alpha (TNF-α) expression was determined by ELISA and IL-10 activity is indicated as the percentage of inhibition of TNF-α expression as compared to the empty vector control (n=4, error bars indicate standard error). Significant difference (P<0.05) between samples is indicated with an asterisk, where biological activity of hIL-10^{mono} was significantly different to all other samples. **B** Yield of human (h) IL-10 in crude extracts 2 to 5 days post infiltration as determined by ELISA (n=3, error bars indicate standard error). **C/D** Western blot analysis under non-reducing conditions of plant produced (p) Fcα-hIL10 and Fcα-hIL-10^{mono} using an antibody raised against human IL-10 and human IgA for visualization, respectively. As controls, empty vector (EV), 50 ng recombinant (r) *E. coli* produced hIL-10 and 10 ng purified hIgA were used. A molecular weight marker is indicated in kDa.

When we studied the cellular fate of human and mouse IL-10 *in planta* using GFP fusions, we revealed that human IL-10 accumulates in granules, while mouse IL-10 does not. The most evident difference between human and mouse IL-10 is the N-glycosylation of mouse IL-10. Both proteins have one potential N-glycosylation site (Asn134) that is not glycosylated, while mouse IL-10 has yet another site that is glycosylated (Asn29). Removal of the Asn29 glycosylation site resulted in increased granulation of mouse IL-10, whereas introduction of this glycosylation site in human IL-10 prevented granulation to a large extent. As glycosylation can aid in protein folding and stabilization, we hypothesized that the granules were in fact very large aggregates, potentially compartmentalised as a cell protective mechanism.

Aggregation can be due to protein instability and/or misfolding resulting in random hydrophobic patch interaction or, in the case of IL-10, due to extensive multimerisation. Literature describes IL-10 to be a 3D domain swapping protein, a term used to describe a process wherein two or more protein chains exchange identical structural elements or "domains" [9, 17]. IL-10 monomers form biologically active dimers by exchanging their E-F helices. However, at high concentrations of IL-10 it would be possible that the domain swapping of IL-10 is not limited to two partners and causes a chain reaction that, in theory, could go on unlimited until a stable form is reached. To determine whether human IL-10 granulation was due to protein instability or due to 3D domain swapping, the behaviour of a previously developed stable monomeric form of human IL-10 [15] was studied *in planta*. This stable monomeric form of human IL-10 has a short glycine-serine (GS)-linker between helices D and E allowing the helices E-F to fold back into its own hydrophobic core created by the A-D helices (human IL-10^{mono}). No granulation was observed when human IL-10^{mono} was fused with GFP. Multimeric or high molecular weight bands of human IL-10^{mono} were also not detected by Western blot. Furthermore, yield was increased on average 16-fold by this minor modification of the human IL-10 protein. This expression level was comparable to the highest yield obtained for human IL-10 so far, however, without retention in the ER or fusion with a stable protein partner. Taken together, we show that not protein instability, but extensive multimerisation of human IL-10, due to its intrinsic 3D domain swapping characteristic, is the major limiting factor for yield.

Bennett and colleagues introduced the term 3D domain swapping [17] and up to date almost 300 proteins have been reported with 3D domain swapping ability [24]. Besides IL-10, the cytokines Granulocyte Macrophage Colony-Stimulating Factor (GM-CSF), Interleukin-5 (IL-5) and Interferon-beta (IFN-β) are also described as 3D domain swapping proteins [9] and have been heterologously expressed in several platforms. Upon expression in *E. coli* the formation of inclusion bodies were described for these four cytokines [25-28]. Although formation of inclusion bodies could be indicative of multimerisation, they are a relatively common feature when expressing eukaryotic proteins in *E. coli* due to aggregation of folding intermediates. However, in the case of IFN-β and IL-5, experiments also demonstrated the need for slow

removal of denaturing agents to prevent aggregation to reoccur after solubilisation of inclusion bodies [25, 27]. To our knowledge, IL-5 has not been produced in plants before and expression of human IFN-β in lettuce leaves resulted in poor yields [29]. On the other hand, many research groups have produced human GM-CSF *in planta*. Interestingly, human GM-CSF secreted by a tobacco cell culture could be stabilized in the growth medium by addition of bovine serum albumin, gelatin or salt, all methods that could inhibit protein aggregation [30, 31]. Next to that, when produced in rice seeds a morphological change in both plant protein body type I and the ER were observed together with high molecular weight variants of human GM-CSF from 50 to 120 kDa [32, 33]. Menassa and colleagues also described a 200 kDa weight variant of human IL-10 when expressed stably in a tobacco species [12]. For both human IL-10 and human GM-CSF yield from rice seeds was relatively high compared to the leaf-based and cell culture systems, however, extraction from rice seeds was only efficient using reducing agents [23, 34]. Unfortunately, use of reducing agents during extraction would increase production costs again, as the protein would need to be chemically refolded. There are many indications that extensive multimerisation has also hampered the production of other 3D domain swapping cytokines, and may represent a more common mechanism affecting yield than thus far anticipated. In addition, it is unclear how 3D domain swapping influences biopharmaceutical proteins during formulation, storage and administration.

Although we could increase the yield of human IL-10 significantly by expression of a stable monomeric form of IL-10, this form was not biologically active, as determined by its ability to reduce TNF-α secretion by lipopolysaccharide-stimulated macrophages. Even though human IL-10^{mono} has all the receptor binding sites, the affinity may be reduced. Josephson and co-workers did show biological activity of human IL-10^{mono}based on its ability to induce proliferation of a B cell line, however, a 10-fold higher amount of human IL-10^{mono} was needed to obtain the same effect as recombinant human IL-10 [15]. Stable monomeric forms of IL-5 were also created and shown to have reduced receptor affinity. Nevertheless, they did show biological activity with concentration dependence close to that of the wild type [35]. Such stabilized biologically active forms of 3D domain swapping cytokines will ensure that the protein stays in its biologically active conformation, which may have great benefit for formulation and administration to patients.

To mimic the natural dimerization of IL-10 while still preventing multimerisation, human IL-10^{mono} was fused to the naturally dimerising Fc portion of IgA2m1 (Fcα), which restored the ability to suppress TNF-α in stimulated macrophages. As well as ensuring dimeric conformation, it is likely that fusion with Fcα increases *in vivo* half-live of IL-10, which could be beneficial for efficacy. Our construct is comparable to that of mouse IL-10 fused to the N-terminus of a non-cytolytic version of the Fc portion of mouse IgG2a, which was shown to be functional *in vivo* with a prolonged efficacy due to its increased circulating half-life [36]. However, an increased half-life of IL-10 could also lead to unwanted systemic side effects,

such as a compromised immune system. By using complete antibodies it may be possible to circumvent these side effects. Through the antigen binding capacity of an antibody with IL-10 fused C-terminally, specific cells or tissues can be targeted enabling IL-10 therapy for a variety of diseases, as was already demonstrated by Schwager and co-workers [37]. In disease therapy, use of a human fusion protein is desired to limit the chance of an immune response against the therapeutic agent. Also, in inflammatory disease therapy the use of IgA may be preferred over other antibody isotypes, as it does not activate the classical complement pathway. Thus, a stable molecule that ensures a biologically active conformation and targeting, such as Fcα-IL-10, may be a more suitable format for medical applications and may prove effective in therapy of inflammatory diseases.

Experimental procedures

Construction of expression cassettes and vector

The complete native open reading frames (ORF) of human and mouse IL-10 were amplified from the MegaMan Human Transcriptome cDNA library (Stratagene) and FirstChoice PCR-Ready Mouse Spleen cDNA library (Ambion), respectively, using oligonucleotides indicated in Table 1. Sense and antisense oligonucleotides included NcoI and KpnI restriction sequences, respectively, for subsequent cloning steps. Introduction of the NcoI restriction site resulted in the addition of two extra amino acids to the signal peptide at the N-terminus. All oligonucleotides used for subsequent reamplification, mutagenesis and insertion can be found in Table 1.

To create hIL-10thk and mIL-10thk, both hIL-10 and mIL-10 ORFs were reamplified, resulting in the removal of the stop codon and replacement of the KpnI by a NotI restrictrion site. The thk coding DNA fragment could be added in frame. To remove and introduce the glycosylation site of mouse and human IL-10 respectively, the QuickChange XL kit (Stratagene) was used. For insertion of the 6 amino acid glycine-serine spacer overlap extension PCR was performed in combination with original hIL-10 oligonucleotides.

The DNA fragment encoding the enhanced green fluorescent protein (eGFP) was reamplified from gateway vector pK7FWG2 [38]. Constructs with eGFP fused C-terminally to IL-10 were composed of the IL-10 fragments ending with NotI (as described for the h/mIL-10thk constructs) the “NotI-NcoI” oligo and the eGFP fragment.

The constant domains 2 and 3 of human immunoglobulin A2m1 heavy chain including a N-terminal IgA signal peptide and a C-terminal linker sequence was synthetically constructed by GeneArt and fused N-terminally to reamplified SpeI-hIL-10 and SpeI-hIL-10mono fragments. All constructs were placed under the control of the 35S promoter of the *Cauliflower mosaic virus* with duplicated enhancer (d35S) and the *Agrobacterium tumefaciens*

nopaline synthase transcription terminator (Tnos). A 5′ leader sequence of the Alfalfa mosaic virus RNA 4 (AlMV) was included between the promoter and construct to boost translation. All elements were present in pRAP35 (or pUCAP35S [39]) from which expression cassettes were digested with AscI and PacI and ligated into a modified version of the expression vector pMDC32 [40], renamed pHYG. The pHYG vector resulted upon removal of the Gateway recombination sequences by digestion with EcoRI and HindIII and replacement with a DNA fragment (oligos "pHYG adaptor") including AscI and PacI restriction sites for insertion of the expression cassettes. All construct sequences were confirmed by sequencing (BaseClear) at the expression vector stage. The expression vectors were subsequently transformed to *Agrobacterium tumefaciens* strain MOG101 for plant expression.

Table 1 ◆ Oligonucleotides for construct re-amplification, mutagenesis and insertion.

Gene fragment	Function	*5′->3′* Sense oligonucleotides	*5′->3′* Antisense oligonucleotides
hIL-10	Isolation hIL-10 NcoI/KpnI	ccatgggcATGCACAGCTCAGCACTGCTCT	cggtaccTCAGTTTCGTATCTTCATTGTCA
mIL-10	Isolation mIL-10 NcoI/KpnI	ccatggccATGCCTGGCTCAGCACTGCTAT	ccggtaccTTAGCTTTTCATTTTGATCATC
pHYG adaptor	MCS including AscI/PacI	aattcggcgcgcctacgcgtaaggacgagctctgaggtacctctagattaattaaa	agcttttaattaatctagaggtacctcagagctcgtccttacgcgtaggcgcgccg
hIL-10-NotI	Removal stop codon, addition NotI	ccatgggcATGCACAGCTCAGCACTGCTCT	ggcggccgcGTTTCGTATCTTCATTGTCATG
mIL-10-NotI	Removal stop codon, addition NotI	ccatggccATGCCTGGCTCAGCACTGCTAT	ggcggccgcGCTTTTCATTTTGATCATCATG
thk	Thrombin, 6xHIS, KDEL	ggccgcattagttcctcgtggttctgctagccatcaccatcaccatcacaaagatgagctatgacgtacgggtac	ccgtacgtcatagctcatctttgtgatggtgatggtgatggctagcagaaccacgaggaactaatgc
eGFP	Isolation GFP NcoI/KpnI	gtcgacggatccATGGTGAGCAAGGGCGAGGAGCTGTTC	aggtaccTTAGCTCATGACTGACTTGTAGAGCTCGTCCATGCCGAGAG
NotI-NcoI	Linker C-terminal GFP fusion	ggccgctgcagtcgacggatc	catggatccgtcgactgcagc
mIL-10^{N29S}	Removal glycosylation site	GCCGGGAAGACAATAgCTGCACCCACTTCCC	GGGAAGTGGGTGCAGcTATTGTCTTCCCGGC
hIL-10^{S29N}	Introduction glycosylation site	CCCAGTCTGAGAACAaCTGCACCCACTTCCCAG	CTGGGAAGTGGGTGCAGtTGTTCTCAGACTGGG
hIL-10^{mono}	Insertion GGGSGG-linker	AAACggtggcggatctgggggtAAGAGCAAGGCCGTGGAGCAGGTGAA	CTTaccccagatccgccaccGTTTTCACAGGGAAGAAATCGATGA
SpeI-hIL-10	Removal SP, addition SpeI	accatggactagtAGCCCAGGCCAGGGCAC	cggtaccTCAGTTTCGTATCTTCATTGTCA
h/m IL-10	Q-PCR analysis	ATGATCCAGTTTTACCTGG	AGAAATCGATGACAGCG
Nb β-actin	Q-PCR analysis	CCAGGTATTGCCGATAGAATG	GAGGGAAGCCAAGATAGAGC

Native sequences in capitals, added/mutated sequences in small and restriction sites are underlined. h; human, IL-10; interleukin-10, MCS; multiple cloning site, m; mouse, Nb; *Nicotiana* benthamiana, SP; signal peptide for secretion, thk; thrombin-6xHIS-KDEL tag.

***Agrobacterium tumefaciens* transient transformation assay**

Agrobacterium tumefaciens clones were cultured overnight (o/n) at 28°C in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH7.0) containing 50 µg/ml kanamycin and 20 µg/ml rifampicin. The optical density (OD) of the o/n cultures was measured at 600 nm and used to inoculate 50 ml of LB medium containing 200 µM acetosyringone and 50 µg/ml kanamycin with x µl of culture using the following formula: x = 80000/(1028*OD). OD was measured again after 16 hours and the bacterial cultures were centrifuged for 15 min at 2800 x*g*. The bacteria were resuspended in MMA infiltration medium (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 µM acetosyringone till an OD of 1 was reached. After 1-2 hours incubation at room temperature, the two youngest fully expanded leaves of 5-6 weeks old *Nicotiana benthamiana* plants were infiltrated completely. Infiltration was performed by injecting the *Agrobacterium* suspension into a *Nicotiana benthamiana* leaf at the abaxial side using a 1 ml syringe. Infiltrated plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) and infiltrated leaves were harvested at selected time points.

Total soluble protein extraction

Leaves were immediately snap-frozen upon harvesting, homogenized in liquid nitrogen and stored at -20C until use. Homogenized plant material was ground in ice-cold extraction buffer (50mM phosphate-buffered saline (PBS) pH=7.4, 100 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 0.1% v/v Tween-20, 2% w/v immobilized polyvinylpolypyrrolidone (PVPP)) using 2 ml/g fresh weight. Crude extract was clarified by centrifugation at 16.000 x*g* for 5min at 4C and supernatant was directly used in an ELISA and BCA protein assay. For western blotting and biological activity assays, above-mentioned protein extraction would be followed by desalting using a G25 Sephadex column and filter sterilization (0.22 µm; Millipore Corporation).

Quantification of human and mouse IL-10 mRNA levels

mRNA was isolated from homogenized plant material using the RNAeasy Plant Mini Kit (Qiagen) according to suppliers protocol. A Turbo DNaseI (Ambion) treatment was included to remove any residual DNA. cDNA was synthesised using the SuperScript·III First-Strand Synthesis System (invitrogen) according to suppliers protocol. Samples were analysed by quantitative PCR in triplo using ABsolute SYBR Green Fluorescein mix (Thermo Scientific). *Nicotiana benthamiana* β-actin was used as a reference gene. Oligonucleotides used can be found in Table 1. Relative transcript levels of IL-10 versus actin were determined by the Pfaffl method [41].

Quantification of human and mouse IL-10 protein levels

IL-10 protein concentration in crude plant extract was determined by ELISA. Human and mouse IL-10 ELISA Ready-SET-Go!® kits (eBioscience) were used according to suppliers protocol using the model 680 plate reader (BioRad) to measure the OD at 450 nm with correction filter of 690 nm. For sample comparison the total soluble protein (TSP) concentration was determined by the BCA method (Pierce) according to supplier's protocol using bovine serum albumin (BSA) as a standard.

Protein analysis by western blot

Soluble plant proteins were separated under reducing or non-reducing conditions by SDS-PAGE on a NuPAGE® 12% Bis-Tris gel (Invitrogen). Recombinant *E. coli* produced human or mouse IL-10 (R&D Systems) or IgA1 (Invivogen) was used as a control. Proteins were transferred to an Invitrolon™ PVDF membrane (Invitrogen) by semi-dry blotting procedure. Thereafter the membrane was blocked in PBST-BL (PBS containing 0.1% v/v Tween-20 and 5% w/v non-fat dry milk powder) for 1 hour at room temperature, followed by overnight incubation with hIL-10 specific goat polyclonal antibody (R&D systems), mIL-10 specific rat monoclonal antibody (BioLegend) or IgA specific HRP conjugated goat polyclonal antibody (Sigma) in PBST at 4°C. The membrane was washed 5 times with 5 min intervals in PBST. For h/mIL-10 specific western blots the procedure was continued with a 1 hour incubation at room temperature with HRP conjugated secondary antibodies (Jackson ImmunoResearch) in PBST and washed again as described before. Finally, the SuperSignal West Femto substrate (Pierce) was used to detect HRP-conjugated antibodies.

Confocal microscopy

Plants were agro-infiltrated with expression cassettes encoding C-terminal fusions of GFP to human and mouse IL-10 as described previously. Leaves were taken from the plant and small sections were examined from the abaxial side using a Zeiss LSM510 confocal laser-scanning microscope in combination with an argon ion laser supplying a 488 nm wavelength.

Biological activity assay

The monocyte THP-1 and monocyte/macrophage RAW264.7 cell lines were purchased from the American Type Culture Collection and cells were maintained at 37°C with 5% CO_2 THP-1 cells were cultured in RPMI-1640 medium containing containing 4 mM L-glutamine, 25 mM HEPES and supplemented with 10% fetal calf serum, 50 U/ml penicillin and 50 µg/ml streptomycin. THP-1 monocytes were differentiated into macrophages for 4 days using 30 ng/ml PMA at a density of $3x10^5$ cells/ml in 96 well plates. Cells were allowed to rest for 2-3 days in medium without PMA prior to bioassays. RAW264.7 cells were cultured in DMEM containing 4 mM L-glutamine, 25 mM HEPES and supplemented with 10% newborn calf

serum, 50 U/ml penicillin and 50 µg/ml streptomycin. Cells were sub-cultured every 2-3 days, whereby the cells were harvested by gently disrupting the monolayer with a cell scraper. For cell-based assays cells were seeded in 96 well plates at a density of $5x10^4$ cells/well and allowed to rest overnight. For bioassays cells were pre-treated with 10-50 ng/ml plant produced or recombinant *E. coli* produced human or mouse IL-10 (R&D Systems) in plant extract for 20 min, and subsequently stimulated with 1 µg/ml of lipopolysaccharide (Sigma). After overnight incubation supernatants were analysed with the mouse TNF-α ELISA Ready-Set-Go! Kit (eBioscience) according to the supplier's protocol.

Data analysis

All data shown in the figures indicate the average of at least three biological replicates (*n*) that were determined by at least three technical replicates. In the figure legends *n* is indicated and error bars indicate the standard error. Significant differences between samples were calculated using the student's t-test and regarded as significant when $P<0.05$. Significant differences in expression levels between constructs were calculated by using the whole data set from dpi 2 to 5 and are indicated in the main text as well as the figure legends. When comparing biological activity between proteins significant differences are indicated in the figure by asterisks.

Acknowledgements

This paper was financially supported in part by Synthon (Nijmegen, The Netherlands). The authors declare no conflict of interest. We would like to thank Gerry Ariaans for all helpful discussions and Kevin van der Eijken, Francisco Marques, Bart Nijland, Debbie van Raaij, Sonja Warmerdam and Edgar Wills for their input in the practical work of this paper.

References

1. Asadullah, K., W. Sterry, and H.D. Volk, *Interleukin-10 Therapy - Review of a New Approach.* Pharmacological Reviews, 2003. 55(2): p. 241-269.
2. O'Garra, A., et al., *Strategies for use of IL-10 or its antagonists in human disease.* Immunological Reviews, 2008. 223(1): p. 114-131.
3. Moore, K.W., et al., *Interleukin-10.* Annual Review of Immunology, 1993. 11(1): p. 165-190.
4. Moore, K.W., et al., *Interleukin-10 and the interleukin-10 receptor.* Annual Review of Immunology,, 2001. 19: p. 683-765.
5. Mocellin, S., et al., *The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle.* Cytokine & Growth Factor Reviews, 2004. 15(1): p. 61-76.
6. Saraiva, M. and A. O'Garra, *The regulation of IL-10 production by immune cells.* Nature Reviews Immunology, 2010. 10(3): p. 170-181.
7. Walter, M.R. and T.L. Nagabhushan, *Crystal-Structure of Interleukin-10 Reveals an Interferon Gamma-Like Fold.* Biochemistry, 1995. 34(38): p. 12118-12125.
8. Zdanov, A., et al., *Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon gamma.* Structure, 1995. 3(6): p. 591-601.
9. Liu, Y. and D. Eisenberg, *3D domain swapping: As domains continue to swap.* Protein Science, 2002. 11(6): p. 1285-1299.
10. Colombel, J.F., et al., *Interleukin 10 (Tenovil) in the prevention of postoperative recurrence of Crohn's disease.* Gut, 2001. 49(1): p. 42-6.
11. Steidler L, V., K, *Genetically modified Lactococcus lactis: novel tools for drug delivery.* International Journal of Dairy Technology, 2006. 59(2): p. 140-146.
12. Menassa R., et al., *A self-contained system for the field production of plant recombinant interleukin-10.* Molecular Breeding, 2001. 8: p. 177-185.
13. Menassa, R., et al., *Therapeutic effectiveness of orally administered transgenic low-alkaloid tobacco expressing human interleukin-10 in a mouse model of colitis.* Plant Biotechnology Journal, 2007. 5: p. 50-59.
14. Patel, J., et al., *Elastin-like polypeptide fusions enhance the accumulation of recombinant proteins in tobacco leaves.* Transgenic Research, 2007. 16(2): p. 239-249.
15. Josephson, K., et al., *Design and Analysis of an Engineered Human Interleukin-10 Monomer.* Journal of Biological Chemistry, 2000. 275(18): p. 13552-13557.
16. Schouten, A., et al., *The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco.* Plant Molecular Biology, 1996. 30(4): p. 781-793.
17. Bennett, M.J., M.P. Schlunegger, and D. Eisenberg, *3D domain swapping: A mechanism for oligomer assembly.* Protein Science, 1995. 4(12): p. 2455-2468.
18. Yang, S., H. Levine, and J.N. Onuchic, *Protein Oligomerization Through Domain Swapping: Role of Inter-molecular Interactions and Protein Concentration.* Journal of Molecular Biology, 2005. 352(1): p. 202-211.
19. Brandizzi, F., et al., *ER quality control can lead to retrograde transport from the ER lumen to the cytosol and the nucleoplasm in plants.* The Plant Journal, 2003. 34(3): p. 269-281.
20. Josephson, K., N.J. Logsdon, and M.R. Walter, *Crystal Structure of the IL-10/IL-10R1 Complex Reveals a Shared Receptor Binding Site.* Immunity, 2001. 15(1): p. 35-46.

21. Conley, A.J., et al., *Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants.* Biotechnology and Bioengineering, 2009a. 9999(9999): p. n/a.
22. Menassa R., et al., *Subcellular targeting of human interleukin-10 in plants.* Journal of Biotechnology, 2004. 108(2): p. 179-83.
23. Fujiwara, Y., et al., *Extraction and purification of human interleukin-10 from transgenic rice seeds.* Protein Expression and Purification, 2010. 72(1): p. 125-130.
24. Shameer, K., et al., *3dswap-pred: prediction of 3D domain swapping from protein sequence using Random Forest approach.* Protein & Peptide Letters, 2011. 18(10): p. 1010-20.
25. Rao, D.V., et al., *Cloning, high expression and purification of recombinant human intereferon-beta-1b in Escherichia coli.* Applied Biochemistry and Biotechnology, 2009. 158(1): p. 140-54.
26. Proudfoot, A.E., et al., *Preparation and characterization of human interleukin-5 expressed in recombinant Escherichia coli.* Biochemical Journal, 1990. 270(2): p. 357-61.
27. Schwanke, R.C., et al., *Molecular cloning, expression in Escherichia coli and production of bioactive homogeneous recombinant human granulocyte and macrophage colony stimulating factor.* International Journal of Biological Macromolecules, 2009. 45(2): p. 97-102.
28. Sun, J., et al., *Cloning, expression and identification of human interleukin-10 gene.* Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi, 2007. 23(3): p. 220-2.
29. Li, J., et al., *Transient expression of an active interferon-beta in lettuce.* Scientia Horticulturae, 2007. 112: p. 258-265.
30. James, E.A., et al., *Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells.* Protein Expression and Purification, 2000. 19(1): p. 131-8.
31. Lee, J.H., et al., *Increased production of human granulocyte-macrophage colony stimulating factor (hGM-CSF) by the addition of stabilizing polymer in plant suspension cultures.* Journal of Biotechnology, 2002. 96(3): p. 205-11.
32. Luo, J., et al., *Proteomic analysis of rice endosperm cells in response to expression of hGM-CSF.* Journal of Proteome Research, 2009. 8(2): p. 829-37.
33. Ning, T., et al., *Oral administration of recombinant human granulocyte-macrophage colony stimulating factor expressed in rice endosperm can increase leukocytes in mice.* Biotechnology Letters, 2008. 30(9): p. 1679-86.
34. Sardana, R., et al., *Biologically active human GM-CSF produced in the seeds of transgenic rice plants.* Transgenic Research, 2007. 16(6): p. 713-721.
35. Li, J., et al., *Monomeric isomers of human interleukin 5 show that 1:1 receptor recruitment is sufficient for function.* PNAS, 1997. 94(13): p. 6694-9.
36. Zheng, X., et al., *Administration of noncytolytic IL-10/Fc in murine models of lipopolysaccharide-induced septic shock and allogeneic islet transplantation.* The Journal of Immunology, 1995. 154(10): p. 5590-5600.
37. Schwager, K., et al., *Preclinical characterization of DEKAVIL (F8-IL10), a novel clinical-stage immunocytokine which inhibits the progression of collagen-induced arthritis.* Arthritis Research & Therapy, 2009. 11(5): p. R142.
38. Karimi, M., D. Inzé, and A. Depicker, *GATEWAY™ vectors for Agrobacterium-mediated plant transformation.* Trends in Plant Science, 2002. 7(5): p. 193-195.
39. van Engelen, F.A., et al., *Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco.* Plant Molecular Biology, 1994. 26(6): p. 1701-10.

40. Curtis, M.D. and U. Grossniklaus, *A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta.* Plant Physiology, 2003. 133(2): p. 462-469.
41. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Research, 2001. 29(9): p. e45.

Chapter 3

Monomeric IgA can be produced *in planta* as efficient as IgG, yet receives different N-glycans

Lotte B. Westerhof, Ruud H. P. Wilbers, Debbie R. van Raaij, Dieu-Linh Nguyen, Aska Goverse, Maurice G. L. Henquet, Cornelis, H. Hokke, Dirk Bosch, Jaap Bakker and Arjen Schots

This work has been published in:
Plant Biotechnology Journal, 2014; doi:10.1111/pbi.12251

Abstract

The unique features of IgA, such as the ability to recruit neutrophils and suppress the inflammatory responses mediated by IgG and IgE, make it a promising antibody isotype for several therapeutic applications. However, in contrast to IgG, reports on plant-production of IgA are scarce. We produced IgA1κ and IgG1κ versions of three therapeutic antibodies directed against pro-inflammatory cytokines in *Nicotiana benthamiana*: Infliximab and Adalimumab, directed against TNF-α, and Ustekinumab, directed against the interleukin-12p40 subunit. We evaluated antibody yield, quality and N-glycosylation. All six antibodies had comparable levels of expression between 3,5 and 9% of total soluble protein content and were shown to have neutralizing activity in a cell-based assay. However, IgA1κ-based Adalimumab and Ustekinumab were poorly secreted compared to their IgG counterparts. Infliximab was poorly secreted regardless of isotype backbone. This corresponded with the observation that both IgA1κ and IgG1κ-based Infliximab were enriched in oligomannose-type N-glycan structures. For IgG1κ-based Ustekinumab and Adalimumab the major N-glycan type was the typical plant complex N-glycan; biantennary with terminal *N*-acetylglucosamine, β1,2-xylose and core α1,3-fucose. In contrast, the major N-glycan on the IgA-based antibodies was xylosylated, but lacked core α1,3-fucose and one terminal *N*-acetylglucosamine. This type of N-glycan occurs usually in marginal percentages in plants and was never shown to be the main fraction of a plant-produced recombinant protein. Our data demonstrate that the antibody isotype may have a profound influence on the type of N-glycan an antibody receives.

Introduction

The use of monoclonal antibodies for the treatment of cancer, autoimmune diseases and inflammatory disorders is rapidly growing. Antibodies of the IgG isotype are most common, but other isotypes such as IgA may offer important advantages. IgA is best known as secretory IgA, a dimeric IgA complex enveloped by the secretory component, that helps protect mucosal surfaces against pathogens and control commensal bacteria. IgA is also the second most predominant isotype in the blood [1]. In primates serum IgA is monomeric, in contrast to other animals where serum IgA is dimeric. The immunotherapeutic potential of serum IgA has only been recognized recently, because the biology of serum IgA was poorly understood [2, 3]. Due to differences in IgA functioning between primates and other animals, a proper animal model was lacking. Most important in this regard is that rodents lack a homolog of the human Fc receptor for IgA (FcαRI or CD89) [3]. This already indicated that rodents lack an important effector function that is mediated by FcαRI in primates.

The creation of a transgenic mouse expressing human FcαRI helped reveal IgA's unique ability to recruit neutrophils, the largest population of white blood cells in the body [4]. This capacity may be a valuable feature in cancer therapy, because IgA induced tumour cell killing was shown to be mediated by neutrophils. IgG does not efficiently recruit neutrophils, but mediates cell lysis predominantly via the complement system and peripheral blood mononucleated cells [5]. In addition, serum IgA has immunosuppressive properties that may be of therapeutic interest. When free IgA associates with FcαRI on immune cells, inhibitory signals are triggered that block pro-inflammatory responses initiated by association of IgG and IgE with their Fc receptors. Thus, IgA may be used as an overall immune suppressor of inflammatory disorders [2]. The ability of IgA to suppress the immune response could be combined with the antigen binding capacity. Many autoimmune and inflammatory disorders are nowadays treated by neutralization of pro-inflammatory cytokines using antibodies. Neutralization of the pro-inflammatory cytokine TNF-α is the most common therapy. Because TNF-α also occurs as a membrane bound variant, use of an antibody directed against TNF-α not only results in neutralization of free TNF-α, but also result in neutrophil-mediated killing of immune cells harbouring TNF-α on their membrane [6, 7]. Use of IgA may enhance killing of these immune cells through its previously mentioned unique ability to recruit neutrophils. As a consequence this will result in better suppression of the immune response required for autoimmune and inflammatory disease treatment.

Most, if not all, therapeutic antibodies are made using mammalian cell lines. However, plants have been shown to be an excellent alternative production platform. Like animal cell lines plants are capable of producing high levels of complex proteins, but are easier to manipulate and more economic in use [8, 9]. Next to that, plant-produced proteins often appear more homogeneously glycosylated compared to proteins produced by mammalian cell

lines [10]. Engineering of the N-glycosylation machinery has been successful in plants and research on humanizing N-glycosylation in plants has advanced over the last decade, allowing production of antibodies and other proteins with human N-glycan types [11-13]. However, most of this work has been based on the plant expression of IgG antibodies and few reports on the plant expression of IgA exist [14]. The few IgA antibodies that have been produced in plants were based on a variety of IgA backbones (murine, chicken and human), were expressed in different plant species (tobacco, maize, rice, tomato and *Arabidopsis*) and demonstrated a great variability in yield [15-20].

To obtain insight in the effect of isotype on the yield, quality and N-glycosylation of antibodies when expressed in plants we compared the expression of monomeric IgA and IgG in *Nicotiana benthamiana*. Thereto we expressed the therapeutic antibodies Infliximab, Adalimumab and Ustekinumab in their original form as IgG1κ antibodies as well as the IgA1κ variants of these three antibodies. Both Infliximab and Adalimumab are directed against TNF-α. Infliximab is a chimeric antibody with mouse variable regions on a human backbone [21] and Adalimumab is a fully human antibody [22]. Ustekinumab is a human antibody directed against the interleukin-12 p40 subunit (IL-12p40) shared by the pro-inflammatory cytokines IL-12 and IL-23 [23]. All three antibodies are used to treat several inflammatory disorders such as rheumatoid arthritis, psoriasis and inflammatory bowel disease. Upon plant expression we observed a similar high yield and quality of the IgA1κ-based antibodies compared to the original IgG1κ variants. The N-glycosylation pattern, however, was different. While IgG-based antibodies predominantly carried the typical plant complex N-glycan for secreted proteins; a biantennary N-glycan with terminal *N*-acetylglucosamine (GlcNAc) with β1,2-xylose and core α1,3-fucose (GnGnXF), IgA-based antibodies lacked core α1,3-fucose and one terminal GlcNAc (GnMX or MGnX).

Results

Antibody isotype and idiotype have little effect on yield

To establish the production potential of human IgA in plants we investigated the expression of the variable regions of the three therapeutic antibodies Infliximab (I), Adalimumab (A) and Ustekinumab (U) on both a human IgA1κ and human IgG1κ backbone *in planta* (Figure 1a). Light chain (κ) and either the alpha (α)-1 or gamma (γ)-1 heavy chain were combined in one expression vector and transiently expressed in *Nicotiana benthamiana*. Yield was determined up to 15 days post infiltration (dpi) with 3-day intervals. Figure 1b shows the average recombinant protein levels of three biological replicates. IgG yields peaked at dpi 6 with 82, 90 and 68 μg of antibody (Ab) per mg of total soluble protein (TSP) for IgG-I, IgG-A and IgG-U, respectively. IgA yields peaked on later time points at dpi 9 for IgA-I and IgA-U and dpi 12 for

IgA-A. Although prolonged accumulation can be indicative of greater protein stability, the overall yield of the IgA's is lower, 35, 70 and 46 µg Ab/mg TSP for IgA-I, IgA-A and IgA-U, respectively. Thus, antibody yield most likely depends more on the expression system than idiotype or isotype.

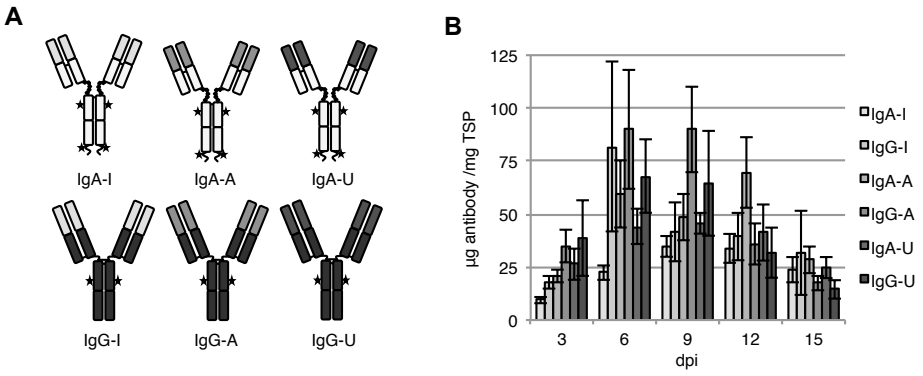


Figure 1 ◆ Plant production of IgA1κ and IgG1κ antibodies. A Cartoon representing the expressed antibodies. **B** Yield of Infliximab (I), Adalimumab (A) or Ustekinumab (U) on an IgA1κ or IgG1κ backbone when transiently expressed in *Nicotiana benthamiana* up to 15 days post infiltration (dpi). Bars indicate average yield and error bars indicate standard error (n=3).

Plants produce high quality IgA antibodies

To investigate protein quality the antibodies were analysed by western blot. Figure 2 shows the visualization of the κ, γ and α chains when samples were run under reduced conditions. For all antibodies the predominant band of the heavy chain migrated at the expected size (~49 and ~51 kDa for the γ and α chain, respectively). For α-A, α-U and the recombinant control a doublet (two bands migrating very close to each other) was observed. As the constant region of α-1 has two N-glycosylation sites these bands can represent α chains with none, one or two N-glycans. Surprisingly, α-I shows only the higher of the two bands observed for α-A, α-U and the recombinant control. This suggests that the α-I chain lacks either the unglycosylated or the one time glycosylated variant of α-I. The κ-A and κ-U chains also had the expected size (~23,5 kDa), while κ-I had a slightly higher molecular weight. Because no N-glycosylation sites are present in the κ chains this difference cannot be explained by differences in N-glycosylation. Comparison of plant-produced κ-I to the murine myeloma cell line produced κ-I showed equal electrophoretic mobility. This suggests that intrinsic properties of the variable domain lead to aberrant running behaviour and does not indicate an actual increase in molecular weight (data not shown).

For the α chains several additional bands were observed that migrated faster than the expected size. These additional bands most likely represent proteolytic degradation products. A different pattern can be seen for α-I compared to α-A and α-U. It is possible α-I is sensitive to other proteases than the other two chains or that IgA-I is exposed to different proteases, which would suggest a different intracellular location of this antibody. Alternatively, the bands of α-I represent different N-glycan variants of the same degradation products. For κ-I it was striking to observe that an additional band migrating around 20 kDa was only seen when co-expressed with the γ heavy chain, but not α heavy chain. It most likely also represents a degradation product that may be the consequence of a different conformation of IgA-I and IgG-I whereby κ-I is more susceptible to proteases when it is associated with the γ-I chain than the α-I chain. Alternatively, IgG-I subcellular targeting may be different to IgA-I and therefore may encounter different proteases. Whatever the cause of the differences in degradation products between the antibody chains, the majority of each antibody chain appears to be intact.

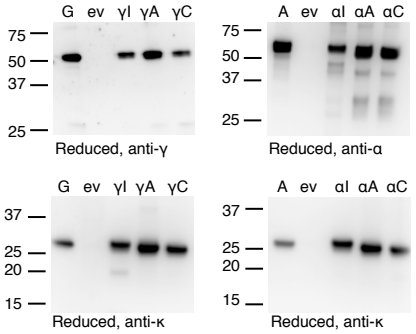


Figure 2 ◆ Evaluation of quality of plant-produced IgA1κ and IgG1κ antibodies. Western blots under reducing conditions of 50 ng Infliximab (I), Adalimumab (A) or Ustekinumab (U) on an IgA1κ or IgG1κ backbone in leaf extracts visualizing the kappa chain (κ), gamma chain (γ) or alpha chain (α), as indicated. Recombinant IgG (G) or IgA (A) and empty vector (ev) control were included.

Efficiency of antibody secretion from plant cells depends on isotype and idiotype

As antibodies are secretory proteins we developed a method to compare secretion efficiency of proteins as an indicator of aberrant subcellular targeting. The amount of secreted antibody was determined by 'washing out' the apoplast fluid from the transformed leaf followed by isolation of intracellular proteins and determination of antibody concentration in both fractions. Figure 3a shows the secretion efficiency of all six antibodies as the means of three biological replicates. Independent of idiotype, IgG secretion was at least three times more efficient than

IgA secretion. Poor secretion of a murine IgA/IgG chimera from plant cells was demonstrated before [24]. This murine IgA/G chimera was targeted to the vacuole due to a cryptic sorting signal in the tailpiece of murine IgA [25]. Human IgA has a similar tailpiece, although the sequence differs from mouse IgA. To reveal whether or not poor secretion of human IgA is also caused by its tailpiece, a variant of IgA-U without tailpiece (IgA-UΔT) was expressed and its secretion efficiency was determined (Figure 3b). While the average yield of IgA-UΔT was similar, the secretion efficiency was enhanced more than two fold. Like the murine tailpiece, the human tailpiece may be responsible for the poor secretion as a consequence of vacuolar targeting.

To evaluate if the degradation products described in the previous section were caused by protease activity inside the cell or in the apoplast, both intracellular and apoplast fractions were evaluated by western blot analysis (figure 3c). As all apoplast fractions were enriched with the degradation products we assume that most, if not all, proteolytic degradation takes place in the apoplast. This explains why the degradation product of κ-I that was only detected upon co-expression with the γ chain, but not the α-chain, as 10% of IgG-I compared to only 1,8% of IgA-I was secreted.

The secretion efficiency between idiotypes also differed. IgG-I secretion was two times less efficient compared to IgG-A and IgG-U secretion. The same result was obtained comparing the IgA antibodies. Apparently, both isotype and idiotype are important for secretion efficiency.

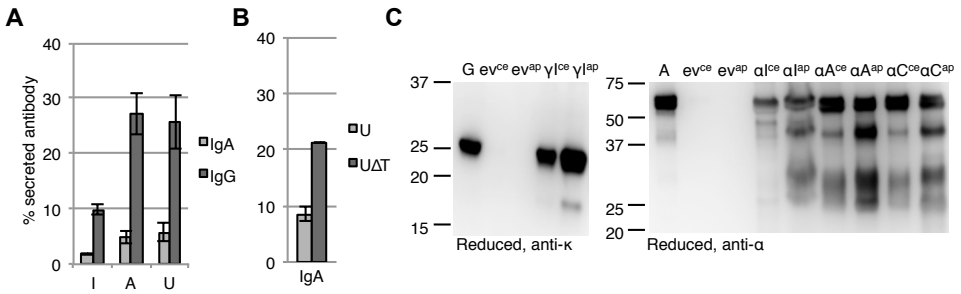


Figure 3 ◆ Secretion and proteolytic degradation of plant-produced IgA1κ and IgG1κ antibodies. A/B Percentage of secretion of IgG1κ and IgA1κ by determining the amount of antibody in apoplast fluids (ap) versus crude extracts (ce) (*n*=3, error bars indicate standard error). **C** Western blot analysis under reducing conditions of 50 ng Infliximab (I), Adalimumab (A) and Ustekinumab (U) with and without tailpiece (ΔT) on an IgA1κ or IgG1κ backbone visualizing the kappa chain (κ) or alpha heavy chain (α), as indicated. Recombinant IgG (G) or IgA (A) and empty vector (ev) control were included.

Differential N-glycosylation of IgA compared to IgG

Because differences in subcellular localization can influence N-glycosylation, we evaluated N-glycan maturity of all antibodies. Thereto the antibodies were treated with the enzymes peptide:N-glycosidase F (PNGase F) or endoglycosidase H (Endo H) followed by western blot analysis. PNGase F only cleaves N-glycans which lack the plant-specific α1,3-fucose, while Endo H only cleaves oligomannose-type N-glycans. Treatment of the IgG antibodies with either enzyme did not reduce heavy chain size, except the recombinant control (G), which became smaller upon PNGase F treatment (Figure 4, top panel). This indicates that the plant-produced IgG antibodies do not carry oligomannose-type N-glycans and confirms the presence of the plant-specific α1,3-fucose. In contrast, when the IgA-A and IgA-U antibodies were treated with PNGase F a major fraction of the α heavy chains was reduced in size (Figure 4, second panel). EndoH did not affect IgA-A and IgA-U. This suggests that a major fraction of IgA-A and IgA-U carry complex type N-glycans that are devoid of core α1,3-fucose. Treatment of IgA-I with Endo H or PNGaseF caused a reduction in size of the α-I chain suggesting that IgA-I carries oligomannose-type N-glycans.

To determine if the atypical N-glycosylation of IgA-A and IgA-U is the result of poor secretion of IgA, both IgA-U and IgA-UΔT from the apoplast or crude leaf extract were evaluated by treatment with PNGase F (Figure 4, third panel). Treatment with PNGase F reduced the major fraction of the α heavy chains in size of both intracellular and secreted IgA-U and IgA-UΔT. This suggests that the absence of a core α1,3-fucose on N-glycans of IgA does not depend on its subcellular targeting. Furthermore, the fact that IgA-UΔT shows only one band that is sensitive to PNGase F suggests that this band represents a glycosylated α chain. Because one of the N-glycan sites of human IgA resides in the tailpiece IgA-UΔT cannot carry more than one N-glycan. The absence of a second band for non-glycosylated IgA-UΔT therefore indicates that the N-glycan site in constant domain 2 must always receive an N-glycan. The doublet in IgA-A and IgA-U must therefore be the result of partial N-glycosylation of the tailpiece.

To assess the presence of β1,2-xylose on the IgA antibodies a western blot was done using a polyclonal rabbit anti-horseradish peroxidase antibody specific for plant α1,3-fucose and β1,2-xylose (Figure 4, bottom panel). For both IgA-A and IgA-U two bands appeared confirming the presence of β1,2-xylose on both N-glycan variants. For IgA-UΔT only one band appeared due to the absence of the N-glycosylation site present in the tailpiece.

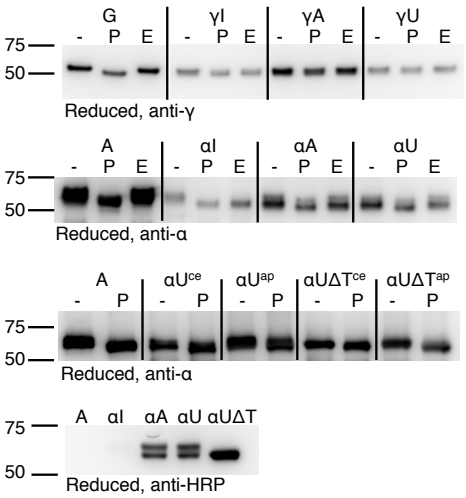


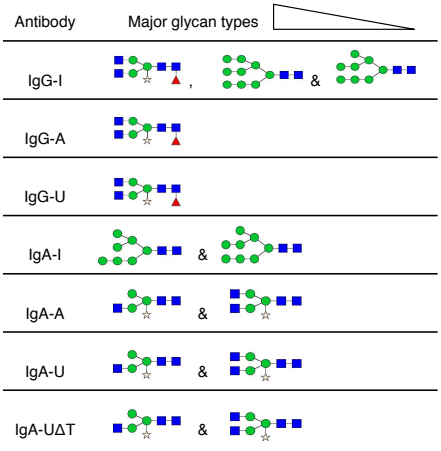
Figure 4 ◆ Evaluation of N-glycan maturity of plant-produced IgA1κ and IgG1κ antibodies. Western blots under reducing conditions of 50 ng purified Infliximab (I), Adalimumab (A) or Ustekinumab (U) on an IgG1κ or IgA1κ backbone with or without tailpiece (ΔT) visualizing the gamma heavy chain (γ), the alpha heavy chain (α) or β1,2xylose/α1,3-fucose (HRP), as indicated. Ustekinumab (U/UΔT) was isolated from apoplast (ap) or crude extract (ce) and were analysed separately by visualizing the alpha heavy chain (α). Recombinant IgG (G) or IgA (A) and an empty vector (ev) were included. Samples were treated with Endo H (H) or PNGase F (P) to screen for the presence of oligomannose type N-glycans or α1,3-fucose, respectively.

To obtain a more detailed picture we analysed the composition of the N-glycans on the plant-produced antibodies using matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) (Table 1). For IgG, tryptic glycopeptides were prepared prior to MALDI-TOF-MS analysis. As predicted, the major N-glycan type present on IgG glycopeptides was identified as the typical plant complex N-glycan for secreted proteins; a biantennary N-glycan with terminal GlcNAc with β1,2-xylose and core α1,3-fucose (GnGnXF) (Supplement 1). Surprisingly, IgG-I carried a small fraction of oligomannose-type N-glycans, which may be indicative of ER-retention. This is in line with the fact that IgG-I was less efficiently secreted compared to IgG-A and IgG-U (Figure 3).

Analysis of IgA N-glycans based on tryptic glycopeptides was not successful. Tryptic digestion of IgA results in relatively large glycopeptides compared to IgG and might therefore not be suitable for this approach. Alternatively, IgA was digested with either trypsin, pepsin or thermolysin, followed by PNGase A release of N-glycans. Unfortunately, this approach did also not result in clear N-glycan profiles. Because we already confirmed that the majority of the N-glycans on plant-produced IgA do not carry α1,3-fucose, we used PNGase F to release *N*-glycans from IgA and this did result in clear N-glycan profiles (Supplement 2). The N-glycan

profiles for IgA revealed that IgA-I carries mainly oligomannose-type N-glycans. The predominant N-glycan type found on IgA-A, IgA-U and IgA-UΔT was similar to IgG, but lacked one terminal GlcNAc. Although we cannot exclude that N-glycans on plant-produced IgA may occasionally carry a core α1,3-fucose, we can assume that the predominant N-glycan on plant-produced IgA-A, IgA-U and IgA-UΔT is GnMX or MGnX. In conclusion, it seems that the N-glycan type that an antibody receives upon expression *in planta* is mainly determined by the antibody isotype.

Table 1 ◆ Major N-glycan types carried by the plant-produced IgA1κ or IgG1κ variants of Infliximab (I), Adalimumab (A) and Ustekinumab (U) as determined using MALDI-TOF-MS.



Antigen binding capacity

To determine the antigen binding capacity of the plant-produced antibodies cell-based assays were performed. For Infliximab and Adalimumab, which are directed against TNF-α, we assessed the capacity of these antibodies to abrogate TNF-α induced apoptosis of L929 cells. For Ustekinumab, which is directed against the p40 subunit of IL-12 and IL-23, its capacity to block IL-23 induced production of IL-17 by murine splenocytes was assessed.

Cell viability of L929 cells increased in a dose dependent manner upon increasing concentration of plant-produced IgG-I, IgA-I, IgG-A and IgA-A, but not plant-produced IgG-U and IgA-U, which were used as negative controls (Figure 5a). No significant difference was observed between the murine myeloma cell (SP2/0) produced Infliximab and plant-produced IgG-I, IgA-I, IgG-A and IgA-A. It is surprising that IgG-I and IgA-I perform equally well when compared to the control in this assay as both antibodies were enriched in oligomannose-type N-glycan structures that may indicate improper protein folding.

IL-17 production reduced in a dose dependent manner upon increasing concentration of both plant-produced IgG-U and IgA-U, but not plant-produced IgG-I and IgA-I, which were used as negative controls (Figure 5b). No significant difference was observed between the murine myeloma cell (SP2/0) produced Ustekinumab and the plant-produced variants.

Thus, regardless the antibody isotype plants produce functional Infliximab, Adalimumab and Ustekinumab.

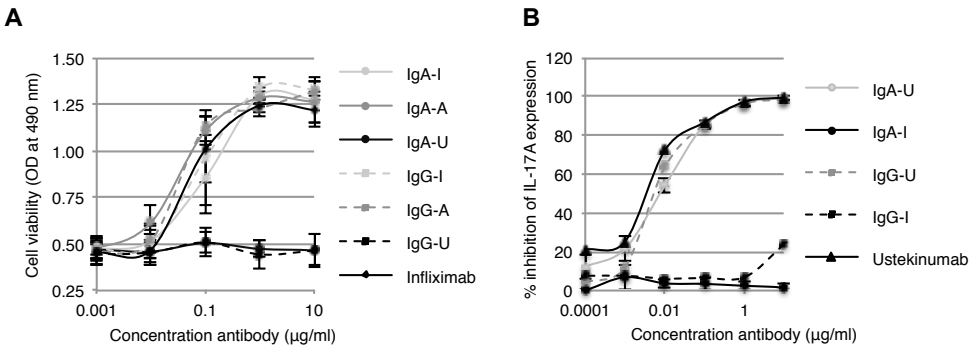


Figure 5 ◆ Biological activity of plant-produced IgA1κ and IgG1κ antibodies. A Cell viability of L929 cells when exposed to TNFα in combination with different concentrations of plant-produced Infliximab (I), Adalimumab (A) and Ustekinumab (U) on an IgG1κ or IgA1κ backbone. Mammalian cell-produced Infliximab (IgG) and a commercially available anti-TNFα IgA were included. **B** IL-17 production by murine splenocytes exposed to IL-23 in combination with different concentrations of plant-produced Infliximab (I) and Ustekinumab (U) on an IgG1κ or IgA1κ backbone. Mammalian cell-produced Ustekinumab (IgG) was included. Bars indicate averages and error bars indicate standard error (n=3).

Discussion

Currently, most therapeutic antibodies are of the IgG isotype. Yet IgA was suggested as a promising alternative isotype that may increase the therapeutic opportunities of several antibody-based therapies [2]. Therefore, we evaluated the plant production of IgA. We expressed IgA1κ variants of the three commercially available therapeutic antibodies, Infliximab, Adalimumab and Ustekinumab in *Nicotiana benthamiana*. The production of the IgA1κ variants was compared to the production of their original form as IgG1κ antibodies. Expression of all six antibodies was successful with high yields ranging from 3,5% to 9% of total soluble protein whereby IgG performed slightly better for all three antibodies. Although the majority of the antibodies remained intact, IgA was found to be somewhat more sensitive to proteolytic enzymes than IgG. Many reports on the plant expression of IgG exist, but these are often based on the expression of a single antibody. Comparing the results of these reports reveals a large heterogeneity in antibody quantity and quality, whereby antibody yield varies several orders of magnitude [14]. However, our data reveal only slight discrepancies in antibody quantity and quality between isotypes and idiotypes indicating that the previously reported heterogeneity may largely be attributed to e.g. differences in transformation method, subcellular targeting, and use of plant tissue and species.

In line with previous reports on N-glycosylation of plant expressed IgG's, all our IgG-based antibodies predominantly carried the typical complex plant N-glycan for secreted proteins; biantennary with terminal GlcNAc and β1,2-xylose and core α1,3-fucose (GnGnXF^3). However, our IgA-based Adalimumab and Ustekinumab predominantly carried a similar N-glycan type as IgG, but lacked core α1,3-fucose and one terminal GlcNAc (GnMX or MGnX). Oligomannose-type *N*-glycans were found on both the IgG and IgA variants of our plant-produced Infliximab. On IgG Infliximab oligomannose was found to a minor extent; however, on IgA Infliximab this was the predominant N-glycan type.

Differences in N-glycosylation may be caused by differences in subcellular targeting. Infliximab was less efficiently secreted into the apoplast compared to Adalimumab and Ustekinumab, regardless of isotype backbone. This coincided with the presence of oligomannose-type N-glycans. Oligomannose-type N-glycans on heterologously expressed proteins are often indicative of retention in the endoplasmic reticulum (ER) and is associated with protein misfolding. Because only our IgG and IgA-based Infliximab showed enrichment of oligomannose-type N-glycans, improper processing of Infliximab chains *in planta* seems a logical explanation for the presence of these N-glycan types. As protein misfolding is coupled with protein degradation, the yield of a misfolded protein is expected to be low. However, we did not find significantly lower yields for both isotypes of Infliximab compared to the yields of Adalimumab and Ustekinumab. Also, both IgG and IgA-based Infliximab showed the same capacity to neutralize TNF-α when compared to murine myeloma cell (SP2/0) produced

Infliximab. Still, oligomannose enrichment was only demonstrated for IgG and IgA infliximab and must therefore depend on idiotype. In conclusion, since folding of Infliximab leads to functional antibodies, the reason for oligomannose enrichment remains to be elucidated.

Evaluation of the secretion efficiency of all antibodies revealed that the IgA variants were secreted less efficiently into the apoplast when compared to IgG. A previous study demonstrated that plant-based expression of a chimeric murine IgG/A antibody resulted in targeting of this antibody to the lytic vacuole. A cryptic targeting signal in the tailpiece of the antibody was responsible for vacuolar targeting [24, 25]. Upon arrival in the vacuole the murine G/A chimera underwent proteolytic degradation. We also found degradation products of our IgA-based antibodies, while no degradation was observed for the IgG-based variants. However, IgA isolated from the apoplast showed more prominent degradation than intracellular IgA. Thus, in contrast to the murine G/A chimeric antibody degradation of human IgA likely resulted from proteolysis in the apoplast. In the C-terminus of the vacuolar proteins tobacco β-glucanase and potato PT20 the dipeptides VS and SV are present [26, 27]. The murine tailpiece contains the sequence 'VSVSV', a repetition of VS and SV and was therefore suggested to be the cryptic vacuolar targeting. The human tailpiece contains the sequence 'VSV' which may fulfil the same role. However, the human tailpiece also contains the dipeptides 'AE', which is a conserved motif in vacuolar targeting signals of several *Graminea* lectins [28], and 'VD', found in the vacuolar targeting signal of tobacco chitinase A [29]. These proteins are supposedly sorted to the protein storage vacuole [30], which may also be the destination of human IgA. Targeting to the protein storage vacuole instead of the lytic vacuole would explain the absence of proteolytic products of intracellular IgA as well as the fact that yield reduction in comparison to IgG was minor.

Golgi dependent and independent vacuolar targeting pathways have been identified [31] and the route taken seems to depend on protein intrinsic properties. Because the attachment of β1,2-xylose occurs in the medial-Golgi [32, 33], our IgA-based antibodies must travel to the Golgi as the N-glycans on our IgA-based Adalimumab and Ustekinumab carried this sugar residue. After attachment of β1,2-xylose the core α1,3-fucose sugar residue is added in the *trans*-Golgi [32, 33]. The core α1,3-fucose sugar residue is absent from the majority of the N-glycans of our IgA-based Adalimumab and Ustekinumab. Thus, the lack of core α1,3-fucose could be explained by vacuolar targeting if IgA is transported to the vacuole from the medial-Golgi. However, the core α1,3-fucose was also absent from the N-glycans of secreted IgA-based Ustekinumab, which must have travelled through the *trans*-Golgi to reach the apoplast.

Another, more plausible, reason for the lack of core fucosylation may be found in the protein intrinsic properties. Petrescu and colleagues (2004) analysed the protein environment of N-glycosylation sites and demonstrated that many N-glycan sites are poorly accessible [34]. As a result N-glycans lacking one or more sugar residues are found, because the core of a N-glycan cannot always be reached by core altering glycosyltransferases. On human serum IgA

less than half of the N-glycans carry the typical mammalian core α1,6-fucose [35, 36], whereas 80-98% of the N-glycans on human serum IgG antibodies are core fucosylated [37]. IgG and IgA are both N-glycosylated in constant domain 2 (Cγ2 and Cα2, respectively) and structurally these domains are similar. The major differences between these domains are the positions of interdomain disulphide bridges and the N-glycans [38]. The N-glycans of Cγ2 of IgG stabilise the interaction between the γ heavy chains, whereas in IgA this is done by a disulphide bridge. Compared to IgG, the N-glycans of the Cα2 domain of IgA are overall more exposed and therefore more accessible for processing [39]. This may explain why these N-glycans have increased sialylation and galactosylation compared to N-glycans of IgG [35]. However, this does not rule out the possibility that the core of these N-glycans have lower accessibility for core altering enzymes compared to IgG.

Alternatively, specific amino acids surrounding the N-glycosylation site may interfere with efficient core fucosylation. While evaluating the direct surroundings of the N-glycosylation sites of human IgG1 and IgA1 (position -3 to +4 from the asparagine in the N-X-S/T site), the only observed consistent difference was the presence of the hydrophobic amino acids leucine or valine on position X of the N-glycosylation sites of IgA1, whereas IgG1 contains a hydrophilic serine. We found this striking as chicken ovalbumin also contains a leucine on position X of its N-glycosylation signal and when we expressed ovalbumin in plants, the core α1,3-fucose was absent from this N-glycan as well (unpublished data). Also plant-produced ovalbumin was insensitive towards PNGase A digestion. We therefore hypothesize that protein intrinsic properties that might interfere with core α1,3-fucosylation, could also interfere with PNGase A release of N-glycans for MALDI-TOF-MS analysis. How amino acids surrounding the N-glycan site can influence core fucosylation or *in vitro* N-glycan release may become apparent when more N-glycan profiles from plant-produced proteins become available in the future.

Next to the core α1,3-fucose a terminal GlcNAc was lacking on IgA-based Adalimumab and Ustekinumab. Although GnM or MGn-type N-glycans occur in marginal percentages in plants, most proteins carry either MM (paucimannosidic) or GnGn-type N-glycans. MM-type N-glycans are the result of the activity of enzymes that remove terminal GlcNAc's. Three enzymes (HEXO1-3) capable of removing terminal GlcNAc residues were identified in *Arabidopsis* and shown to localize to the vacuole (HEXO1) and plasma membrane (HEXO2/3) [40, 41]. Therefore, N-glycans on secreted as well as vacuolar proteins can be modified in this manner. All three enzymes were demonstrated to remove the terminal GlcNAc residues from both α1,3- and α1,6-branched mannoses without strict preference. Thus, subcellular targeting cannot explain why only one terminal GlcNAc would be removed and not the other. The underlying mechanism controlling GlcNAc removal is unclear, but seems to depend on protein intrinsic properties. It is possible that the GlcNAc's are inefficiently removed from the N-glycans of IgA resulting in a mix of GnMX and MGnX-type glycans. Alternatively, it is also possible that one of the terminal GlcNAc's is not added. If this

is the case, the α1,6-branched mannose must not have received its GlcNAc, as β1,2-xylose was present and the terminal GlcNAc on the α1,3-branched mannose is required for β1,2-xylosyltransferase activity [42]. *N*-acetylglucosaminyltransferase II (GnT-II) is the enzyme responsible for the addition of a GlcNAc to the α1,6-branched mannose. Thus, for unknown reasons it is possible that GnT-II does not act on N-glycans of IgA. This would, however, be a matter of efficiency, as we did find a proportion of N-glycans with terminal GlcNAc's on both arms on our IgA-based Adalimumab and Ustekinumab. Galactosylation of IgA *in planta* could reveal if the GlcNAc residue is not added or removed. The presence of terminal galactose would prevent removal of GlcNAc's. Humanising the N-glycans on IgA would be a necessary step in any case for future application of IgA as a biopharmaceutical.

Monoclonal IgG antibodies are expressed in several other expression platforms, like yeast, insect cells and mammalian cells, however production of IgA based antibodies has also gained less attention in these platforms. In yeast for instance, only IgG-based antibodies have been expressed and were shown to carry high mannose-type N-glycans, which is typical for yeast [43]. Insect cells also add high mannose-type N-glycans to IgG or paucimannosidic type N-glycans carrying core α1,3-fucose and/or α1,6-fucose [43, 44]. Only one IgA based antibody has been expressed in insect cells and this antibody carried high-mannose type N-glycans [45]. Our Infliximab also carried high mannose-type N-glycans and like the insect cell produced IgA it is a chimeric antibody with mouse variable regions on a human backbone. In both yeast and insect cells attempts to humanise N-glycosylation of IgG have been successful [44, 46], however, have never been evaluated for IgA. Finally, expression of IgG and IgA in different types of mammalian cells (CHO and Sp2/0) have revealed differences in N-glycan composition, but main differences were found in the amount and type of sialylation or galactosylation [39]. As galactosylation would prevent the removal of GlcNAc's, we can only conclude that GlcNAc is efficiently added in these mammalian expression platforms. On the other hand, core α1,6-fucosylation of CHO cell produced IgA1 was less efficient compared to IgG, but was still comparable to that of serum IgA1 [39]. Due to large differences in N-glycosylation pathways of these different expression hosts and the limited experience with N-glycan analysis of IgA expressed in insect and mammalian cells, it is difficult to conclude whether or not the differences in N-glycosylation between plant-expressed IgG and IgA are plant specific.

In conclusion, we have demonstrated that plants can be used equally well for the production of IgA as IgG in terms of yield and functionality. However, the N-glycan an antibody receives upon expression in *Nicotiana benthamiana* is determined by the isotype. Whether or not the N-glycans on plant produced human IgA can be humanized is a next step that will also yield insight in the protein intrinsic properties that influence N-glycosylation.

Experimental procedures

Construct design

All genes required for expression of the 3 different antibodies (Infliximab (I), Adalimumab (A) and Ustekinumab (U) on a human IgG1κ or IgA1κ backbone), were synthetically made by GeneArt (Bleiswijk, the Netherlands). The constant domains of human immunoglobulin alpha-1 (ACC82528.1), gamma-1 (AJ294730.1) and kappa (AGH70219.1) chains were ordered in their native sequence. All variable regions and the signal peptide of the *Arabidopsis thaliana* chitinase gene (AAM10081.1) were recoded from the amino acid sequence using codons preferred by in-house codon optimisation. These gene fragments were flanked by the following restriction sites at the 5′ and 3′-end for subsequent cloning steps: NcoI-EagI, EagI-NheI, NheI-KpnI, EagI-BsiWI, BsiWI-KpnI for the signal peptide, heavy chain variable regions, heavy chain constant domains, light chain variable regions and the light chain constant domain, respectively. None of the restriction sites introduced additional amino acids except NcoI, which introduced an alanine after the first methionine. Fragments were assembled using standard cloning procedures and ligated into the shuttle vector pRAPa, a pRAP (or pUCAP35S)[47] derivative modified to include an AsiSI restriction site by introduction of the self-annealed oligo 5′- AGCTGGCGATCGCC-3′ into a HindIII linearized pRAP. The kappa chain expression cassette was digested from pRAPa with AscI and PacI and ligated into the expression vector pHYG [48]. Subsequently, the heavy chain cassettes were combined with the kappa chain expression vector using the restriction sites AscI and AsiSI, of which the latter creates the same overhang as PacI. Vectors were transformed to *Agrobacterium tumefaciens* strain MOG101 for plant expression.

Agrobacterium tumefaciens transient transformation assay

Agrobacterium tumefaciens clones were cultured overnight (o/n) at 28°C in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH7.0) containing 50 μg/ml kanamycin and 20 μg/ml rifampicin. The optical density (OD) was measured at 600 nm and used to inoculate 50 ml of LB medium containing 200 μM acetosyringone and 50 μg/ml kanamycin with x μl of culture using the following formula: x = 80000/(1028*OD). OD was measured again after 16 hours and the bacterial cultures were centrifuged for 15 min at 2800 *g*. The bacteria were resuspended in MMA infiltration medium (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 μM acetosyringone till an OD of 1 was reached. All constructs were co-expressed with the tomato bushy stunt virus silencing inhibitor p19 by mixing *Agrobacterium* cultures 1:1. After 1-2 hours incubation at room temperature, the two youngest fully expanded leaves of 5-6 weeks old *Nicotiana benthamiana* plants were infiltrated at the abaxial side using a 1 ml syringe. Infiltrated plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) and infiltrated leaves were harvested at selected time points.

Apoplast soluble protein extraction

Leaves were vacuum infiltrated with ice-cold extraction buffer (50mM phosphate-buffered saline (PBS) pH=7.4, 100 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 0.1% v/v Tween-20) for 10 min. Leaves were dried from the outside and placed in a 10-ml syringe hanging in a blue cap and centrifuged at 2000 x *g* for 15min at 4°C. The volume of apoplast fluid was determined and directly used in an ELISA and BCA protein assay.

Total soluble protein extraction

Leaves were snap-frozen, homogenized in liquid nitrogen (after apoplast extraction or immediately upon harvesting) and stored at -80°C until use. Homogenized plant material was ground in ice-cold extraction buffer (50mM phosphate-buffered saline (PBS) pH=7.4, 100 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 0.1% v/v Tween-20, 2% w/v immobilized polyvinylpolypyrrolidone (PVPP)) using 2 ml/g fresh weight. Crude extract was clarified by centrifugation at 16.000 x *g* for 5min at 4°C and supernatant was directly used in an ELISA and BCA protein assay.

IgA and IgG quantification

IgA and IgG concentrations in apoplast fluids and crude extracts were determined by sandwich ELISA using goat polyclonal anti-human kappa antibody (Sigma-Aldrich; Zwijndrecht, the Netherlands) as capture antibody. HRP-conjugated goat polyclonal antibodies (Sigma-Aldrich) directed against the constant domains of IgG and IgA were used for detection. For sample comparison the total soluble protein (TSP) concentration was determined using the BCA Protein Assay (Pierce) according to supplier's protocol using bovine serum albumin (BSA) as a standard.

Protein analysis by western blot

Fifty ng of antibody in crude extract was separated under reducing conditions by SDS-PAGE on a NuPAGE® 12% Bis-Tris gel (Life Technologies; Bleiswijk, the Netherlands) and transferred to an Invitrolon™ PVDF membrane (Life Technologies) by a wet blotting procedure. Recombinant IgG1κ (Sigma-Aldrich) or IgA1κ (InvivoGen) was used as a control. For visualization goat polyclonal antibodies (Sigma-Aldrich) directed against human kappa or human IgA followed by incubation with a HRP-conjugated secondary antibody (Jackson ImmunoResearch; Suffolk, UK) or HRP-conjugated antibody directed against human IgG were used. SuperSignal West Femto substrate (Thermo Fisher Scientific; Etten-Leur, the Netherlands) was used to detect HRP-conjugated antibodies. Pictures were taken using a G:BOX Chemi System device (SynGene; Cambridge, UK).

Antibody purification

For purification of plant-produced antibodies crude extracts were first desalted over G25 Sephadex columns. Plant-produced IgG was then purified using Protein G affinity matrix (Thermo Fisher Scientific) and plant-produced IgA was purified using IgA CaptureSelect affinity matrix (Life Technologies).

N-glycan analysis

IgA or IgG (in apoplast fluids, crude extracts or purified) was deglycosylated with PNGase F or Endo H (both from Bioké; Leiden, the Netherlands) to screen for the presence of plant-specific α1,3-fucosylated or oligomannose-type N-glycans, respectively. A western blot using a rabbit polyclonal anti-horseradish peroxidase antibody (Jackson ImmunoResearch) was used to detected plant specific α1,3-fucose and β1,2-xylose, followed by a HRP-conjugated donkey anti-rabbit IgG (Jackson ImmunResearch).

For analysis of N-glycan composition, glycopeptides were prepared from 10 µg of antibody by in-gel trypsin digestion using the Trypsin Profile IGD kit (Sigma-Aldrich). Tryptic glycopeptides were purified over ZipTip$_{C18}$ (Merck; Amsterdam, the Netherlands), eluted with 18% v/v acetonitrile and analysed by MALDI-TOF-MS. For IgA N-glycan analysis proteolytic digestion with trypsin (Sigma-Aldrich), pepsin or thermolysin (both from Promega; Leiden, the Netherlands) was performed on 10-100 µg IgA prior to the release of N-glycans with 0.5 mU PNGase A (Roche; Woerden, The Netherlands). Released N-glycans were purified by C18 Bakerbond™ SPE cartridges (VWR; Amsterdam, The Netherlands) and subsequent Extract Clean™ Carbo SPE columns (Grace; Breda, the Netherlands). N-glycans were labelled with anthranilic acid (AA) (Sigma-Aldrich), desalted over Biogel P10 (BioRad) and analysed by MALDI-TOF-MS. Alternatively, N-glycans were released from 20 µg IgA by PNGase F digestion (Bioké). Released N-glycans were purified on Extract-Clean™ Carbograph Columns (Grace), AA-labelled and purified over ZipTip$_{C18}$ (Merck) prior to MALDI-TOF-MS analysis.

TNFα neutralization assay

L929 cells (DSMZ; Braunschweig, Germany) were cultured in RPMI-1640 medium containing 4 mM L-glutamine, 25 mM HEPES and supplemented with 10% fetal calf serum, 50 U/ml, 50 µg/ml streptomycin and 50 µM β-mercaptoethanol at 37°C with 5% CO_2. Cells were seeded in 96 well plates at a density of $5x10^4$cells/well and allowed to rest overnight. The cells were then treated with a 10-fold serial dilution of purified plant-produced antibodies from 10 µg/ml to 0.01 µg/ml in combination with 1 ng/ml recombinant *E. coli* produced human TNF-α and 1 µg/ml Actinomycin D (both from R&D Systems; Abingdon, UK). As positive controls mammalian cell produced infliximab and an anti-TNF-α-IgA (InvivoGen) were used. After overnight incubation cell viability was determined using the CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega) according to the supplier's protocol.

IL-23 neutralization assay

Splenocytes were isolated from spleens of C57BL/6 mice by passing them through a 70 µm nylon cell strainer (BD Biosciences; Breda, the Netherlands). Red blood cells were lysed using the Mouse Erythrocyte Lysing Kit (R&D Systems) and obtained splenocytes were seeded in 96 well plates at a density of $5x10^5$ cells/well in RPMI-1640 medium containing 4 mM L-glutamine, 25 mM HEPES and supplemented with 10% fetal calf serum, 50 U/ml penicillin, 50 µg/ml streptomycin and 50 µM β-mercaptoethanol. Splenocytes were treated with a 10-fold serial dilution of purified plant-produced antibodies from 10 µg/ml to 0.001 µg/ml in combination with 1 ng/ml recombinant *E. coli* produced human IL-23 and 10 ng/ml PMA. As positive control mammalian cell produced Ustekinumab was used. After 5 days supernatants were analysed for IL-17 expression with the mouse IL-17 Ready-Set-Go! ELISA Kit (eBioscience) according to the supplier's protocol.

Acknowledgments

This paper was financially supported in part by Synthon (Nijmegen, The Netherlands) and a grant from the Dutch Ministry of Economic Affairs (PID07124). The authors declare no conflict of interest. We would also like to thank Gerry Ariaans for all helpful discussions and Tim Warbroek, Aleksandra Syta and Bob Engelen for their input in the experimental work of this paper.

References

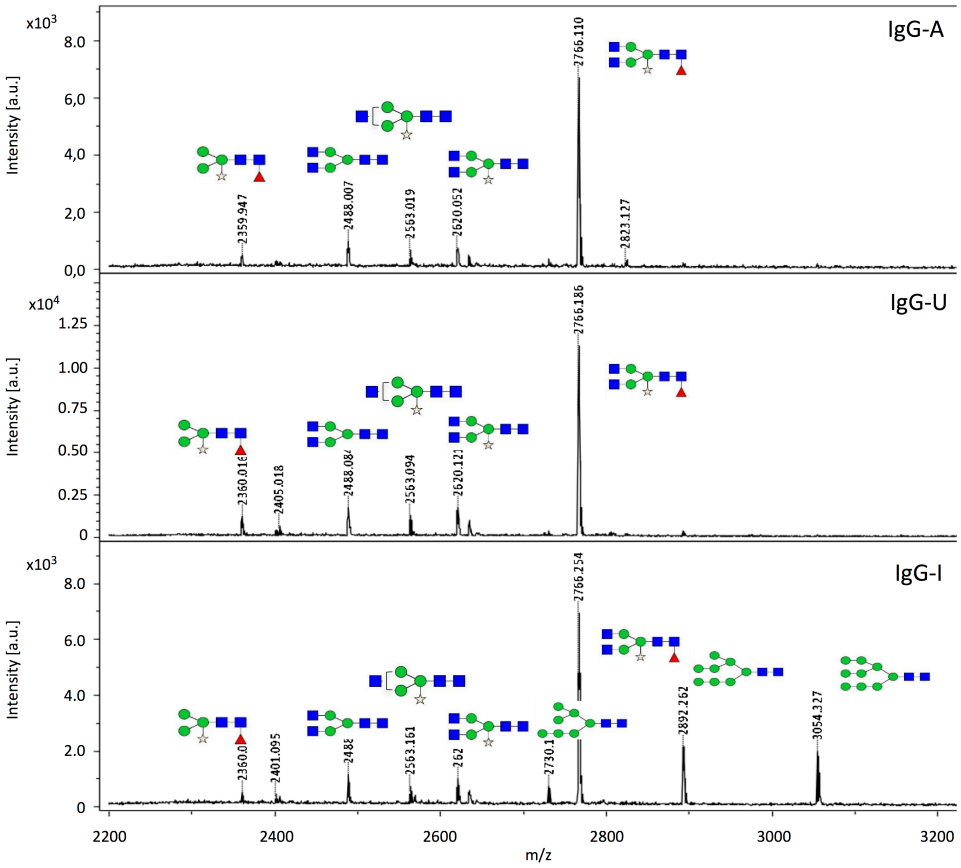
1. Stoop, J.W., et al., *Serum immunoglobulin levels in healthy children and adults.* Clinical & Experimental Immunology, 1969. 4(1): p. 101-12.
2. Bakema, J.E. and M. van Egmond, *Immunoglobulin A: A next generation of therapeutic antibodies?* MAbs, 2011. 3(4): p. 352-61.
3. Snoeck, V., I.R. Peters, and E. Cox, *The IgA system: a comparison of structure and function in different species.* Veterinary Research, 2006. 37(3): p. 455-467.
4. van der Steen, L., et al., *Immunoglobulin A: Fc(alpha)RI interactions induce neutrophil migration through release of leukotriene B4.* Gastroenterology, 2009. 137(6): p. 2018-29 e1-3.
5. Huls, G., et al., *Antitumor immune effector mechanisms recruited by phage display-derived fully human IgG1 and IgA1 monoclonal antibodies.* Cancer Research, 1999. 59(22): p. 5778-84.
6. Shen, C., et al., *Adalimumab induces apoptosis of human monocytes: a comparative study with infliximab and etanercept.* Alimentary Pharmacology & Therapeutics, 2005. 21(3): p. 251-8.
7. Van den Brande, J.M., et al., *Infliximab but not etanercept induces apoptosis in lamina propria T-lymphocytes from patients with Crohn's disease.* Gastroenterology, 2003. 124(7): p. 1774-85.
8. Schillberg, S., R. Fischer, and N. Emans, *Molecular farming of recombinant antibodies in plants.* Cellular and Molecular Life Sciences, 2003. 60(3): p. 433-45.
9. Hiatt, A. and M. Pauly, *Monoclonal antibodies from plants: A new speed record.* PNAS, 2006. 103(40): p. 14645-14646.
10. Cox, K.M., et al., *Glycan optimization of a human monoclonal antibody in the aquatic plant Lemna minor.* Nature Biotechnology, 2006. 24(12): p. 1591-7.
11. Bosch, D., et al., *N-Glycosylation of plant-produced recombinant proteins.* Current Pharmceutical Design, 2013.
12. Castilho, A., et al., *Rapid High Yield Production of Different Glycoforms of Ebola Virus Monoclonal Antibody.* Plos One, 2011. 6(10).
13. Strasser, R., et al., *Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous beta1,4-galactosylated N-glycan profile.* Journal of Biolgical Chemistry, 2009. 284(31): p. 20479-85.
14. De Muynck, B., C. Navarre, and M. Boutry, *Production of antibodies in plants: status after twenty years.* Plant Biotechnol Journal, 2010. 8(5): p. 529-63.
15. Ma, J.K.-C., et al., *Assembly of monoclonal antibodies with IgG1 and IgA heavy chain domains in transgenic tobacco plants.* European Journal of Immunology, 1994. 24(1): p. 131-138.
16. Karnoup, A.S., V. Turkelson, and W.H. Anderson, *O-linked glycosylation in maize-expressed human IgA1.* Glycobiology, 2005. 15(10): p. 965-81.
17. Nicholson, L., et al., *A recombinant multimeric immunoglobulin expressed in rice shows assembly-dependent subcellular localization in endosperm cells.* Plant Biotechnology Journal, 2005. 3: p. 115-127.
18. Wieland, W.H., et al., *Plant expression of chicken secretory antibodies derived from combinatorial libraries.* Journal of Biotechnology, 2006. 122(3): p. 382-91.
19. Juarez, P., et al., *Neutralizing antibodies against rotavirus produced in transgenically labelled purple tomatoes.* Plant Biotechnology Journal, 2012. 10: p. 341-352.
20. Nakanishi, K., et al., *Production of Hybrid-IgG/IgA Plantibodies with Neutralizing Activity against Shiga Toxin 1.* Plos One, 2013. 8(11): p. e80712.

21. Ebert, E.C., *Infliximab and the TNF-alpha system.* American Journal of Physioliology - Gastrointestinal and Liver Physiology, 2009. 296(3): p. G612-20.
22. Bain, B. and M. Brazil, *Adalimumab.* Nature Reviews Drug Discovery, 2003. 2(9): p. 693-94.
23. Gandhi, M., E. Alwawi, and K.B. Gordon, *Anti-p40 Antibodies Ustekinumab and Briakinumab: Blockade of Interleukin-12 and Interleukin-23 in the Treatment of Psoriasis.* Seminars in Cutaneous Medicine and Surgery, 2010. 29(1): p. 48-52.
24. Frigerio, L., et al., *Assembly, secretion, and vacuolar delivery of a hybrid immunoglobulin in plants.* Plant Physiology, 2000. 123(4): p. 1483-1493.
25. Hadlington, J.L., et al., *The C-terminal extension of a hybrid immunoglobulin A/G heavy chain is responsible for its golgi-mediated sorting to the vacuole.* Molecular Biology of the Cell, 2003. 14(6): p. 2592-2602.
26. Koide, Y., et al., *The N-terminal propeptide and the C terminus of the precursor to 20-kilo-dalton potato tuber protein can function as different types of vacuolar sorting signals.* Plant and Cell Physiology, 1999. 40(11): p. 1152-1159.
27. Vitale, A. and N.V. Raikhel, *What do proteins need to reach different vacuoles?* Trends in Plant Science, 1999. 4(4): p. 149-155.
28. Bednarek, S.Y., et al., *A Carboxyl-Terminal Propeptide Is Necessary for Proper Sorting of Barley Lectin to Vacuoles of Tobacco.* Plant Cell, 1990. 2(12): p. 1145-1155.
29. Neuhaus, J.M., et al., *A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole.* PNAS, 1991. 88(22): p. 10362-6.
30. Neuhaus, J.M. and J.C. Rogers, *Sorting of proteins to vacuoles in plant cells.* Plant Molecular Biology, 1998. 38(1-2): p. 127-144.
31. Park, M., et al., *Identification of the protein storage vacuole and protein targeting to the vacuole in leaf cells of three plant species.* Plant Physiology, 2004. 134(2): p. 625-639.
32. Fitchettelaine, A.C., et al., *Distribution of Xylosylation and Fucosylation in the Plant Golgi-Apparatus.* Plant Journal, 1994. 5(5): p. 673-682.
33. Rayon, C., P. Lerouge, and L. Faye, *The protein N-glycosylation in plants.* Journal of Experimental Botany, 1998. 49(326): p. 1463-1472.
34. Petrescu, A.J., et al., *Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding.* Glycobiology, 2004. 14(2): p. 103-114.
35. Mattu, T.S., et al., *The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc alpha receptor interactions.* Journal of Biological Chemistry, 1998. 273(4): p. 2260-2272.
36. Field, M.C., et al., *Structural analysis of the N-glycans from human immunoglobulin A1: comparison of normal human serum immunoglobulin A1 with that isolated from patients with rheumatoid arthritis.* Biochemistry Journal, 1994. 299 (Pt 1): p. 261-75.
37. Pucic, M., et al., *High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations.* Molecular & Cellular Proteomics, 2011. 10(10).
38. Herr, A.B., E.R. Ballister, and P.J. Bjorkman, *Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc.* Nature, 2003. 423(6940): p. 614-20.
39. Yoo, E.M., et al., *Differences in N-glycan structures found on recombinant IgA1 and IgA2 produced in murine myeloma and CHO cell lines.* Mabs, 2010. 2(3): p. 320-334.
40. Strasser, R., et al., *Enzymatic properties and subcellular localization of Arabidopsis beta-N-acetylhexosaminidases.* Plant Physiology, 2007. 145(1): p. 5-16.

41. Liebminger, E., et al., *beta-N-Acetylhexosaminidases HEXO1 and HEXO3 Are Responsible for the Formation of Paucimannosidic N-Glycans in Arabidopsis thaliana.* Journal of Biological Chemistry, 2011. 286(12): p. 10793-10802.
42. Kajiura, H., et al., *Arabidopsis beta 1,2-xylosyltransferase: Substrate specificity and participation in the plant-specific N-glycosylation pathway.* Journal of Bioscience and Bioengineering, 2012. 113(1): p. 48-54.
43. Cerutti, M. and J. Golay, *Lepidopteran cells, an alternative for the production of recombinant antibodies?* MAbs, 2012. 4(3): p. 294-309.
44. Palmberger, D., et al., *Insect cells for antibody production: evaluation of an efficient alternative.* Journal of Biotechnology, 2011. 153(3-4): p. 160-6.
45. Carayannopoulos, L., E.E. Max, and J.D. Capra, *Recombinant human IgA expressed in insect cells.* Proc Natl Acad Sci U S A, 1994. 91(18): p. 8348-52.
46. Hamilton, S.R. and T.U. Gerngross, *Glycosylation engineering in yeast: the advent of fully humanized yeast.* Current Opinion Biotechnology, 2007. 18(5): p. 387-92.
47. van Engelen, F.A., et al., *Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco.* Plant Molecular Biology, 1994. 26(6): p. 1701-10.
48. Westerhof, L.B., et al., *3D domain swapping causes extensive multimerisation of human interleukin-10 when expressed in planta.* PLoS One, 2012. 7(10): p. e46460.

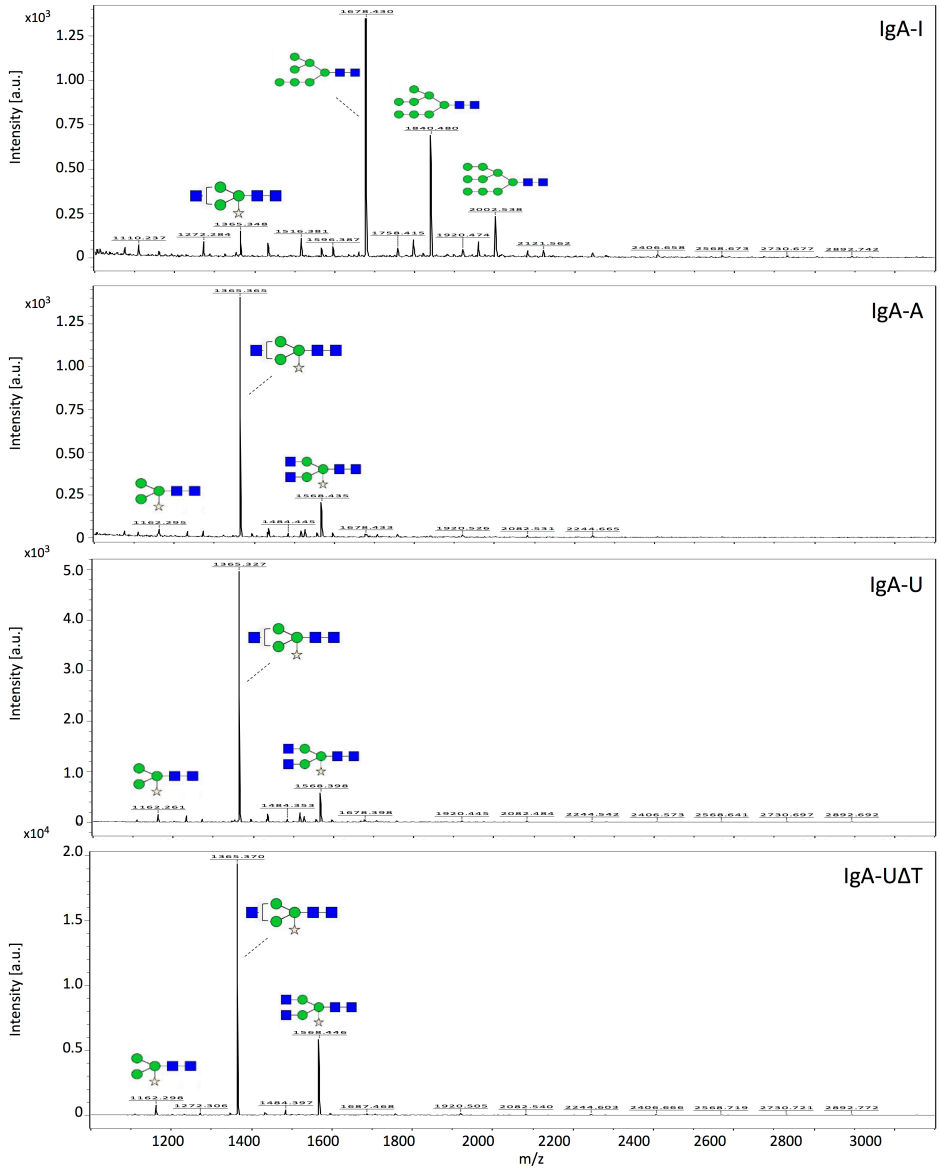
Supplementary Figures

Supplementary Figure 1



Supplementary Figure 1 ◆ *N*-glycan types carried by the plant-produced IgG1κ variants of Infliximab (I), Adalimumab (A) and Ustekinumab (U) as determined using MALDI-TOF-MS.

Supplementary Figure 2



Supplementary Figure 2 ◆ *N*-glycan types carried by the plant-produced IgA1κ variants of Infliximab (I), Adalimumab (A) and Ustekinumab (U) as determined using MALDI-TOF-MS.

Chapter 4

Evaluation of transient expression of the heteromultimeric protein complex sIgA in plants reveals J-chain incorporation as a production bottleneck

Lotte B. Westerhof, Ruud H. P. Wilbers, Debbie R. van Raaij, Christina Z. van Wijk, Aska Goverse, Jaap Bakker and Arjen Schots

Abstract

Secretory IgA (sIgA) is the most prominent antibody at mucosal surfaces playing a key role in the first line of defence. It is a promising antibody isotype for a variety of therapeutic settings such as passive vaccination or treatment of chronic inflammations by administering sIgA directed against pro-inflammatory cytokines. However, heterologous expression of sIgA is still suboptimal. The challenge for sIgA expression lies in the fact that it is a heteromultimeric protein complex for which assembly of four polypeptides; the alpha heavy chain, the light chain, the joining chain and part of the polymeric-Ig-receptor (called the secretory component) in a 4:4:1:1 ratio is required. We evaluated the transient expression *in planta* of sIgA variants of the clinical antibody Ustekinumab harbouring the heavy chain isotypes α1, α2m1 or α2m2. Ustekinumab is directed against the p40 subunit shared by the pro-inflammatory cytokines interleukin(IL)-12 and IL-23. Of the three different variants we obtained the highest yield with sIgA1. Combining all four required genes in one multi-gene expression vector increased yield three fold compared to co-infiltration of four separate expression vectors. However, not the expression strategy, but assembly turned out to be the limiting step for sIgA production, with the incorporation of the joining chain as the bottleneck. Our data demonstrate that different strategies of transient expression *in planta* are suitable for the production of heteromultimeric protein complexes such as sIgA, but can even be further enhanced by improving sIgA assembly.

Introduction

Secretory (s)IgA is the most predominant antibody type in external secretions of the human body and plays a key role in the first line of defence against mucosal pathogens. While human serum IgA is primarily monomeric, B cells in mucosa-associated lymphoid tissue secrete IgA in a dimeric form through incorporation of the joining chain. Dimeric (d)IgA can bind the polymeric immunoglobulin receptor on the basolateral surface of epithelial cells and upon binding dIgA is transcytosed to the luminal side of the cell. Here, the receptor is cleaved and dIgA is released in association with a part of the receptor, called the secretory component (SC). This complex is referred to as secretory IgA. When dIgA binds antigen just before, during (e.g. intracellular viruses in the epithelial cells) or after secretion, the immune system is not stimulated due to the limited presence of FcαRI positive immune cells in the mucosal area and the reduced affinity of sIgA to FcαRI. On the luminal side of the mucosa the glycans on the SC facilitate binding of sIgA to bacteria as well as to the mucus followed by sIgA (and bacterial) clearance by mucus release. This concept of pathogen binding and clearing without immune stimulation is referred to as immune exclusion and is believed not only to play a role in combating mucosal pathogens, but also in controlling commensal bacteria. In order to fulfil this role the human body secretes 40-60 mg sIgA per kg body weight each day [1].

IgA is found in two isotypes, IgA1 and IgA2 of which the latter occurs in two allotypes, IgA2m1 and IgA2m2. All three alpha heavy chains consist of one variable domain, three constant domains and an extended tailpiece allowing IgA to dimerise by incorporation of the joining chain. There are three major differences between these IgA variants. Firstly, IgA1 has an extended hinge region. This extension is heavily *O*-glycosylated, which is suspected to play a conformational role [2], allowing the binding of more distantly spaced antigens. Secondly, alpha heavy chains of isotypes 1 and 2m2 are covalently linked to their light chains via a disulphide bridge upon IgA assembly. No such linkage exists in IgA2m1, however, this allows the formation of a intermolecular disulphide bridge between the free cysteines within each of the two light chains. Finally, all variants differ from each other in the number of N-glycosylation sites they carry (2, 4 and 5 for IgA1, IgA2m1 and IgA2m2, respectively). The ratio wherein sIgA isotypes are present depends on the mucosal area and is most likely the result of the presence of specific bacteria as the hinge of IgA1 is sensitive to bacterial proteases [3-6].

The use of recombinant sIgA in passive mucosal immunotherapy in humans and livestock has been suggested as a good alternative for antibiotics. Several *in vivo* studies demonstrated (s)IgA's potential to locally control mucosal pathogens, such as *Mycobacterium tuberculosis* in the lungs [7], *Streptococcus mutans* in the mouth [8] and *Salmonella typhimurium*, *Vibrio cholera*, and *Cryptosporidium parvum* in the gastrointestinal tract [9]. When sIgA would be directed against pro-inflammatory cytokines and administered to the gut, it may relieve symptoms and induce remission in patients suffering from inflammatory bowel diseases (IBDs).

Currently treatment of IBD often includes anti-TNF-α therapy. However, systemic neutralization of this cytokine has been associated with the onset of tuberculosis [10]. Non-systemic cytokine neutralization may reduce infection risks and other side effects. Because sIgA is stable in the gut, sIgA-based therapy could be localized. Luminal administration of an anti-TNF-α antibody has been demonstrated effective in mouse models of IBD [11] and local drug administration has been suggested to be important for efficacy of IBD therapies [12].

Plants are a great production platform for pharmaceutical proteins. As eukaryotes they are able to correctly fold complex proteins and assemble protein complexes such as virus like particles [13] and secretory IgA [14]. Compared to mammalian cells, plants are a more economic production platform, as they do not require expensive culture conditions. Furthermore, the N-glycosylation pathway has been engineered to facilitate expression of proteins with humanized N-glycans [15]. Also, the engineering of mammalian mucin-type *O*-glycans has been achieved in plants [16, 17].

Plant expression of a murine sIgA/G hybrid was achieved by stable transformation of the four individual genes required for sIgA assembly followed by crossing of the highest producers [14]. A drawback of this strategy is that it is a lengthy and laborious process. Transient expression is much faster and almost always results in higher yields, as there is no constraint on expression imposed by the site of insertion. To achieve transient expression of more than one gene simultaneously, *Agrobacterium* cultures harbouring vectors for the expression of the individual genes need to be co-infiltrated (co-expression) or a multi-cassette vector facilitating expression of all genes should be used. Transient expression of chicken sIgA was achieved by co-infiltration [18] and human sIgA was transiently expressed with a multi-cassette vector assembled using the GoldenBraid system [19]. The risk with co-infiltration is that cells may not be transformed with all genes, but use of a multi-cassette vector may be impractical if expression of the individual proteins needs to be adjusted to reflect the stoichiometry of the protein complex. While all studies achieved sIgA expression, they also showed the presence of a large proportion of monomeric IgA as well as other assembly intermediates. The presence of these assembly intermediates complicates down-stream processing.

The objective of this study was to determine the factors that limit sIgA assembly. Thereto we studied the expression and assembly of a sIgA variant of the clinical antibody Ustekinumab (CNTO1275), which has specificity for the p40 subunit shared by interleukin-12 and interleukin-23. First, we evaluated the transient expression of all individual genes required for sIgA assembly whereby all three known alpha heavy chain types (1, 2m1 and 2m2) were included. Next we compared co-expression with the use of multi-cassette vectors. Although the use of a multi-cassette vector including all genes increased sIgA expression slightly compared to co-infiltration, the bottleneck for sIgA expression turned out not to be inefficient co-ordinate expression, but assembly of the complex. We conclude that the inefficient incorporation of the joining chain that limits sIgA assembly may be a consequence of

inefficient N-glycosylation of the IgA tailpiece and/or joining chain. Thus, improved N-glycosylation may be the key to enhance sIgA assembly.

Results

Expression of individual components required for sIgA assembly

In order to achieve *in vivo* assembly of a heteromultimeric protein complex all required genes should be expressed simultaneously in the same cell at the right level. The sIgA protein complex comprises four proteins and an overview of the individual proteins and the protein complexes IgA, dimeric (d)IgA and sIgA based on alpha heavy chain isotype 1 are given in figure 1a. Expression cassettes and the combinations thereof to achieve expression of these protein complexes are given in figure 1b.

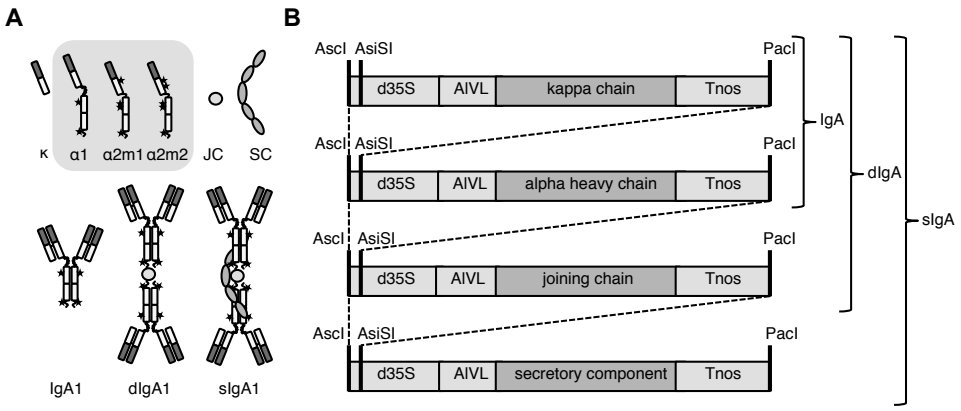


Figure 1 ◆ Overview of sIgA assembly and expression. A Individual proteins, alpha heavy chain 1 (α^1), 2m1 (α^{21}) or 2m2 (α^{22}), kappa chain (κ), joining chain (JC) and secretory component (SC), and protein complexes IgA1, dimeric (d)IgA1 and secretory (s)IgA1. **B** Expression cassettes with 35S promoter of the *Cauliflower mosaic virus* with duplicated enhancer (d35S), *Agrobacterium tumefaciens* nopaline synthase transcription terminator (Tnos), leader sequence of the Alfalfa mosaic virus RNA 4 (AlMV) and AscI, AsiSI and PacI restriction sites are indicated. Co-expression is required for IgA, dIgA and sIgA assembly as indicated by the accolades.

To evaluate the level and course of expression of the individual genes required for sIgA assembly we first expressed them individually and monitored expression over time using western blot analysis (Figure 2). All three known alpha heavy chain variants were evaluated. Upon expression of the alpha heavy chains and kappa chain bands were detected the expected size for these proteins and are assumed to represent the intact proteins. Furthermore, several bands >100 kDa were observed for the heavy chains and most likely represent multimers. A few faint bands below 50 kDa were detected, which most likely represent products of proteolytic degradation. Upon expression of the secretory component a band was detected that migrates ~10 kDa lower as compared to the secretory component of the sIgA control. This may be explained by differences in number and/or type of N-glycans received by the secretory component when expressed in plants. The secretory component has seven potential glycosylation sites, of which some may not be glycosylated in plants. Furthermore, the most common N-glycans of plant-secreted proteins are 0,7-1,1 kDa smaller than the most typical N-glycan found on human secretory component [20] which could account for the ~10 kDa difference. Upon expression of the joining chain many bands were detected, but all migrate higher as expected for a single joining chain and most of them most likely represent multimers.

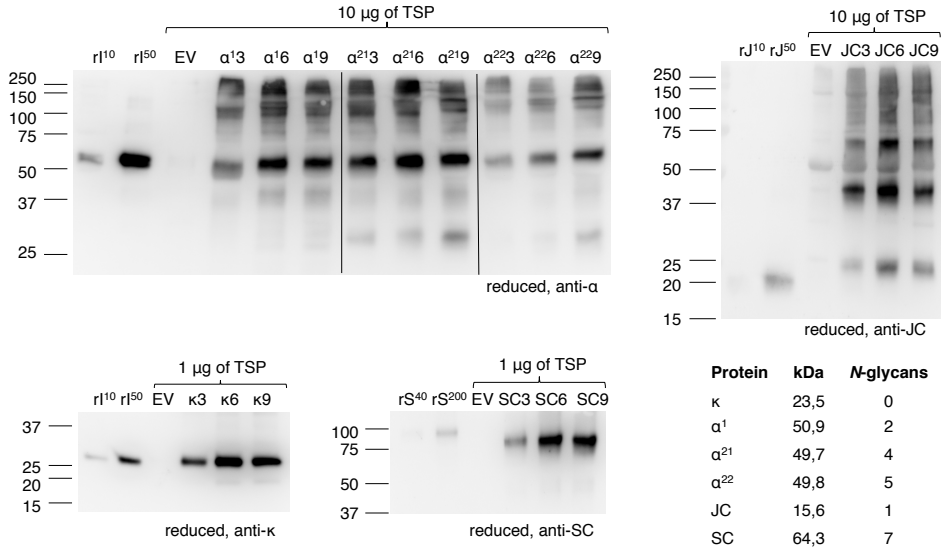


Figure 2 ◆ Plant expression of individual genes required for sIgA assembly. Leafs were agro-infiltrated for the expression of alpha heavy chain 1 (α^{1}), 2m1 (α^{21}), 2m2 (α^{22}), kappa chain (κ), joining chain (JC) or secretory component (SC) and harvested 3, 6 and 9 days post infiltration. Total soluble protein (TSP) was extracted and separated by SDS-PAGE under reducing conditions followed by visualization on a Western blot using protein specific antibodies as indicated. As controls 10 or 50 ng of recombinant IgA1κ ($rI^{10/50}$), recombinant joining chain ($rJ^{10/50}$) or 40 or 200 ng purified sIgA ($rS^{40/200}$, containing ~10 or 50 ng SC) and an empty vector (EV) sample were used.

The course of expression of the individual proteins is similar and peaks at 6 days post infiltration (dpi), except for alpha heavy chain 2m2, which peaks at 9 dpi. However, yield of the individual proteins varies. By comparing the band intensity of each individual protein to the recombinant control we estimate yields between 1 and 5 µg/mg TSP for the alpha heavy chains, between 5 and 20 µg/mg TSP for the joining chain and between 50 and 200 µg/mg TSP for the kappa chain and the secretory component at dpi 6. Yield estimation of the heavy chains and joining chain is solely based on the intact monomeric proteins and is most likely underestimated due to the presence of multimers. Nonetheless, as the stoichiometric ratio between the heavy chain, the kappa chain, the joining chain and the secretory component is 4:4:1:1, we assume heavy chain expression to be the limiting factor for sIgA assembly if stabilization of individual proteins upon co-expression does not occur.

Co-infiltration and use of a multi-cassette vector yield similar amounts of sIgA

Co-expression can be achieved by transient transformation in two ways, either by co-infiltration of *Agrobacterium* cultures each harbouring a vector for individual gene expression or by using multi-cassette vectors facilitating expression of all genes. With co-infiltration it is possible that not all cells are transformed with all expression vectors. The use of a multi-cassette vector would ensure that a transformed cell receives all genes, but transformation may be less efficient due to the larger size of the T-DNA. We successfully constructed several multi-cassette vectors for expression of IgA1κ (alpha heavy chain 1 and kappa chain), dIgA1κ (alpha heavy chain 1, kappa chain and joining chain) and sIgA1κ (all four genes) whereby all genes are under control of the same promoter and terminator (Figure 1b). Subsequently, sIgA was expressed by co-infiltration of all genes individually (4-vector system), co-infiltration of the secretory component and the joining chain with the multi-cassette vector for IgA1κ expression (3-vector system), co-infiltration of the secretory component and the multi-cassette vector for dIgA1κ expression (2-vector system) and the infiltration of the multi-cassette vector for sIgA1κ expression (1-vector system) (Figure 3b). Figure 3a shows the average sIgA yield of three biological replicates. To correct for the lower OD of *Agrobacterium* in the final mix of the 3, 2 and 1-vector system compared to the 4-vector system, an *Agrobacterium* culture carrying an empty vector was used to increase the OD (grey bars) or the OD of the *Agrobacterium* carrying a multi-cassette vector was increased (black bars). Combining all genes in the same vector results in only a slight increase in yield compared to the 2, 3 and 4-vector system. Apparently there is no clear optimal transient expression strategy for sIgA in terms of yield. Co-infiltration of different *Agrobacterium* cultures might still be preferred as the expression of individual genes can be adjusted to reflect the stoichiometric ratio of sIgA by adjusting the concentration of bacteria to be infiltrated.

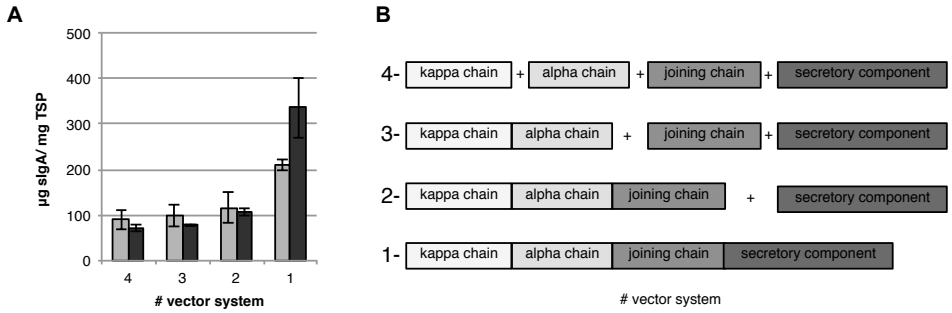


Figure 3 ◆ sIgA yield upon plant transformation using co-infiltration and multi-cassette vectors. A Average sIgA yield 6 days post infiltration (n=3) in leaf extracts when co-infiltrating all genes individually, co-infiltrating the secretory component and joining chain with the multi-cassette vector for IgA1κ expression, co-infiltrating the secretory component and the multi-cassette vector for dIgA expression and the infiltration of the multi-cassette vector for sIgA expression as indicated as the 4, 3, 2 and 1-vector system, respectively, as depicted in **B**. The final OD of the Agrobacterium culture was set to 2,5 using either Agrobacterium carrying an empty vector (light grey bars) or the OD of the Agrobacterium carrying the multi-cassette vector (dark grey bars) was increased. Error bars indicate standard error.

Stabilization of alpha heavy chains and kappa light chain upon co-expression

To determine the limiting factor in sIgA assembly we first evaluated the efficiency of IgA assembly in the absence of the joining chain and secretory component. Thereto, we co-expressed the alpha heavy chains and the kappa chain using the dual-cassette expression vectors for all three IgA variants and compared it with the individual expression of the alpha heavy chains and kappa chains. Leaf extracts were analysed by western blot under reducing and non-reducing conditions (Figure 4). Visualization of the alpha heavy chains and kappa chain under reducing conditions demonstrated that all proteins accumulate to a higher level upon co-expression. As the accumulation of the alpha heavy chains increases upon co-expression with the light chain the presence of a degradation product just above 25 kDa also becomes clear. Considering the size of this degradation product it is most likely the result of cleavage in the hinge region. The hinge regions of alpha heavy chain 1 and alpha heavy chains 2m1 and 2m2 seem all sensitive to proteolysis.

While analysing co-expression of alpha heavy chains with kappa light chain under non-reducing conditions, a band around the expected size for IgA (~150 kDa) was detected on blots either treated with anti-alpha heavy chain or anti-kappa chain specific antibodies. We therefore assumed that this band represents the intact IgA complex. Next to the 150 kDa band, several bands migrating >250 kDa were detected. As these bands were also seen in the recombinant IgA control we assume they represent multimers of IgA. Also, several bands <150 kDa were detected, which may represent assembly intermediates, degradation products and/or individual polypeptide chains. Proteolytic cleavage in the hinge region results in Fc and Fab

fragments of the same size as an intact alpha heavy chain (~50 kDa). A band of 50 kDa is clearly detected on both alpha heavy chain and the kappa chain specific blots and therefore most likely represents Fab fragments. Two bands only detected on the kappa chain specific blot just below 25 and 50 kDa most likely represent un-associated monomeric and dimeric kappa chains. Assuming only un-associated kappa chain is present, we conclude that accumulation of the alpha heavy chain is the limiting factor for IgA yield, despite the fact that the alpha heavy chains stabilize upon co-expression with the kappa chain.

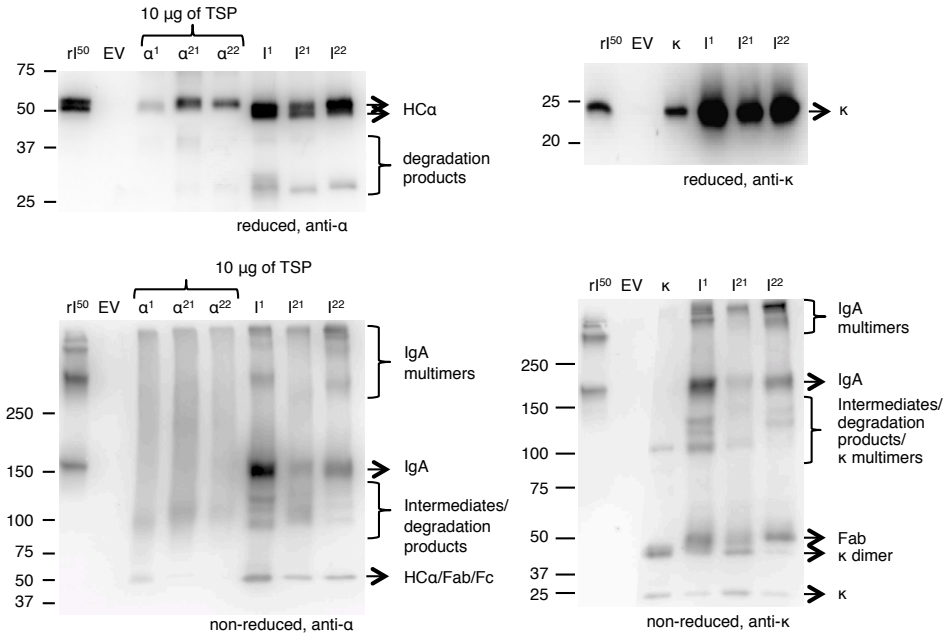


Figure 4 ◆ IgA assembly using multi-cassette expression vectors. Leafs were agro-infiltrated for the expression of alpha heavy chain 1 (α^1), 2m1 (α^{21}), 2m2 (α^{22}) and kappa chain (κ) individually or co-expressed (I) using a multi-cassette vector and harvested 6 days post infiltration. Total soluble proteins were extracted and 1 µg, unless stated otherwise, was separated by SDS-PAGE under reducing conditions followed by visualization using protein specific antibodies as indicated. As controls 50 ng of recombinant IgA1κ (rI^{50}) and an empty vector (EV) sample were used.

IgA dimerization is the limiting step in sIgA assembly

Next we evaluated the efficiency of sIgA assembly. We used the multi-cassette vectors to express IgA, dIgA and sIgA (Figure 3b) with all three alpha heavy chain variants. Leaf extracts were analysed on western blot under non-reducing conditions (Figure 5). Upon dIgA1, dIgA2m1 and dIgA2m2 expression two bands around 300 and >420 kDa were detected on the joining chain specific blot (third panel from the top). These bands can also be seen on the alpha heavy chain and kappa chain specific blots (first and second panel respectively). We assume that these bands represent dIgA and multimerized dIgA. The 150 kDa band representing monomeric IgA in the alpha heavy chain and kappa chain specific blots is still present upon co-expression of the joining chain. This could imply that dimerization of IgA is not 100% efficient or that the expression of the joining chain is limiting. The presence of free joining chains was, however, not detected.

Upon sIgA1 expression monomeric IgA (~150 kDa) is still detected on the alpha heavy chain and kappa chain specific blots. Next to that, bands of <100 kDa, ~220 kDa, ~400 kDa and >420 kDa were detected on the secretory component specific blot (bottom panel). Because the band <100 kDa was only detected on the secretory component specific blot, this band most likely represents loose secretory component. The 220 kDa band was also detected on the alpha heavy chain and kappa chain specific blot, but not on the joining chain specific blot and therefore most likely represents secretory component associated with monomeric IgA. Both the ~400 kDa and >420 kDa bands were also detected on the alpha heavy chain, kappa chain and joining chain blot and therefore must represent sIgA and multimers of sIgA. Upon sIgA2m1 and sIgA2m2 expression only the band <100 kDa representing the loose secretory component and the band >420 kDa representing multimeric sIgA was clearly distinguished. It may be that alpha heavy chains 2m1 and 2m2 are more sensitive to multimerisation. No dIgA was detected upon expression of all four genes using any of the alpha heavy chains. Apparently sIgA assembly is equally efficient for all three alpha heavy chains and the expression of the secretory component is not limiting. The latter can also be concluded by the ample presence of loose secretory component in the sIgA samples.

Both IgA and sIgA yield were determined using a sandwich ELISA. While yields of the IgA variants ranged between 10-40 µg/mg TSP, sIgA yield was at least 9 fold higher for all variants (Figure 6). The higher sIgA yield, compared to IgA yield, may be explained by the fact that sIgA assembly prevents proteolytic degradation of IgA. However, sIgA yield may also be overestimated due to the presence of monomeric IgA associated with the secretory component and/or the presence of multimers. Taken together our data show that dimerization of IgA is the limiting step for sIgA assembly. Improving dIgA assembly could further increase sIgA yield and simplify down-stream processing.

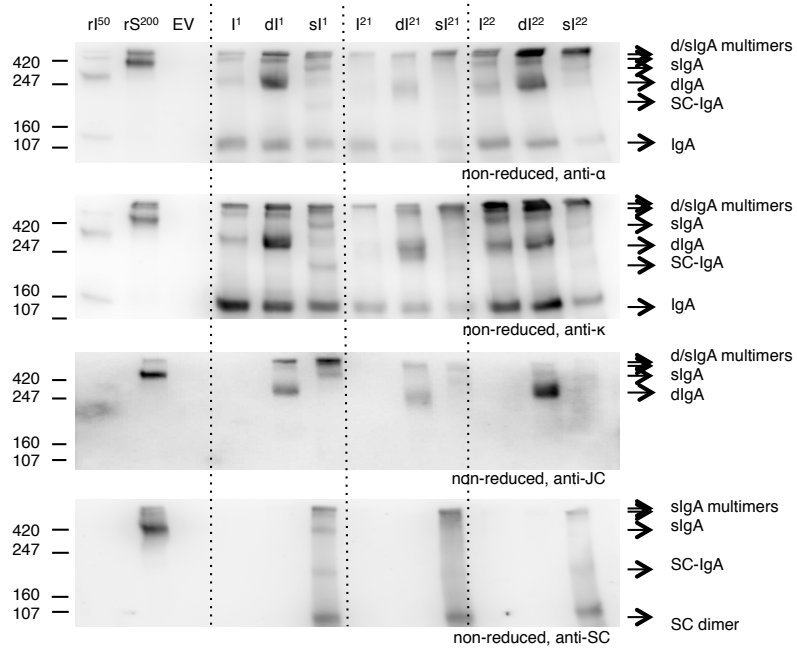


Figure 5 ◆ sIgA assembly using multi-cassette expression vectors. Leafs were agro-infiltrated for the expression of IgA1 (I^1), 2m1 (I^{21}) or 2m2 (I^{22}) together with joining chain (dI) and secretory component (sI) using multi-cassette vectors (vector system 1 in figure 3b) and harvested 6 days post infiltration. Total soluble proteins were extracted and 1 µg was separated by SDS-PAGE under reducing conditions followed by visualization using protein specific antibodies as indicated. As controls 50 ng of recombinant IgA1κ (rI^{50}) and 200 ng purified sIgA (rS^{200}) and an empty vector (EV) sample were used.

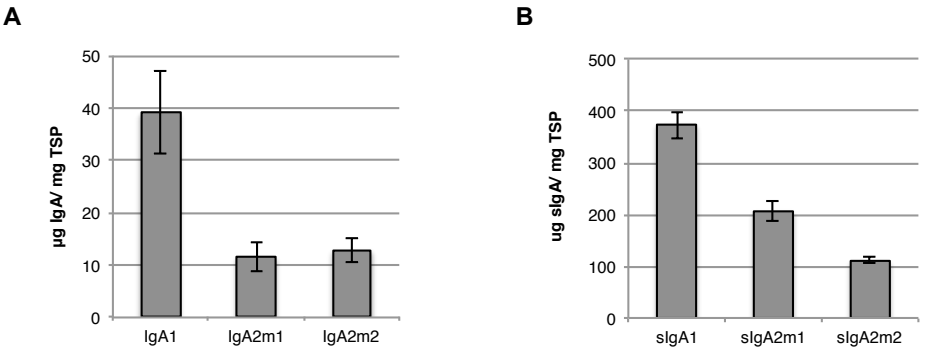


Figure 6 ◆ IgA and sIgA yield using multi-cassette expression vectors. A Average IgA yield (n=3) upon expression of IgA1, IgA2m1 and IgA2m2 using the multi-cassette vectors (vector system 1 in figure 3b) and harvested 6 days post infiltration. IgA yield in leaf extracts was determined by sandwich ELISA using a kappa chain specific capture and alpha heavy chain specific detection antibody. Error bars indicate standard error. **B** Average sIgA yield (n=3) upon expression of sIgA1, sIgA2m1 and sIgA2m2 using the multi-cassette vectors and harvested 6 days post infiltration. sIgA yield in leaf extracts was determined by sandwich ELISA using a kappa chain specific capture and secretory component specific detection antibody. Error bars indicate standard error.

Joining chain incorporation is the limiting factor for sIgA yield

Next, we investigated whether joining chain expression is the limiting factor for sIgA assembly. Thereto, we attempted to increase the expression of the joining chain by increasing the OD of the *Agrobacterium* culture harbouring the vector for joining chain expression using the 3-vector system. sIgA yield was determined by sandwich ELISA and joining chain expression was evaluated using western blot analysis (Figure 7a and 7b). While the expression of the joining chain increased by using a higher OD of the *Agrobacterium* culture harbouring the vector for the joining chain expression, sIgA yield did not increase. We therefore assume that not joining chain expression, but its incorporation in the dIgA complex is the limiting factor for sIgA yield.

On a side note, the band that most likely represents monomeric joining chain (~25 kDa) appears to be a doublet (two bands migrating very close to each other). A doublet may represent twice the same protein with a different number of N-glycans. Because the joining chain only harbours one N-glycosylation site, this doublet should represent a non-glycosylated and glycosylated joining chain.

To evaluate if lowering IgA expression would influence the efficiency of dIgA assembly, we also attempted to decrease the amount of IgA by reducing the OD of the *Agrobacterium* culture harbouring the vector for IgA expression using the 3-vector system. Again sIgA yield was determined by sandwich ELISA and sIgA assembly was evaluated using western blot analysis (Figure 7c and 7d). Upon reduction of the OD of the *Agrobacterium* culture harbouring the vector for IgA expression the amount of monomeric and dimeric IgA reduced. However, also sIgA yield is reduced when the OD of the *Agrobacterium* culture that harbours the vector for IgA expression becomes lower than 0.2. At this OD, significant amounts of monomeric IgA with and without associated secretory component and dimeric IgA are still present. Thus, dIgA assembly cannot be improved by adjusting the expression of its individual components. It is therefore likely that not the capacity of the plant cell to assemble the protein complex, but intrinsic properties of the individual proteins determine dIgA assembly efficiency.

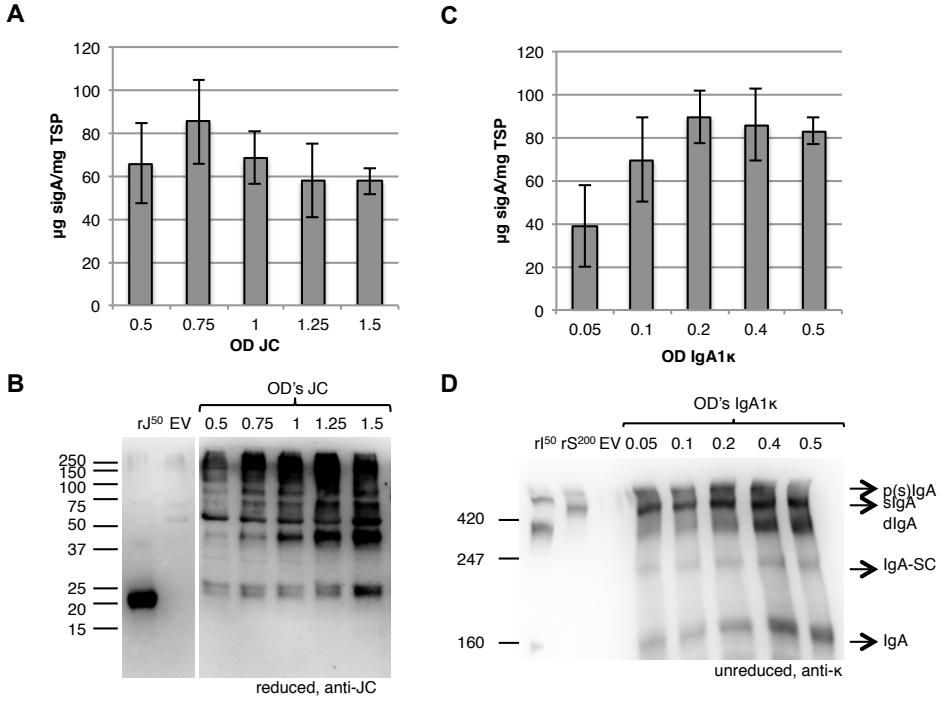


Figure 7 ◆ Incorporation of the joining chain is the limiting factor for sIgA yield. Leafs were agro-infiltrated for the expression of IgA1, joining chain (JC) and secretory component (SC) by co-infiltration (vector system 3 in figure 3b) and harvested 6 days post infiltration. **A** Average sIgA yield (n=3, error bars indicate standard error) and **B** western blot analysis upon increasing the OD of the *Agrobacterium* culture facilitating joining chain expression. As controls recombinant joining chain (rJ^{50}) and an empty vector (EV) sample were used. **C** Average sIgA yield (n=3, error bars indicate standard error) and **D** western blot analysis of sIgA assembly upon decreasing OD of the *Agrobacterium* culture facilitating IgA expression. As controls 50 ng of recombinant IgA1κ (rI^{50}) and 200 ng purified sIgA (rS^{200}) and an empty vector (EV) sample were used.

Discussion

We have studied the limiting factors for sIgA production upon transient expression in *N. benthamiana*. Because sIgA is a heteromultimeric protein complex transient expression can be achieved in several ways. *Agrobacterium* cultures each harbouring expression vectors that facilitate expression of the individual components can be co-infiltrated or a multi-cassette expression vector facilitating expression of several or all components may be used. Also, a combination of these two strategies may be used. The risk with co-infiltration is that not all cells are transformed with all genes. This may result in the presence of un-associated components or assembly intermediates that may complicate downstream processing. Use of a multi-cassette vector would ensure that each transformed cell expresses each gene, however, the much larger T-DNA may be less efficiently transferred into the plant cells. We first evaluated the course of expression of all individual proteins followed by evaluation of four expression strategies to achieve sIgA production. We co-infiltrated all genes individually (4-vector system), co-infiltrated the secretory component and joining chain with the multi-cassette vector for IgA1κ expression (3-vector system), co-infiltrated the secretory component and the multi-cassette vector for dIgA1κ expression (2-vector system) and finally also used a single multi-cassette vector for sIgA1κ expression (1-vector system). Although the T-DNA of the single multi-cassette vector for sIgA1κ expression was ~10,3 kbp in size, it provided slightly higher sIgA yield compared to the 4, 3 and 2-vector systems. Thus, in our study increased size of the T-DNA does not reduce transformation efficiency. Furthermore, the 4, 3 and 2-vector systems all gave comparable sIgA yield. Therefore, most cells must have been transformed repeatedly, whereby most cells received all T-DNAs required to facilitate sIgA expression. Despite the fact that the same promoter and terminator sequences were used to facilitate expression of all genes, loss of vector parts during cloning or loss of sIgA expression upon repeated plant transformation was never observed. We therefore assume that our multi-cassette expression vectors are stable and recombination did not occur.

Although the use of the single multi-cassette expression vector did result in accumulation of a significant amount of fully assembled sIgA, various assembly intermediates were also detected. Regardless of the expression strategy used a significant proportion of monomeric IgA was observed next to sIgA. In three other studies on expression of sIgA in plants a significant proportion of monomeric IgA was observed as well. Stable expression of a murine IgA/G hybrid with murine joining chain resulted in a 1:1 ratio of IgA and dIgA [14]. Upon transient co-expression of a chicken IgA and chicken joining chain just over half of the IgA dimerized [18]. In a study on transient expression of human sIgA, the use of a multi-cassette vector resulted in a mix of IgA, dIgA and sIgA of which 33% was confirmed to be sIgA [19]. Follow up studies on the expression of the murine IgA/G hybrid demonstrated that this antibody was targeted to the vacuole due to a cryptic targeting signal in the tailpiece of murine IgA [21, 22]. In our previous publication on expression of monomeric human IgA we demonstrate that

human IgA is also poorly secreted from plant cells [23]. The tailpieces of both human and chicken IgA contain similar sequences as the suggested cryptic targeting signal of murine IgA. Thus, it may well be that also chicken and human IgA are targeted to the vacuole. If IgA is targeted to the vacuole, it is possible that a proportion of IgA is transported to the vacuole before the joining chain can be incorporated. Unfortunately the tailpiece of IgA cannot be removed, as the penultimate cysteine residue forms a disulphide bond with a free cysteine of the joining chain [24]. An investigation if mutations in the cryptic vacuolar targeting signal can abolish the vacuolar targeting without influencing the complex assembly may provide a solution.

Reports on the expression of sIgA in Chinese hamster ovary (CHO) cells also identify dIgA assembly as the yield-limiting step in sIgA expression [25, 26]. Because mammalian cells are devoid of vacuoles, vacuolar targeting cannot explain the lack of sIgA assembly. Unfortunately, both studies did not determine joining chain expression, thus it is unclear if sIgA assembly in CHO cells is caused by limited joining chain expression or inefficient incorporation of the joining chain. To evaluate if either joining chain expression or incorporation was the limiting step for the yield of our sIgA we increased the bacterial OD of the *Agrobacterium* culture facilitating joining chain expression and demonstrated that while joining chain expression was increased, sIgA yield was not, nor had the proportion of monomeric IgA diminished. We therefore assumed that joining chain incorporation was the limiting step for sIgA assembly. We also, observed that the band assumed to represent monomeric joining chain migrated as a doublet. Because the joining chain harbours only one N-glycosylation site it is possible that this doublet represents a non-glycosylated and a once glycosylated version of the joining chain. It was demonstrated that incorporation of the joining chain is reduced if the asparagine 48 of the joining chain or asparagine 549 of the alpha heavy chain is not glycosylated [24, 27]. In a previous study we already confirmed that the N-glycosylation of the asparagine 549 of the alpha heavy chain is partial [23]. Partial N-glycosylation as a reason for inefficient incorporation of the joining chain coincides with the fact that also reduced expression of IgA did not alter the IgA:sIgA ratio. We therefore assume that not the capacity of the plant cell to assemble dIgA is limiting, but that partial N-glycosylation of IgA and/or joining chain is the reason for inefficient sIgA assembly both for plant as well as CHO cell produced sIgA.

Because N-glycosylation is co-translational, limited access to the glycosylation site cannot explain inefficient N-glycosylation. However, in a large-scale analysis of glycoproteins it was demonstrated that the glycosylation signals N-X-T and N-X-S are N-glycosylated in 70% and 30% of the cases, respectively [28]. This suggests that the signal N-X-T is more efficiently glycosylated. Both the N-glycosylation site in the tailpiece of the alpha heavy chains and the joining chain are of the N-X-S type. Mutation of the serine to a threonine may enhance glycosylation and therefore also sIgA assembly. Despite the fact that sIgA yield was limited by the efficiency of joining chain incorporation, sIgA yield was already high, in the order of

hundreds of µg sIgA per mg TSP. If sIgA assembly can be improved by mutation of the N-glycosylation signals of the alpha heavy chain and/or the joining chain, thereby reducing the presence of sIgA intermediates, plants could allow the economic production of sIgA for clinical use.

Experimental procedures

Construct design

GeneArt (Bleiswijk, the Netherlands) synthesized all below-mentioned gene fragments except the constant domains of the human immunoglobulin alpha-2m1, which was amplified from the human transcriptome library MegaMan (Agilent Technologies, Middelburg, the Netherlands). First, undesired restriction sites were removed from the sequences of the constant domains of human immunoglobulin alpha-1 (ACC82528.1), alpha-2m1 (AL928742.3) and kappa (AGH70219.1) chains, joining chain (AK312014.1) and secretory component (1-764 codons of the polymeric immunoglobulin receptor; AAB23176.1). To obtain the sequence for the constant domains of the human immunoglobulin alpha-2m2 the sequence of the human immunoglobulin alpha-2m1 was adapted (as indicated in P01877). The variable regions of the commercial antibody Ustekinumab (CNTO-1275) and the signal peptide of the *Arabidopsis thaliana* chitinase gene (AAM10081.1) were recoded from the amino acid sequence using codons preferred by *Nicotiana benthamiana*. For subsequent cloning and assembly of the full alpha heavy chain genes, the gene fragments were flanked by the following restriction sites at the 5′ and 3′-end: NcoI-EagI, EagI-NheI, NheI-KpnI, for the signal peptide, the heavy chain variable and alpha heavy chain constant regions, respectively. For subsequent cloning and assembly of the kappa chain, the gene fragments were flanked by the following restriction sites at the 5′ and 3′-end: NcoI-EagI, EagI-BsiWI, BsiWI-KpnI for the signal peptide the kappa chain variable and constant region, respectively. For subsequent cloning of the joining chain and secretory component the sequences were flanked by NcoI-KpnI at the 5′ and 3′-end, respectively. None of the restriction sites used introduced extra amino acids except NcoI, which in some cases introduced an extra alanine after the start methionine. Genes were ligated into the shuttle vector pRAPa, a pRAP (or pUCAP35S)[29] derivative modified to include an AsiSI restriction site by introduction of the self-annealed oligo 5′-AGCTGGCGATCGCC-3′ into a HindIII linearized pRAP. In pRAPa all open reading frames are placed under the control of the 35S promoter of the *Cauliflower mosaic virus* with duplicated enhancer (d35S) and the *Agrobacterium tumefaciens* nopaline synthase transcription terminator (Tnos). A 5′ leader sequence of the Alfalfa mosaic virus RNA 4 (AlMV) is also included between the promoter and gene to boost translation. From pRAPa the expression cassettes were digested with AscI and PacI, confirmed by sequencing, and ligated into the

expression vector pHYG [30]. Use of the restriction sites AscI and AsiSI allowed subsequent introduction of expression cassettes as AsiSI creates the same overhang as PacI (Figure 2a). Expression cassettes were transformed to *Agrobacterium tumefaciens* strain MOG101 for plant expression.

Plant transformation

Agrobacterium clones were cultured overnight (o/n) at 28°C in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH7.0) containing 50 µg/ml kanamycin and 20 µg/ml rifampicin. The optical density (OD) of the o/n cultures was measured at 600 nm and used to inoculate 50 ml of LB medium containing 200 µM acetosyringone and 50 µg/ml kanamycin. After 16 hours the bacterial cultures were centrifuged for 15 min at 2800 *g* and the bacteria were resuspended in MMA infiltration medium (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 µM acetosyringone. For co-expression *Agrobacterium* cultures were either mixed or multi-cassette vectors were used. The final OD of each *Argobacterium* culture in an infiltration mix was 0.5 unless otherwise indicated. The total OD of the infiltration mix was kept the same within an experiment by use of an *Agrobacterium* culture harboring an empty vector if needed. The *Tomato bushy stunt virus* (TBSV) silencing inhibitor p19 was always co-expressed. After 1-2 hours incubation of the infiltration mix at room temperature, the two youngest fully expanded leaves of 5-6 weeks old *Nicotiana benthamiana* plants were infiltrated completely. Infiltration was performed by injecting the *Agrobacterium* suspension into a *Nicotiana benthamiana* leaf at the abaxial side using a needleless 1 ml syringe. Infiltrated plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) and infiltrated leaves were harvested at selected time points.

Total soluble protein extraction

Leaf disks were cut from infiltrated leafs and immediately snap-frozen upon harvesting and homogenized in liquid nitrogen using a TissueLyser II (Qiagen, Venlo, the Netherlands) and stored at -20°C until use. Homogenized plant material was mixed in ice-cold extraction buffer (50mM phosphate-buffered saline (PBS) pH=7.4, 100 mM NaCl, 10 mM EDTA, 0.1% v/v Tween-20, 2% w/v immobilized polyvinylpolypyrrolidone (PVPP)) using 2 ml/g fresh weight in the TissueLyser II. Crude extract was clarified by centrifugation at 16.000 x*g* for 5min at 4°C.

IgA and sIgA quantification

IgA and sIgA concentrations in crude extracts were determined by sandwich ELISA. ELISA plates (Greiner Bio One; Alphen aan den Rijn, the Netherlands) were coated with a goat polyclonal anti-human kappa antibody (Sigma-Aldrich-Aldrich; Zwijndrecht, the Netherlands) in coating buffer (eBioscience, Vienna, Austria) o/n at 4°C in a moist environment. After this and each following step the plate was washed 5 times with 30 sec intervals in PBST (1x PBS, 0,05% Tween-20) using an automatic plate washer model 1575 (BioRad; Veenendaal, the

Netherlands). The plate was blocked with assay diluent (eBioscience) for 1 h at room temperature. Samples and a standard line were loaded in serial dilutions and incubated for 1 h at room temperature. For IgA determination recombinant human IgA1κ (InvivoGen; Toulouse, France) was used as a standard in a 2-fold dilution series from 100 to 0,31 ng/ml in assay diluent. For sIgA determination colostrum purified sIgA (Sigma-Aldrich) was used as a standard in a 2-fold dilution series from 1000 to 3,1 ng/ml in assay diluent. Hereafter a HRP-conjugated goat polyclonal antibody directed against the constant domains of human IgA (Sigma-Aldrich) or a biotinylated goat polyclonal antibody directed against the human secretory component (Sigma-Aldrich) was used for detection of IgA and sIgA, respectively. Avidin-HRP (eBioscience) conjugate was used to bind the biotin of the anti-secretory component antibody. 3,3′,5,5′-Tetramethylbenzidine (TMB) substrate (eBioscience) was added and colouring reaction was stopped using 0.18M sulphuric acid after 1-30 min. OD read outs were performed using the model 680 microplate reader (BioRad) at 450 nm with correction filter of 690 nm. For sample comparison the total soluble protein (TSP) concentration was determined using the BCA Protein Assay Kit (Pierce) according to supplier's protocol using bovine serum albumin (BSA) as a standard.

Protein analysis by Western blot

For western blot analysis clarified protein extracts were desalted using a Sephadex G25 (VWR International; Amsterdam, The Netherlands) column prior to BCA analysis. One µg (unless otherwise indicated) of total soluble protein was separated under reducing conditions by SDS-PAGE on in house made 6 or 12% Bis-Tris gels. Recombinant IgA1κ (InvivoGen) and/or colostrum purified sIgA (Sigma-Aldrich) or recombinant joining chain (Sino Biological; Cologne, Germany) were used as controls. Proteins were transferred to an Invitrolon™ PVDF membrane (Invitrogen) by a wet blotting procedure (Life technologies; Bleiswijk, the Netherlands). Thereafter the membrane was blocked in PBST-BL (PBS containing 0.1% v/v Tween-20 and 5% w/v non-fat dry milk powder) for 1 hour at room temperature, followed by overnight incubation with an anti-human kappa (Sigma-Aldrich), anti-human immunoglobulin alpha (Sigma-Aldrich), anti-joining chain (Nordic Immunological Laboratories; Tilburg, the Netherlands) or HRP-conjugated anti-secretory component (Sigma-Aldrich) antibody. The membrane was washed 5 times in PBST (PBS containing 0.1% v/v Tween-20). For detection of human kappa, human immunoglobulin alpha and joining chain a HRP-conjugated anti-goat IgG antibody (Jackson ImmunoResearch; Suffolk, UK) was used where after washing steps were repeated. Finally, the SuperSignal West Dura substrate (Thermo Fisher Scientific; Etten-Leur, the Netherlands) was used as HRP substrate. Pictures were taken using a G:BOX Chemi System device (SynGene; Cambridge, UK).

Acknowledgments

This chapter was financially supported in part by Synthon (Nijmegen, The Netherlands) and a grant from the Dutch Ministry of Economic Affairs (PID07124). The authors declare no conflict of interest. We would like to thank Gerard Rouwendal for his help in designing the multi-cassette vector system and Tim Warbroek, Aleksandra Syta and Bob Engelen for their input in the experimental work.

References

1. Conley, M.E. and D.L. Delacroix, *Intravascular and Mucosal Immunoglobulin A: Two Separate but Related Systems of Immune Defense?* Annals of Internal Medicine, 1987. 106(6): p. 892-899.
2. Narimatsu, Y., et al., *Effect of glycosylation on cis/trans isomerization of prolines in IgA1-hinge peptide.* Journal of the American Chemical Society, 2010. 132(16): p. 5548-9.
3. Stoop, J.W., et al., *Serum immunoglobulin levels in healthy children and adults.* Clinical & Experimental Immunology, 1969. 4(1): p. 101-12.
4. Vaerman, J.-P., J.F. Heremans, and C.B. Laurell, *Distribution of alpha chain subclasses in normal and pathological IgA-globulins.* Immunology, 1968. 14: p. 425-432.
5. Delacroix, D.L., et al., *IgA subclasses in various secretions and in serum.* Immunology, 1982. 47: p. 383-385.
6. Kerr, A., *The structure and function of human IgA.* Biochemical Journal, 1990. 271: p. 285-296.
7. Williams, A., et al., *Passive protection with immunoglobulin A antibodies against tuberculous early infection of the lungs.* Immunology, 2004. 111(3): p. 328-33.
8. Ma, J.K., et al., *Characterization of a recombinant plant monoclonal secretory antibody and preventive immunotherapy in humans.* Nature Medicine, 1998. 4(5): p. 601-6.
9. Enriquez, F.J. and M.W. Riggs, *Role of immunoglobulin A monoclonal antibodies against P23 in controlling murine Cryptosporidium parvum infection.* Infection and Immunity, 1998. 66(9): p. 4469-73.
10. Keane, J., et al., *Tuberculosis associated with infliximab, a tumor necrosis factor (alpha)-neutralizing agent.* New England Journal of Medicine, 2001. 345(15): p. 1098-1104.
11. Bhol, K.C., et al., *AVX-470: a novel oral anti-TNF antibody with therapeutic potential in inflammatory bowel disease.* Inflammatory Bowel Diseases, 2013. 19(11): p. 2273-81.
12. Neurath, M.F., *Cytokines in inflammatory bowel disease.* Nature Reviews Immunology, 2014. 14(5): p. 329-42.
13. Thuenemann, E.C., et al., *A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles.* Plant Biotechnology Journal, 2013. 11(7): p. 839-846.
14. Ma, J.K., et al., *Generation and assembly of secretory antibodies in plants.* Science, 1995. 268(5211): p. 716-9.
15. Bosch, D., et al., *N-Glycosylation of plant-produced recombinant proteins.* Current Pharmaceutical Design, 2013.
16. Yang, Z., et al., *Engineering mammalian mucin-type O-glycosylation in plants.* Journal Biological Chemistry, 2012. 287(15): p. 11911-23.
17. Castilho, A., et al., *Engineering of sialylated mucin-type O-glycosylation in plants.* Journal Biological Chemistry, 2012. 287(43): p. 36518-26.
18. Wieland, W.H., et al., *Plant expression of chicken secretory antibodies derived from combinatorial libraries.* Journal of Biotechnology, 2006. 122(3): p. 382-91.
19. Juarez, P., et al., *Combinatorial Analysis of Secretory Immunoglobulin A (sIgA) Expression in Plants.* International Journal of Molecular Sciences, 2013. 14(3): p. 6205-22.

20. Royle, L., et al., *Secretory IgA N- and O-glycans provide a link between the innate and adaptive immune systems.* Journal of Biological Chemistry, 2003. 278(22): p. 20140-20153.
21. Frigerio, L., et al., *Assembly, secretion, and vacuolar delivery of a hybrid immunoglobulin in plants.* Plant Physiology, 2000. 123(4): p. 1483-1493.
22. Hadlington, J.L., et al., *The C-terminal extension of a hybrid immunoglobulin A/G heavy chain is responsible for its golgi-mediated sorting to the vacuole.* Molecular Biology of the Cell, 2003. 14(6): p. 2592-2602.
23. Westerhof, L.B., et al., *Monomeric IgA can be produced in planta as efficient as IgG, yet receives different N-glycans* Plant Biotechnol Journal, 2014. In press - see chapter 3 this thesis.
24. Atkin, J.D., et al., *Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly.* Journal of Immunology, 1996. 157(1): p. 156-9.
25. Berdoz, J., et al., *In vitro comparison of the antigen-binding and stability properties of the various molecular forms of IgA antibodies assembled and produced in CHO cells.* PNAS, 1999. 96(6): p. 3029-3034.
26. Li, C., et al., *Construction of a Chimeric Secretory IgA and Its Neutralization Activity against Avian Influenza Virus H5N1.* Journal of Immunology Research, 2014.
27. Krugmann, S., et al., *Structural requirements for assembly of dimeric IgA probed by site-directed mutagenesis of J chain and a cysteine residue of the alpha-chain CH2 domain.* Journal of Immunology, 1997. 159(1): p. 244-9.
28. Petrescu, A.J., et al., *Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding.* Glycobiology, 2004. 14(2): p. 103-114.
29. van Engelen, F.A., et al., *Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco.* Plant Molecular Biology, 1994. 26(6): p. 1701-10.
30. Westerhof, L.B., et al., *3D domain swapping causes extensive multimerisation of human interleukin-10 when expressed in planta.* PLoS One, 2012. 7(10): p. e46460.

Chapter 5

Expression of *Schistosoma mansoni* omega-1 with diantennary N-glycans carrying Lewis X motifs in plants

Ruud H.P. Wilbers☯, Lotte B. Westerhof☯, Debbie R. van Raaij, Eva Capuder, Alja van der Schuren, Jose L. Lozano-Torres, Dieu-Linh Nguyen, Leonie Hussaarts, Nicole N. Driessen, Maria Yazdanbakhsh, Dirk Bosch, Geert Smant, Jaap Bakker, Cornelis H. Hokke and Arjen Schots

☯Equal contribution

Abstract

Schistosoma mansoni is a parasitic trematode that, like other helminths, secretes immunomodulatory proteins to sustain infection and to complete its life cycle. These secreted proteins are main topics of research as they are possible vaccine candidates or may have therapeutic potential to treat inflammatory disorders. Omega-1 is secreted by eggs of *S. mansoni* and is a key factor for the induction of Th2 cell polarisation via instruction of dendritic cells. *In vitro* induction of Th2 responses by omega-1 is dependent on its RNase activity as well as its N-glycans that are required for the internalisation by dendritic cells. However, the exact role(s) of N-glycans on the immunomodulatory properties of helminth secreted glycoproteins remains to be elucidated. As the purification of a single glycoprotein from *S. mansoni* egg extracts is inefficient and unsustainable, a platform is required that enables production of omega-1 with its native N-glycans. Here we show that *S. mansoni*-derived omega-1 can be efficiently produced in *Nicotiana benthamiana* plants by means of agroinfiltration. Omega-1 was purified from the apoplast (intercellular space) of *N. benthamiana* leaves and was shown to have RNase activity. Omega-1 produced in wild-type plants carried paucimannosidic type N-glycans, which are typical for plant-secreted proteins. By the controlled co-expression of two glycosyltransferases in wild-type plants, omega-1 received N-glycans that carried terminal Lewis X motifs, like *S. mansoni*-derived omega-1. Our results demonstrate that plants are an excellent platform for the expression of helminth glycoproteins carrying engineered N-glycans, which opens up a new field of research.

Introduction

Over the last few decades the incidence of immune-mediated diseases, like inflammatory bowel disease, asthma, multiple sclerosis and autoimmune (type 1) diabetes, has strongly increased in industrialised countries. Interestingly, there is an inverse correlation between the occurrences of these immune-mediated diseases with parasitic helminth infections. Helminths are complex multicellular organisms that upon infection of their host are master regulators of their host's immune system. The trematode *Schistosoma mansoni* is one of the best-studied helminths and the ability of *S. mansoni* eggs to induce strong Th2 responses in humans and experimental animal models is well documented [1, 2]. The protective effect of infection with *S. mansoni* against immune-mediated disease was tested in several animal models for human diseases. For instance, *S. mansoni* was shown to protect mice against TNBS-induced colitis [3], allergen-induced airway hyperreactivity [4], autoimmune encephalitis [4, 5] and autoimmune diabetes [6, 7]. Protection against immune-mediated diseases was associated with the down-regulation of Th1-type cytokines (IFN-γ and IL-12) and increased expression of regulatory and Th2-type cytokines (IL-4, IL-10 and/or IL-13). However, more information is required to understand exactly how helminths, like *S. mansoni*, modulate their host's immune system. Knowledge on helminth secreted proteins that play a key role in immune modulation could ultimately lead to the identification of new therapeutic targets for the treatment of several immune-mediated diseases.

Immune modulation during helminth infection is mediated by a variety of proteins that are secreted by the cercariae (larvae), worms and/or their eggs [8]. The mix of soluble egg antigens from *S. mansoni* (SEA) is well known for its capacity to promote Th2 responses, however, *S. mansoni* eggs secrete several proteins and not all these proteins are functionally characterised [9, 10]. The purification and functional characterisation of a single protein from Schistosome egg extracts is difficult, because the amount of SEA that can be obtained is limited. Therefore, an expression platform is required that enables the production of helminth secreted proteins in large quantities.

Recently omega-1 was identified as the major component in *S. mansoni* eggs that is responsible for conditioning dendritic cells (DC's) for Th2 polarisation [11-13]. Omega-1 is a glycoprotein secreted by the eggs of *S. mansoni* and has T2 ribonuclease activity. Site-directed mutagenesis of the catalytic domain or the two potential N-glycosylation sites of omega-1 demonstrated that the priming of Th2 responses by omega-1 treated DC's was dependent on its RNase activity as well as its N-glycosylation [14]. The major N-glycans identified on *S. mansoni*-derived omega-1 are core-difucosylated diantennary N-glycans with one or two terminal Lewis X motifs [15]. Omega-1 was shown to bind to several C-type lectin receptors, but the mannose receptor (CD206) was shown to be crucial for internalisation into DC's and subsequent polarisation of Th2 cells [14]. Besides omega-1 several SEA components have been

shown to interact with antigen presenting cells via a range of C-type lectin receptors and in many cases this interaction is thought to be important for shaping immune responses during schistosomiasis [16]. As the interaction with distinct C-type lectin receptors can influence the outcome of the immune response that is initiated by a certain SEA protein, it is crucial that for the functional characterisation of such proteins that they carry appropriate N-glycan structures. Therefore, an expression platform is required that not only enables high expression of such helminth proteins, but it should also be capable of producing native helminth N-glycan structures.

In the last two decades, plants have emerged as a promising expression platform for the production of recombinant proteins. Engineering of the N-glycosylation machinery has been successful in part because plants have demonstrated a high degree of tolerance towards changes in the glycosylation pathway, allowing the modification of recombinant glycoproteins in a specific and controlled manner [17]. Notably research on humanising N-glycans in plants has advanced over the last couple of years. Transgenic plants lacking non-human α1,3-fucose and β1,2-xylose sugar residues on their N-glycans have been generated (ΔXT/FT plants) and typical mammalian core α1,6-fucose can be engineered in plants as well [18, 19]. Lewis X motifs have been synthesised before in *Nicotiana tabacum* by stably expressing a hybrid β1,4-galactosyltransferase and a hybrid α1,3-fucosyltransferase IXa that were targeted to the medial-Golgi compartment [20]. However, these N-glycans were of a hybrid type as they only contained one antenna with a Lewis X motif and another antenna containing mannoses. The formation of hybrid N-glycans was attributed to the incorrect targeting of the glycosyltransferases. The efficiency of galactosylation in plants has proven to be a challenge. Expression of human β1,4-galactosyltransferase resulted in the formation of incompletely processed N-glycans, mainly lacking typical plant β1,2-xylose and/or α1,3-fucose or a single β1,4-galactose [21]. In contrast, targeting human β1,4-galactosyltransferase to the *trans*-Golgi compartment in ΔXT/FT plants allowed the generation of monoclonal antibodies with diantennary N-glycans containing two terminal galactose residues [22]. The authors postulated that the reason for inefficient galactosylation was that human β1,4-galactosyltransferase competes for the same acceptor substrate as the endogenous enzymes in the plant Golgi network.

In this study we investigated the suitability of plants as expression platform for *S. mansoni*-derived omega-1 and we investigated the possibility to engineer the glycosylation machinery such that diantennary N-glycans carrying Lewis X motifs were synthesized. Thereto, we used co-expression of β1,4-galactosyltransferase from *Danio rerio* (zebrafish) and α1,3-fucosyltransferase IXa from *Tetraodon nigriviridus* (pufferfish), which were targeted to the *trans*-Golgi network with the cytoplasmic-transmembrane-stem (CTS) domain of rat α2,6-sialyltransferase. Our work demonstrates that plants are an excellent expression platform for the production of omega-1 in large quantities and due to its efficient secretion is easily purified. Furthermore, we show that controlled expression of β1,4-galactosyltransferase is

required for efficient engineering of diantennary N-glycans that carry Lewis X motifs in wild-type plants.

Results

Plants are an excellent expression host for *S. mansoni*-derived Omega-1

To express biologically active omega-1 from *Schistosoma mansoni* we designed a codon-optimised gene of omega-1. Codon optimisation was applied as the GC-content of the native omega-1 gene is only 40% and plants generally prefer a higher GC-content in their highly expressed genes [23]. The mature omega-1 protein sequence was preceded by the *Arabidopsis thaliana* chitinase signal peptide for secretion followed by a N-terminal 6x histidine-tag and FLAG-tag (H6F). Expression was achieved by agroinfiltration of 5-6 weeks-old *Nicotiana benthamiana* plants, while co-expressing the p19 silencing suppressor of tomato bushy stunt virus. Expression of omega-1 was analysed in crude plant extracts at 3, 5 and 7 days post infiltration using western blots probed with an anti-FLAG antibody (Fig. 1a). Omega-1 migrated as two bands between 25 and 37 kDa, which were detected at all three time points with greatest band intensity at dpi 5. The calculated molecular weight of omega-1 is ~27 kDa, but omega-1 harbours two N-glycosylation sites. Thus, the two bands most likely represent intact omega-1, but with a different number of N-glycans.

Next we tested whether omega-1 is efficiently secreted from plant cells into the apoplast. At 5 dpi we isolated apoplast fluids from empty vector and omega-1 infiltrated leaves and used the remaining leaf material to make crude extracts. Both samples were analysed by SDS-PAGE and Coomassie staining (Figure 1b). Strikingly, the majority of the plant-produced omega-1 (>90%) was recovered from the apoplast. This means that omega-1 is secreted with high efficiency. Only few endogenous proteins are present in apoplast fluids and upon expression of omega-1, it became the predominant protein constituting >50% of total soluble protein. Apoplast fluids would therefore be an excellent starting point for further purification of omega-1.

Omega-1 was purified from the plant apoplast fluid using Ni-NTA affinity chromatography with a yield of approximately 150 µg omega-1 per ml of apoplast fluid. Purified omega-1 was analysed with SDS-PAGE followed by Coomassie staining and revealed two dominant bands that were previously detected with the anti-FLAG western blot (Figure 1c). Coomassie staining also revealed a faint band of ~25 kDa, which likely represents non-glycosylated omega-1.

Since omega-1 was characterised as a T2 ribonuclease (RNase), purified omega-1 was analysed for its RNase activity. Purified omega-1 was used in a RNA break down assay using total RNA from mouse bone marrow-derived dendritic cells. A mutant of omega-1, which

should lack RNase activity (omega-1^{H58F}) [14], was also expressed in plants, purified, and tested for activity. Figure 1d shows that purified omega-1 from plants is able to break down total RNA from mouse dendritic cells, whereas the omega-1^{H58F} mutant did not. This indicates that plant produced omega-1 has RNase activity.

Taken together our data show that plants offer an excellent expression platform for the production of large quantities of *S. mansoni*-derived omega-1. Furthermore, omega-1 is secreted into the apoplast with high efficiency, thereby enabling easy purification of active omega-1.

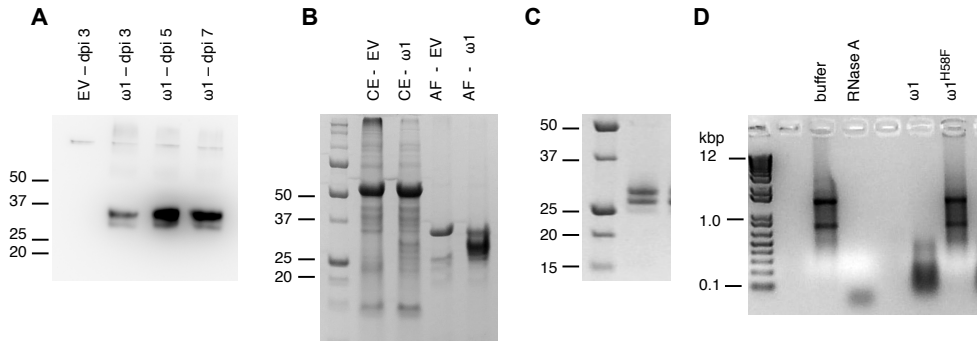


Figure 1 ◆ Analysis of Omega-1 (*ω*1) expressed in *N. benthamiana* plants. (a) Anti-FLAG western blot analysis under reducing conditions of 1 *μ*g total protein from crude plant extracts for omega-1 or empty vector (EV) infiltrated plants 3, 5 and 7 days post infiltration (dpi). (b). SDS-PAGE analysis under reducing conditions and Coomassie staining of 2.5 *μ*l of crude extract (CE) or apoplast fluid (AF) from omega-1 or EV infiltrated plants. (c) SDS-PAGE analysis under reducing conditions and Coomassie staining of 2 *μ*g purified omega-1. (d) RNase breakdown assay of total RNA from mouse bone marrow-derived dendritic cells using 500 ng purified omega-1 and omega-1^{H58F}. RNase A was used a positive control.

Glyco-engineering of omega-1 results in hybrid Lewis X-type N-glycans

A major fraction of *S. mansoni*-derived omega-1 naturally harbours diantennary N-glycans that carry Lewis X motifs. We set out to engineer Lewis X type N-glycans in *N. benthamiana*. Lewis X carrying N-glycans were previously synthesised in *N. tabacum*, but resulted in the formation of hybrid Lewis X-type N-glycans due to targeting to the medial-Golgi [20]. In our study we used α1,3-fucosyltransferase IXa from *Tetraodon nigriviridus* (pufferfish) and β1,4-galactosyltransferase 1 from *Danio rerio* (zebrafish). Both genes were genetically fused with the *trans*-Golgi targeting CTS domain of rat α2,6-sialyltransferase from here forward named sialFucT and sialGalT, respectively. We co-expressed both glycosyltransferases together with omega-1 and the p19 silencing suppressor in plants and purified omega-1 from the apoplast.

For N-glycan analysis tryptic peptides were prepared from purified omega-1 and N-glycans were released by PNGase A digestion. The N-glycosylation profile was determined by matrix

assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS). Representative N-glycan profiles are shown in Figure 2 for omega-1 isolated from wild-type *N. benthamiana* plants with and without co-expression of sialFucT and sialGalT (Figure 2a and 2b, respectively). Without co-expression of the glycosyltransferases the predominant N-glycan type found on omega-1 is a typical plant complex glycan for secreted proteins with β1,2-xylose and core α1,3-fucose. The N-glycans lack terminal GlcNAc residues putatively due to β-hexosaminidase activity in the apoplast (paucimannosidic type N-glycan; $MMXF^3$).

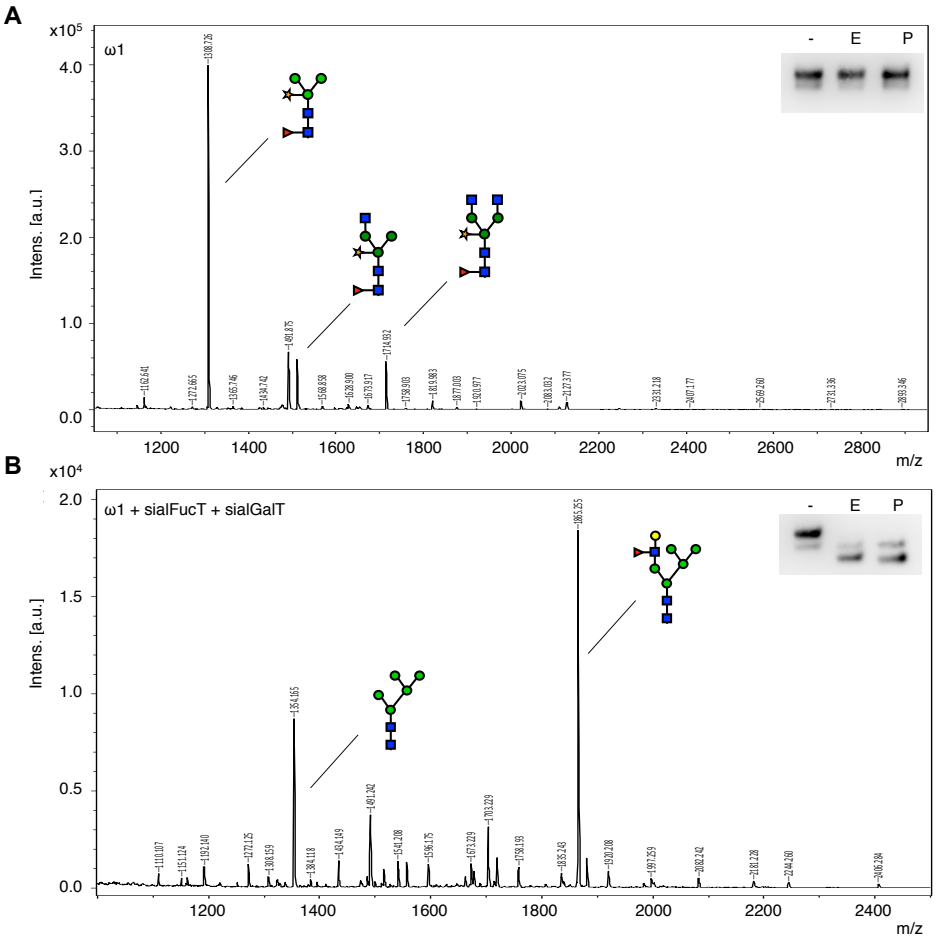


Figure 2 ◆ N-glycan profiles for omega-1 expressed in *N. benthamiana* plants. N-glycans were analysed using MALDI-TOF-MS analysis upon tryptic digestion and PNGase A release of N-glycans. **A** N-glycan profile for omega-1 from wild-type *N. benthamiana* plants. **B** N-glycan profile for omega-1 from wild-type *N. benthamiana* plants upon co-expression of sialFucT and sialGalT to engineer Lewis X carrying N-glycans. Optical densities of 0.5 were used for each construct. Each glycan profile is accompanied by an image of *in vitro* digestion of omega-1 with PNGase F (P) and Endo H (E) as analysed by anti-FLAG western blotting

The major N-glycan found on omega-1 upon co-expression of sialFucT and sialGalT consisted of one antenna carrying Lewis X, while the other antenna carried two additional mannoses (hybrid Lewis X). Furthermore, plant specific β1,2-xylose and core α1,3-fucose were absent. Besides the hybrid Lewis X-type N-glycan we also observed a smaller fraction of mannose-enriched N-glycans (Man5). To complement the N-glycan profiles, omega-1 was deglycosylated with PNGase F and Endo H to confirm the presence or absence of core α1,3-fucose or oligomannose/hybrid type N-glycans, respectively. In the upper right corner of both N-glycan profiles, the deglycosylation of omega-1 is shown upon anti-FLAG western blotting. In both cases, deglycosylation with PNGase F or Endo H confirms the obtained N-glycan profile. N-glycans carrying Lewis X motifs are formed upon co-expression of sialFucT and sialGalT, but are still of the hybrid-type as reported by Rouwendal and co-workers (2009).

Over-expression of sialGalT causes hybrid Lewis X-type N-glycans

Over-expression of a glycosyltransferase might result in overflow of the glycosyltransferase into other Golgi sub-compartments. In our case, our glycosyltransferases might therefore not be restricted to the *trans*-Golgi only. Activity of β1,4-galactosyltransferase in the medial-Golgi may interfere with α-mannosidase II activity in this compartment resulting in the formation hybrid N-glycans [24]. As Endo H was capable of cleaving off hybrid Lewis X-type N-glycans, we employed Endo H digestion to investigate if over-expression of β1,4-galactosyltransferase causes the formation of these hybrid N-glycans. We co-expressed omega-1 with sialFucT and sialGalT, but used different ratio's for the *Agrobacterium* cultures of these two glycosyltransferases. By scaling down optical densities of *Agrobacterium* cultures (testing a range from 0.5 to 0.05) we assume the expression of either sialFucT or sialGalT can be reduced. Apoplast fluids were harvested at 5 dpi and were treated with Endo H. The cleavage of hybrid N-glycans by Endo H was then analysed using anti-FLAG western blots probed with an anti-FLAG antibody (Figure 3a). When using an optical density of 0.5 for both glycosyltransferases, omega-1 migrates slightly higher in the SDS-PAGE gel than omega-1 expressed without co-expression of both glycosyltransferases, indicating a higher N-glycan mass. Upon treatment with Endo H the majority of omega-1 is deglycosylated, which confirms the presence of oligomannose- and/or hybrid-type N-glycans. When the optical density for sialGalT was lowered to 0.1 or 0.05, Endo H was only in part capable of cleaving the N-glycans from omega-1. Lowering the expression of sialFucT on the other hand did not have this effect. This observation supports the hypothesis that the over-expression of sialGalT and not sialFucT is linked to the formation of hybrid Lewis X N-glycans.

We also tested whether controlled expression of sialGalT by using a weaker promoter could prevent the formation of hybrid N-glycans completely. We chose the promoter region of the potato resistance gene *Gpa2* to control the expression of sialGalT, because resistance genes are constitutively expressed at a low level. Co-expression of omega-1 with sialFucT and sialGalT, the latter being controlled by the *Gpa2* promoter region, resulted in the production of

omega-1 that was hardly sensitive to Endo H (Figure 3b). Next, western blot analysis with a rabbit anti-HRP antibody, which specifically binds to plant β1,2-xylose and core α1,3-fucose, was performed as hybrid glycans do not carry these typical plant sugar residues (Figure 3c). As expected the anti-HRP antibody bound strongly to omega-1 from wild-type plants, but not to omega-1 co-expressed with the *d35S*:sialGalT construct. When the *Gpa2*:sialGalT construct was used, the N-glycans on omega-1 did carry typical plant sugars again. We therefore conclude that by co-expressing the *Gpa2*:sialGalT construct omega-1 does not receive hybrid-type N-glycans. However, the question is whether sialGalT expression controlled by the *Gpa2* promoter is still high enough to facilitate efficient galactosylation.

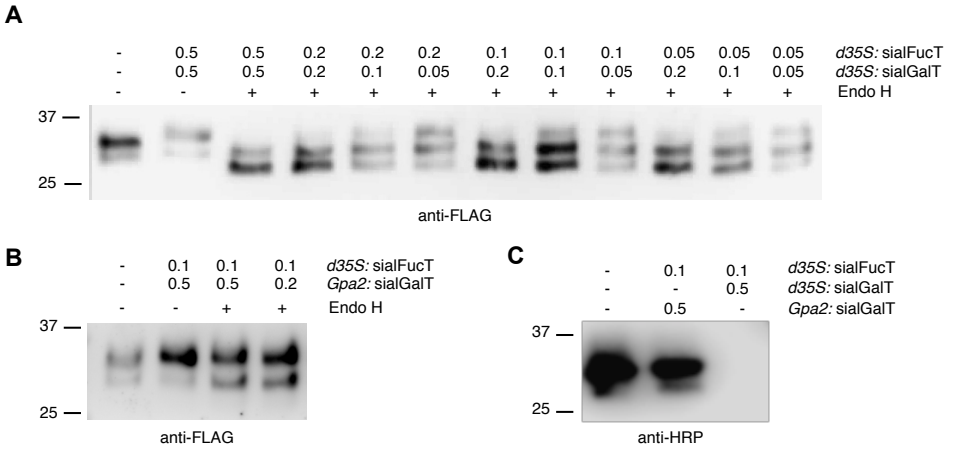


Figure 3 ◆ Screening for Lewis X type N-glycans. A Anti-FLAG western blot analysis of Endo H treated apoplast fluids for omega-1 expressed in wild-type plants while co-expressing *35S:*sialFucT and *35S:*sialGalT at different ratio's of *Agrobacterium* cultures. Optical densities (OD) of 0.5-0.05 were used. **B** Anti-FLAG western blot analysis of Endo H treated apoplast fluids for omega-1 expressed in wild-type plants while co-expressing *35S:*sialFucT and *Gpa2:*sialGalT (using an OD of 0.5-0.1) to lower sialGalT expression. **C** Anti-HRP western blot analysis of purified omega-1 expressed in wild-type plants or engineered omega-1 while co-expressing *35S:*sialFucT (OD=0.1) with either *d35S:*sialGalT or *Gpa2:*sialGalT (using an OD of 0.5).

Controlled expression of sialGalT enables Lewis X synthesis in wild-type *N. benthamiana*

To investigate if omega-1 carries diantennary N-glycans with terminal Lewis X motifs upon co-expression with *35S:*sialFucT and *Gpa2:*sialGalT we analysed the N-glycan composition. N-glycans were prepared as described above and a representative N-glycan profile is shown in Figure 4a. The predominant N-glycan type found on omega-1 upon Lewis X engineering is still the typical plant paucimannosidic N-glycan. However, by using this engineering strategy a significant proportion of diantennary N-glycans carrying Lewis X motifs was also found. On

top of that, several intermediate N-glycans containing antennae with a single Lewis X motif and/or a terminal galactose residue were found as well. Strikingly, all N-glycan structures contained typical plant β1,2-xylose. None of them were of the hybrid type carrying additional mannose residues on the α1,6-mannosyl branch. The presence of Lewis X was also confirmed by a binding assay with an agglutinin from *Ricinus communis* (RCA-I). As RCA-I does not bind to Lewis X motifs, omega-1 was first treated with α1,3/4-fucosidase from *Xanthomonas manihotis*. Figure 4b shows that RCA-I binds only to glyco-engineered omega-1 after removal of α1,3-linked fucose. Therefore, the N-glycans lacking a fucose residue must be missing this residue from the core, and not from the Lewis X motif. Altogether, we conclude that synthesis of diantennary N-glycans carrying Lewis X motifs can be achieved in wild-type *N. benthamiana* plants, but depends on the tightly controlled expression of sialGalT.

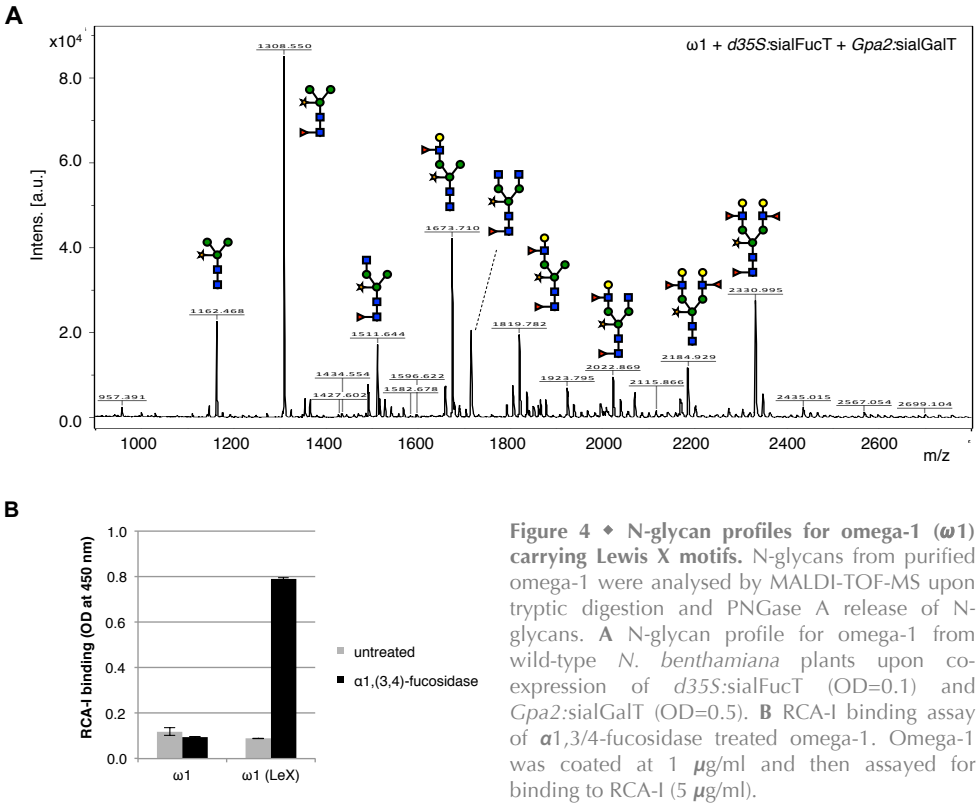


Figure 4 ◆ N-glycan profiles for omega-1 (ω1) carrying Lewis X motifs. N-glycans from purified omega-1 were analysed by MALDI-TOF-MS upon tryptic digestion and PNGase A release of N-glycans. **A** N-glycan profile for omega-1 from wild-type *N. benthamiana* plants upon co-expression of *d35S:*sialFucT (OD=0.1) and *Gpa2:*sialGalT (OD=0.5). **B** RCA-I binding assay of α1,3/4-fucosidase treated omega-1. Omega-1 was coated at 1 μg/ml and then assayed for binding to RCA-I (5 μg/ml).

Discussion

Schistosoma mansoni soluble egg antigens (SEA) play a crucial role in immune modulation during helminth infection and more detailed knowledge on the activity of its individual components can ultimately lead to the identification of new therapeutic targets for treatment of inflammatory disorders. Functional studies on individual components from this egg antigen mixture are challenged by the limited amount of SEA proteins that can be isolated. In our study we investigated the possibility to use *Nicotiana benthamiana* plants as a production platform for large quantities of the biologically active *S. mansoni*-derived egg antigen omega-1. Furthermore, we investigated whether the *N*-glycosylation machinery of plants could be engineered to synthesise diantennary N-glycans carrying terminal Lewis X motifs, which naturally occur on omega-1. Omega-1 was highly expressed and, most remarkable, >90% of omega-1 was secreted into the apoplast. The efficient secretion of omega-1 enabled easy purification as only few endogenous proteins reside in this extracellular compartment. In addition, purification from the apoplast fluid would ensure that all the proteins have passed through the entire secretory pathway and most likely results in a more homogeneous N-glycan composition.

N-glycan analysis revealed that omega-1 expressed in wild-type *N. benthamiana* plants predominantly carries paucimannosidic N-glycans containing the plant specific sugar residues β1,2-xylose and core α1,3-fucose ($MMXF^3$). The lack of terminal GlcNAc residues on the α1,3- and α1,6-mannosyl branches of N-glycans of plant proteins can be attributed to β-hexosaminidase activity in the apoplast as postulated by Strasser and co-workers [25]. They identified three β-hexosaminidases (HEXO1-3) in *Arabidopsis thaliana,* which localised in either the vacuole (HEXO1) or the apoplast (HEXO2 and HEXO3) that were able to remove terminal GlcNAc residues from N-glycans. Especially HEXO3 seemed responsible for the formation of paucimannosidic N-glycans in leaves of *A. thaliana* [26]. Up to 7 genes with high similarity to β-hexosaminidases from *Arabidopsis* can be identified within the genome of *N. benthamiana*, but their function and localisation are unknown.

In a first attempt to engineer terminal Lewis X motifs, a hybrid β1,4-galactosyltransferase from *Danio rerio* (zebrafish) and a hybrid α1,3-fucosyltransferase IXa from *Tetraodon nigriviridus* (pufferfish) were co-expressed with omega-1. The CTS domain of rat α2,6-sialyltransferase was used for *trans*-Golgi targeting as it was previously shown to enable efficient galactosylation [22]. In contrast, medial-Golgi targeting of galactosyltransferase with the CTS domain of *Arabidopsis* β1,2-xylosyltransferase resulted in the formation of hybrid N-glycans [20, 27]. These authors suggested that hybrid N-glycans are formed when early galactosylation in the medial-Golgi prevents trimming of mannose residues on the α1,6-mannosyl branch of a N-glycan by α-mannosidase-II (Man-II). In our experiments, co-expression of *trans*-Golgi targeted glycosyltransferases surprisingly resulted in the formation of

Lewis X hybrid N-glycans as well. Like previously observed, our hybrid Lewis X N-glycans are also devoid of plant specific β1,2-xylose and core α1,3-fucose. The lack of β1,2-xylose on Lewis X hybrid N-glycans would be in line with the suggestion that β1,2-xylosyltransferase only adds xylose after removal of the two mannoses on the α1,6-mannosyl branch, but before the addition of galactose to terminal GlcNAc residues [28]. This coincides with the fact that β1,2-xylosyltransferase from *Arabidopsis thaliana* is inhibited by mannoses on the α1,6-mannosyl branch or terminal galactose on the α1,3-mannosyl branch [29]. Thus, early galactosylation in the medial-Golgi of plants blocks the addition of β1,2-xylose, but may also inhibit Man-II activity leading to hybrid N-glycans. This would infer that within our experiments the use of a *trans*-Golgi targeting CTS domain is not sufficient to retain zebrafish galactosyltransferase within the *trans*-Golgi compartment of plants. Perhaps, over-expression of zebrafish galactosyltransferase causes overflow of the *trans*-Golgi compartment, resulting in medial-Golgi localised galactosyltransferase.

In order to lower the transient expression of sialGalT we used the *Gpa2* promoter region. This controlled fashion of expressing sialGalT resulted in the complete lack of hybrid Lewis X N-glycans. Furthermore, this engineering strategy is the first report that describes the synthesis of diantennary N-glycans carrying Lewis X motifs. Besides complex N-glycans carrying two Lewis X motifs, several intermediate N-glycans containing antennae with a single Lewis X motif and/or terminal galactose residues were found. Still, a major fraction of paucimannosidic N-glycans is also found. This indicates that perhaps the efficiency of galactosylation is not yet optimal and it has to be investigated whether this efficiency can be increased by fine-tuning sialGalT expression. Still, our findings demonstrate another layer of complexity to glyco-engineering, as it is not sufficient to only choose a proper CTS domain for targeting of a glycosyltransferase. Controlled expression of the glycosyltransferase is also important to obtain correct Golgi-targeting.

Alternatively, β-galactosidase activity may also interfere with obtaining a homogenous N-glycan composition. As the majority of the N-glycans, including some of the Lewis X intermediate N-glycans, lack terminal GlcNAc residues, we conclude that omega-1 is still sensitive to β-hexosaminidases in the leaf apoplast. This could be caused by inefficient galactosylation, as galactose would prevent GlcNAc removal by β-hexosaminidase. The presence of ungalactosylated N-glycan structures could also be due to the result of β-galactosidase activity, also leading to sensitivity to β-hexosaminidase. However, Strasser and co-workers were able to stably express a *trans*-Golgi targeted human GalT (ST-GalT) in ΔXT/FT *N. benthamiana* plants, which enabled the transient expression of a monoclonal antibody with homogeneous galactosylation on both antennae [22]. If β-galactosidases would be able to remove galactose from N-glycans in plants, this efficiency of galactosylation would most likely not have been achieved. However, a difference in sensitivity towards β-galactosidases between proteins may exist. The N-glycans of IgG are not surface exposed and stabilise the interaction between the γ heavy chains [30] and might therefore be less accessible for β-galactosidases

compared to the N-glycans of omega-1. However, detailed structural information on omega-1 does not exist and makes it hard to draw conclusions about differences in sensitivity to β-galactosidases.

In the case of Lewis X engineering, the addition of the α1,3-linked fucose to the Lewis X motif would protect the terminal galactose residues against β-galactosidases [20]. In our study, the sialFucT enzyme seems to be efficient as the predominant N-glycan found on omega-1 upon co-expression with sialFucT and sialGalT (both controlled with the 35S promoter) carried complete Lewis X, although of the hybrid type.,Besides this hybrid Lewis X-type N-glycan we did observe a smaller fraction of mannose-enriched N-glycans (Man5), which could arise after inefficient fucosylation of the Lewis X motif and subsequent trimming down by β-galactosidases and β-hexosaminidases. In subsequent experiments we reduced the optical density of the *Agrobacterium* culture that harbours the expression vector for sialFucT, which should result in reduced expression of this enzyme. Future experiment should therefore be performed where sialFucT expression is kept as high as possible to protect terminal galactose residues. Again this states that controlled and coordinate expression of glycosyltransferases is crucial for efficient glyco-engineering. The efficiency of fucosylation of the Lewis X motif could also be reduced by the presence of fucosidases in the plant apoplast. An α1,3/4-fucosidase was cloned from *A. thaliana* and shown to be able to hydrolyse fucose in 3- and 4-linkage to GlcNAc in Lewis determinants [31]. Within the genome of *N. benthamiana* several sequences with high similarity to this *Arabidopsis* α1,3/4-fucosidase can be found as well, but whether they have the same specificity and reside in the apoplast still has to be determined.

From our study we can conclude that Lewis X glyco-engineering results in a heterogeneous mixture of N-glycans containing Lewis X motifs. However, the Lewis X containing N-glycans nicely resemble the heterogeneity of the N-glycans naturally found on *S. mansoni*-derived omega-1 [15]. Both N-glycans of *S. mansoni*-derived omega-1 are diantennary and display similar heterogeneity. The major motif found on the N-glycans of omega-1 is Lewis X, but Lewis X is not always found on both antennae. Antennae with terminal galactose and/or mannose are found as well as some tandem Lewis X motifs, N-actylgalactosamine containing motifs (LDN-F), and difucosylated terminal motifs. In the future, efforts have to be made to shift the N-glycan composition on plant-produced omega-1 from paucimannosidic to Lewis X containing N-glycans. However, it is not necessary to obtain a homogenous N-glycan composition. Only when more detailed studies on the exact role of a specific N-glycan (like Lewis X) are planned, a more homogenous N-glycan composition is preferred. The use of ΔXT/FT or ST-GalT plants [22] may then be more suitable.

Our study demonstrates that plants are an excellent platform for the production of *S. mansoni*-derived omega-1 carrying engineered N-glycans. This highlights the fact that plants are a promising research tool to produce helminth glycoproteins with immunomodulatory functions and study their function. The production of uncharacterised helminth glycoproteins could enable their functional characterisation and the role of their N-glycans in

immunomodulation. The use of plants as a production platform for helminth glycoproteins opens up a new field of research and might ultimately lead to the identification of new therapeutic targets for treatment of inflammatory disorders.

Experimental procedures

Construction of expression vectors

The complete sequence encoding mature *Schistosoma mansoni* omega-1 protein was codon optimised according to an in-house optimisation procedure based on codons frequently used in highly expressed genes in plants. The mature protein sequence was preceded by a signal peptide from the *Arabidopsis thaliana* chitinase gene (cSP) and a N-terminal 6x histidine-FLAG tag (H6F). The amino acid sequence VEDAS was placed in-between the signal peptide and tags to ensure proper signal peptide cleavage and to prevent cleavage of the N-terminal histidine tag. The full cSP-H6F-omega-1 sequence was flanked by NcoI and KpnI restriction sequences at 5' and 3' ends respectively and synthetically constructed by GeneArt (Life Technologies; Bleiswijk, the Netherlands). The H58F mutation in the catalytic site of omega-1 to remove RNase activity was introduced by means of overlap extension PCR. Both omega-1 and omega-1^{H58F} were cloned into the pHYG expression vector using NcoI/KpnI [32].

Expression vectors for hybrid α1,3-fucosyltransferase IXa (Fut9a) from *Tetraodon nigriviridus* (pufferfish) and hybrid β1,4-galactosyltransferase (GalT) from *Danio rerio* (zebrafish) to engineer N-glycans carrying terminal Lewis X motifs were used [20, 33]. The N-terminal domain of *Arabidopsis thaliana* β1,2-xylosyltransferase in the Fut9a hybrid gene was replaced by the CTS domain of rat α2,6-sialyltransferase enabling *trans*-Golgi targeting (from now on referred to as sialFucT). The GalT gene already contained this CTS domain for *trans*-Golgi targeting (referred to as sialGalT). Expression cassettes for sialFucT and sialGalT were then transferred to the plant expression vector pBINPLUS [34]. For controlled expression of sialGalT the full gene was reamplified to introduce NcoI/KpnI restriction at the 5' and 3' ends respectively. sialGalT was then used to replace YFP in the pRAP-pGPAII::Gpa2-3'UTR-YFP vector, thereby placing the expression of sialGalT under the control of the promoter region of the *Gpa2* gene [35]. The entire *Gpa2:*sialGalT expression cassette was then transferred to pBINPLUS with AscI/PacI.

Expression of all constructs (unless stated differently) was driven by the 35S promoter of the *Cauliflower mosaic virus* with duplicated enhancer (d35S) and the *Agrobacterium tumefaciens* nopaline synthase transcription terminator (Tnos). A 5′ leader sequence of the Alfalfa mosaic virus RNA 4 (AlMV) was included between the promoter and construct to boost translation. In all experiments the silencing suppressor p19 from tomato bushy stunt virus in pBIN61 was co-infiltrated to enhance expression [36].

Agroinfiltration of Nicotiana benthamiana

Agrobacterium tumefaciens clones were cultured overnight (o/n) at 28°C/250 rpm in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH=7.0) containing 50 µg/ml kanamycin and 20 µg/ml rifampicin. The optical density (OD) of the o/n cultures was measured at 600 nm and used to inoculate 50 ml of LB medium containing 200µM acetosyringone and 50 µg/ml kanamycin with x µl of culture using the following formula: x = 80000/(1028*OD). OD was measured again after 16 hours and the bacterial cultures were centrifuged for 15 min at 4000 rpm. The bacteria were suspended in MMA infiltration medium (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 µM acetosyringone till a final OD of 0.5 was reached. For co-infiltration experiments with *Agrobacterium* harbouring the expression vectors for glycosyltransferases were used at a final OD of 0.05-0.5. After 1-2 hours incubation at room temperature, the two youngest fully expanded leaves of 5-6 weeks old *Nicotiana benthamiana* plants were infiltrated completely. Infiltration was performed by injecting the *Agrobacterium* suspension into a *Nicotiana benthamiana* leaf at the abaxial side using a 1 ml syringe. *Nicotiana benthamiana* plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) and infiltrated leaves were harvested at selected time points.

Total soluble protein extraction

Leaves were immediately snap-frozen upon harvesting, homogenized in liquid nitrogen and stored at -20°C until use. Homogenized plant material was ground in ice-cold extraction buffer (50mM phosphate-buffered saline (pH=7.4), 100 mM NaCl, 0.1% v/v Tween-20 and 2% w/v immobilized polyvinylpolypyrrolidone (PVPP)) using 2 ml/g fresh weight. Crude extracts were clarified by centrifugation at 16.000 rpm for 5 min at 4°C and supernatants were directly analysed for total protein content by a BCA protein assay (Life Technologies).

Apoplast wash

Leaves were submerged in ice-cold extraction buffer (50mM phosphate-buffered saline (pH=7.4), 100 mM NaCl and 0.1% v/v Tween-20) after which vacuum was applied for 10min. Vacuum was released slowly to ensure infiltration of the apoplastic space. Leaves were placed in 10-ml syringes and centrifuged for 20 min. at 2000*xg*. Apoplast washes were clarified by centrifugation at 16.000 rpm for 5 min at 4°C and supernatants were stored at -80°C.

Purification of omega-1

Plant produced omega-1 was purified from apoplast fluids using Ni-NTA Sepharose (IBA GmbH; Goettingen, Germany). Apoplast fluid was transferred over G25 Sephadex columns to exchange for Ni-NTA binding buffer (50 mM phosphate buffered saline (pH 8) containing 100 mM NaCl). Ni-NTA Sepharose was incubated with the apoplast fluid for 30 min. followed by 3

washing steps with binding buffer. Omega-1 was eluted from the Ni-NTA Sepharose with 0.5M imidazole (in binding buffer). Finally, purified omega-1 was dialysed against PBS.

Protein analysis by western blot

Omega-1 was separated under reducing conditions by SDS-PAGE on a 12% Bis-Tris gel. To analyse purity of the proteins the gel was stained with Coomassie brilliant blue staining. For western blotting proteins were transferred to a PVDF membrane by a wet blotting procedure. Thereafter the membrane was blocked in PBST-BL (PBS containing 0.1% v/v Tween-20 and 5% w/v non-fat dry milk powder) for 1 hour at room temperature, followed by overnight incubation with a HRP conjugated anti-FLAG M2 antibody (Sigma-Aldrich; Zwijndrecht, the Netherlands) in PBST at 4°C. The membrane was washed 5 times with 5 min intervals in PBST. Finally, the SuperSignal West Femto substrate (Thermo Fisher Scientific; Etten-Leur, the Netherlands) was used to detect HRP-conjugated antibodies in the G:BOX Chemi System (Syngene; Cambridge, UK).

Characterisation of Omega-1 N-glycosylation

Omega-1 (100 ng) was deglycosylated with PNGase F or Endo H (both from Bioké; Leiden, the Netherlands) to screen for the presence of plant-specific α1,3-fucose or oligomannose/hybrid-type N-glycans respectively. Deglycosylated omega-1 was then analysed by anti-FLAG western blot. The synthesis of Lewis X structures on omega-1 was analysed by ELISA with biotinylated agglutinin from *Ricinus communis* (RCA-I) (Bio-Connect; Huissen, The Netherlands). For this 1 µg/ml of purified omega-1 was treated overnight with 0.5 mU α1,3/4-fucosidase from *Xanthomonas manihotis* at 37°C. Omega-1 was then coated overnight at 4°C on ELISA plates. Plates were blocked for 1 hour at room temperature with PBST with 1% w/v BSA and all subsequent steps were performed in blocking buffer. Biotinylated lectins (5 µg/ml) were incubated for 1 hour at room temperature and after washing the plate with PBST, avidin-HRP (eBioscience; Vienna, Austria) was incubated for 30 min. at room temperature. TMB substrate (eBioscience) was used for detection.

Analysis of N-glycan composition by MALDI-TOF-MS

1-2 µg of purified omega-1 was reduced for 10 min at 95°C in PBS containing 1.3% w/v SDS and 0.1% v/v β-mercaptoethanol. SDS was neutralized with NP-40 prior to overnight digestion at 37°C with trypsin (Sigma-Aldrich) using trypsin that was immobilized to NHS-activated Sepharose (GE Healthcare; Zwolle, The Netherlands). Trypsin-beads were removed from the digestion mix by centrifugation and the pH of the mix was adjusted to 5 using 1M sodium acetate. 0.5 mU of PNGase A (Roche; Woerden, The Netherlands) was used to release N-glycans from omega-1 while incubating overnight at 37°C. Released N-glycans were purified by C18 Bakerbond™ SPE cartridges (VWR; Amsterdam, The Netherlands) and subsequent Extract Clean™ Carbo SPE columns (Grace; Breda, the Netherlands). N-glycans were labeled

with anthranilic acid (Sigma) and desalted over Biogel P10 (BioRad; Veenendaal, The Netherlands). Samples in 75% acetonitrile were mixed with 1 µl of matrix solution (20 mg/ml 2,5-dihydroxybenzoic acid in 50% acetonitrile, 0.1% TFA) and were dried under a stream of warm air. Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectra (MS) were obtained using an Ultraflex II mass spectrometer (Bruker Daltonics; Wormer, The Netherlands).

Bone marrow-derived dendritic cells

Wild-type C57BL/6J mice were bred and maintained under specific pathogen-free conditions in the animal facilities at Wageningen University. Bone marrow was isolated from the femur and tibia of 8-12 week old mice. Bone marrow derived dendritic cells (BMDC's) were differentiated as described [37] using 10% spent medium from murine GM-CSF transfected X63 cells [38]. X63-GM-CSF cells were kindly provided by dr. M. Lutz (University of Erlangen-Nuremberg) with approval of dr. B. Stockinger (MRC National Institute for Medical Research). Briefly, bone marrow cells were plated at 2×10^5 cells/ml in bacteriological petri dishes and incubated at 37°C/5% CO_2. At day 3, 6 and 8 medium was refreshed and at day 10 both adherent and non-adherent cells were harvested.

RNase activity assay

Total RNA was isolated from $5.0x10^6$ BMDC's using the RNeasy Mini Kit (Qiagen; Venlo, The Netherlands) as described by the supplier's protocol. 1 µg of total RNA was incubated with 500 ng purified omega-1 for 1 hour at 37°C using PBS as buffer. PBS only was used as negative control and 500 ng RNase A as positive control. RNA breakdown was then analysed on a 1.5% agarose gel and visualised with UV light.

Acknowledgements

We would like to thank Tam Nguyen, Simone Oostindie and Kim van Noort for their input in the practical work of this paper.

References

1. Pearce, E.J., et al., *Th2 response polarization during infection with the helminth parasite Schistosoma mansoni.* Immunological Reviews, 2004. 201: p. 117-126.
2. Pearce, E.J. and A.S. MacDonald, *The immunobiology of schistosomiasis.* Nature Reviews Immunology, 2002. 2(7): p. 499-511.
3. Elliott, D.E., et al., *Exposure to schistosome eggs protects mice from TNBS-induced colitis.* American Journal of Physiology - Gastrointestestinal and Liver Physiology, 2003. 284(3): p. G385-91.
4. Mangan, N.E., et al., *Helminth-modified pulmonary immune response protects mice from allergen-induced airway hyperresponsiveness.* Journal of Immunology, 2006. 176(1): p. 138-47.
5. Sewell, D., et al., *Immunomodulation of experimental autoimmune encephalomyelitis by helminth ova immunization.* Int Immunol, 2003. 15(1): p. 59-69.
6. Cooke, A., et al., *Infection with Schistosoma mansoni prevents insulin dependent diabetes mellitus in non-obese diabetic mice.* Parasite Immunology, 1999. 21(4): p. 169-76.
7. Zaccone, P., et al., *Schistosoma mansoni antigens modulate the activity of the innate immune response and prevent onset of type 1 diabetes.* European Journal of Immunology, 2003. 33(5): p. 1439-49.
8. Allen, J.E. and R.M. Maizels, *Diversity and dialogue in immunity to helminths.* Nature Reviews Immunology, 2011. 11(6): p. 375-388.
9. Cass, C.L., et al., *Proteomic analysis of Schistosoma mansoni egg secretions.* Molecular and Biochemical Parasitology, 2007. 155(2): p. 84-93.
10. Mathieson, W. and R.A. Wilson, *A comparative proteomic study of the undeveloped and developed Schistosoma mansoni egg and its contents: the miracidium, hatch fluid and secretions.* International Journal of Parasitology, 2010. 40(5): p. 617-28.
11. Fitzsimmons, C.M., et al., *Molecular characterization of omega-1: A hepatotoxic ribonuclease from Schistosoma mansoni eggs.* Molecular and Biochemical Parasitology, 2005. 144(1): p. 123-127.
12. Everts, B., et al., *Omega-1, a glycoprotein secreted by Schistosoma mansoni eggs, drives Th2 responses.* Journal of Experimental Medicine, 2009. 206(8): p. 1673-80.
13. Steinfelder, S., et al., *The major component in schistosome eggs responsible for conditioning dendritic cells for Th2 polarization is a T2 ribonuclease (omega-1).* Journal of Experimental Medicine, 2009. 206(8): p. 1681-90.
14. Everts, B., et al., *Schistosome-derived omega-1 drives Th2 polarization by suppressing protein synthesis following internalization by the mannose receptor.* Journal of Experimental Medicine, 2012. 209(10): p. 1753-1767.
15. Meevissen, M.H.J., et al., *Structural Characterization of Glycans on Omega-1, a Major Schistosoma mansoni Egg Glycoprotein That Drives Th2 Responses.* Journal of Proteome Research, 2010. 9(5): p. 2630-2642.
16. Meevissen, M.H.J., M. Yazdanbakhsh, and C.H. Hokke, *Schistosoma mansoni egg glycoproteins and C-type lectins of host immune cells: Molecular partners that shape immune responses.* Experimental Parasitology, 2012. 132(1): p. 14-21.
17. Bosch, D., et al., *N-Glycosylation of Plant-produced Recombinant Proteins.* Current Pharmaceutical Design, 2013. 19(31): p. 5503-5512.
18. Strasser, R., et al., *Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure.* Plant Biotechnology Journal, 2008. 6(4): p. 392-402.
19. Castilho, A., et al., *Rapid High Yield Production of Different Glycoforms of Ebola Virus Monoclonal Antibody.* Plos One, 2011. 6(10).

20. Rouwendal, G.J., et al., *Synthesis of Lewis X epitopes on plant N-glycans.* Carbohydrate Research, 2009. 344(12): p. 1487-93.
21. Bakker, H., et al., *Galactose-extended glycans of antibodies produced by transgenic plants.* PNAS, 2001. 98(5): p. 2899-904.
22. Strasser, R., et al., *Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous beta1,4-galactosylated N-glycan profile.* Journal oc Biological Chemistry, 2009. 284(31): p. 20479-85.
23. Wang, L. and M. Roossinck, *Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants.* Plant Molecular Biology, 2006. 61(4-5): p. 699-710.
24. Palacpac, N.Q., et al., *Stable expression of human β1,4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns.* PNAS, 1999. 96(8): p. 4692-4697.
25. Strasser, R., et al., *Enzymatic properties and subcellular localization of Arabidopsis beta-N-acetylhexosaminidases.* Plant Physiology, 2007. 145(1): p. 5-16.
26. Liebminger, E., et al., *beta-N-Acetylhexosaminidases HEXO1 and HEXO3 Are Responsible for the Formation of Paucimannosidic N-Glycans in Arabidopsis thaliana.* Journal of Biological Chemistry, 2011. 286(12): p. 10793-10802.
27. Bakker, H., et al., *An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes.* PNAS, 2006. 103(20): p. 7577-82.
28. Zeng, Y.C., et al., *Purification and specificity of beta 1,2-xylosyltransferase, an enzyme that contributes to the allergenicity of some plant proteins.* Journal of Biological Chemistry, 1997. 272(50): p. 31340-31347.
29. Kajiura, H., et al., *Arabidopsis beta 1,2-xylosyltransferase: Substrate specificity and participation in the plant-specific N-glycosylation pathway.* Journal of Bioscience and Bioengineering, 2012. 113(1): p. 48-54.
30. Herr, A.B., E.R. Ballister, and P.J. Bjorkman, *Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc.* Nature, 2003. 423(6940): p. 614-20.
31. Zeleny, R., et al., *Molecular cloning and characterization of a plant alpha1,3/4-fucosidase based on sequence tags from almond fucosidase I.* Phytochemistry, 2006. 67(7): p. 641-8.
32. Westerhof, L.B., et al., *3D Domain Swapping Causes Extensive Multimerisation of Human Interleukin-10 When Expressed In Planta.* Plos One, 2012. 7(10).
33. Hesselink, T., et al., *Expression of natural human beta1,4-GalT1 variants and of non-mammalian homologues in plants leads to differences in galactosylation of N-glycans.* Transgenic Res, 2014.
34. van Engelen, F.A., et al., *pBINPLUS: an improved plant transformation vector based on pBIN19.* Transgenic Research, 1995. 4(4): p. 288-90.
35. Sacco, M.A., et al., *The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death.* PLoS Pathogen, 2009. 5(8): p. e1000564.
36. Voinnet, O., et al., *An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus.* Plant Journal, 2003. 33(5): p. 949-956.
37. Lutz, M.B., et al., *An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow.* Journal of Immunological Methods, 1999. 223(1): p. 77-92.
38. Zal, T., A. Volkmann, and B. Stockinger, *Mechanisms of tolerance induction in major histocompatibility complex class II-restricted T cells specific for a blood-borne self-antigen.* Journal of Experimental Medicine, 1994. 180(6): p. 2089-99.

Chapter 6

Codon use and mRNA structure analyses across kingdoms indicates selection on both mRNA stability and translatability

Lotte B. Westerhof, J.C. Pjotr Prins, Mark G. Sterken, Ruud H.P. Wilbers, Debbie R. van Raaij, L. Basten Snoek, Martijn H.M. Holterman, Hans Helder, Jan Kammenga, Geert Smant, Erik J. Slootweg, Aska Goverse, Jaap Bakker and Arjen Schots

Abstract

To boost heterologous protein production the codon use of a gene of interest is often adapted to reflect the expression host's codon use in highly expressed genes (optimal codons). However, the results obtained with this strategy are variable. A comparison between the overall codon use and the codon use in highly expressed genes of several plant species revealed that optimal codons are not always the codons of which the use is increased most with expression. Although the codon composition of highly expressed genes differs between monocots and dicots, the same codons often rise in frequency with increasing expression levels (expression codons) and are in many cases C-ending. We used these conserved expression codons to optimise the codon composition of three genes, which enhanced protein yield significantly upon stable and transient expression in plants. Upon stable transformation both transcript levels and protein yield per transcript had increased. Next, we analysed whether this expression-linked codon bias found in plants also extends to other kingdoms of life. Thereto, expression-linked codon use was investigated in *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabditis elegans* (Animalia) and *Mus musculus* (Animalia) using more than 250 microarrays per species. We found that in all species, except *M. musculus,* C-ending codons are most often positively correlated with expression. In addition, computational analyses of various mRNA characteristics revealed a similar selection pressure across kingdoms that increases both stability and translatability. Combining gene expression data with available protein abundance data showed that an increased number of stem-loop transitions together with a reduction of stem size increases translation efficiency. Further studies are required to develop a codon optimisation strategy that involves the design of an ideal mRNA structure that maximizes mRNA stability and translatability, and the role of expression codons herein to boost heterologous protein expression.

Introduction

Most amino acids are encoded by multiple synonymous codons and the frequency wherein synonymous codons are used is not equal within a given species. Also, within species a bias in codon use in highly expressed genes can be observed, linking codon use to gene expression. The codons used most frequently in highly expressed genes (optimal codons) have been shown to correspond to genomic G+C content [1] and often match the most abundant tRNAs in many species [2-7]. It is assumed that codons that match more abundant tRNAs would be translated faster as tRNA availability for translation occurs via diffusion and the chance of encountering a more abundant tRNA is larger than encountering a rarer tRNA [8]. An increase in translation rate would allow ribosomes to finish translation sooner to restart translation. Also, the chance that a ribosome initially loads a non-matching tRNA is smaller when a codon matches a more abundant tRNA. This creates an energetic advantage as three-quarter of the energy to incorporate an amino acid is lost if a non-matching tRNA has to be rejected after proofreading [8]. Thus, the use of optimal codons in highly expressed genes was hypothesized to provide a fitness gain by increasing translational efficiency.

As an increased translation efficiency may boost protein yield, codon optimisation of heterologously expressed genes using optimal codons of the production host has been a common strategy that has met varying success [9]. A study on heterologous expression of 154 variants of GFP differing only in synonymous codon use in *E. coli* demonstrated that the use of optimal codons was positively correlated with bacterial growth, but not protein yield [10]. About half of the variation in GFP protein levels was explained by folding energy of the first third of the mRNA. Therefore the authors argued that while translation rate may have been increased by the use of optimal codons, protein yield was not influenced because translation initiation must have been the rate-limiting step for translation. This may be a general feature as suggested by ribosomal density studies. Most ribosomes are found at the beginning of mRNAs and the overall packing density of nearly all mRNAs is below maximum [11, 12].

In this study we present an alternative codon optimisation strategy that led to a significant increase in both mRNA stability and mRNA translatability (i.e higher mRNA levels and more proteins per mRNA molecule). Because we used an identical sequence to code for the signal peptide that preceded our genes we assumed translation initiation to be identical. Hence, an increase in mRNA translatability can only be explained by an increase in the rate of translation of the remainder of the messenger. This indicates that free ribosomes may be limiting during heterologous protein expression [13]. If free ribosomes are limited, the rate-limiting step for translation would no longer be the time it takes for translation to initiate, but the time it takes for a ribosome to become available for translation. Thus, while translation initiation may be the rate-limiting step, an increased translation rate of the remaining mRNA will increase protein yield by enabling ribosomes to become available sooner.

Our codon optimisation strategy involved the selection of codons of which the use is increased most with expression (expression codons) across plant species. Wang and Roossinck [14] compared the codon use of highly expressed genes to the overall codon use in several plant species. They found that optimal codons and expression codons are not always the same. Also, while monocots and dicots differ significantly in their G+C content and have different optimal codons, they often have the same expression codons, which are C-ending in most cases. These conserved expression codons were used to optimize the codon use of three genes and boosted protein yield significantly upon stable and transient expression in plants. We demonstrate that upon stable transformation both transcript levels and protein yield per transcript increases. Next, we analysed whether the occurrence of this expression-linked codon bias is also observed in other kingdoms of life. Thereto, expression data were collected of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabditis elegans* (Animalia) and *Mus musculus* (Animalia) and more than 250 microarrays per species were analysed covering a wide range of strains/ecotypes, treatments, culture conditions, tissues and developmental stages. We found that in all species, except *M. musculus,* C-ending codons are most often positively correlated with expression.

Studying the correlation between expression level and mRNA characteristics including gene length, minimal folding energy, number of bound nucleotides, mean stem and loop sizes (stretches of bound and unbound nucleotides, respectively) and number of stem-loop transitions uncovered a general trend across kingdoms. The number of bound nucleotides and the number of stem-loop transitions were positively correlated with expression levels, while loop size was negatively correlated with expression. Combining the gene expression data with available protein abundance data demonstrated that protein:mRNA ratio (proxy for translation efficiency) is positively correlated with the number of stem-loop transitions and negatively correlated with stem and loop size. This general pattern across kingdoms reveals a selection pressure created by gene expression on both mRNA stability and translatability. An increase in the number of nucleotide bonds favours stability, while a more even distribution of these bonds enhances translatability. Altogether, our data indicate that a successful codon optimisation strategy should focus on computational models that calculate the ideal mRNA structure whereby both stability and translatability are enhanced.

Results and discussion

Codon optimisation improves mRNA stability and translatability

Wang and Roossinck determined of which codons the use was increased most comparing overall codon use to the codon use in highly expressed genes in 11 plant species [14]. Although the codons used most frequently in highly expressed genes (optimal codons) differed between monocots and dicots, the use of the same codons often increases with expression (expression codons). We selected one codon per amino acid that was most often identified as an expression codon across these 11 plant species. Strikingly, most of these codons were C-ending, except for the amino acids Arg (CGT) and Gly (GGT). The codons of the amino acids Gln, Glu and Lys, that can only be encoded by A or G-ending codons, were G-ending. To investigate the effect of these codons on heterologous protein production in plants, we recoded the gene sequence of three genes with these codons. We chose the genes of *Aequorea victoria* green fluorescent protein (GFP), *Gallus gallus* ovalbumin (OVA) and *Mus musculus* interleukin-10 (IL-10) because of their variation in codon use (Figure 1a). To eliminate differences caused by translation initiation all genes were preceded by the signal peptide of *Arabidopsis thaliana* chitinase. GFP was also expressed without this signal peptide, as it is normally not secreted.

The native and optimised variants of these four constructs were used to transform *Arabidopsis thaliana* using the floral dip method and their expression in seedlings was evaluated by determining mRNA transcript and protein levels (Figure 1b; Table 1). An increased protein yield found upon optimisation could be partly explained by an increase in mRNA transcript levels, i.e. increased mRNA stability (Table 1). Comparing protein:mRNA ratios of transformants within a similar mRNA expression range showed that codon optimisation resulted in more protein per mRNA transcript. Thus, codon optimisation also resulted in increased mRNA translatability.

Upon transient transformation transcript levels are always much higher. An increase in mRNA stability and translatability may than no longer improve protein yield. Therefore, we also determined protein yield upon transient expression of the three genes in *Nicotiana benthamiana* with and without co-expression of the gene silencing inhibitor p19 of tomato bushy stunt virus (Figure 1c; Table 2). Also upon transient expression codon optimisation lead to higher protein yield on all days for all genes, except for OVA unless p19 was co-expressed. In most cases co-expression of p19 had a favourable effect on protein yield independent of optimisation. This is not surprising as, mRNA transcript levels are always high in transient expression, which increases the risk of gene silencing. Thus, the mRNA of the optimised variant of OVA must have been more sensitive to gene silencing compared to the native variant.

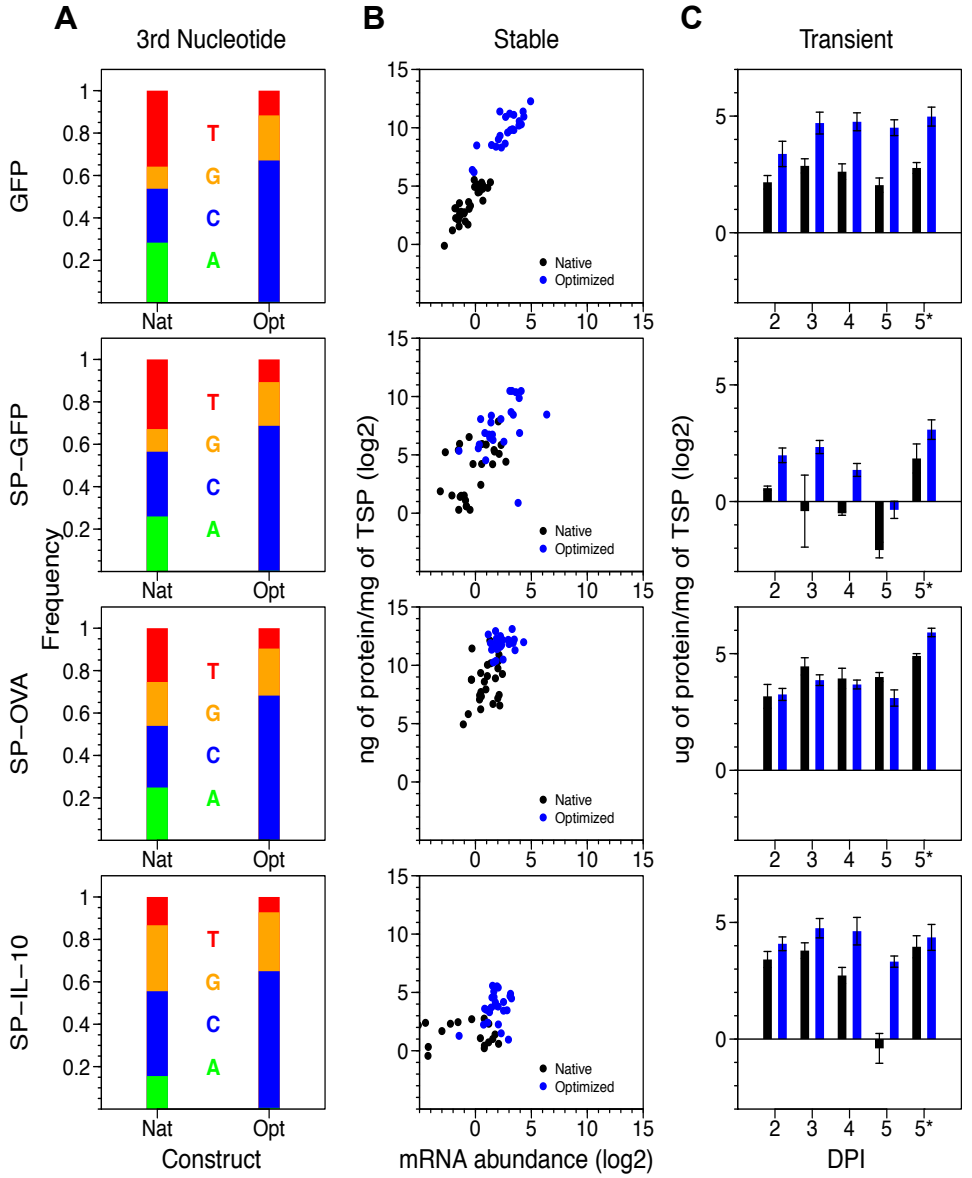


Figure 1 ◆ Codon optimisation boosts protein yield by increasing mRNA stability and translatability. A Nucleotide content of the third codon position in the constructs for *Aequorea victoria* green fluorescent protein (GFP), *Gallus gallus* ovalbumin (OVA) and *Mus musculus* interleukin-10 (IL-10) with additional chitinase signal peptide (SP) expression. GFP was also expressed without SP. **B** Expression in transformed *Arabidopsis thaliana* seedlings. For each plant analysed the protein yield in ng per mg total soluble protein (TSP) is plotted against the relative mRNA transcript concentration as compared to the *A. thaliana* household gene TIP-41. **C** Transient expression in *Nicotiana benthamiana* leaves (native and optimised in black and blue bars, respectively). Protein yield in µg per mg TSP at 2 to 5 days post infiltration (DPI), * indicates co-expression with the silencing inhibitor p19 of tomato bushy stunt virus. *n*=3, error bars indicate standard error.

Table 1 ◆ Codon optimisation of GFP, interleukin-10 and ovalbumin genes boosts expression in *Arabidopsis thaliana*.

		n=	Relative mRNA conc.	Fold change	Protein yield	Fold change	Relative mRNA conc. range	*n*=	Protein: mRNA ratio	Fold change
GFP	N	32	0.88	11***	17.03	75***	0.8-2.7	4	22.8±2.70	7.1
	O	23	9.25		1276		0.9-2.5	14	161±58.5	
SP-GFP	N	26	1.63	5.8*	33.28	12**	1.4-4.9	11	18.0±5.16	3.5*
	O	24	9.53		399.5		1.2-4.8	12	63.9±14.5	
SP-OVA	N	26	2.37	2.4***	717.3	5.5***	2.0-5.3	12	356.2±142.5	2.8**
	O	30	5.62		3937		2.2-5.5	23	1014±121.7	
SP-IL-10	N	17	1.37	3.1***	3.30	5.5***	1.7-4.2	8	1.26±0.43	5.3***
	O	25	4.23		17.9		1.7-4.1	16	6.68±1.02	

Average relative mRNA transcript concentration as compared to the *A. thaliana* household gene TIP-41 and protein yield in µg per mg total soluble protein (TSP) determined in *A. thaliana* seedlings upon stable transformation of native (N) and optimised (O) sequences of *Aequorea victoria* green fluorescent protein (GFP), *Gallus gallus* ovalbumin (OVA) and *Mus musculus* interleukin-10 (IL-10) with additional chitinase signal peptide (SP). GFP was also expressed without signal peptide. Protein:mRNA ratios were calculated. Because translatability may be lower with a higher mRNA concentration due to the limited number of free ribosomes, the protein:mRNA ratios were calculated of samples within the same mRNA concentration range, as indicated. The fold change when comparing the optimised to the native variant was calculated for the relative mRNA concentration, protein yield and protein:mRNA ratio. For each average the number of included seedlings is indicated (*n*). Significance of fold changes were calculated with a Welch's *t*-test: * P<0.05, ** P<0.01, ***P<0.001.

Table 2 ◆ Codon optimisation boosts protein yield in transient expression in *Nicotiana benthamiana*.

		dpi 2-5		dpi 5 + p19	
		Protein yield	Fold change	Protein yield	Fold change
GFP	N	5.7	4***	7.0	4.8
	O	23		34	
SP-GFP	N	1.0	3.2**	4.4	2.1
	O	3.2		9.2	
SP-OVA	N	17	0.7	30	2.0*
	O	12		61	
SP-IL-10	N	8.4	2.5**	17	1.4
	O	21		24	

Average protein yield in µg per mg total soluble protein determined in *N. benthamiana* leaves upon transient transformation of native (N) and optimised (O) sequences of *Aequorea victoria* green fluorescent protein (GFP), *Gallus gallus* ovalbumin (OVA) and *Mus musculus* interleukin-10 (IL-10) with additional chitinase signal peptide (SP) (GFP was also expressed without SP) at 2 to 5 days post infiltration (dpi) (*n*=12) or 5 dpi whereby tested genes were co-expressed with the viral silencing inhibitor p19 of tomato bushy stunt virus. (*n*=3). Significance of fold change in protein yield were calculated with a Welch's *t*-test: * P<0.05, ** P<0.01, ***P<0.001.

Evaluating the average yield from dpi 2-5 or with co-expression of p19 on dpi 5 revealed a lower yield increase upon codon optimisation compared to stable expression in *A. thaliana*. This is not surprising as at least some of the gain in mRNA stability due to the codon optimisation is compensated by the increased transcription in transient expression. Whether this gain in protein yield is predominantly the result of an increase in mRNA translatability or a combination of a gain in mRNA stability and translatability remains to be determined.

To explain the differences found in mRNA stability we first calculated the thermodynamic stability of the predicted secondary mRNA structures. Upon codon optimisation the minimum free folding energy had decreased, indicative for a more stable mRNA, from -0.25 to -0.35 and -0.31 to -0.33 kcal/mol/nt for GFP and OVA, respectively. However, for IL-10, the minimum free folding energy increased from -0.31 to -0.28 kcal/mol/nt indicating a less stable mRNA. Thus, an overall increase in physical stability could not explain the increased mRNA transcript levels of IL-10. However, it is still possible that instable regions of IL-10 were removed upon codon optimisation, while the overall stability decreased.

In vivo mRNA half-life is predominantly controlled by other factors than physical stability, namely; the occurrence of a splicing event, through AU-rich destabilizing elements in the UTRs, and the presence of sequences that are targets for microRNA [15-17]. In our experiments, all genes were expressed using the same expression controlling components, thus contained the same UTRs and did not contain introns. However, the sequences of the ORFs varied greatly between the native and optimised variants (78, 76 and 83% homology for GFP, OVA and IL-10, respectively). Therefore, there could be a difference in the presence of microRNA targets and also a difference in the occurrence of stretches of double stranded (ds)RNA between the native and optimised variants. The dsRNA stretches could be processed to small interfering RNAs and, like binding of microRNAs, can trigger gene silencing [18]. In stable expression, gene silencing can also be due to gene methylation, but this always results in the complete absence of transcripts, and therefore transformants without detectable expression were not considered. In our transient expression experiment co-expression of the silencing inhibitor p19 gave comparable results. Taken together, differences in mRNA decay based on above mentioned sequence features are unlikely to explain the differences in mRNA stability in our experiments.

Translation has also been linked to mRNA decay. Ribosomes can shield nuclease target sites [19], however, in large-scale *in vivo* studies mRNA half-life could not be linked to the number of nuclease target sites [20, 21] or ribosomal density [12]. When translation initiation is equal, as is expected in our experiments, an increase in translatability should result in a lower density of ribosomes. Thus, there would have been less ribosomes on the optimised variants compared to their native counterparts, and the optimised variants would be less protected against nucleases.

While translation *per se* may not influence mRNA half-life, errors in translation have been proven to lead to mRNA degradation by mRNA surveillance mechanisms. Three mRNA

surveillance mechanisms have been identified: I) nonsense mediated decay by the recognition of a premature stop codon, II) non-stop decay by the lack of a stop codon and III) no-go decay by stalled ribosomes [21]. Occurrence of a premature stop codon or the lack of a stop codon can be caused by a mutation or a ribosomal slip causing a frame-shift. Frame-shifts can be caused by a 'slippery' sequence that may be found in proximity of a strong mRNA structure [22, 23]. A ribosome may also stall at a strong stem-loop structure without slipping [24] and trigger degradation [25, 26]. It is possible that the native and optimised variants differ in the presence of 'slippery' sequences and/or strong mRNA structures. Thus, differences in level of translation-linked mRNA decay may explain the difference in mRNA transcript levels in our experiment. In addition, ribosomes have intrinsic helicase activity [27] and recently it was shown that strong mRNA structures such as pseudoknots and hairpins can stall translation only temporarily [28]. Therefore, Qu *et al* argue that the mRNA structure provides a mechanical basis for cellular regulation of translation rate [28]. Thus, increased mRNA translatability of the optimised genes may be explained by an increased translation rate caused by differences in the mRNA structure.

General codon bias extends to other kingdoms of life

To investigate if the general codon bias found by Wang and Roossinck [14] among plants transcends kingdoms of life we acquired expression data of *Escherichia coli (Bacteria), Saccharomyces cerevisiae (Fungi), Caenorhabdites elegans (Animalia)*, *Arabidopsis thaliana (Plantae)* and *Mus musculus (Animalia)*. Per species >250 microarrays originating from several studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues were used (Supplementary Table 1A-F). First, the expression was ranked and the average rank was used as a measure of overall expression. Subsequently, the correlation between expression and nucleotide content was analysed per species. The relation between the species based on this correlation was visualized in a heat map (Figure 2).

Across all kingdoms a strong positive correlation between expression and overall G content, in particular G on the first codon position was found and a negative correlation between expression and A and T on the first codon position. Next, the correlation between expression and codon use was evaluated (Figure 3). Across kingdoms the use of CGT (Arg/R), AAG (Lys/K), GGT (Gly/G), GTT (Val/V) and GCT (Ala/A) is positively correlated with expression. However, the fact that the nucleotide contents of the first and second codon position are correlated with expression indicates that there is a correlation between amino acid usage and expression. Unsurprisingly, highly expressed genes are relatively rich in the amino acids encoded by G-starting triplets: Ala, Gly, and Val (Figure 4).

First, to uncouple the amino acid bias from the codon use bias, we calculated the relative synonymous codon use. Subsequently, we made a comparison between high- and low-expressed genes, as a correlation between codon use and expression may only be found in genes expressed above a certain threshold. We started grouping genes based on expression

from the centre (50% highest versus 50% lowest) until, with 1% steps, the pools with 5% highest and 5% lowest expressed genes were reached. With each step the codon use frequencies in both high- and low-expressed gene pools were calculated together with the difference in codon use frequency between the high- versus the low-expressed gene pool. Finally, the difference in codon use frequency was correlated (Spearman) to the expression defining percentage. The relation between the species based on this correlation was visualized in a heat map (Figure 5; Supplementary Table 2A-E show codon use frequencies of all, the bottom 5% low- and top 5% high-expressed genes and fold codon use change (top/bottom) per species).

Strikingly, when clustering the correlations between the 5 species, *E. coli*, *S. cerevisiae*, *C. elegans* and *A. thaliana* group together well. *M. musculus* seems to have an overall lower codon bias and in ~50% of the cases selects for other codons compared to the overall selection of the other species. Excluding *M. musculus*, 13 codons are positively correlated with expression for all species. These 13 codons encode 11 different amino acids and a termination of translation (twice a codon for Thr/T). Comparable to the general codon bias found in plants [14], 8 of these 13 codons are C-ending. Furthermore, 18 codons are consistently negatively correlated with expression in these four species. Of these codons most are A-ending (8), while none of them are C-ending.

Taken together our data suggest that a conserved selection pressure influences expression across kingdoms of life. Our heterologous protein expression experiments suggested a role for the mRNA structure in translation rate. As the translational machinery does not vary greatly across kingdoms, the mRNA structure is a likely candidate to be the driving force behind this selection pressure.

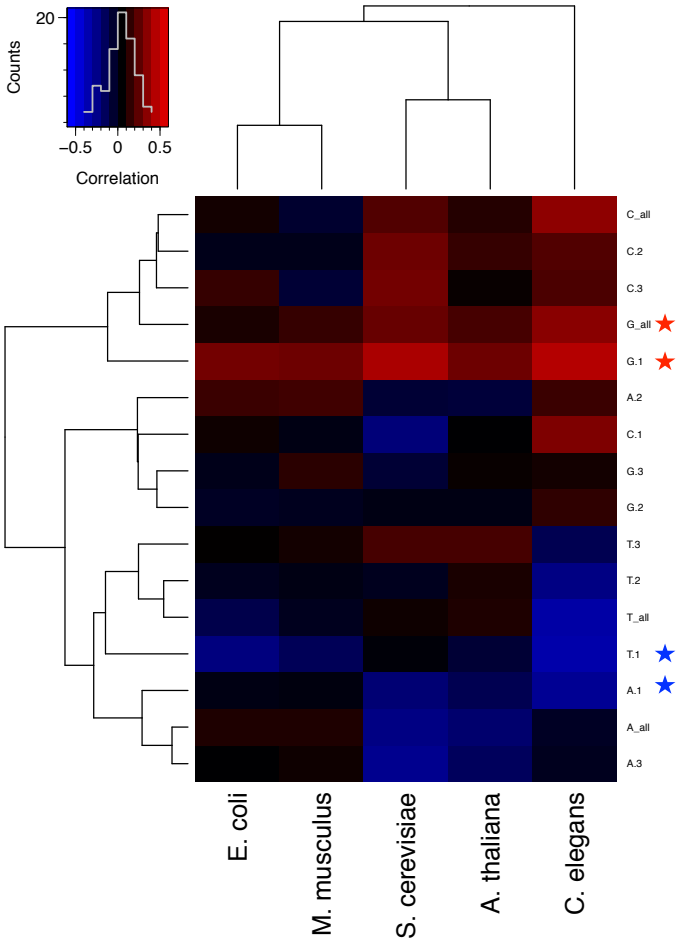


Figure 2 ◆ Heat map displaying the relation between species of several kingdoms of life based on expression-linked nucleotide use. Expression data of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) (>250 microarrays per species) originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) were rank-normalized and averaged. Subsequently, correlations (Spearman) between expression and nucleotide use (overall and for each codon position) were calculated per species and used to generate this heat map. Consistent positive and negative correlations across species are indicated with a red and blue star, respectively.

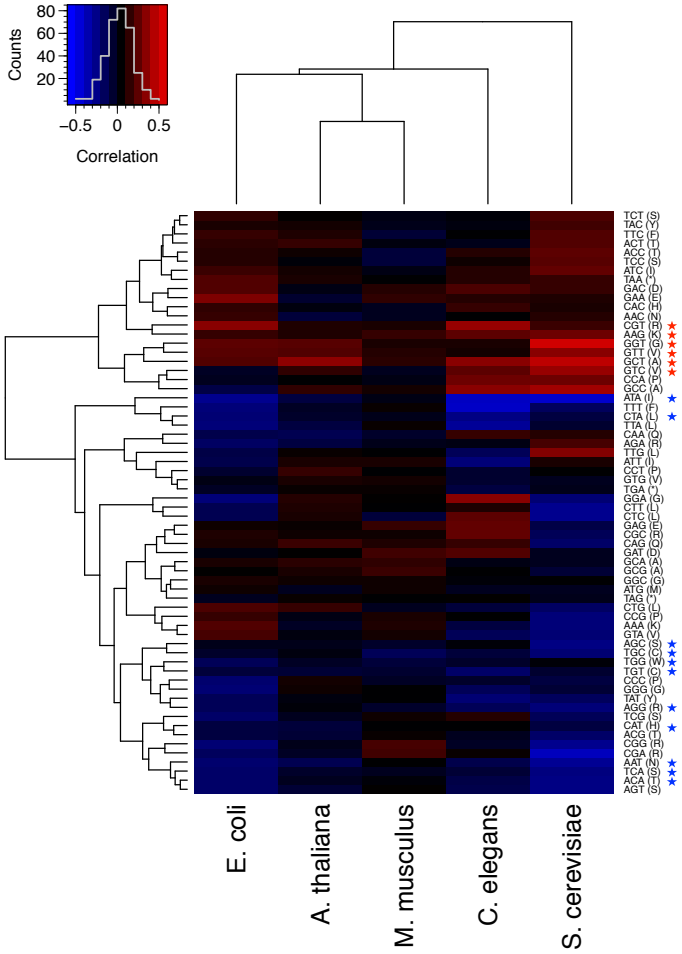


Figure 3 ◆ Heat map displaying the relation between species of several kingdoms of life based on expression-linked codon use. Expression data of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) (>250 microarrays per species) originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) were rank-normalized and averaged. Subsequently, correlations (Spearman) between expression and codon use were calculated per species and used to generate this heat map. Consistent positive and negative correlations across species are indicated with a red and blue star, respectively.

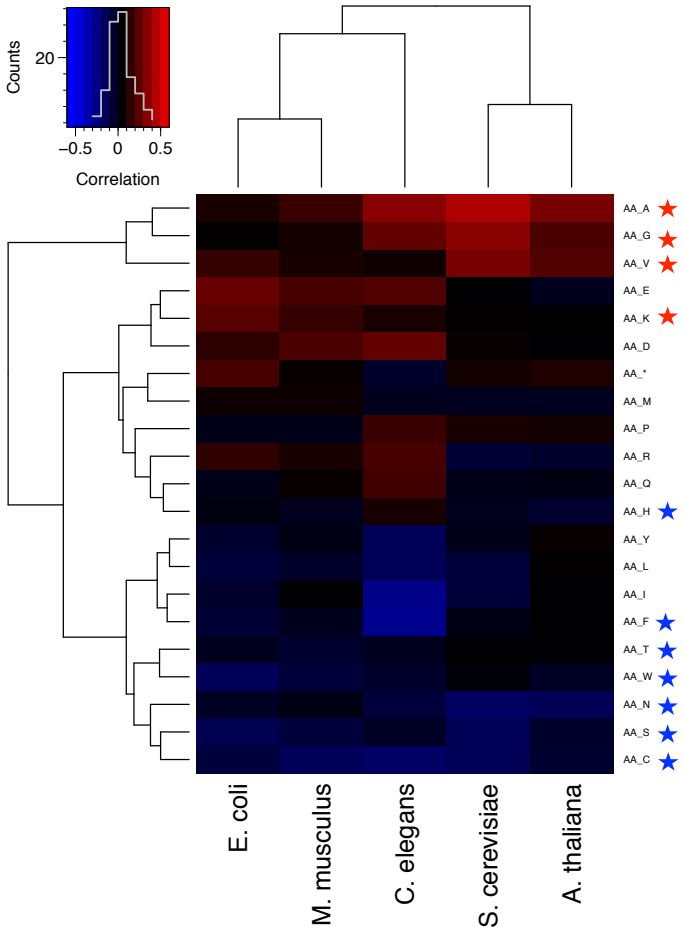


Figure 4 ◆ Heat map displaying the relation between species of several kingdoms of life based on expression-linked nucleotide use. Expression data of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) (>250 microarrays per species) originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) were rank-normalized and averaged. Subsequently, correlations (Spearman) between expression and amino acid use were calculated per species and used to generate this heat map. Consistent positive and negative correlations across species are indicated with a red and blue star, respectively.

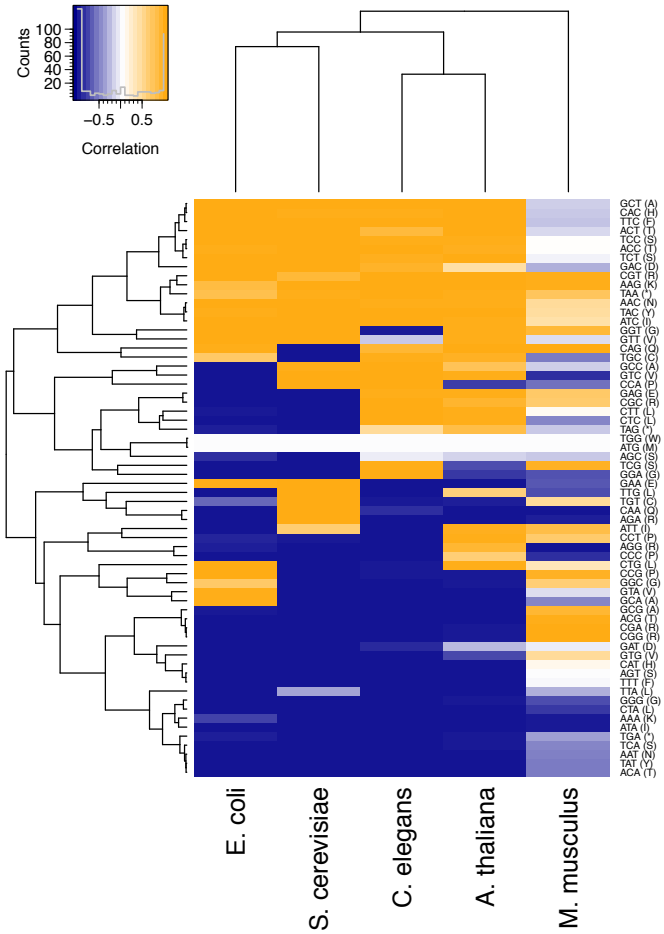


Figure 5 ◆ Heat map displaying the relation between species of several kingdoms of life based on expression-linked codon bias. Expression data of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) (>250 microarrays per species) originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) was rank-normalized and averaged. Subsequently, we started grouping genes based on expression from the centre (50% highest versus 50% lowest) until, with 1% steps, the extremes (5% highest versus 5% lowest) were reached. With each step the synonymous codon use frequencies in both high- and low-expressed gene pool were calculated together with the difference in codon use frequency between the high- versus the low-expressed gene pool. Finally, the difference in codon use frequency was correlated to the expression defining percentage (Spearman). The relation between the species based on this correlation is visualized in this heat map.

Highly expressed genes prefer a stable, but 'airy' mRNA structure

To evaluate if the mRNA structure could be the driver of selection that gives rise to the observed general codon bias, we evaluated if expression is linked to mRNA structure characteristics. Thereto, we predicted the mRNA structures of all genes and determined gene length, minimal free folding energy, number of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of the number of stem/loop transitions and plotted these against expression (Figure 6; Table 3). Also a heat map displaying the relation between the species based on the correlation (Spearman) between these structure characteristics and expression was generated (Figure 7; Supplementary Table 3). This heat map demonstrates that the number of bound nucleotides and the number of stem/loop transitions was consistently positively correlated and mean loop size consistently negatively correlated with expression across all species.

The positive correlation with the number of bound nucleotides indicates a general adaptation towards a more stable mRNA molecule. Also, a low folding energy (more stable) is correlated with high expression in *S. cerevisiae*, *C. elegans* and *A. thaliana*, but not in *E. coli* and *M. musculus*. Still, in *E. coli* there seems to be a relation between expression and folding energy, as is demonstrated by the trend line that indicates an optimum (Figure 6). An optimum in mRNA stability may indicate a trade-off between stability and translatability in this species. A trade-off in stability and translatability may also explain why there is a correlation between mRNA folding energy and expression in *S. cerevisiae*, *C. elegans* and *A. thaliana*. These species have an overall lower G+C content resulting in on average weaker mRNAs (Table 3) and have therefore more to gain in terms of stability before translatability is affected.

The number of stem-loop transitions and mean loop size are also correlated with expression (positive and negative, respectively) in all species, which suggests that there is a general adaptation towards dividing nucleotide bonds equally over the mRNA molecule. In other words, highly expressed genes prefer a stable, but 'airy' mRNA molecule. This again indicates a trade-off between mRNA stability and translatability. It is striking that while folding energy in *S. cerevisiae*, *C. elegans* and *A. thaliana* is on average much higher (less stable mRNA) (6-10%) compared to *E. coli* and *M. musculus*, the fraction of bound nucleotides, mean stem and loop size and number of transitions do not differ that much (Table 3). This means that while the mRNA folding energy may differ between species with different G+C content, the overall mRNA structure characteristics are more similar across species.

Taken together our data suggest that there is a general selection towards an optimal folding energy across kingdoms of life whereby number and type of nucleotide bonds (e.g. A-U and G-U bonds are weaker than G-C bonds) are balanced with short loops to facilitate efficient translation. This is in line with the observation that translation rate is greatly influenced by G+C content and strong mRNA structures [28].

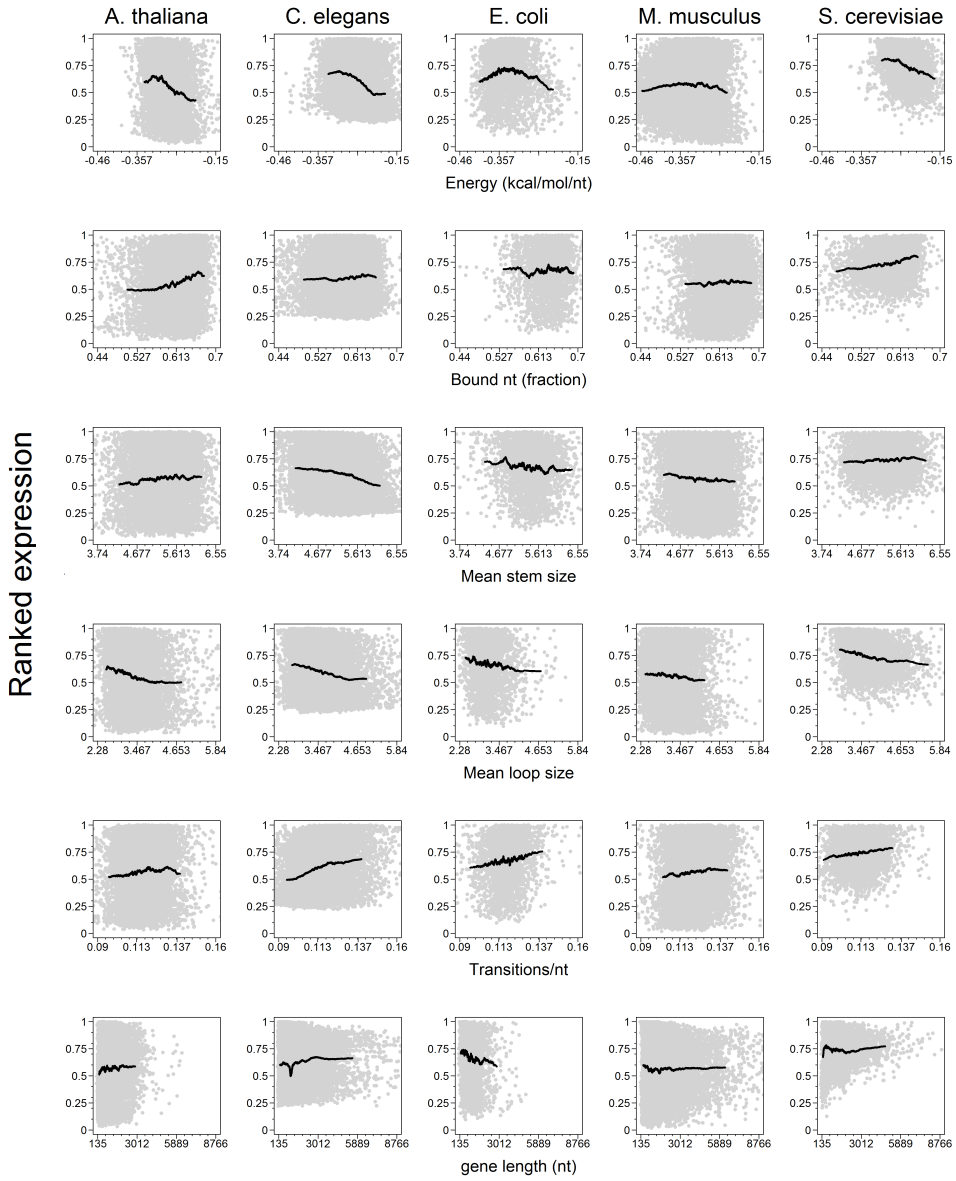


Figure 6 ◆ mRNA structure features plotted against ranked expression with moving average (black line). The mRNA structures of all genes of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) were predicted and gene length, minimal free folding energy (kcal/mol/nucleotide), fraction of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of stem/loop transitions per nucleotide were determined. Previously mentioned mRNA characteristics plotted against expression.

Table 3 ◆ mRNA characteristics of highly expressed genes per species.

	E. coli	*S. cerevisiae*	*C. elegans*	*A. thaliana*	*M. musculus*
G+C content	0.52	0.40	0.43	0.45	0.52
Gene length (nt)	1004	1490	1359	1097	1752
Energy (kcal/mol/nt)	-0.34	-0.24	-0.25	-0.28	-0.34
Bound nt (fraction)	0.63	0.59	0.60	0.62	0.62
Mean stem size	5.42	5.29	5.18	5.29	5.18
Mean loop size	3.12	3.72	3.48	3.26	3.10
Transitions / nt	0.117	0.111	0.116	0.117	0.121

Averages of mRNA characteristics of the top 5% high-expressed genes of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia).

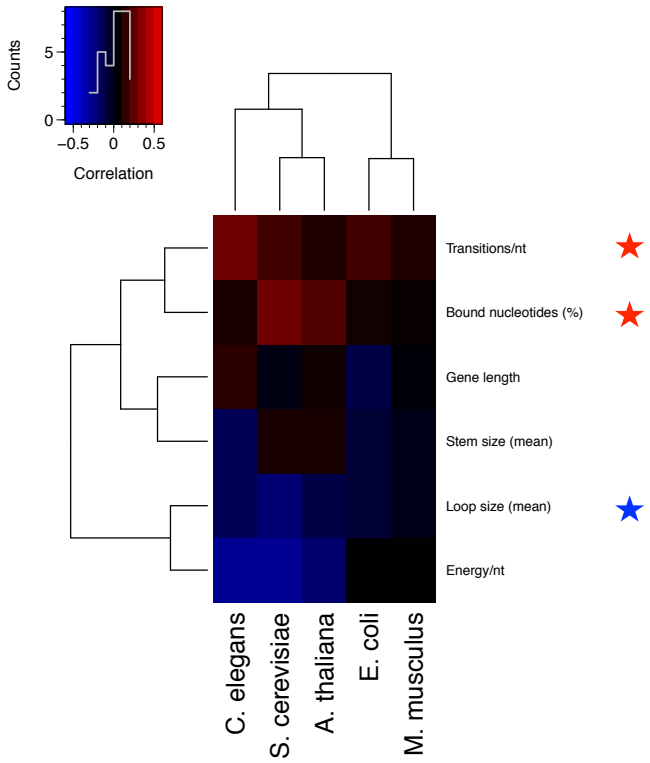


Figure 7 ◆ Across kingdoms highly expressed genes prefer a stable, but 'airy' mRNA structure. The mRNA structures of all genes of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) were predicted and gene length, minimal free folding energy (kcal/mol/nucleotide), fraction of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of stem/loop transitions per nucleotide were determined and correlated with expression (Spearman) (Supplementary Table 3). Consistent positive and negative correlations across species are indicated with a red and blue star, respectively.

A link between mRNA structure and expression, may explain the increase in mRNA stability and translatability in our heterologous protein expression experiments. Therefore we predicted and evaluated the mRNA structures of the native and optimised variants of the expressed genes (Supplementary Figure 1; Supplementary Table 4)**.** Optimised variants of GFP and OVA had an increased folding energy indicative of a more stable mRNA. All optimised variants had an increased number of stem-loop transitions (except SP-GFP), which is in line with a more 'airy' mRNA molecule. Thus, although changes in the mRNA structure upon optimisation differ from gene to gene, an improved mRNA structure could be the basis of increased protein yield in our experiments.

A more 'airy' mRNA increases translation rate

On a cellular level translation efficiency was demonstrated to be the most important factor in controlling protein abundance whereas protein turnover plays only a minor role [29]. Therefore, protein:mRNA ratio is a good proxy of translation rate. To evaluate if the mRNA structure characteristics we found to be linked to expression are also linked to translation rate we combined our expression data with large-scale protein abundance data retrieved from PaxDB. To evaluate to what extent our expression data predict protein abundance we calculated the correlation (Spearman) between our expression data and the protein abundance: *E. coli* 0.59, *S. cerevisiae* 0.67, *C. elegans* 0.59, *A. thaliana* 0.62 and *M. musculus* 0.36. When we evaluated the relation between the protein:mRNA ratio and the previously mentioned mRNA structure characteristics we obtained a similar picture as when using the expression data (Figure 8; Supplementary Table 5; Supplementary Figure 2-4 heat maps demonstrate the relation between species based on correlations of protein:mRNA ratio and nucleotide content, codon use and amino acid use).

As with expression, the number of stem-loop transitions is positively correlated with protein:mRNA ratio and mean loop size is negatively correlated across all species. Also, the folding energy is negatively correlated (more stable mRNA) for *S. cerevisiae*, *C. elegans* and *A. thaliana*, but not for *E. coli* and *M. musculus*. However, in contrast to the expression data, gene length is consistently negatively correlated with protein:mRNA ratio. This is in line with the fact that the packing density of ribosomes was shown to decrease with mRNA transcript length [12]. Also, a negative correlation with mean stem size is found for all species and the fraction of bound nucleotides is not correlated, except for *S. cerevisiae*. Thus, small stem size must be important for an increased translation rate. This again highlights the trade-off between mRNA stability and translatability.

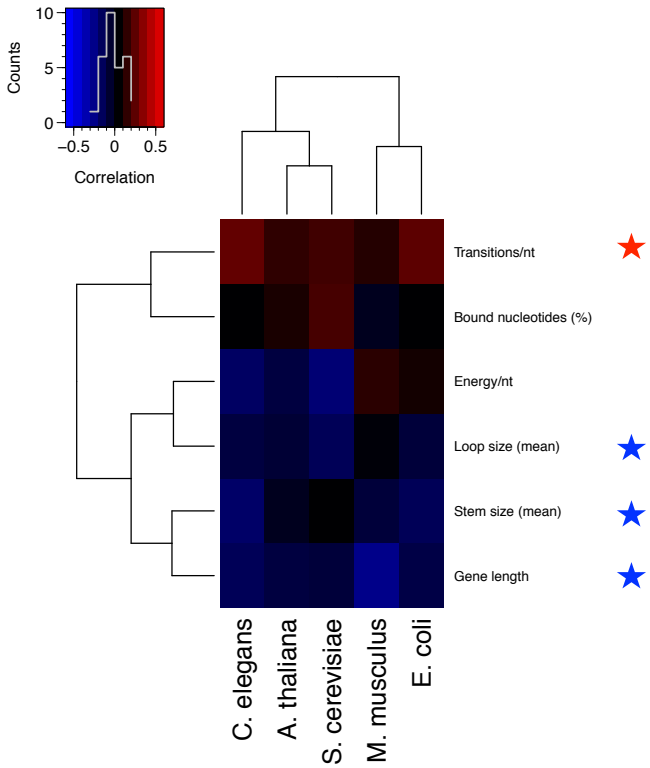


Figure 8 ◆ Across kingdoms high translated transcripts are 'airy'. Correlation (Spearman) between mRNA structure characteristics and protein:mRNA ratios per species (Supplementary Table 5). The mRNA structures of all genes of *Escherichia coli (Eubacteria), Saccharomyces cerevisiae (Fungi), Caenorhabdites elegans (Animalia), Arabidopsis thaliana (Plantae)* and *Mus musculus (Animalia)* were predicted and gene length, minimal free folding energy, percentage of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of stem/loop transitions were determined and correlated (Spearman) with protein:mRNA ratios. Rank-normalized mRNA levels were divided by protein abundance (retrieved from PaxDB). Consistent positive and negative correlations across species are indicated with a red and blue star, respectively.

Concluding paragraph

Our data indicate that across kingdoms of life mRNA structures are exposed to a selection pressure that increases both stability and translatability. However, a trade-off between stability and translatability most likely exists. An increase in the number of nucleotide bonds favours stability, but reduces translatability. This trade-off may explain the more even distribution of stems and loops in mRNAs coding for genes with high translational efficiency. Translatability is enhanced by an increase in the number stem-loop transitions, while loop and stem sizes are negatively correlation with translation efficiency. In *E. coli, S. cerevisiae, C. elegans* and *A. thaliana* a general expression-linked codon bias was found towards C-ending codons. A codon optimisation strategy that employed predominantly C-ending codons boosted the protein yield of three genes upon expression in plants by increasing mRNA transcript levels and protein yield per transcript. It is therefore likely that mRNA stability and translatability are increased by the use of by C-ending codons in these species. Only *M. musculus* demonstrated a weaker and different codon bias, while the same mRNA structural features were linked to expression. In the case of *M. musculus* other codons must therefore facilitate an improvement of the mRNA structure in this species, perhaps because *M. musculus* already has a high C content in its coding sequences. To develop a codon optimisation strategy that is robustly successful, computational models should be developed that determine the effects of expression codons on mRNA structures. Further analyses are required to pinpoint the effects of expression codons on the number of stem-loop transitions and size of stems and loops. Ideally, large-scale heterologous gene expression experiments should be carried out to verify the effectiveness and robustness of this novel codon optimization strategy to increase protein yield.

Codon optimisation is worthwhile, as increased translation rate of a heterologously expressed gene boost protein yield, even when the transcript levels are already high. In an extensive ribosomal density study in yeast [11] and in a global analysis on translation in mammalian cells [29] a large variation in translation rate was found. Both studies also revealed that translation rate was the major contributor in controlling protein abundance. Although many studies have shown evidence that translation initiation is the rate-limiting step of translation [11, 12], an increase in translation rate may still influence protein abundance if the number of free ribosomes is limiting. If free ribosomes are limiting increased translation rate will make translating ribosomes faster available for other messengers. Especially in heterologous gene expression, whereby the gene of interest is significantly over-expressed (often several orders of magnitude above household genes) it is likely that free ribosomes are limited. Increased translational efficiency of the heterologous gene can therefore increase protein yield significantly.

Experimental procedures

Construct design

The native and optimised sequences coding for *Aequorea victoria* green-fluorescent protein (GFP) (L29345.1; nt 7-807) *Gallus gallus* ovalbumin (OVA) (NM_205152.2; nt 4-1161) and *Mus musculus* interleukin-10 (IL-10) (NM_010548.2; nt63-537) together with the optimised sequence for the *Arabidopsis thaliana* basic chitinase signal peptide (cSP) (BAA82810.1; nt15-33) were synthetically made by GeneArt (Thermo Fisher Scientific, Breda, the Netherlands). Optimisation was performed by recoding the protein sequences using the C-ending codons for all amino acids (TCC in the case of Ser), except Arg and Gly, for which the T-ending codons were used, and Gln, Glu and Lys, for which the G-ending codons were used. Synonymous mutations to either native or optimised sequences were sometimes introduced to remove undesired restriction and the cryptic splice sites in native GFP[30]. Gene fragments were flanked with sequences including the restriction sites NcoI (5′) and EagI-BspHI (3′) for cSP, EagI (3′) and KnpI (5′) for IL-10 and OVA and NcoI (3′) and KpnI (5′) for GFP to allow fragment assembly and subsequent in frame cloning into the plant expression vector pHYG[31]. Fragment assembly was accomplished by the in frame ligation of cSP with IL-10 and OVA using the EagI site and cSP with GFP using the BspHI (cSP) and NcoI (GFP) sites. ORFs were confirmed by sequencing in expression vector stage. All vectors were transformed to *Agrobacterium tumefaciens* strain GV3101 for stable transformation of *Arabidopsis thaliana* or MOG101 for agroinfiltration in *Nicotiana benthamiana*.

Stable transformation of *Arabidopsis thaliana*

Agrobacterium tumefaciens clones were cultured overnight (o/n) at 28°C in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH7.0) containing 50 µg/ml kanamycin. Bacterial cultures were centrifuged for 15 min at 2800 *g* and resuspended in MMA (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 µM acetosyringone and 0.03% silwet-L77 till an OD of 0.5 was reached. *Arabidopsis thaliana* plants were submerged in the bacterial suspension for 1 min and kept in a moist environment for 2 days. Plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) until seeds could be collected. Seeds were sterilized by 4-hour exposure to chlorine gas and plated on basic agar plates (8g/l Bacto Agar, 0.101g/l KNO^3) containing 30 µg/ml hygromycin and 100 µg/ml cefotaxim. Plates were kept in the dark at 4°C for 2 days, then placed in artificial light for 7 hours at 24°C, again kept in the dark at rt for 5 days and finally placed in a climate chamber with 12 hour light regime at 24°C for 2 days. At this stage 10 to 40 seedlings per transformant plant were selected and placed in individual pots with Knop agar (1x Knop, 1% sucrose, 8g/l Plant Agar pH6.4) containing 30 µg/ml hygromycin and 100 µg/ml cefotaxim. Seedlings that showed good growth and root formation after 10 days were transferred to fresh

pots and allowed to grow for 2 more weeks. Thereafter plants were harvested and snap-frozen. Plant material was homogenized using a TissueLyser II (Qiagen) and stored at -80°C until further use.

Transient transformation of *Nicotiana benthamiana*

Agrobacterium tumefaciens clones were cultured overnight (o/n) at 28°C in LB medium (10g/l pepton140, 5g/l yeast extract, 10g/l NaCl with pH7.0) containing 50 µg/ml kanamycin and 20 µg/ml rifampicin. The optical density (OD) of the o/n cultures was measured at 600 nm and used to inoculate 50 ml of LB medium containing 200 µM acetosyringone and 50 µg/ml kanamycin with x µl of culture using the following formula: x = 80000/(1028*OD). OD was measured again after 16 hours and the bacterial cultures were centrifuged for 15 min at 2800 *g*. The bacteria were resuspended in MMA infiltration medium (20g/l sucrose, 5g/l MS-salts, 1.95g/l MES, pH5.6) containing 200 µM acetosyringone till an OD of 1 was reached. All constructs were co-expressed with the tomato bushy stunt virus silencing inhibitor p19 by mixing Agrobacterium cultures 1:1. After 1-2 hours incubation at room temperature, the two youngest fully expanded leaves of 5-6 weeks old *Nicotiana benthamiana* plants were infiltrated completely. Infiltration was performed by injecting the *Agrobacterium* suspension into a *Nicotiana benthamiana* leaf at the abaxial side using a 1 ml syringe. Infiltrated plants were maintained in a controlled greenhouse compartment (UNIFARM, Wageningen) and infiltrated leaves were harvested at selected time points.

Determination of heterologous gene expression

Total RNA was isolated from homogenized plant material using the RNAeasy Plant Mini Kit (Qiagen) according to supplier's protocol. A Turbo DNaseI (Ambion) treatment was included to remove any residual DNA. cDNA was synthesised using the SuperScript III First-Strand Synthesis System (invitrogen) according to suppliers protocol using an oligo(dT) primer. Samples were analysed by quantitative PCR in triplo using ABsolute SYBR Green Fluorescein mix (Thermo Scientific). *Arabidopsis thaliana* TIP-41 (AY074349.1) was used as a reference gene. The oligonucleotides used for amplification of both native and optimised IL-10, OVA and GFP and TIP-41were 5'-AACCTCTTCCTCTTCCTC-3' / 5'-GGAAGTGGGTGCAGTT-3'; 5'-AACCTCTTCCTCTTCCTC-3' / 5'-GGGCAGTAGAAGATGTTC-3'; 5'-GACGGTAACTACAA-GACC-3' / 5'-TTGTCGGCCATGATGTA-3'; and 5'-GCTCATCGGTACGCTCTTTT-3' / 5'-TCCATCAGTCAGAGGCTTCC-3', respectively. Relative transcript levels of the genes versus TIP-41 were determined by the Pfaffl method[32].

Determination of heterologous protein expression

Homogenized plant material was ground in ice-cold extraction buffer (50mM phosphate-buffered saline (PBS) pH=7.4, 100 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA),

0.1% v/v Tween-20, 2% w/v immobilized polyvinylpolypyrrolidone (PVPP)) using 2 ml/g fresh weight. Crude extract was clarified by centrifugation at 16.000x*g* for 5 min at 4°C and supernatant was directly used in an ELISA and BCA protein assay. Mouse IL-10 expression levels were determined using the Mouse IL-10 ELISA Ready-SET-Go! kit (eBioscience) according to the supplier's protocol. For the quantification of OVA and GFP, a rabbit anti-ovalbumin or a chicken anti-GFP (both from Rockland Immunochemicals Inc.) was used to coat ELISA plates o/n at 4°C in a moist environment. After this and each following step the plate was washed 5 times with 30 sec intervals in PBST (1x PBS, 0,05% Tween-20) using an automatic plate washer (BioRad model 1575). The plate was blocked with assay diluent (eBioscience) for 1 h at room temperature. Samples and standard lines were loaded in serial dilutions and incubated for 1 h at room temperature. Standard lines were made from purified chicken ovalbumin (Sigma) or recombinant GFP (Roche). For detection of OVA and GFP a rabbit anti-ovalbumin:HRP antibody or a rabbit anti-GFP:HRP antibody (both from Rockland Immunochemicals Inc.) were used, respectively. A 3,3',5,5'-Tetramethylbenzidine (TMB) substrate (eBioscience) was added and colouring reaction was stopped using stop solution (0.18M sulphuric acid) after 1-15 min. Read outs were performed using the model 680 microplate reader (BioRad) to measure the OD at 450 nm with correction filter of 690 nm. For sample comparison total soluble protein (TSP) concentration was determined using the BCA Protein Assay Kit (Pierce) according to supplier's protocol using bovine serum albumin (BSA) as a standard.

Gene expression datasets

Gene expression datasets of 5 species (*Escherichia coli, Arabidopsis thaliana*, *Saccharomyces cerevisiae, Caenorhabditis elegans,* and *Mus musculus*) were downloaded from Gene Expression Omnibus (GEO). Gene-expression sets were selected based on platform (Affimetrix), release date (not earlier than 2008), publication linked to the GEO set and number of samples in the study. In total 2067 gene-expression profiles were collected, representing 8 or 9 different studies per organism. An overview can be found in Supplementary Table 1A-F.

Protein abundance datasets

Protein abundance datasets were retrieved from PaxDb [33], where the integrated datasets of *Escherichia coli, Arabidopsis thaliana*, *Saccharomyces cerevisiae, Caenorhabditis elegans,* and *Mus musculus* were downloaded.

Gene expression normalization

Gene expression was normalized based on rank. Per species one array platform was used and per species probes were ranked according to their intensities. The average rank per probe was used as a measure of overall gene expression to distinguish genes with overall low and high expression levels for each species.

mRNA Sequences

The coding sequences (CDS) of all genes of 5 species were downloaded from sequence/genome repositories. For *Escherichia coli*, the CDS of strains CFT073, EDL933, MG1655 and Sakai were obtained from NCBI, accesscions NC_004431.1, NC_002655.1, NC_U00096.3 and NC_002695.1 respectively. For *Arabidopsis thaliana*, the CDS of the 20101108 release were obtained from TAIR [34]. For *Saccharomyces cerevisiae*, the open reading frames (without UTR, introns, etc.) of the 20110203 release were obtained from the Saccharomyces genome database [35]. For *Caenorhabditis elegans*, the CDS of WS241 were obtained from WormBase [36]. For *Mus musculus*, the CDS of the 20130508 release (GRCm38.p1) were obtained from the NCBI CCDS database [37].

mRNA folding

The mRNAs of all species were folded using Vienna RNA fold [38], at 20^{o}C, using the parameters of Andronescu, 2007 [39]. The *M. musculus* mRNA was also folded at 37^{o}C and the *S. cerevisiae* also at 30^{o}C, but all the reported comparisons are based on 20^{o}C.

mRNA sequence and structure statistics

Several statistics were taken from the mRNA sequence: gene length, codon usage, and nucleotide usage. Also from the predicted mRNA structure several statistics were taken: number of bound nucleotides, number of free nucleotides, average stem size, average loop size, variation in stem size, variation in loop size, and energy of the structure.

Gene expression and mRNA folding statistics

The correlations (Spearman) between gene expression and the various mRNA-based statistics were calculated by Spearman correlation (in R 3.0.2 x64). For some of the factors a correction was applied for gene-length, these were: number of bound nucleotides, number of unbound nucleotides, energy of the structure, number of stems, number of loops, triplet usage, nucleotide usage, and amino acid usage.

For expression codon analysis, the frequencies of use of synonymous codons was calculated. This was done over a receding window, from 50% highest versus 50% lowest until 5% highest versus 5% lowest, in increments of 1%.

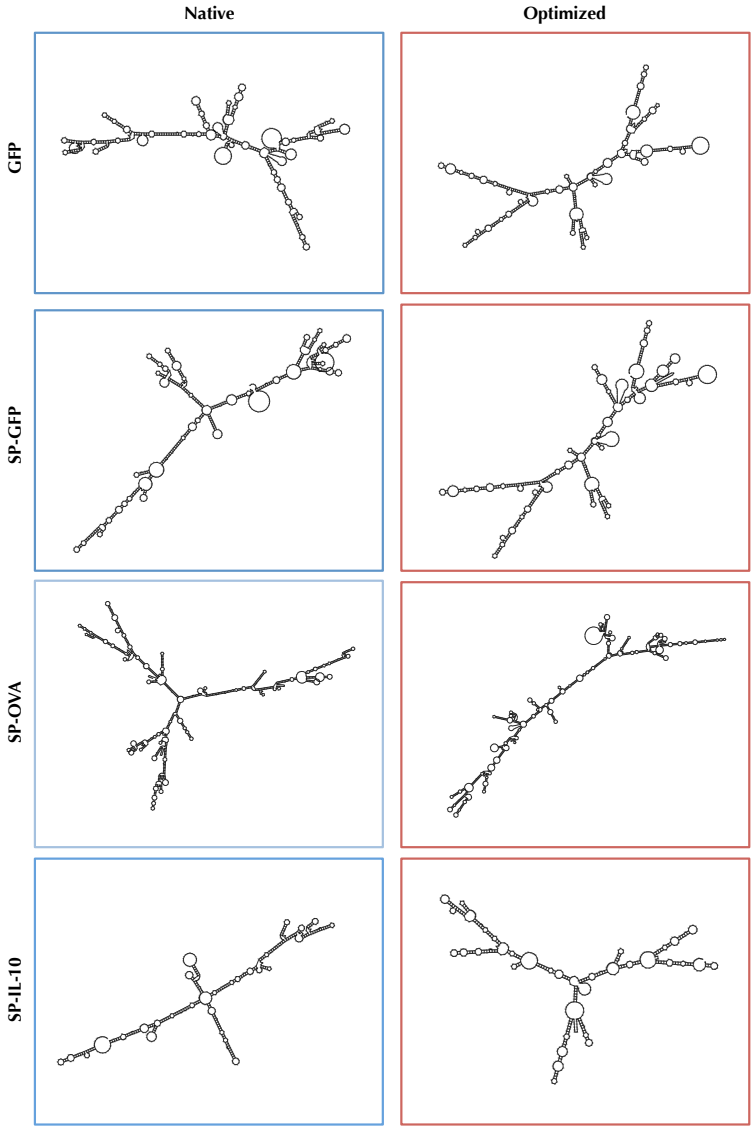
References

1. Hershberg, R. and D.A. Petrov, *General rules for optimal codon choice.* PLoS Genetics, 2009. 5(7): p. e1000556.
2. Post, L.E., et al., *Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in Escherichia coli.* PNAS, 1979. 76(4): p. 1697-1701.
3. Powell, J.R. and E.N. Moriyama, *Evolution of codon usage bias in Drosophila.* PNAS, 1997. 94(15): p. 7784-7790.
4. Ikemura, T., *Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes.* Journal of Molecular Biology, 1981. 146(1): p. 1-21.
5. Stenico, M., A.T. Lloyd, and P.M. Sharp, *Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases.* Nucleic Acids Research, 1994. 22(13): p. 2437-46.
6. Wright, S.I., et al., *Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata.* Molecular Biology and Evolution, 2004. 21(9): p. 1719-26.
7. Kanaya, S., et al., *Codon Usage and tRNA Genes in Eukaryotes: Correlation of Codon Usage Diversity with Translation Efficiency and with CG-Dinucleotide Usage as Assessed by Multivariate Analysis.* Journal of Molecular Evolution, 2001. 53(4): p. 290-298.
8. Bulmer, M., *The Selection-Mutation-Drift Theory of Synonymous Codon Usage.* Genetics, 1991. 129(3): p. 897-907.
9. Jackson, M.A., et al., *Design rules for efficient transgene expression in plants.* Plant Biotechnology Journal, 2014.
10. Kudla, G., et al., *Coding-Sequence Determinants of Gene Expression in Escherichia coli.* Science, 2009. 324(5924): p. 255-258.
11. Ingolia, N.T., et al., *Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling.* Science, 2009. 324(5924): p. 218-223.
12. Arava, Y., et al., *Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae.* PNAS, 2003. 100(7): p. 3689-3894.
13. Shah, P., et al., *Rate-limiting steps in yeast protein translation.* Cell, 2013. 153(7): p. 1589-601.
14. Wang, L. and M. Roossinck, *Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants.* Plant Molecular Biology, 2006. 61(4): p. 699-710.
15. Wang, Y., et al., *Precision and functional specificity in mRNA decay.* PNAS, 2002. 99(9): p. 5860-5865.
16. Yang, E., et al., *Decay Rates of Human mRNAs: Correlation With Functional Characteristics and Sequence Attributes.* Genome Research, 2003. 13: p. 1863-1872.
17. Narsai, R., et al., *Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana.* Plant Cell, 2007. 19(11): p. 3418-3436.
18. Zamore, P.D. and B. Haley, *Ribo-gnome: The Big World of Small RNAs.* Science Special Section, 2005. 309: p. 1519-1524.
19. Iost, I. and M. Dreyfus, *The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation.* EMBO Journal, 1995. 14(13): p. 3252-3261.

20. Bernstein, J.A., et al., *Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays.* PNAS, 2002. 99(15): p. 9697-9702.
21. Garneau, N.L., J. Wilusz, and C.J. Wilusz, *The highways and byways of mRNA decay.* Nature Reviews Molecular Cell Biology, 2007. 8(2): p. 113-126.
22. Lin, Z., R.J.C. Gilbert, and I. Brierley, *Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U6A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting.* Nucleic Acids Research, 2012: p. 1-16.
23. Hansen, T.M., et al., *Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting.* PNAS, 2007. 104(14): p. 5830-5835.
24. Doma, M.K. and R. Parker, *Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.* Nature 2006. 440(23): p. 561-564.
25. Onouchi, H., et al., *Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of Arabidopsis.* Genes & Development, 2005. 19: p. 1799-1810.
26. Hosoda, N., et al., *Translation Termination Factor eRF3 Mediates mRNA Decay through the Regulation of Deadenylation.* The Journal of Biological Chemistry, 2003. 278(40): p. 38287-38291.
27. Takyar, S., R.P. Hickerson, and H.F. Noller, *mRNA Helicase Activity of the Ribosome.* Cell, 2005. 120: p. 49-58.
28. Qu, X., et al., *The ribosome uses two active mechanisms to unwind messenger RNA during translation.* Nature, 2011. 475: p. 118-121.
29. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature 2011. 473: p. 337-342.
30. Reichel, C., et al., *Enhanced green fluorescence by the expression of an Aequorea victoria green fluorescent protein mutant in mono- and dicotyledonous plant cells.* PNAS, 1996. 93(12): p. 5888-93.
31. Westerhof, L.B., et al., *3D Domain Swapping Causes Extensive Multimerisation of Human Interleukin-10 When Expressed In Planta.* Plos One, 2012. 7: p. e46460.
32. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Research, 2001. 29(9): p. e45.
33. Wang, M., et al., *PaxDb, a database of protein abundance averages across all three domains of life.* Mol Cell Proteomics, 2012. 11(8): p. 492-500.
34. Lamesch, P., et al., *The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools.* Nucleic Acids Research, 2012. 40(Database issue): p. D1202-10.
35. Cherry, J.M., et al., *Saccharomyces Genome Database: the genomics resource of budding yeast.* Nucleic Acids Research, 2012. 40(Database issue): p. D700-5.
36. Yook, K., et al., *WormBase 2012: more genomes, more data, new website.* Nucleic Acids Research, 2012. 40(Database issue): p. D735-41.
37. Farrell, C.M., et al., *Current status and new features of the Consensus Coding Sequence database.* Nucleic Acids Research, 2014. 42(Database issue): p. D865-72.
38. Lorenz, R., et al., *ViennaRNA Package 2.0.* Algorithms for Molecular Biology, 2011. 6: p. 26.
39. Andronescu, M., et al., *Efficient parameter estimation for RNA secondary structure prediction.* Bioinformatics, 2007. 23(13): p. i19-28.

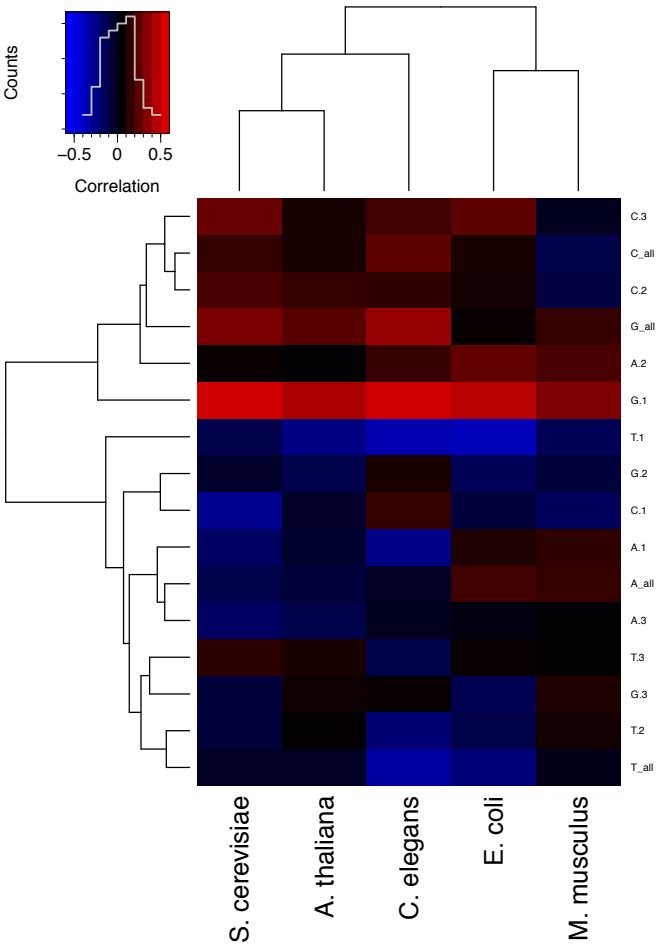
Supplementary figures

Supplementary Figure 1



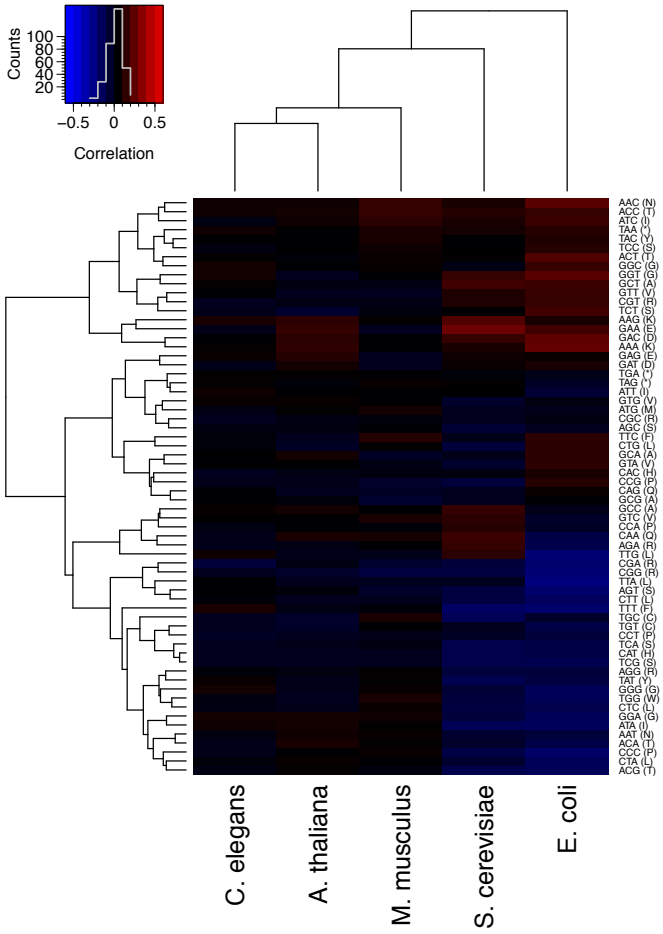
Supplementary Figure 1 ◆ mRNA structure predictions of the constructs used for heterologous protein expression. Sequences of the native and optimised variants of *Aequorea victoria* green fluorescent protein (GFP), *Gallus gallus* ovalbumin (OVA) and *Mus musculus* interleukin-10 (IL-10) with additional signal peptide (SP) and GFP without SP flanked by the 5′ and 3′-UTRs asexpected from our expression cassette were used to predict the mRNA secondary structure.

Supplementary Figure 2



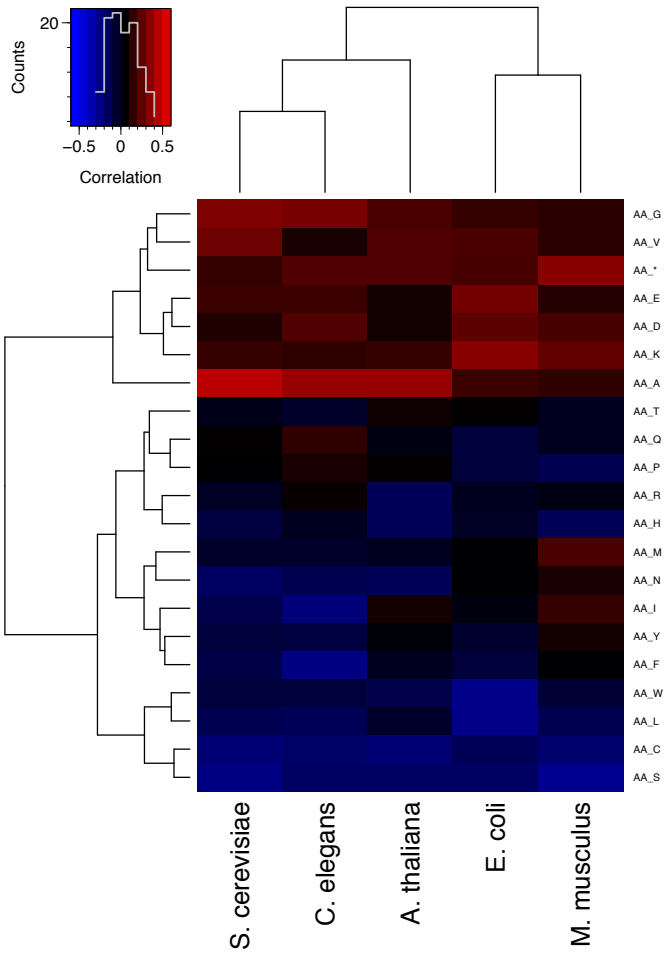
Supplementary Figure 2 ◆ Heat maps displaying the relation between species of several kingdoms of life based on translation rate-linked nucleotide use. Correlation (Spearman) between mRNA:protein ratios (proxy for translation rate) and nucleotide content (overall and for each codon position) for the species *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia). For each species >250 microarrays originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) was rank-normalized, averaged and divided by protein abundance (retrieved from PaxDB) before correlations (Spearman) between protein:mRNA ratios and nucleotide use were calculated.

Supplementary Figure 3



Supplementary Figure 3 ◆ Heat maps displaying the relation between species of several kingdoms of life based on translation rate-linked codon use. Correlation (Spearman) between mRNA:protein ratios (proxy for translation rate) and codon use for the species *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia). For each species >250 microarrays originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) was rank-normalized, averaged and divided by protein abundance (retrieved from PaxDB) before correlations (Spearman) between protein:mRNA ratios and nucleotide use were calculated.

Supplementary Figure 4



Supplementary Figure 4 ◆ Heat maps displaying the relation between species of several kingdoms of life based on translation rate-linked amino acid use. Correlation (Spearman) between mRNA:protein ratios (proxy for translation rate) and amino acid use for the species *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia). For each species >250 microarrays originating from multiple studies covering a wide range of strains/ecotypes, culturing conditions, developmental stages and tissues (Supplementary Table 1A-F) was rank-normalized, averaged and divided by protein abundance (retrieved from PaxDB) before correlations (Spearman) between protein:mRNA ratios and nucleotide use were calculated.

Supplementary tables

Supplementary Table 1A-F ◆ Gathered expression data and references

Supplementary Table 1A ◆ Overview of the gathered expression data per species.

	E. coli	*S. cerevisiae*	*C. elegans*	*A. thaliana*	*M. musculus*
Strains/ecotypes	7	13	14	8	9
Samples	168	316	391	415	777
Controls	105	211	109	101	565
Papers	8	9	9	9	9
Treatments	20	14	29	73	21
Tissues	1	1	3	11	28

Treatments: > Different strains/mutants and tissues receiving the same experimental treatment are counted as a single treatment, all measurements in a time series are counted as a single treatment

Additional remarks:
> *M. musculus* data sets Thorrez *et al.,* 2009 and Xue *et al.,* 2013 do not include the control spot on the slide in their datasets
> *E. coli* expression values from the Dong and Schellhorn 2009 dataset off to a single decimal and from Ito *et al.,* 2009 dataset to two decimals

Supplementary Table 1B ◆ Description of the gathered *E. coli* expression data.

Study	Short experiment description	strain	Growth temp.	Growth medium	Time series	# Samples	# Control samples	GEO accession	Remarks
Laubacher and Ades 2008	Transcription profiling of Escherichia coli treated with cefsulodin and mecillinam affecting the peptidoglycan layer	K12-MG1655	37°C	LB	yes	18	8	GSE10160 and GSE10159 (identical)	samples taken before cell growth slowed due to antibiotics, **control: no antibiotics added**
Dong and Schellhorn 2009	Global effect of RpoS on gene expression in pathogenic Escherichia coli	O157:H7-EDL933	37°C	LB	no	12	6	GSE17420	wt and rpoS mutant, sampled in stationary and exponential phase, **control: wildtype, note: values in GSE file rounded to a single decimal**
Ito *et al.*, 2009	Gene expression of E. coli at the outside and inside of biofilms	K12-MG1655 (pOX38Km)	37°C	MOPS minimal medium	no	12	12	GSE13589	4 stadia: inside and outside biofilm, stationary and exponential planktonic, **note: values in GSE file rounded to two decimals**
Jandu *et al.*, 2009	Transcription profiling of E. coli growth in the absence or presence of HEp-2 epithelial cells	O157:H7-CL56	37°C	minimal medium / Penassay broth	no	18	9	GSE14238	4 growth conditions: Pen. Broth, min. med, min. med with epithelial cells, min. med. With 5% CO2, **control: Pen. Medium en minimal medium**
Potrykus *et al.*, 2010	Transcription profiling of E. coli cells overexpressing GraL for short and long periods of time	K12-MG1655	37°C	LB	no	24	6	GSE15406 (GSE15404 and GSE15405 contain parts of this dataset)	samples taken 15 minutes after induction of GraL overexpression by IPTG but half the strains continuously overexpress GraL, wt-like strain and a ppGpp deficient cell (renders cell vulnerable to global regulators) both transformed with plasmids causing overexpression after induction or continuously, control plasmids with empty vectors, **controls: wt-like strain with empty vector plasmids**
Mensa *et al.*, 2011	E.coli response to Antimicrobial Arylamides	K12-D31	37°C	LB	yes	18	6	GSE31140	exposed to PMX 10070 or polymyxin, sampled after 20 and 60 min., **control: no antimicrobial treatment**
Sharma *et al.*, 2011	Design of an improved host platform for the over expression of recombinant proteins in Escherichia coli	K12-DH5α, BL-21 (DE3)	37°C	Terrific Broth (TB)	yes	27	19	GSE28412 + GSE29439 + GSE29440	A CG2 strain was created from DH5α by knocking in vhb and knocking out ribB, BL21 transformed with plasmids for the production of rhIFN-β and xylanase, DH5α (wt and CG2) for GFP, samples taken for multiple specific growth rates before induction and 2, 4 and 6 hrs after induction, **control: BL21 and DH5α wt, note: CG2 strain not mentioned at all in paper**
Le Gac *et al.*, 2012	Evolution of gene expression during long term coexistence in a bacterial evolution experiment	E. coli B str. REL606	37°C	Davis minimal medium	no	39	39	GSE30639	2 strains (L&S) from Ara-2 diverged from one another by generation 6,500 and then coexisted until at least generation 40,000, sampled four clones from each lineage from frozen samples obtained at three time points during the evolution experiment: generations 6,500 (6.5K), 17,000 (17K), and 40,000 (40K), ancestral strain sampled as well

Supplementary Table 1C ◆ Description of the gathered *S. cerevisiae* expression data.

Study	Short experiment description	Strain	Growth temperature	Growth medium	Time series	# Samples	# Control samples	GEO accession	Remarks
Li and Klevecz 2006	effect of phenelzine on transcriptional oscillation	IFO0233	30°C	glucose medium	yes	48	11	GSE9302	samples were taken every 4 minutes after inoculation of phenelzine for 200 minutes, **control: time series without phenelzine**
Gibson *et al.*, 2008 a	transcriptional change during full-scale industrial brewing processes	CB11	20°C	wort (85% malt, 15% sugar adjunct)	yes	17	17	GSE9423	samples taken in a time series from a propagation vessel and a 4th generation fermentation cycle
Gibson *et al.*, 2008 b	transcription during brewery fermentation	CB11	unknown (same as above study?)	wort (85% malt, 15% sugar adjunct)	yes	14	14	GSE10205	samples taken in a time series during a fermentation cycle
Rossouw *et al.*, 2008; Rossouw and bauer 2009; Rossouw *et al.*, 2009; Rossouw *et al.*, 2010	transcription during fermation of wine yeast strains	VIN13, DV10, BM45, EC1118, 285	22°C	synthethic wine must MS300	yes	43	43	GSE11651	samples taken at days 2, 5 and 14 during exponential, early and late logarithmic growth phases respectively
Fendt *et al.*, 2010	transcription factors controlling metabolic pathways	FY4	30°C	minimal medium	no	23	12	GSE24057	wild type and gcr2, pho2, bas1, gcn4 deletions, **control: wild type**
Bester *et al.*, 2012	transcription in mss11 and flo8 mutants involved in regulation of cell-wall encoding proteins	Σ1278b, BY4742	30°C	minimal SCD medium	no	16 and 12	6: 4 Σ1278b, 2 BY4742 5: 4 Σ1278b, 1 BY4742	GSE17716 and GSE29371	wildtype strains and mss11 and flo8 deletion and overexpression mutants, reference strains transformed with empty vectors, **control: wildtype**
Chin *et al.*, 2012	transcriptional oscillation	CEN.PK 113-7D	30°C	fermentor medium	yes	81	81	GSE30053 (GSE30051 and GSE30052 contain subsets)	samples taken every 4 minutes from the 2 hr redox cycle population and every 5 minutes from the 4 hr redox cycle population
Simmons Kovacs *et al.*, 2012	individual contributions of a transcription factor network and cyclin-CDKs to the maintenance of cell-cycle oscillations	BF264-D	37°C	YEP medium	yes	48	15	GSE32974	mutant strain lacking CDK1 activity at 37+°C, samples taken during a 350min time series, **control: wildtype time series**
Bessonov *et al.*, 2013	testing the Interdependent Correlation Clustering (ICC) method to analyze the relationships among genes conditioned to expression of NSF1 during wine fermentation	M2	18°C	Riesling grape must	yes	14	7	GSE34117	wt and nsf1 mutant sampled after 24 hr and 20 and 85% glucose fermented, **control: wt time series**

Supplementary Table 1D ◆ Description of the gathered *C. elegans* expression data.

Study	Short experiment description	Strain	Nematode part	Age	Growth temp.	Time series	# Samples	# Control samples	GEO accession	Remarks
Kirienko and Fay 2007	role of LIN35 in cell-proliferation of larvae	N2 wt, lin-35(n745)	whole larvae or embryo	L1, L4 and embryonic stage	20°C	no	18	9: 3/stage	GSE6547	**control: N2 wt**
Von Stetina *et al.*, 2007	gene expression in neurons	N2, SD1241, NC694	all cells or neurons	embryo and L2	20°C	no	24	14: 4 embryo 10 L2	GSE8004	mutants used to identify neuron cells, **control: wildtype embryo and L2 all cells, L2 underwent non-specific control IP**
Jans *et al.*, 2009	gene regulation by the dosage compensation complex	N2 and DCC mutant strains	embryo	embryonic	unknown	no	26	12	GSE14640-GPL200 & GSE14649 (identical)	Control spots do not include GEO file, 2 dosage compensation mutants and a X0 mutant, each mutant and their control grown on a different medium, **control: wildtypes**
Grompone *et al.*, 2012	gene expression in worms fed on lactic acid bacteria	N2	whole worm	adult	20°C	yes	18	6: 3/feeding time	GSE42192	3 or 10 day feeding period on OP50 or lactobacillus strains, **control: OP50 fed**
Pietsch *et al.*, 2012	transcriptomes of nematodes exposed to quercetin and tannic acid	N2	whole worm	L4	20°C	no	24	6	GSE35354	different conc. Of quercetin or tannic acid in the growth medium, **control: no querc. or TA**
Macneil *et al.*, 2013	transcriptome of nematodes reared on E. coli or Comamonas	N2	whole worm	young adult and gravid adult	20°C	no	18	6: 3/stage	GSE43959	nematodes fed with 2 E. coli strains, a Comamonas strain and a mixture, **control: fed on OP50**
Sarkies *et al.*, 2013	gene expression upon viral (Orsay) infection	N2, JU1580, RDE-1	whole worm	mixed stage	20°C	no	20	7: 4 N2, 3 JU1580	GSE41058-GPL200	N2 wt relatively resistant to Orsay, the wild strain JU1580 is sensitive, rde-1 mutants of N2 are sensitive as well, **control: uninfected N2 and JU1580**
Uno *et al.*, 2013	transcriptomes of nematodes after fasting-stimulus	N2 wt, 3 mutants	whole worm	adult	unknown	yes	46	9	GSE27677	2 experiments: 1)N2 fed and fasting time series transcriptome analysis, 2) transcriptome analysis of N2 and 3 mutants after 2 days of fasting, **control: N2 fed, all time points**
Baugh *et al.*, 2005	Transcription profiling of time series of C.elegans mutants lacking or with extra C-blastomeres and wild type	N2, JJ532, JJ518	embryo	embryo	22°C	yes	123	34	GSE2180	4 genotypes: N2 wt, pie-1 and pie-1;pal-1 mutants of JJ532 and mex-3;skn-1 mutant of JJ518, pal-1 and skn-1 mutants created with RNAi, sampled from 0-186 minutes after the 4-cell embryo stage, **control: wildtype time series**
Yanai *et al.*, 2008	transcription of RNAi knock down of 13 transcription factors early in the C lineage of C.elegans mex-3 (zu155) mutant worms	JJ518	embryo	embryo	22°C ?	no	74	6	GSE9665	fed with 12 different RNAi's, **control: no RNAi**

Supplementary Table 1E ◆ Description of the gathered *A. thaliana* expression data.

Study	Short experiment description	Ecotype	Plant part	Age	Growh temperature	Time series	# samples	# Control samples	GEO accession	Remarks
Goda *et al.*, 2008	gene expression in response to plant hormones	Columbia	seedling	7 days	23°C	no	72	6	GSE39384	part of the samples mutants
Goda *et al.*, 2008	gene expression in response to plant hormone inhibitors	Columbia	seedling	7 days	23°C	no	60	6	GSE39385	Several of the controls mentioned in the paper are not included in this dataset
Delker *et al.*, 2010	natural variation in transcriptional response to auxin	7 (incl. Col)	seedling	7 days	20°C	yes	83	21: 3/ecotype	GSE18975	samples taken 0, 0.5, 1 and 3h after induction, continuous illumination, **control: 0h post induction**
Nozue *et al.*, 2011	gene expression during hypocotyl growth	Columbia	seedling	4-6 days	20°C	yes	48	21	GSE21684	wildtype, arythmic circadian clock (CCA1-OX) and pif4pif5 mutants, grown under short-day conditions for 3 days and then transferred to short light/dark cycles and sampled in a time series after 120 to 4120 minutes
Hewezi *et al.*, 2012	gene expression regulated by miR396-GRF	Wassilewskija, Columbia	roots	14 days	23°C	no	15	6: 3/ecotype	GSE31593	also triple mutant grf1/grf2/grf3 and trangenic plants overexpressing rGRF1 and rGRF2, **control: wild types**
Reeves *et al.*, 2012	mutant analysis of gene expression in flowers	Columbia	stage 12 buds, stage 13 flowers	unknown	unknown	no	18	6: 3/stage	GSE32193	myb and arf mutants, **control: wildtype**
Ruckle *et al.*, 2012	impact of light and plastid signaling on the transcriptome	Columbia	seedling	7 days	21°C	yes	47	8	GSE24517	seedlings grown in presence or absence of lincomyn, after 6 days subjected to shift in fluence rate of light, seedlings collected after 0, 0.5, 1, 4 and 24 h, **control: 0h without lincomyn**
Xu *et al.*, 2012	gene expression during callus initiation	Columbia	root explants and aerial organ explants	10 days	22°C	yes	30	6: 3/plant part	GSE29543	samples 0, 12, 24, 48 and 96 hrs after transfer to callus inducing medium, **control: 0h**
Bargmann *et al.*, 2013	tissue specific gene expression in auxin response	Columbia	stele, pericycle, epidermis, columella, intact root	7 days	22°C	no	30	15: 3/plant part	GSE35580	
Rugnone *et al.*, 2013	gene expression response to light pulses given at the middle of the day and night	Columbia	aerial rosette	15 days	22°C	no	12	6: 3/time point	GSE46761 GPL198 & GSE46621 (identical)	1hr light pulse given at middle of day or middle of night, controls without light pulse were kept in the dark for 1 hr

Supplementary Table 1F ◆ Description of the gathered *M. musculus* expression data.

Study	Short experiment description	Strain	Mouse tissue	Age	Time series	# Samples	# Control samples	GEO accession	Remarks
Finkielstain *et al.*, 2009	gene expression in multiple tissues during somatic growth in early postnatal life	C57BL/6	kidney, liver, heart, lung	1, 4, 8 weeks	no	40	40	GSE38754	male mice, castrated at 3 weeks
Henrichsen *et al.*, 2009	effect of copy number variation on transcriptomes	7BL/6J, 129S2, DBA2/J, AKR/J, A/J, C3HeB/FeJ	lung, kidney, brain, heart, testis, liver	11-14 weeks	no	108	108	GSE10744	inbred lines, male mice
Schmelzer *et al.*, 2010	genome-wide transcript profiling in liver, heart, brain and kidney of senescensce accelerated mice prone 1 (SAMP1) mice supplemented with the reduced or oxidized form of Q10	SAMP1	brain, heart, kidney, liver	6 or 14 months	no	72	24	GSE15129	female mice, **control: no coenzyme supplemented**
Chaignat *et al.*, 2011	effect of copy number variation on transcriptomes	129S2, C57BL/6J, DBA/2J	brain, liver	embryonic stage E14.5, 1 day post-natal, 7days, 11-14 weeks	yes	72	72	GSE16675	male mice
Stevens *et al.*, 2011; Vartanian *et al.*, 2011	gene expression in mice preconditioned with lipopolysaccharides (LPS) or oligodeoxynucleotide (CpG) after an induced stroke	C57Bl/6J	brain, blood	8-12 weeks	yes	224	60	GSE32529	male mice received either preconditioning alone, preconditioning plus ischemic challenge (45 min MCAO) or ischemic challenge alone. Preconditioning included: LPS, CpG, saline, brief ischemia (MCAO, 12 min) or sham surgery. **control: saline or sham surgery preconditioning without MCAO and unhandeled mice, all time points and tissues**
Taher *et al.*, 2011	gene expression during early limb development	FVB	whole embryo, forelimb, hindlimb	embryo stage 9.5-13.5	yes	51	51	GSE30138	
Thorrez *et al.*, 2011	tissue-specific repression of house-keeping genes	C57BL6	23 tissues	10-12 weeks	no	73	73	GSE24207	No control spots in the GEO file,
Seita *et al.*, 2012	gene expression map of mouse hematopoiesis	C57BL/6	bone marrow, spleen, thymus, thymocyte	2-3 months	no	101	101	GSE34723	male mice, multiple hematopoiesis populations per tissue
Xue *et al.*, 2013	Expression data for mouse embryogenesis from oocyte to newborn	C57BL/6J	dechorionated embryo	unfertilized egg to newborn	yes	36	36	GSE39897	No control spots in the GEO file

Supplementary Tables 2A-E ◆ Relative synonymous codon use frequency averages of all genes and gene subsets of *Escherichia coli, Saccharomyces cerevisiae, Caenorhabdites elegans*, *Arabidopsis thaliana* and *Mus musculus*.

Supplementary Tables 2A ◆ Relative synonymous codon use frequency averages of all genes and gene subsets based on expression for *E. coli*.

AA	Triplet	All	Top 5%	Bottom 5%	Top/Bottom
*	TAA	0.648	0.871	0.643	1.355
	TAG	0.067	0.021	0.064	0.328
	TGA	0.285	0.107	0.293	0.365
A	GCT	0.160	0.332	0.154	2.156
	GCC	0.266	0.139	0.275	0.505
	GCA	0.209	0.258	0.216	1.194
	GCG	0.365	0.271	0.356	0.761
C	TGT	0.435	0.385	0.466	0.826
	TGC	0.565	0.615	0.534	1.152
D	GAT	0.617	0.429	0.649	0.661
	GAC	0.383	0.571	0.351	1.627
E	GAA	0.693	0.760	0.681	1.116
	GAG	0.307	0.240	0.319	0.752
F	TTT	0.562	0.290	0.615	0.472
	TTC	0.438	0.710	0.385	1.844
G	GGT	0.343	0.527	0.327	1.612
	GGC	0.413	0.406	0.395	1.028
	GGA	0.098	0.031	0.116	0.267
	GGG	0.146	0.036	0.162	0.222
H	CAT	0.557	0.295	0.591	0.499
	CAC	0.443	0.705	0.409	1.724
I	ATT	0.503	0.302	0.530	0.570
	ATC	0.434	0.688	0.380	1.811
	ATA	0.063	0.010	0.090	0.111
K	AAA	0.770	0.768	0.793	0.968
	AAG	0.230	0.232	0.207	1.121
L	TTA	0.124	0.042	0.155	0.271
	TTG	0.124	0.059	0.130	0.454
	CTT	0.100	0.065	0.112	0.580
	CTC	0.103	0.068	0.106	0.642
	CTA	0.035	0.008	0.041	0.195
	CTG	0.515	0.758	0.457	1.659
M	ATG	1.000	1.000	1.000	1.000
N	AAT	0.432	0.182	0.486	0.374
	AAC	0.568	0.818	0.514	1.591
P	CCT	0.152	0.140	0.165	0.848
	CCC	0.115	0.025	0.137	0.182
	CCA	0.185	0.134	0.199	0.673
	CCG	0.547	0.702	0.498	1.410
Q	CAA	0.337	0.213	0.369	0.577
	CAG	0.663	0.787	0.631	1.247
R	CGT	0.396	0.636	0.363	1.752
	CGC	0.410	0.332	0.410	0.810
	CGA	0.058	0.010	0.071	0.141
	CGG	0.089	0.011	0.094	0.117
	AGA	0.030	0.007	0.044	0.159
	AGG	0.016	0.004	0.019	0.211
S	TCT	0.150	0.323	0.132	2.447
	TCC	0.155	0.256	0.136	1.882
	TCA	0.117	0.058	0.123	0.472
	TCG	0.155	0.057	0.171	0.333
	AGT	0.143	0.060	0.158	0.380
	AGC	0.280	0.247	0.280	0.882
T	ACT	0.168	0.328	0.167	1.964
	ACC	0.449	0.508	0.409	1.242
	ACA	0.120	0.048	0.154	0.312
	ACG	0.263	0.116	0.270	0.430
V	GTT	0.257	0.436	0.258	1.690
	GTC	0.214	0.113	0.219	0.516
	GTA	0.152	0.225	0.153	1.471
	GTG	0.377	0.226	0.370	0.611
W	TGG	1.000	1.000	1.000	1.000
Y	TAT	0.555	0.331	0.582	0.569

Gene subsets were defined by expression in terms of percentage; top 5% high-, bottom 5% low-expressed. The fold change in codon use comparing high to low expressed genes (Top/Bottom) was also calculated.

Supplementary Tables 2B ◆ Relative synonymous codon use frequency averages of all genes and gene subsets based on expression for *S. cerevisiae*.

AA	Triplet	All	Top 5%	Bottom 5%	Top/Bottom
*	TAA	0.480	0.731	0.403	1.814
	TAG	0.225	0.117	0.290	0.403
	TGA	0.295	0.152	0.307	0.495
A	GCT	0.367	0.593	0.339	1.749
	GCC	0.223	0.280	0.215	1.302
	GCA	0.296	0.105	0.319	0.329
	GCG	0.113	0.023	0.127	0.181
C	TGT	0.627	0.829	0.594	1.396
	TGC	0.373	0.171	0.406	0.421
D	GAT	0.656	0.526	0.642	0.819
	GAC	0.344	0.474	0.358	1.324
E	GAA	0.701	0.854	0.699	1.222
	GAG	0.299	0.146	0.301	0.485
F	TTT	0.593	0.353	0.616	0.573
	TTC	0.407	0.647	0.384	1.685
G	GGT	0.455	0.823	0.387	2.127
	GGC	0.197	0.093	0.197	0.472
	GGA	0.224	0.051	0.279	0.183
	GGG	0.124	0.033	0.138	0.239
H	CAT	0.643	0.440	0.617	0.713
	CAC	0.357	0.560	0.383	1.462
I	ATT	0.463	0.522	0.469	1.113
	ATC	0.258	0.430	0.236	1.822
	ATA	0.280	0.048	0.295	0.163
K	AAA	0.581	0.299	0.639	0.468
	AAG	0.419	0.701	0.361	1.942
L	TTA	0.279	0.216	0.244	0.885
	TTG	0.283	0.567	0.251	2.259
	CTT	0.127	0.057	0.163	0.350
	CTC	0.057	0.014	0.086	0.163
	CTA	0.142	0.103	0.143	0.720
	CTG	0.112	0.043	0.113	0.381
M	ATG	1.000	1.000	1.000	1.000
N	AAT	0.598	0.303	0.594	0.510
	AAC	0.402	0.697	0.406	1.717
P	CCT	0.310	0.227	0.305	0.744
	CCC	0.160	0.053	0.164	0.323
	CCA	0.407	0.701	0.401	1.748
	CCG	0.123	0.018	0.129	0.140
Q	CAA	0.686	0.893	0.663	1.347
	CAG	0.314	0.107	0.337	0.318
R	CGT	0.140	0.201	0.131	1.534
	CGC	0.058	0.017	0.078	0.218
	CGA	0.068	0.001	0.088	0.011
	CGG	0.040	0.002	0.064	0.031
	AGA	0.478	0.724	0.420	1.724
	AGG	0.217	0.055	0.218	0.252
S	TCT	0.261	0.452	0.246	1.837
	TCC	0.157	0.289	0.147	1.966
	TCA	0.211	0.108	0.218	0.495
	TCG	0.097	0.036	0.096	0.375
	AGT	0.163	0.063	0.172	0.366
	AGC	0.111	0.051	0.121	0.421
T	ACT	0.343	0.482	0.333	1.447
	ACC	0.210	0.352	0.213	1.653
	ACA	0.307	0.133	0.325	0.409
	ACG	0.140	0.034	0.129	0.264
V	GTT	0.389	0.511	0.368	1.389
	GTC	0.201	0.347	0.210	1.652
	GTA	0.216	0.060	0.226	0.265
	GTG	0.195	0.082	0.196	0.418
W	TGG	1.000	1.000	1.000	1.000
Y	TAT	0.568	0.302	0.558	0.541
	TAC	0.432	0.698	0.442	1.579

Gene subsets were defined by expression in terms of percentage; top 5% high-, bottom 5% low-expressed. The fold change in codon use comparing high to low expressed genes (Top/Bottom) was also calculated.

Supplementary Tables 2C ◆ Relative synonymous codon use frequency averages of all genes and gene subsets based on expression for *C. elegans*.

AA	Triplet	All	Top 5%	Bottom 5%	Top/Bottom
*	TAA	0.496	0.694	0.439	1.581
	TAG	0.179	0.141	0.162	0.870
	TGA	0.325	0.165	0.399	0.414
A	GCT	0.354	0.423	0.325	1.302
	GCC	0.199	0.302	0.157	1.924
	GCA	0.314	0.198	0.385	0.514
	GCG	0.133	0.077	0.134	0.575
C	TGT	0.555	0.447	0.588	0.760
	TGC	0.445	0.553	0.412	1.342
D	GAT	0.679	0.631	0.693	0.911
	GAC	0.321	0.369	0.307	1.202
E	GAA	0.621	0.534	0.671	0.796
	GAG	0.379	0.466	0.329	1.416
F	TTT	0.481	0.261	0.605	0.431
	TTC	0.519	0.739	0.395	1.871
G	GGT	0.204	0.168	0.214	0.785
	GGC	0.124	0.086	0.134	0.642
	GGA	0.592	0.711	0.544	1.307
	GGG	0.080	0.035	0.109	0.321
H	CAT	0.611	0.513	0.649	0.790
	CAC	0.389	0.487	0.351	1.387
I	ATT	0.534	0.470	0.538	0.874
	ATC	0.314	0.478	0.226	2.115
	ATA	0.152	0.052	0.236	0.220
K	AAA	0.588	0.381	0.665	0.573
	AAG	0.412	0.619	0.335	1.848
L	TTA	0.110	0.049	0.169	0.290
	TTG	0.234	0.212	0.258	0.822
	CTT	0.249	0.306	0.214	1.430
	CTC	0.174	0.280	0.116	2.414
	CTA	0.091	0.042	0.112	0.375
	CTG	0.142	0.112	0.133	0.842
M	ATG	1.000	1.000	1.000	1.000
N	AAT	0.625	0.484	0.655	0.739
	AAC	0.375	0.516	0.345	1.496
P	CCT	0.178	0.126	0.220	0.573
	CCC	0.088	0.054	0.100	0.540
	CCA	0.532	0.691	0.494	1.399
	CCG	0.202	0.130	0.186	0.699
Q	CAA	0.651	0.650	0.679	0.957
	CAG	0.349	0.350	0.321	1.090
R	CGT	0.217	0.350	0.150	2.333
	CGC	0.096	0.175	0.067	2.612
	CGA	0.236	0.146	0.231	0.632
	CGG	0.091	0.046	0.098	0.469
	AGA	0.288	0.250	0.357	0.700
	AGG	0.071	0.032	0.097	0.330
S	TCT	0.206	0.235	0.214	1.098
	TCC	0.130	0.177	0.112	1.580
	TCA	0.257	0.205	0.273	0.751
	TCG	0.156	0.169	0.125	1.352
	AGT	0.149	0.104	0.173	0.601
	AGC	0.102	0.109	0.103	1.058
T	ACT	0.324	0.346	0.329	1.052
	ACC	0.175	0.297	0.144	2.062
	ACA	0.345	0.249	0.383	0.650
	ACG	0.156	0.108	0.143	0.755
V	GTT	0.388	0.413	0.407	1.015
	GTC	0.220	0.320	0.168	1.905
	GTA	0.158	0.097	0.191	0.508
	GTG	0.234	0.170	0.234	0.726
W	TGG	1.000	1.000	1.000	1.000
Y	TAT	0.559	0.414	0.631	0.656
	TAC	0.441	0.586	0.369	1.588

Gene subsets were defined by expression in terms of percentage; top 5% high-, bottom 5% low-expressed. The fold change in codon use comparing high to low expressed genes (Top/Bottom) was also calculated.

Supplementary Tables 2D ◆ Relative synonymous codon use frequency averages of all genes and gene subsets based on expression for *A. thaliana*.

AA	Triplet	All	Top 5%	Bottom 5%	Top/Bottom
*	TAA	0.345	0.371	0.263	1.411
	TAG	0.204	0.194	0.194	1.000
	TGA	0.451	0.435	0.543	0.801
A	GCT	0.432	0.498	0.383	1.300
	GCC	0.161	0.171	0.174	0.983
	GCA	0.263	0.221	0.278	0.795
	GCG	0.144	0.110	0.164	0.671
C	TGT	0.593	0.561	0.591	0.949
	TGC	0.407	0.439	0.409	1.073
D	GAT	0.674	0.644	0.662	0.973
	GAC	0.326	0.356	0.338	1.053
E	GAA	0.511	0.442	0.523	0.845
	GAG	0.489	0.558	0.477	1.170
F	TTT	0.502	0.427	0.515	0.829
	TTC	0.498	0.573	0.485	1.181
G	GGT	0.334	0.398	0.316	1.259
	GGC	0.141	0.119	0.152	0.783
	GGA	0.371	0.367	0.387	0.948
	GGG	0.154	0.115	0.145	0.793
H	CAT	0.606	0.526	0.612	0.859
	CAC	0.394	0.474	0.388	1.222
I	ATT	0.400	0.429	0.375	1.144
	ATC	0.363	0.432	0.373	1.158
	ATA	0.236	0.139	0.252	0.552
K	AAA	0.490	0.385	0.517	0.745
	AAG	0.510	0.615	0.483	1.273
L	TTA	0.135	0.082	0.148	0.554
	TTG	0.220	0.233	0.229	1.017
	CTT	0.257	0.290	0.248	1.169
	CTC	0.181	0.207	0.172	1.203
	CTA	0.105	0.080	0.121	0.661
	CTG	0.102	0.108	0.082	1.317
M	ATG	1.000	1.000	1.000	1.000
N	AAT	0.502	0.430	0.489	0.879
	AAC	0.498	0.570	0.511	1.115
P	CCT	0.381	0.407	0.353	1.153
	CCC	0.106	0.112	0.109	1.028
	CCA	0.327	0.336	0.351	0.957
	CCG	0.186	0.146	0.186	0.785
Q	CAA	0.564	0.465	0.648	0.718
	CAG	0.436	0.535	0.352	1.520
R	CGT	0.168	0.241	0.161	1.497
	CGC	0.070	0.077	0.068	1.132
	CGA	0.118	0.087	0.120	0.725
	CGG	0.092	0.059	0.086	0.686
	AGA	0.352	0.301	0.363	0.829
	AGG	0.199	0.234	0.202	1.158
S	TCT	0.280	0.303	0.253	1.198
	TCC	0.129	0.147	0.127	1.157
	TCA	0.204	0.178	0.212	0.840
	TCG	0.108	0.100	0.114	0.877
	AGT	0.151	0.139	0.158	0.880
	AGC	0.127	0.134	0.135	0.993
T	ACT	0.334	0.374	0.300	1.247
	ACC	0.207	0.260	0.213	1.221
	ACA	0.302	0.253	0.313	0.808
	ACG	0.157	0.114	0.175	0.651
V	GTT	0.400	0.432	0.372	1.161
	GTC	0.193	0.219	0.199	1.101
	GTA	0.145	0.095	0.157	0.605
	GTG	0.262	0.253	0.271	0.934
W	TGG	1.000	1.000	1.000	1.000
Y	TAT	0.504	0.418	0.508	0.823
	TAC	0.496	0.582	0.492	1.183

Gene subsets were defined by expression in terms of percentage; top 5% high-, bottom 5% low-expressed. The fold change in codon use comparing high to low expressed genes (Top/Bottom) was also calculated.

Supplementary Tables 2E ◆ Relative synonymous codon use frequency averages of all genes and gene subsets based on expression for *M. musculus*.

AA	Triplet	All	Top 5%	Bottom 5%	Top/Bottom
*	TAA	0.258	0.351	0.323	1.087
	TAG	0.235	0.222	0.253	0.877
	TGA	0.507	0.427	0.424	1.007
A	GCT	0.289	0.320	0.316	1.013
	GCC	0.377	0.331	0.340	0.974
	GCA	0.232	0.246	0.266	0.925
	GCG	0.101	0.103	0.078	1.321
C	TGT	0.476	0.516	0.507	1.018
	TGC	0.524	0.484	0.493	0.982
D	GAT	0.450	0.521	0.500	1.042
	GAC	0.550	0.479	0.500	0.958
E	GAA	0.412	0.466	0.495	0.941
	GAG	0.588	0.534	0.505	1.057
F	TTT	0.445	0.507	0.499	1.016
	TTC	0.555	0.493	0.501	0.984
G	GGT	0.175	0.208	0.197	1.056
	GGC	0.332	0.319	0.287	1.111
	GGA	0.257	0.272	0.313	0.869
	GGG	0.236	0.201	0.204	0.985
H	CAT	0.410	0.468	0.472	0.992
	CAC	0.590	0.532	0.528	1.008
I	ATT	0.343	0.404	0.362	1.116
	ATC	0.495	0.448	0.419	1.069
	ATA	0.162	0.148	0.219	0.676
K	AAA	0.398	0.407	0.471	0.864
	AAG	0.602	0.593	0.529	1.121
L	TTA	0.068	0.089	0.095	0.937
	TTG	0.132	0.152	0.152	1.000
	CTT	0.132	0.154	0.154	1.000
	CTC	0.194	0.169	0.176	0.960
	CTA	0.079	0.079	0.092	0.859
	CTG	0.396	0.357	0.331	1.079
M	ATG	1.000	1.000	1.000	1.000
N	AAT	0.436	0.481	0.501	0.960
	AAC	0.564	0.519	0.499	1.040
P	CCT	0.306	0.335	0.316	1.060
	CCC	0.298	0.250	0.275	0.909
	CCA	0.288	0.310	0.323	0.960
	CCG	0.108	0.105	0.086	1.221
Q	CAA	0.253	0.258	0.350	0.737
	CAG	0.747	0.742	0.650	1.142
R	CGT	0.084	0.105	0.080	1.312
	CGC	0.170	0.153	0.122	1.254
	CGA	0.123	0.145	0.104	1.394
	CGG	0.194	0.179	0.128	1.398
	AGA	0.213	0.232	0.318	0.730
	AGG	0.216	0.186	0.249	0.747
S	TCT	0.193	0.222	0.220	1.009
	TCC	0.211	0.195	0.188	1.037
	TCA	0.143	0.149	0.170	0.876
	TCG	0.054	0.057	0.039	1.462
	AGT	0.156	0.171	0.174	0.983
	AGC	0.243	0.206	0.209	0.986
T	ACT	0.249	0.273	0.275	0.993
	ACC	0.345	0.313	0.312	1.003
	ACA	0.295	0.314	0.328	0.957
	ACG	0.111	0.099	0.085	1.165
V	GTT	0.174	0.225	0.217	1.037
	GTC	0.245	0.215	0.241	0.892
	GTA	0.119	0.138	0.146	0.945
	GTG	0.461	0.423	0.395	1.071
W	TGG	1.000	1.000	1.000	1.000
Y	TAT	0.423	0.481	0.498	0.966
	TAC	0.577	0.519	0.502	1.034

Gene subsets were defined by expression in terms of percentage; top 5% high-, bottom 5% low-expressed. The fold change in codon use comparing high to low expressed genes (Top/Bottom) was also calculated.

Supplementary Table 3 ◆ Correlation between mRNA structure characteristics and gene expression per species.

	E. coli	*S. cereviciae*	*C. elegans*	*A. thaliana*	*M. musculus*
Gene length	-0.146	-0.041	0.093	0.030	-0.016
Energy (kcal.mol/nt)	-0.006	-0.319	-0.316	-0.229	0.006
Bound nt (fraction)	0.038	0.236	0.061	0.172	0.015
Mean stem size	-0.111	0.054	-0.182	0.053	-0.055
Mean loop size	-0.115	-0.241	-0.179	-0.155	-0.046
Transitions /nt	0.140	0.144	0.227	0.071	0.069

The mRNA structures of all genes of *Escherichia coli (*Bacteria*), Saccharomyces cerevisiae (*Fungi*), Caenorhabdites elegans (*Animalia*), Arabidopsis thaliana (*Plantae*)* and *Mus musculus (*Animalia*)* were predicted and gene length, minimal free folding energy, percentage of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of stem/loop transitions were determined and correlated (Spearman) with expression. Basis of the heat map presented in Figure 7.

Supplementary Table 4 ◆ Calculated mRNA structure characteristics of the constructs used for heterologous protein expression.

		Energy kcal/mol/nt	Bound nt's (fraction)	Mean stem size	Mean loop size	Transitions
GFP	N	-0.21	0.56	5.74	4.48	0.097
	O	-0.33	0.57	5.15	3.85	0.111
SP-GFP	N	-0.22	0.57	5.21	3.89	0.109
	O	-0.32	0.54	5.22	4.31	0.104
SP-OVA	N	-0.29	0.61	5.28	3.34	0.116
	O	-0.31	0.55	4.38	3.56	0.126
SP-IL-10	N	-0.29	0.60	5.02	3.29	0.120
	O	-0.27	0.54	4.08	3.47	0.131

Analysis of the mRNA secondary structure predictions given in Supplementary Figure 1. Folding energy, bound nucleotides and number of transitions are corrected for gene length. Stem and loop sizes are mean values.

Supplementary Tables 5 ◆ Correlation between mRNA structure characteristics and protein:mRNA ratios per species.

	E. coli	*S. cerevisiae*	*C. elegans*	*A. thaliana*	*M. musculus*
Gene length	-0.146	-0.116	-0.180	-0.139	-0.288
Energy (kcal/mol/nt)	0.043	-0.237	-0.212	-0.138	0.087
Bound (fraction)	-0.009	0.148	-0.006	0.062	-0.058
Mean stem size	-0.193	-0.011	-0.216	-0.058	-0.121
Mean loop size	-0.121	-0.182	-0.139	-0.105	-0.015
Transitions /nt	0.199	0.140	0.213	0.104	0.081

Correlations (Spearman) between mRNA structure characteristics and mRNA:protein ratios per species. The mRNA structures of all genes of *Escherichia coli* (Bacteria), *Saccharomyces cerevisiae* (Fungi), *Caenorhabdites elegans* (Animalia), *Arabidopsis thaliana* (Plantae) and *Mus musculus* (Animalia) were predicted and gene length, minimal free folding energy, percentage of bound nucleotides, mean stem and loop (stretches of bound and unbound nucleotides, respectively) size and number of stem/loop transitions were determined and correlated (Spearman) with mRNA:protein ratios. Rank-normalized mRNA levels were divided by protein abundance (retrieved from PaxDB). Basis of the heat map presented in Figure 8.

References of gathered expression data

Bargmann, B.O., et al., *A map of cell type-specific auxin responses.* Mol Syst Biol, 2013. 9: p. 688.

Baugh, L.R., et al., *The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the C. elegans embryo.* Development, 2005. 132(8): p. 1843-54.

Bessonov, K., et al., *Functional analyses of NSF1 in wine yeast using interconnected correlation clustering and molecular analyses.* PLoS One, 2013. 8(10): p. e77192.

Bester, M.C., D. Jacobson, and F.F. Bauer, *Many Saccharomyces cerevisiae Cell Wall Protein Encoding Genes Are Coregulated by Mss11, but Cellular Adhesion Phenotypes Appear Only Flo Protein Dependent.* G3 (Bethesda), 2012. 2(1): p. 131-41.

Chaignat, E., et al., *Copy number variation modifies expression time courses.* Genome Res, 2011. 21(1): p. 106-13.

Chin, S.L., et al., *Dynamics of oscillatory phenotypes in Saccharomyces cerevisiae reveal a network of genome-wide transcriptional oscillators.* FEBS J, 2012. 279(6): p. 1119-30.

Delker, C., et al., *Natural variation of transcriptional auxin response networks in Arabidopsis thaliana.* Plant Cell, 2010. 22(7): p. 2184-200.

Dong, T. and H.E. Schellhorn, *Global effect of RpoS on gene expression in pathogenic Escherichia coli O157:H7 strain EDL933.* BMC Genomics, 2009. 10: p. 349.

Fendt, S.M., et al., *Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast.* Mol Syst Biol, 2010. 6: p. 432.

Finkielstain, G.P., et al., *An extensive genetic program occurring during postnatal growth in multiple tissues.* Endocrinology, 2009. 150(4): p. 1791-800.

Gibson, B.R., et al., *Carbohydrate utilization and the lager yeast transcriptome during brewery fermentation.* Yeast, 2008. 25(8): p. 549-62.

Gibson, B.R., et al., *The oxidative stress response of a lager brewing yeast strain during industrial propagation and fermentation.* FEMS Yeast Res, 2008. 8(4): p. 574-85.

Goda, H., et al., *The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access.* The Plant Journal, 2008. 55(3): p. 526-542.

Henrichsen, C.N., et al., *Segmental copy number variation shapes tissue transcriptomes.* Nat Genet, 2009. 41(4): p. 424-9.

Hewezi, T., et al., *The Arabidopsis microRNA396-GRF1/GRF3 regulatory module acts as a developmental regulator in the reprogramming of root cells during cyst nematode infection.* Plant Physiol, 2012. 159(1): p. 321-35.

Jandu, N., et al., *Enterohemorrhagic Escherichia coli O157:H7 gene expression profiling in response to growth in the presence of host epithelia.* PLoS One, 2009. 4(3): p. e4889.

Jans, J., et al., *A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome.* Genes Dev, 2009. 23(5): p. 602-18.

Kirienko, N.V. and D.S. Fay, *Transcriptome profiling of the C. elegans Rb ortholog reveals diverse developmental roles.* Dev Biol, 2007. 305(2): p. 674-84.

Laubacher, M.E. and S.E. Ades, *The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance.* J Bacteriol, 2008. 190(6): p. 2065-74.

Le Gac, M., et al., *Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli.* Proc Natl Acad Sci U S A, 2012. 109(24): p. 9487-92.

Li, C.M. and R.R. Klevecz, *A rapid genome-scale response of the transcriptional oscillator to perturbation reveals a period-doubling path to phenotypic change.* Proc Natl Acad Sci U S A, 2006. 103(44): p. 16254-9.

MacNeil, L.T., et al., *Diet-induced developmental acceleration independent of TOR and insulin in C. elegans.* Cell, 2013. 153(1): p. 240-52.

Mensa, B., et al., *Antibacterial mechanism of action of arylamide foldamers.* Antimicrob Agents Chemother, 2011. 55(11): p. 5043-53.

Nozue, K., S.L. Harmer, and J.N. Maloof, *Genomic analysis of circadian clock-, light-, and growth-correlated genes reveals PHYTOCHROME-INTERACTING FACTOR5 as a modulator of auxin signaling in Arabidopsis.* Plant Physiol, 2011. 156(1): p. 357-72.

Pietsch, K., et al., *Meta-Analysis of Global Transcriptomics Suggests that Conserved Genetic Pathways are*

Responsible for Quercetin and Tannic Acid Mediated Longevity in C. elegans. Front Genet, 2012. 3: p. 48.

Potrykus, K., et al., *Imprecise transcription termination within Escherichia coli greA leader gives rise to an array of short transcripts, GraL.* Nucleic Acids Res, 2010. 38(5): p. 1636-51.

Reeves, P.H., et al., *A regulatory network for coordinated flower maturation.* PLoS Genet, 2012. 8(2): p. e1002506.

Rossouw, D. and F.F. Bauer, *Comparing the transcriptomes of wine yeast strains: toward understanding the interaction between environment and transcriptome during fermentation.* Appl Microbiol Biotechnol, 2009. 84(5): p. 937-54.

Rossouw, D., et al., *Comparative transcriptomic and proteomic profiling of industrial wine yeast strains.* Appl Environ Microbiol, 2010. 76(12): p. 3911-23.

Rossouw, D., et al., *Comparative transcriptomic approach to investigate differences in wine yeast physiology and metabolism during fermentation.* Appl Environ Microbiol, 2009. 75(20): p. 6600-12.

Rossouw, D., T. Naes, and F.F. Bauer, *Linking gene regulation and the exo-metabolome: a comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast.* BMC Genomics, 2008. 9: p. 530.

Ruckle, M.E., et al., *Plastids are major regulators of light signaling in Arabidopsis.* Plant Physiol, 2012. 159(1): p. 366-90.

Rugnone, M.L., et al., *LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator.* Proc Natl Acad Sci U S A, 2013. 110(29): p. 12120-5.

Seita, J., et al., *Gene Expression Commons: an open platform for absolute gene expression profiling.* PLoS One, 2012. 7(7): p. e40321.

Sharma, A.K., et al., *Comparative transcriptomic profile analysis of fed-batch cultures expressing different recombinant proteins in Escherichia coli.* AMB Express, 2011. 1(1): p. 33.

Simmons Kovacs, L.A., et al., *Cyclin-dependent kinases are regulators and effectors of oscillations driven by a transcription factor network.* Mol Cell, 2012. 45(5): p. 669-79.

Stevens, S.L., et al., *Multiple preconditioning paradigms converge on interferon regulatory factor-dependent signaling to promote tolerance to ischemic brain injury.* J Neurosci, 2011. 31(23): p. 8456-63.

Taher, L., et al., *Global gene expression analysis of murine limb development.* PLoS One, 2011. 6(12): p. e28358.

Thorrez, L., et al., *Tissue-specific disallowance of housekeeping genes: the other face of cell differentiation.* Genome Res, 2011. 21(1): p. 95-105.

Uno, M., et al., *A fasting-responsive signaling pathway that extends life span in C. elegans.* Cell Rep, 2013. 3(1): p. 79-91.

Vartanian, K.B., et al., *LPS preconditioning redirects TLR signaling following stroke: TRIF-IRF3 plays a seminal role in mediating tolerance to ischemic injury.* J Neuroinflammation, 2011. 8: p. 140.

Von Stetina, S.E., et al., *Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C. elegans nervous system.* Genome Biol, 2007. 8(7): p. R135.

Xu, K., et al., *A genome-wide transcriptome profiling reveals the early molecular events during callus initiation in Arabidopsis multiple organs.* Genomics, 2012. 100(2): p. 116-24.

Xue, L., et al., *Global expression profiling reveals genetic programs underlying the developmental divergence between mouse and human embryogenesis.* BMC Genomics, 2013. 14: p. 568.

Yanai, I., et al., *Pairing of competitive and topologically distinct regulatory modules enhances patterned gene expression.* Mol Syst Biol, 2008. 4: p. 163.

Chapter 7

General discussion

Lotte B. Westerhof

The development and application of biopharmaceuticals has benefited from the revolution in life sciences that took place over the last three decades. Today many biopharmaceuticals have been approved for medical use. Most are produced in heterologous production systems that make use of the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae* or mammalian cell lines. All production systems have their advantages and disadvantages and the choice of system depends on requirements of the produced protein. As the cost of biopharmaceuticals is becoming increasingly important, plants may provide an alternative as they combine the advantage of *E. coli* being economic with the advantage of mammalian cells being able to fold, assemble and glycosylate complex proteins. This thesis describes the production of a variety of proteins *in planta* and focuses on increasing yield and biological activity. In this chapter parameters influencing yield and activity, i.e. transformation, transcription, translation and protein behaviour, will be discussed.

Transformation efficiency and transcription

The promoter of a gene predominantly determines the level of its transcription. However, in heterologous expression not only the choice of the promoter, but also the manner of transformation has a profound influence on the transcript level of the gene(s) of interest. The two aspects that are relevant in this matter are the number of gene copies available for transcription and whether or not the gene is integrated into the genome.

Transient transformation often results in higher expression levels than stable transformation. Even when both transient and stable transformation are mediated by *Agrobacterium* and when a binary vector system is used [1]. Stable transformation results in the integration of one or a few copies of the gene into the plant genome. Subsequently, transcription depends on the type of promoter as well as its accessibility. The promoter will be readily accessible if the gene is integrated into a part of the genome that is actively used for transcription. This is more often the case with *Agrobacterium*-mediated transformation compared to indirect transformation methods. However, transcription does not depend on genomic integration. As long as the expression cassette is double stranded, transcription can occur. Although the T-DNA fragment that *Agrobacterium* transfers to the plant cell is single stranded a double strand is formed to allow genomic integration. From the moment that the T-DNA is double stranded, transcription is possible [2-4]. *Agrobacterium* transfers many more T-DNA copies to the plant cell than are eventually integrated. T-DNAs that are not integrated into the genome are eventually broken down, explaining the transient nature of initial expression [4]. Thus, in *Agrobacterium*-mediated transient expression many more copies of T-DNA are transcriptionally active independent of integration. This results in higher transcript levels when compared to stable transformation.

High copy numbers also explain the success of transient expression using viral vectors. Upon using a viral vector, transcription of the gene of interest is also not limited by genomic integration. And as the virus replicates and spreads throughout the plant the copy number of the gene of interest increases. An example of a superior transient expression system using viral vectors is the magnICON® system [5]. This system is based on the use of two vectors each carrying (parts of) the genome of the noncompeting viruses tobacco mosaic virus and potato virus X. This system is suited for the expression of heterodimeric proteins, such as antibodies, and has frequently resulted in high yields.

Before the research described in this thesis, heterologous gene expression in plants was mainly used in our laboratory to investigate the role of resistance genes and virulence factors. For these studies, high expression levels are usually not preferred. *Agrobacterium*-mediated transient transformation using a binary vector system with pBINPLUS [6] as the expression vector was always sufficient to facilitate gene expression in plants for this purpose. Upon our first effort to produce interleukin (IL)-10 in plants we achieved a very low yield. Initially, we attributed this to IL-10's supposed inherent instability. However, when we needed another antibiotic resistance gene for stable transformation we obtained another expression vector known as pMDC32 [7]. To suit our cloning needs this vector was modified and re-named pHYG for its hygromycin resistance gene (Chapter 2). We noticed that pHYG has a higher copy number in *E. coli*. Because of this and the smaller size of pHYG (~3 kbp difference) it became easier to manipulate expression cassettes at this stage. Therefore we started to use this vector for our transient transformations. To our surprise protein yield also increased tremendously using pHYG in comparison to pBINPLUS. As we used identical expression cassettes, we concluded this must be due to a higher copy number of pHYG in *Agrobacterium* as well, resulting in more T-DNA copies being transferred to the plant cell.

pBINPLUS carries two origins of replication (ori); one originated from the plasmid pRK2 of *Klebsiella aerogenes* (RK2) and one from the plasmid pMB1 of *Bifidobacterium longum* (pMB1 is the parental plasmid of pBR322 and its ori is known as ColE1). ColE1 was added to boost replication in *E. coli*, however, this was not sufficient to facilitate easy cloning in this vector [6]. pHYG is a derivative of pCambia1300 and also carries ColE1 together with the ori isolated from the plasmid pVS1 of *Pseudomonas aeruginosa* (VS1). VS1 does not facilitate replication in *E. coli*. Thus, although pHYG uses ColE1 and pBINPLUS uses ColE1 and RK2 to facilitate replication in *E. coli*, pHYG has a higher copy number in *E. coli* than pBINPLUS. This may be due to the smaller size of pHYG and/or the replication facilitated by ColE1 is limited when combined with RK2 in pBINPLUS. As mentioned, we suspect that pHYG also has a higher copy number in *Agrobacterium*. Again, this may be due to the smaller size of pHYG and/or the possibility that VS1 leads to higher replication levels in *Agrobacterium* when compared to RK2.

mRNA transcript levels of human and mouse interleukin (IL)-10 that were expressed using pHYG exceeded the mRNA transcript level of the household gene actin by ~1000-fold

(Chapter 2). When we expressed three IgG and three IgA antibodies average yield was ~50 µg antibody per milligram (or 5% of) total soluble protein (TSP) (Chapter 3). This translates to ~1 g of antibody per kg fresh weight. Similar expression levels were described by Giritch and co-workers using the magnICON® system [8]. Thus, although we are still using an 'old school' binary vector system for expression, we concluded that transformation efficiency and subsequent transcription were not limiting with regard to protein yield using our plant expression system.

Transcript half-life

The use of the pHYG vector had improved yield of human and mouse IL-10. Next, we further improved yield of human IL-10 30-fold by preventing its extensive multimerization. Thereto we inserted a small glycine-serine linker between helices E and D of IL-10, which allowed IL-10 to fold into itself abolishing the need to dimerise. As a result the yield of human IL-10 increased to ~2 µg/mg TSP, which was comparable to the yield of mouse IL-10 (Chapter 2). Nevertheless, this was still ~25-fold less compared to the yield of IgG and IgA (Chapter 3). We concluded that the difference between IL-10 and antibody yield are most likely not caused by differences in transformation efficiency or level of transcription. Therefore differences in transcript half-life, translation efficiency and/or protein half-life must exist. First we will evaluate the impact of transcript half-life.

Thanks to many studies on mRNA decay we now know that *in vivo* mRNA half-life is connected to gene function whereby transcripts of proteins that have to be tightly regulated have generally short half-lives and transcripts of household genes have generally longer half-lives [9-11]. The features that predominantly determine mRNA half-life are an intron splicing event, sequence elements in the 3' UTR and the presence of microRNA targets [9]. In heterologous expression most of these features are controlled. The expression cassette determines the UTRs a transcript receives and a choice can be made regarding the use of introns. Sequences that are targets for microRNAs can be present and vary from gene to gene. Binding of microRNAs and overexpression of a gene can trigger gene silencing. To prevent gene silencing the co-expression of a viral silencing inhibitor is often used in transient expression [12]. A less potent variant of the silencing inhibitor p19 of tomato bushy stunt virus can even be considered in stable transformation [13].

Initially we did not use a viral silencing inhibitor when expressing IL-10, while we did when expressing antibodies. In chapter 6 we describe the expression of mouse IL-10 with p19, which boosted protein yield significantly at 5 dpi (~1 and ~17 µg IL-10 per mg TSP without and with p19, respectively). However, this was not significantly higher when compared to mouse IL-10 yield on dpi 3 without p19 (~15 µg IL-10 per mg TSP). Strikingly, mouse IL-10

yields were overall much higher in this experiment than previously reported in chapter 2 (~2 µg IL-10 per mg TSP on dpi 3 without p19). The only difference was the use of a non-native signal peptide derived from the *Arabidopsis thaliana* chitinase gene in this mouse IL-10 construct during the later experiments. Although, improper processing of the native signal peptide was not suspected [14], a different signal peptide may change the translation initiation efficiency, which will be discussed in the section 'Translation efficiency'. Differences in yield may also be due to the fickle nature of transient expression. We found that protein yield in biological replicates of experiments performed in consecutive weeks is often similar when plants and leaves of the same age are used. However, protein yield can differ ~2 to 10-fold when performed at different times of the year as a consequence of seasonal effects (unpublished data). This illustrates the difficulty with drawing conclusions about protein yield upon heterologous expression. Biological replicates can best be done during the same season using as identical as possible experimental conditions. These findings also illustrate the difficulty of comparing results between different studies, as conditions will almost certainly differ. In any case, we could conclude that the transcript of mouse IL-10 was sensitive to silencing.

The calculated minimal free folding energy of mRNA, which predicts physical stability, has not been linked to mRNA half-life *in vivo*. However, it may well be that mRNA stability plays a significant role in determining mRNA half-life in heterologous expression. In fact, a recent publication showed that removal of instable mRNA sequences reduced transcript degradation significantly upon stable heterologous expression of isomaltulose in sugarcane [15]. However, whether or not differences in mRNA stability result in differences in protein yield upon transient transformation, whereby mRNA transcript levels are generally much higher, remains to be investigated. In a study whereby 154 variants of GFP differing only in synonymous codon use were expressed in *E. coli*, overall mRNA folding energy was not significantly correlated to protein yield [16]. Yet, this does not prove that one or more regions of poor stability that reside in the transcript of an overall stable mRNA do not affect mRNA half-life. When we expressed native and codon optimised variants of GFP, chicken ovalbumin (OVA) and mouse IL-10 upon stable transformation in *Arabidopsis thaliana* and upon transient transformation in *N. benthamiana* co-expressed with p19, the optimised variants resulted in higher protein yields in all cases (Chapter 6). Transcript levels were also determined in the stable *A. thaliana* transformants and demonstrated that the increase in protein yield upon codon optimisation could be partly explained by an increased concentration of transcript. Because gene silencing in stable transformation leads to a complete absence of transcripts and recombinant protein, differences in gene silencing could not explain the differences in transcript levels. Thus, increased mRNA stability may result in increased transcript half-lives. Yet, while the mRNA transcripts of GFP and OVA had lower folding energies (more stable mRNA) upon codon optimisation, the mouse IL-10 mRNA transcript did not. As previously mentioned, it is still possible that unstable regions of IL-10 were removed upon codon

optimisation, while the overall stability decreased. Furthermore, not only transcript levels had increased. Also the amount of protein per mRNA transcript was higher for the optimised variants. This can only be explained by the fact that codons optimisation increased translation efficiency. An increase in translation efficiency may have also influenced transcript levels. It has been demonstrated that ribosomes can stall or slip causing a frame-shift and both can lead to the degradation of the mRNA transcript and its product. All in all, while mRNA stability is not linked to mRNA half-life *in vivo*, mRNA stability most likely has significant influence on mRNA half-life and thus protein yield in heterologous protein production.

Translation efficiency

A higher protein yield per mRNA molecule can only be explained by an increase in translation rate. The efficiency of translation initiation and the speed of translation itself both determine translation rate. Ribosomal density studies have shown that more ribosomes are found at the beginning of a transcript compared to the rest of the transcript [17, 18]. This observation has lead to the assumption that translation initiation is the rate-limiting step of translation. A high folding energy of the 5' end of a mRNA is believed to facilitate efficient translation initiation as this is a general feature of mRNA structures [19].

As mentioned, when mouse IL-10 was expressed with a different signal peptide, yield was found to be ~7,5-fold higher, which may be due to an increase in translation initiation. When we calculated the folding energy of the first 75 nucleotides (positions -36 to +39) of both the mouse IL-10 transcript with the native or *A. thaliana* chitinase signal peptide, we found that the latter signal peptide increased the folding energy 7,7-fold (making it less stable). An increased translation initiation may therefore explain the yield increase of mouse IL-10.

Even though translation initiation may be the rate-limiting step of translation, an increase in translation rate may still influence protein abundance in a situation where the number of free ribosomes is limiting. An increased translation rate will make translating ribosomes faster available for a next round of translation. Especially in heterologous expression, whereby the gene of interest is significantly over-expressed (with mRNA transcript levels often several orders of magnitude above household genes; Chapter 2), it is likely that free ribosomes are limiting. An increased translation rate of a heterologous gene can therefore also increase protein yield significantly through a faster 'cycling' of ribosomes.

It is assumed that translation rate is predominantly controlled by tRNA availability and the mRNA structure. Availability of a cognate tRNA determines the speed with which a codon is translated because translation occurs via diffusion and arrival of the cognate tRNA is the rate-limiting step for elongation of the amino acid sequence [20]. In addition strong mRNA structures have been shown to hamper translation [21, 22], because ribosomes use their

intrinsic helicase activity [23] to dissociate the nucleotide bonds that occur during mRNA folding before translation can progress. Thus, a less stable mRNA structure can be translated faster. However, there must be a trade off between mRNA translatability and stability, because native mRNA sequences have a lower folding energy (more stable mRNA) than randomized mRNA sequences with the same base composition and length [24].

Highly expressed genes are biased in their codon use and the codons they use most frequently (the so-called optimal codons) correspond with genomic G+C content [25] and have been matched to abundant tRNAs in many species [26-31]. In heterologous protein production codon optimisation of the gene of interest using optimal codons of their production host has shown great variability of success [15]. Optimal codons differ between species, which is in line with differences in G+C content and abundance of tRNA species between species. However, it was striking that a general codon bias was found across plant species, including monocots and dicots that differ significantly in their G+C content [32]. We designed a strategy making use of codons that are increased most with expression across plants (expression codons). Using the same coding sequence for the signal peptide that preceded all genes we ensured that translation initiation of genes is the same. This codon optimisation strategy resulted in higher transcript levels and higher protein yield, which we believe to be a result of an improved mRNA structure that balances mRNA stability with translatability. Because the translation machinery is quite conserved across kingdoms of life we evaluated whether this general codon bias found in plants could also be extended to other kingdoms. Using large-scale gene expression and protein abundance data we observed that this general codon bias exists across kingdoms and conclude that codons in highly expressed genes are selected to facilitate a more stable mRNA structure (more and different type of nucleotide bonds) that is also sufficiently 'airy' (higher frequency of alternating bound (stems) and unbound nucleotides (loops)) to allow efficient translation. Because this codon use trend exists across kingdoms, the strategy of employing expression codons to optimise genes can be used to enhance protein yield in virtually any production platform.

Protein half-life

Transient expression of a codon optimised variant of mouse IL-10 without p19 resulted in a yield of 29 µg IL-10 per mg TSP on 2 dpi (Chapter 6). This has been the highest yield we have been able to obtain for mouse IL-10. Considering that mouse IL-10 is ~8 times smaller compared to an antibody, this level of expression in molar terms is even higher compared to what we obtained for antibodies (Chapter 3; ~50 µg IgG per mg TSP). Regardless of the expression strategy used, the peak of protein accumulation for interleukin-10 was always between 2 and 4 days post infiltration (dpi) (Chapter 2 and 6), whereas the peak of

accumulation of the antibodies was between 6 and 9 dpi (Chapter 3). This may be explained by the differences in protein half-life. It is expected that a less stable protein has its peak in expression earlier when compared to a more stable protein. Expression in a transient system is lost over time. Consequently, a more stable protein can accumulate for a longer period before the level of degradation becomes higher than the level of production as compared to an unstable protein. Thus, *in planta* interleukin-10 half-life is likely shorter than the half-life of antibodies. This is in line with their *in vivo* half-lives. Recombinant interleukin-10 is cleared from human blood within 2.5 hours after intravenous administration [33], while recombinant IgG1 could still be detected after 30 days [34]. IL-10 yield may be improved further by increasing protein stability. Fusion of inherently unstable proteins to more stable proteins has been reported to increase expression levels [35, 36]. The fusion of IL-10 to the Fc portion of IgG increased its *in vivo* half-life to 30 hours [37]. However, on dpi 2 or 3 we did not observe an increase in human IL-10 yield when we fused it to the Fc portion of IgA (Chapter 2). A fusion partner can aid in protein folding by increasing interaction with chaperones and protect against proteolytic degradation and ubiquitination. However, in our case the turnover of IL-10 did not change. It is possible that IL-10 is not unstable in the sense that it is actively targeted for degradation, but merely quickly unfolds and is subsequently degraded. If this is the case, IL-10 yield could only be improved by modifying intramolecular interactions, disulphide bridges, hydrophobicity and/or charge of the protein. It would be a tremendous challenge to achieve this without disrupting biological activity or increasing immunogenicity.

Another striking observation we made concerning antibody yield was that we found little variation in yield between idiotypes. This was unexpected as many reports on plant expression of antibodies exist and demonstrate a variation in yield of several orders of magnitude [38]. Although our experience is limited to two isotypes and three idiotypes, this suggests that the expression system determines yield to a greater extent than protein intrinsic properties. We used the same coding sequence for the signal peptide that preceded all genes and used expression codons to encode the variable regions. The native sequences of the constant domains of the human kappa chain, alpha and gamma heavy chains were used, as these already used expression codons at high frequency. Thus, in our case mRNA stability, translation initiation and translation rate between the antibodies should be comparable, leaving protein intrinsic properties as the main variables for yield.

A unique opportunity to shed more light on variability in antibody yield lies in the hands of the company Icon Genetics that expresses antibodies for treatment of non-Hodgkin lymphoma. They will express a unique antibody idiotype for each patient and could establish a database with expression data of these antibodies. Comparing the expression of idiotypes could help determine the variation caused by the variable domains and link codon use of the variable regions or protein intrinsic properties, such as hydrophobicity, ubiquitination and amino acid content/sequences, to protein yield. Such findings may be translated to other proteins and may result in insight into possibilities to enhance protein stability and quality.

Protein secretion

It has often been suggested that plants are poor protein secretors, likely based on the fact that plants do not secrete many endogenous proteins. We found that secretion efficiency varied between antibodies. Evaluation of the secretion efficiency of three idiotypes on both an IgG and IgA backbone revealed that secretion was reduced for one of the idiotypes and for IgA in general (Chapter 3). The reduction of secretion of IgA may be explained by the presence of a cryptic signal peptide in the tailpiece of the alpha heavy chain that leads to vacuolar targeting. The lack of secretion based on the idiotype can possibly be attributed to improper protein folding. This antibody was enriched in immature N-glycans (high mannose structures), an indication of protein misfolding. Misfolded proteins do not leave the ER and therefore do not undergo maturation of N-glycans in the Golgi. To our surprise the secretion of the *Schistosoma mansoni* egg antigen omega-1 was estimated to be over 90% (Chapter 5). This was in sharp contrast to the secretion of IgG, which was maximum 28%, while the *A. thaliana* chitinase signal peptide was used for expression of all antibody genes as well as omega-1. While cryptic subcellular targeting sequences or improper processing may explain some of the variation in secretion efficiency, one or more other intrinsic protein properties, such as hydrophobicity or charge, must also play a role.

Judging by the efficiency of omega-1 secretion, plants are not per definition poor protein secretors. Increased knowledge on secretion efficiency may provide a solution for purification, which is one of the most important bottlenecks for plant-produced heterologous proteins. Upon extraction of the apoplastic proteins from *N. benthamiana* leaves expressing omega-1 more than 50% of the isolated proteins was omega-1 as revealed by SDS-PAGE. Apoplast extracts are therefore a great basis for purification, facilitating the use of robust traditional purification methods such as ion exchange chromatography that do not rely on the use of protein tags that may hinder protein activity.

N-glycosylation

Not only protein intrinsic properties that influence secretion are largely unknown, but also those properties underlying differences in N-glycosylation. Although it is a well-known fact that the N-glycan type a protein receives depends on species, tissue, environment and developmental stage, it can also differ from protein to protein under identical circumstances. The predominant N-glycan types on plant-produced IgG, IgA and omega-1 all differed from each other while they were all expressed in 5-6 weeks old *N. benthamiana* leaves. The typical N-glycan found on plant proteins is biantennary with terminal *N*-acetylglucosamine (GlcNAc or Gn) residues and the typical plant sugar residues β1,2-xylose (X) and core α1,3-fucose (F),

abbreviated as GnGnXF^3 [39]. This was found to be the predominant N-glycan type on IgG (Chapter 3). The second most common N-glycan in plants lacks the terminal GlcNAcs and is called paucimannosidic and is abbreviated to MMXF^3 [39]. This was found to be the predominant N-glycan type on omega-1 (Chapter 5). The predominant N-glycan on IgA lacked one terminal GlcNAcs and the core α1,3-fucose (GnMX or MGnX) (Chapter 3), a glycan type that occurs in marginal percentages in plants [39], and was never found to be predominant on a particular protein. The absence of the α1,3-fucose may be explained by the fact that the core of the N-glycans on IgA may not be accessible for core altering glycosyltransferases. Petrescu and colleagues (2004) analysed the protein environment of N-glycosylation sites of a wide variety of glycoproteins and demonstrated that many N-glycan sites are poorly accessible [40]. This could explain why less than half of the N-glycans on human serum IgA carry a typical mammalian core α1,6-fucose [41, 42], whereas 80-98% of N-glycans on human serum IgG antibodies are core fucosylated [43].

Thus, accessibility may explain variation in the presence of sugar residues that are attached to the N-glycan core. The cause of the variation in the presence of terminal GlcNAcs is more elusive. Three β-*N*-acetylhexosaminidases (HEXO1-3) were identified in *Arabidopsis thaliana* and their enzymatic activity explains the occurence of paucimannosidic N-glycan in this plant species [44]. In total 7 genomic sequences are present in *N. benthamiana* that have significant homology to HEXO1-3 [45]. It is therefore likely that also the paucimannosidic N-glycan fraction in *N. benthamiana* is the result of β-*N*-acetylhexosaminidase activity and is not caused by a lack of GlcNAc attachment. Upon transient expression of *A. thaliana* HEXO-fluorescent protein fusions in *N. benthamiana* HEXO1 localized to the vacuole and HEXO2/3 to the plasma membrane [46]. This could imply that only vacuolar and secreted proteins are modified to bear paucimannosidic N-glycans. It could well be that the lack of one of the GlcNAc residues of the predominant N-glycan of IgA may be the result of β-*N*-acetylhexosaminidase activity in either the vacuole or the apoplast. *A. thaliana* HEXO1-3 were all demonstrated to remove the terminal GlcNAc residues from both α1,3- and α1,6-branched mannoses without strict preference. However, *N. benthamiana* may harbour a β-*N*-acetylhexosaminidase that does have branch specificity. Alternatively, attachment or removal of terminal GlcNAc residues of N-glycans on IgA is inefficient and the predominant N-glycan form found is an intermediate.

The efficient secretion of omega-1 may explain the lack of both GlcNAc residues on the predominant N-glycan of omega-1. Nevertheless, another *Schistosoma mansoni* egg antigen expressed in *N. benthamiana* in our laboratory demonstrated similar secretion efficiency as omega-1 and carried N-glycans of the GnGnXF^3 type (Wilbers, unpublished data). Thus, next to subcellular targeting other unknown protein intrinsic properties must influence GlcNAc removal.

Intrinsic protein properties may not only influence the type of N-glycan a protein receives, but may also influence N-glycosylation efficiency. N-glycosylation of a specific site is not always 100% efficient. We observed that the N-glycosylation site found in the joining chain and the tailpiece of the alpha heavy chain do not always receive a N-glycan when expressed in *N. benthamiana* (Chapter 3 and 4). Because N-glycosylation is co-translational, limited access to the N-glycosylation site cannot explain inefficient N-glycosylation. However, in a large-scale analysis of glycoproteins it was demonstrated that the glycosylation signals N-X-T and N-X-S are N-glycosylated in 70% and 30% of the cases, respectively [40]. Also, the amino acid sequence requirements around the glycosylation sites were less biased around N-X-T sites as compared to N-X-S sites [40]. This suggests that N-X-T sites are more efficiently glycosylated compared to N-X-S sites, perhaps through improved site recognition. Both the N-glycosylation site in the tailpiece of the alpha heavy chains and the joining chain are of the N-X-S type. Mutation of the serine to a threonine in N-glycosylation sites and/or mutations in the surrounding amino acids may increase N-glycosylation efficiency. Whether or not differences in N-glycosylation efficiency depend on production platform has not been extensively investigated. More knowledge on the protein intrinsic properties that influence N-glycosylation efficiency and the N-glycan type is valuable for the field of glyco-engineering, notably when combined with the production of biopharmaceuticals.

Concluding remarks

Protein intrinsic properties that hamper yield, secretion or N-glycosylation efficiency limit protein yield and quality in all platforms and deserve to be the main focus of research into heterologous protein production. Most production platforms, including plants, have most likely reached their limit in terms of transformation and transcription efficiency. However, an increase in translation efficiency may still increase protein yield in some cases. Transient expression in *N. benthamiana* is still unmatched in speed, ease of transformation and low costs. Nevertheless, it provides the same advantages as other eukaryotic production platforms. It is an excellent platform to study these properties and apply improvements to better produce interesting (glyco)proteins for research and possibly clinical purposes. When purification from plants is improved and when regulatory hurdles are taken it will be a matter of time before *N. benthamiana* and other plant species become an established production platform for biopharmaceutical proteins.

References

1. Wroblewski, T., A. Tomczak, and R. Michelmore, *Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis.* Plant Biotechnology Journal, 2005. 3(2): p. 259-73.
2. Narasimhulu, S.B., et al., *Early transcription of Agrobacterium T-DNA genes in tobacco and maize.* Plant Cell, 1996. 8(5): p. 873-86.
3. Tzfira, T., et al., *Site-specific integration of Agrobacterium tumefaciens T-DNA via double-stranded intermediates.* Plant Physiology, 2003. 133(3): p. 1011-23.
4. Janssen, B.J. and R.C. Gardner, *Localized transient expression of GUS in leaf discs following cocultivation with Agrobacterium.* Plant Molecular Biology, 1990. 14(1): p. 61-72.
5. Roberts-Thomson, I.C., et al., *Cells, cytokines and inflammatory bowel disease: a clinical perspective.* Expert Review of Gastroenterology and Hepatology, 2011. 5(6): p. 703-16.
6. van Engelen, F.A., et al., *pBINPLUS: An improved plant transformation vector based on pBIN19.* Transgenic Research, 1995. V4(4): p. 288-290.
7. Curtis, M.D. and U. Grossniklaus, *A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta.* Plant Physiology, 2003. 133(2): p. 462-469.
8. Giritch, A., et al., *Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors.* PNAS, 2006. 103(40): p. 14701-14706.
9. Narsai, R., et al., *Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana.* Plant Cell, 2007. 19(11): p. 3418-3436.
10. Friedel, C.C., et al., *Conserved principles of mammalian transcriptional regulation revealed by RNA half-life.* Nucleic Acids Research, 2009. 37(17): p. e115-.
11. Bernstein, J.A., et al., *Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays.* PNAS, 2002. 99(15): p. 9697-9702.
12. Voinnet, O., et al., *An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus.* Plant Journal, 2003. 33(5): p. 949-56.
13. Saxena, P., et al., *Improved foreign gene expression in plants using a virus-encoded suppressor of RNA silencing modified to be developmentally harmless.* Plant Biotechnology Journal, 2011. 9(6): p. 703-12.
14. Bortesi, L., et al., *Viral and murine interleukin-10 are correctly processed and retain their biological activity when produced in tobacco.* BMC Biotechnology, 2009. 9(1): p. 22.
15. Jackson, M.A., et al., *Design rules for efficient transgene expression in plants.* Plant Biotechnol J, 2014.
16. Kudla, G., et al., *Coding-Sequence Determinants of Gene Expression in Escherichia coli.* Science, 2009. 324(5924): p. 255-258.
17. Arava, Y., et al., *Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae.* PNAS, 2003. 100(7): p. 3689-3894.
18. Ingolia, N.T., et al., *Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling.* Science, 2009. 324(5924): p. 218-223.
19. Tuller, T., et al., *Translation efficiency is determined by both codon bias and folding energy.* PNAS, 2009. 107(8): p. 3645-3650.

20. Varenne, S., et al., *Translation is a Non-uniform Process; Effect of tRNA availability on the Rate of Elongation of Nascent Polypeptide Chains.* Journal of Molecular Biology, 1984. 180: p. 549-576.
21. Landick, R. and C. Yanofsky, *Stability of an RNA Secondary Structure Affects in Vitro Transcription Pausing in the trp Operon Leader Region.* The Journal of Biological Chemistry, 1984. 259: p. 11550-11555.
22. Qu, X., et al., *The ribosome uses two active mechanisms to unwind messenger RNA during translation.* Nature, 2011. 475: p. 118-121.
23. Takyar, S., R.P. Hickerson, and H.F. Noller, *mRNA Helicase Activity of the Ribosome.* Cell, 2005. 120: p. 49-58.
24. Seffens, W. and D. Digby, *mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences.* Nucleic Acids Research, 1999. 27(7): p. 1578-1584.
25. Hershberg, R. and D.A. Petrov, *General rules for optimal codon choice.* PLoS Genetics, 2009. 5(7): p. e1000556.
26. Post, L.E., et al., *Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1979. 76(4): p. 1697-1701.
27. Powell, J.R. and E.N. Moriyama, *Evolution of codon usage bias in Drosophila.* PNAS, 1997. 94(15): p. 7784-7790.
28. Ikemura, T., *Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes.* Journal of Molecular Biology, 1981. 146(1): p. 1-21.
29. Stenico, M., A.T. Lloyd, and P.M. Sharp, *Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases.* Nucleic Acids Research, 1994. 22(13): p. 2437-46.
30. Wright, S.I., et al., *Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata.* Mol Biol Evol, 2004. 21(9): p. 1719-26.
31. Kanaya, S., et al., *Codon Usage and tRNA Genes in Eukaryotes: Correlation of Codon Usage Diversity with Translation Efficiency and with CG-Dinucleotide Usage as Assessed by Multivariate Analysis.* Journal of Molecular Evolution, 2001. 53(4): p. 290-298.
32. Wang, L. and M. Roossinck, *Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants.* Plant Molecular Biology, 2006. 61(4): p. 699-710.
33. van Deventer, S.J., C.O. Elson, and R.N. Fedorak, *Multiple doses of intravenous interleukin 10 in steroid-refractory Crohn's disease. Crohn's Disease Study Group.* Gastroenterology, 1997. 113(2): p. 383-9.
34. Mankarious, S., et al., *The half-lives of IgG subclasses and specific antibodies in patients with primary immunodeficiency who are receiving intravenously administered immunoglobulin.* Journal of Laboratory and Clinical Medicine, 1988. 112(5): p. 634-40.
35. Sainsbury, F., et al., *Tomato cystatin SlCYS8 as a stabilizing fusion partner for human serpin expression in plants.* Plant Biotechnology Journal, 2013. 11(9): p. 1058-1068.
36. Jang, I.C., et al., *Enhancing protein stability with retained biological function in transgenic plants.* Plant Journal, 2012. 72(2): p. 345-54.
37. Zheng, X., et al., *Administration of noncytolytic IL-10/Fc in murine models of lipopolysaccharide-induced septic shock and allogeneic islet transplantation.* The Journal of Immunology, 1995. 154(10): p. 5590-5600.

38. De Muynck, B., C. Navarre, and M. Boutry, *Production of antibodies in plants: status after twenty years.* Plant Biotechnology Journal, 2010. 8(5): p. 529-63.
39. Strasser, R., et al., *Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure.* Plant Biotechnology Journal, 2008. 6(4): p. 392-402.
40. Petrescu, A.J., et al., *Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding.* Glycobiology, 2004. 14(2): p. 103-114.
41. Mattu, T.S., et al., *The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc alpha receptor interactions.* Journal of Biological Chemistry, 1998. 273(4): p. 2260-2272.
42. Field, M.C., et al., *Structural analysis of the N-glycans from human immunoglobulin A1: comparison of normal human serum immunoglobulin A1 with that isolated from patients with rheumatoid arthritis.* Biochemical Journal, 1994. 299 (Pt 1): p. 261-75.
43. Pucic, M., et al., *High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations.* Molecular & Cellular Proteomics, 2011. 10(10).
44. Liebminger, E., et al., *beta-N-Acetylhexosaminidases HEXO1 and HEXO3 Are Responsible for the Formation of Paucimannosidic N-Glycans in Arabidopsis thaliana.* Journal of Biological Chemistry, 2011. 286(12): p. 10793-10802.
45. Bombarely, A., et al., *A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research.* Molecular Plant Microbe Interactions, 2012. 25(12): p. 1523-30.
46. Strasser, R., et al., *Enzymatic properties and subcellular localization of Arabidopsis beta-N-acetylhexosaminidases.* Plant Physiology, 2007. 145(1): p. 5-16.

Summary

Current treatments of inflammatory disorders are often based on therapeutic proteins. These proteins, so-called biopharmaceuticals, are isolated from a natural resource or, more often, made using cell based fermentation systems. The most common production platforms are based on the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae* or mammalian cell lines (mainly Chinese hamster ovarian (CHO) and murine myeloma (SP2/0) cells). Each platform has advantages and disadvantages and the protein to be produced largely dictates the choice of platform. Plants could provide a unique alternative production platform, as they combine the advantage of *E. coli* being economic with the advantages of mammalian cell lines being able to fold complex proteins, assemble heteromultimeric protein complex and, upon glyco-engineering, provide proteins with human type N-glycans. Furthermore, plant production is easily scaled up as the infrastructure is already in place due to our need for food and feed and plants have a limited risk of contamination with human pathogens. Transient transformation of *N. benthamiana* is a valuable plant production platform, as it is unmatched in terms of speed (matter of weeks).

This thesis describes the production of a variety of proteins and protein complexes *in planta* that are or may be used as biopharmaceuticals. Since 1982, plants can be genetically manipulated, which has lead to the production of many proteins in a variety of plants. Initially, the plants greatest drawback was limited protein yield. However, significant increases in yield have been achieved in the last two decades, predominantly by increasing transformation efficiency and/or level of transcription. Nowadays several plant expression systems exist that facilitate high protein production levels. Most experience is based on the production of antibodies, mainly of the IgG isotype, as these are often used as biopharmaceuticals. In general, IgG antibodies are produced on a scale of several grams per kilogram fresh weight. However, production levels of antibodies have shown to be variable and production levels of particular proteins, such as cytokines, have lagged behind. Cytokines form a large group of immune-signalling molecules and several cytokines have promising therapeutic potential. Their short *in vivo* half-life suggests an inherent instability, which is regarded the major production bottleneck.

In chapter 2 we describe the production of interleukin (IL)-10, a cytokine with immunosuppressive properties. We show that in contrast to mouse IL-10, human IL-10 multimerises extensively *in planta*. Both human and mouse IL-10 form homodimeric protein

complexes through a mechanism referred to as 3D domain swapping. Only for human IL-10 3D domain swapping results in extensive multimerisation. By fusing IL-10 to green fluorescent protein (GFP) multimerisation was visualised, demonstrating that human IL-10 forms granules of organelle size. We discovered that for mouse IL-10 granule formation was prevented by N-glycosylation of the N-terminus. Introduction of this N-glycosylation site in human IL-10 partly prevented granule formation. The insertion of a glycine-serine linker between alpha helices D and E of human IL-10, allowing IL-10 to swap its own domains and form a stable monomer. This prevented granule formation completely and boosted protein yield 30-fold. However, the now comparable yields of human and mouse IL-10 were still low when compared to for instance IgG.

In the next two chapters we shifted the focus to the production of antibodies. Thus far the IgA isotype received little attention as candidate biopharmaceutical. However, the unique features of IgA, such as the ability to recruit neutrophils and suppress the inflammatory responses mediated by IgG and IgE, make it a promising antibody isotype for several therapeutic applications. Therefore, we compared the plant-based expression of IgA to IgG (Chapter 3). The variable regions of three antibodies commonly used in the clinic to treat inflammatory disorders (Infliximab, Adalimumab and Ustekinumab) were grafted on either an IgG1κ or IgA1κ backbone. Surprisingly, we achieved comparable high expression levels for all antibodies. The large variation in antibody yield found in literature is therefore likely due to differences in expression systems and experimental conditions, but not to the antibody idiotype used.

Evaluation of the secretion efficiency and N-glycosylation profiles revealed compelling differences between IgG and IgA antibodies when expressed *in planta*. Compared to IgG, IgA was poorly secreted. Poor secretion is most likely due to vacuolar targeting of IgA, as it was previously demonstrated that murine IgA contained a cryptic vacuolar targeting signal in the tailpiece. IgA also carried different N-glycans compared to IgG, which were never found to be a main fraction of a plant-produced protein. The predominant N-glycan on IgA lacked both the typical plant core α1,3-fucose and one terminal *N*-acetylglucosamine (GlcNAc). The core α1,3-fucose is most likely not added due to low accessibility of the core of the N-glycans. The lack of one GlcNAc may be due to inefficient addition of this sugar residue or (partial) β-hexosaminidase activity that may occur in both the vacuole and the apoplast. Whether the vacuolar targeting plays a role in the lack of the GlcNAc and what protein intrinsic properties influence partial GlcNAc addition or removal is unclear. We also showed that the N-glycosylation site in the IgA tailpiece does not always receive a N-glycan, which may be significant for secretory IgA formation as described in chapter 4.

In chapter 4 we describe the expression of this large heteromultimeric protein complex. Secretory IgA consists of two IgA complexes that are connected via the joining chain and associate with a part of the polymeric-Ig-receptor (called the secretory component). The challenge for sIgA expression is that assembly of four polypeptides (the alpha heavy chain, the

light chain, the joining chain and the secretory component) in a 4:4:1:1 ratio is required. Previous studies on stable and transient expression of sIgA demonstrated that sIgA formation in plants is possible. However, sIgA was always accompanied by a large proportion of monomeric IgA as well as other assembly intermediates. The presence of assembly intermediates may be caused be unequal expression and/or stability of the individual components. However, we demonstrated that not the expression strategy, but protein complex assembly through impaired incorporation of the joining chain was the limiting step for sIgA production. Because joining chain incorporation depends on N-glycosylation of the IgA tailpiece we hypothesise that the partial N-glycosylation of the IgA tailpiece is the cause of inefficient joining chain incorporation. The efficiency of N-glycosylation may be improved by mutation of the glycosylation site N-X-S to an N-X-T site, as the latter glycosylation motif was shown to be more often occupied in a large-scale glycosylation site analysis.

Successful engineering of the plant glycosylation pathway to provide biopharmaceuticals with human N-glycan types has been achieved. Glyco-engineering also provides the opportunity to generate potentially interesting N-glycan types with regard to immunogenicity for vaccination purposes or protein trafficking for *in vivo* targeting of biopharmaceuticals. In chapter 5 we describe the expression of the *Schistosoma mansoni* egg antigen omega-1, an immunomodulatory protein with therapeutic potential. Its biological activity depends on its RNAse activity and its N-glycans that enable internalisation by human dendritic cells. Omega-1 cannot be isolated from natural resources in sufficient quantities to study its *in vivo* biology *i.e.* which N-glycan types facilitate full activity of omega-1. Therefore, we set up the plant-based production of this protein. *S. mansoni*-derived omega-1 predominantly carries diantennary N-glycans with a Lewis X motif on one or both antennae. Lewis X consists of a β1,4-linked galactose and α1,3-linked fucose residue attached to a terminal GlcNAc. This attachment is performed by β1,4-galactosyltransferase and α1,3-fucosyltransferase IXa that naturally do not occur in plants. Co-expression of these two glycosyltransferases with omega-1 resulted in a N-glycan profile comparable to *S. mansoni*-derived omega-1. However, it was necessary to control the expression of the β1,4-galactosyltransferase to contain this enzyme in the *trans*-Golgi compartment. Overexpression most likely leads to overflow of β1,4-galactosyltransferase to the medial-Golgi and galactosylation at this stage disturbs the activity of other glycosyltransferases resulting in hybrid N-glycan types.

We also observed that more than 90% of omega-1 was secreted to the apoplast, which facilitated efficient purification. This demonstrated that plants, as previously suggested, are not poor protein secretors. However, the underlying protein properties controlling secretion efficiency are currently unknown. Both omega-1 and IgG are very efficiently secreted under natural circumstances. Yet, despite the fact that we used the same signal peptide to facilitate secretion of IgG and omega-1, IgG was only secreted from plant cells to a maximum of 28%. More knowledge on protein secretion efficiency in plants may overcome cumbersome

purification, currently the greatest bottleneck in the downstream processing of plant produced proteins.

Finally in chapter 6, we shifted focus from the expression of a particular protein to the effect of codon use on protein yield. We designed a codon optimisation strategy that, unlike other strategies, turned out to be surprisingly robust. This strategy is based on a general codon bias found in plants. Because this codon bias could be found among plant species, including monocots and dicots, and resulted in an increase in mRNA stability and translatability, we suspected that this codon bias arises from a selection pressure on the mRNA structure. We extended this general codon bias to representative species of other kingdoms of life and demonstrate that there is a selection pressure increasing mRNA stability and translatability (more protein per mRNA molecule). Stability was the result of an increase in the number of nucleotide bonds. However, there is a trade off between mRNA stability and translatability, and the nucleotide bonds in an mRNA should be well balanced over the entire molecule, making it 'airy', to ensure efficient translation.

Altogether we conclude that transformation efficiency and level of transcription are no longer limiting factors for protein yield upon plant-based expression. However, an increase in translation initiation and translation rate dictated by codon use may still provide an increase in protein production. Furthermore, we show that many proteins demonstrated specific production bottlenecks. The yield of human IL-10 was hampered by its extensive multimerisation, sIgA assembly is most likely limited by inefficient N-glycosylation of the tailpiece of IgA, both IgG and IgA are inefficiently secreted compared to omega-1 and IgA displayed an aberrant N-glycosylation profile. Therefore, more knowledge on how protein intrinsic properties influence protein yield and/or quality of heterologous produced proteins in plants should now be generated. Transient expression in *N. benthamiana* is an ideal tool to study these protein intrinsic properties that limit protein yield in heterologous expression.

Samenvatting

Farmaceutische eiwitten spelen een steeds grotere rol in de behandeling van auto-immuunziekten en chronische ontstekingen. Deze eiwitten vallen onder de noemer 'biofarmaceutica' en worden geïsoleerd uit menselijk of dierlijk bloed of weefsel, of, wat tegenwoordig vaker het geval is, gemaakt met behulp van levende cellen. Een bekend voorbeeld hiervan is insuline, een eiwit dat gebruikt wordt voor de behandeling van diabetes (suikerziekte). Insuline werd oorspronkelijk uit de alvleesklier van geslachte varkens en runderen gehaald, maar wordt nu gemaakt met behulp van de bacterie *Escherichia coli*. Het voordeel hiervan is dat in tegenstelling tot insuline uit varkens en koeien, de insuline gemaakt door *E. coli* identiek is aan die van de mens. Dit voorkomt dat ons immuunsysteem het insuline als lichaamsvreemd herkent en vernietigd wordt. Daarnaast is door *E. coli*-geproduceerde insuline vele malen goedkoper. Er bestaan verschillende productiesystemen en veel biofarmaceutica worden met behulp van *E. coli* gemaakt, maar ook de gist *Saccharomyces cerevisiae* en dierlijke cellijnen worden vaak gebruikt. Elk productiesysteem heeft voor- en nadelen en het gewenste eiwit bepaald welk systeem gebruikt kan en zal worden. Onderzoek naar het gebruik van planten heeft in de laatste jaren meer aandacht gekregen omdat planten net als *E. coli* goedkoop in gebruik zijn en net als dierlijke cellen in staat zijn complexe eiwitten te maken. Het telen van planten is ook makkelijk op te schalen aangezien de infrastructuur al aanwezig is ten behoeve van de voedselproductie.

Dit proefschrift beschrijft het gebruik van planten voor de productie van een variatie aan eiwitten en eiwitcomplexen die gebruikt worden of in de toekomst mogelijk gebruikt kunnen worden als biofarmaceutica. Sinds 1982 kunnen planten genetisch gemodificeerd worden, waardoor men in staat is elk gewenst eiwit in planten te maken. De genetische modificatie kan op verschillende manieren bereikt worden, maar vaak wordt gebruik gemaakt van een bacterie genaamd *Agrobacterium tumefaciens,* die planten van nature kan 'transformeren'. Dat doet deze bacterie door enkele van zijn eigen genen in het genoom van de plant in te bouwen die er voor zorgen dat de bacterie van nutriënten wordt voorzien. De genen die de bacterie overdraagt, kunnen vervangen worden door een gen dat codeert voor een door ons gewenst eiwit. Een veel gebruikte plant voor productie van eiwitten is de tabaksplant *Nicotiana benthamiana*. De bladeren van deze tabaksplant kunnen geïnfiltreerd worden met een suspensie van *Agrobacterium tumefaciens,* zoals te zien op de voorkant van

dit proefschrift (de donkere delen). Na enkele dagen is de productie van het gewenste eiwit waar te nemen.

De eerste pogingen om biofarmaceutica in planten te maken resulteerden in een te laag productieniveau om economisch interessant te zijn. In de laatste twee decennia is de efficiëntie van transformatie sterk verbetert en heeft het aanpassen van regulerende elementen voor genexpressie voor een sterke verbetering van de productie gezorgd. Tegenwoordig bestaan er verschillende productiestrategieën die een hoog productieniveau in planten kunnen bewerkstelligen. De meeste ervaring is opgedaan met de productie van antilichamen. Iedereen heeft miljarden verschillende antilichamen in zijn/haar lichaam. Elk antilichaam is in staat één specifiek molecuul te binden. Vaak zijn antilichamen gericht tegen moleculen afkomstig van ziekteverwekkers, maar technieken zijn ontwikkeld om antilichamen te maken die een gewenst molecuul kunnen binden. Vanwege deze unieke eigenschap worden antilichamen vaak ingezet als biofarmaceutica om ongewenste moleculen in ons lichaam te neutraliseren. Elk antilichaam bestaat uit variabele domeinen die de bindingsspecificiteit bepalen en constante domeinen die voor interactie met het immuunsysteem zorgen. Er bestaan verschillende constante domeinen die het zogenaamde isotype bepalen. De verschillende isotypen hebben ieder een specifieke interactie met het immuunsysteem. Antilichamen van het IgG isotype worden het meest gebruikt als biofarmaceutica. Over het algemeen kunnen IgG antilichamen gemaakt worden in een orde van grootte van grammen per kilogram plantmateriaal, maar het productieniveau van antilichaam tot antilichaam kan erg variëren.

Het productieniveau van andere biofarmaceutica, zoals bijvoorbeeld die van de cytokines, blijft vaak achter in vergelijking met de productie van antilichamen. Cytokine is de verzamelnaam van een grote groep eiwitten die een rol spelen in de signalering tussen cellen van het immuunsysteem. Vele cytokines zouden mogelijk gebruikt kunnen worden voor de behandeling van auto-immuunziekten en chronische ontstekingen. Hun korte levensduur suggereert dat ze van nature instabiel zijn, wat wordt aangenomen als de hoofdreden voor de achterblijvende productie.

In hoofdstuk 2 beschrijven wij de productie van interleukine-10 (IL-10), een cytokine met ontstekingsremmende eigenschappen. We laten zien dat humaan IL-10, in tegenstelling tot muis IL-10, multimeriseert in planten. Zowel humaan als muis IL-10 vormen onder normale omstandigheden homodimeren (twee dezelfde eiwitmoleculen die aan elkaar binden). Voor IL-10 gaat dit als in een yin-yang teken; het eerste deel van het IL-10 eiwit wordt uitgewisseld met het tweede deel van een ander IL-10 eiwit en visa versa, een mechanisme dat 3D 'domain swapping' wordt genoemd. In tegenstelling tot muis IL-10 beperkt zich dit voor humaan IL-10 echter niet tot twee eiwitten, maar wordt een soort ketting gevormd. Hierbij bind het eerste domein van eiwit 1 met het tweede domein van eiwit 2, maar bind het eerste domein van eiwit 2 met het tweede domein van eiwit 3 etc. Door IL-10 te fuseren met groen fluorescent eiwit werd de multimerisatie gevisualiseerd en hiermee konden we aantonen dat humaan IL-10 granules vormt ter grote van organellen in de cel (tot wel 0.005 millimeter groot). We

ontdekten dat de formatie van granules van muis IL-10 werd voorkomen door een suikermolecuul (N-glycaan) die op het eiwit gezet wordt tijdens de productie (een proces wat N-glycosylering wordt genoemd). Door er voor te zorgen dat ook humaan IL-10 een suikermolecuul kreeg, werd de formatie van granules grotendeels voorkomen. Door het invoegen van een extra stuk flexibele linker tussen het eerste en tweede domein van humaan IL-10 kon het eerste domein uitgewisseld worden met het tweede domein van hetzelfde eiwit. Op deze manier vormt zich een stabiele monomeer. Deze strategie voorkwam de formatie van granules volledig en resulteerde in een 30 keer hogere opbrengst van humaan IL-10.

In de volgende twee hoofdstukken hebben we onze focus verschoven naar de productie van antilichamen. De productie van het IgA isotype heeft vooralsnog niet veel aandacht gekregen als potentieel biofarmaceutisch eiwit. Desalniettemin maken de unieke eigenschappen van IgA het een interessant eiwit voor therapeutische doeleinden. IgA is in staat neutrofielen (de meest voorkomende witte bloedcellen) te rekruteren en deze cellen zouden ingezet kunnen worden om bijvoorbeeld kanker te bestrijden. Daarnaast kan IgA door IgG en IgE geïnduceerde immuunreacties onderdrukken en zou dus als ontstekingsremmer kunnen functioneren. Daarom hebben wij de productie van humaan IgA in planten vergeleken met die van humaan IgG (hoofdstuk 3). Drie antilichamen (Infliximab, Adalimumab en Ustekinumab) die veelvuldig gebruikt worden in de kliniek om chronische ontstekingen te behandelen (zoals de ziekte van Crohn en reumatoïde artritis) werden als IgG en IgA geproduceerd. Alle antilichamen bereikten een vergelijkbaar hoog productieniveau, dit in tegenstelling tot wat we op basis van de literatuur hadden verwacht. De grote variatie in opbrengst beschreven in de literatuur wordt daarom waarschijnlijk vaker veroorzaakt door verschillen in productiesystemen en experimentele condities, en minder vaak door intrinsieke verschillen van antilichamen.

Daarentegen hebben we wel wezenlijke verschillen tussen IgG en IgA gevonden met betrekking tot de uitscheiding en N-glycosylering. Vergeleken met IgG werd IgA aanzienlijk minder uitgescheiden door plantencellen. Dit is wellicht een gevolg van het feit dat IgA waarschijnlijk naar de vacuole wordt gedirigeerd. Voor muis IgA is aangetoond dat het eiwit naar de vacuole gaat door de aanwezigheid van een signaal in het staartstuk van het eiwit. Humaan IgA heeft een vergelijkbaar signaal en zou dus ook in de vacuole terecht kunnen komen. Daarnaast draagt plant-geproduceerd IgA ook hoofdzakelijk andere N-glycanen vergeleken met plant-geproduceerd IgG. De voornaamste N-glycaan op IgA miste het voor planten kenmerkende suikerresidu α1,3-fucose en het eindstandige suikerresidu *N*-acetylglucosamine (GlcNAc). Dat er geen α1,3-fucose toegevoegd wordt, komt wellicht door slechte bereikbaarheid van de N-glycaan voor het enzym om dit suikerresidu er aan te koppelen. De afwezigheid van een van de eindstandige GlcNAcs is mogelijk ook een gevolg van inefficiënte koppeling van dit suikerresidu of het gevolg van (gedeeltelijke) afsplitsing door het enzym β-hexosaminidase. Het is nog onduidelijk of humaan IgA, net als muis IgA, naar de vacuole wordt gestuurd, of dit een rol speelt in de afwezigheid van de eindstandige GlcNAc

en welke inherente eiwiteigenschappen hier aan ten grondslag liggen. Wel hebben we aangetoond dat N-glycosylering van het staartstuk van humaan IgA niet altijd plaats vindt, wat de vorming van het eiwitcomplex secretoir (s)IgA zou kunnen beperken (hoofdstuk 4).

In hoofdstuk 4 beschrijven we de productie van sIgA. Dit grote eiwitcomplex bestaat uit twee IgA moleculen die met elkaar verbonden zijn via een J-keten met daaromheen gevouwen een eiwit dat de secretoire component wordt genoemd. De uitdaging voor het maken van sIgA is dat de assemblage van vier verschillende eiwitten (de zware keten, de lichte keten (samen IgA), de J-keten en het secretoire component) een 4:4:1:1 ratio behoeft. Voorgaande studies hebben aangetoond dat productie van sIgA in planten mogelijk is, maar de productie ging altijd gepaard met de aanwezigheid van een grote hoeveelheid monomeer IgA en andere tussenvormen van het complex. Een ongelijke productie of stabiliteit van de individuele eiwitten zouden de aanwezigheid van deze tussenvormen kunnen verklaren. Wij hebben echter laten zien dat de koppeling van de J-keten aan de IgA moleculen de limiterende stap is voor de vorming van het complex en niet de productiestrategie. We veronderstellen dat de eerder waargenomen partiële N-glycosylering van het staartstuk van IgA ten grondslag ligt aan de inefficiënte koppeling van de J-keten.

In hoofdstuk 5 besteden we aandacht aan het aanpassen van de N-glycosylering in planten. Als model eiwit maken we gebruik van omega-1, een eiwit dat uitgescheiden wordt door de eieren van de humaan parasitaire worm *Schistosoma mansoni*. De biologische activiteit van omega-1 is afhankelijk van zijn N-glycanen die opname door humane dendritische cellen bewerkstelligen. Omega-1 is een immuun-modulatoir eiwit met therapeutische potentie, maar kan niet in voldoende hoeveelheden geïsoleerd worden uit natuurlijke bronnen. Daarom hebben wij een productiesysteem opgezet om omega-1 met zijn natuurlijke N-glycaan te maken in planten. De N-glycanen van omega-1 dragen 'Lewis X' motieven die bestaan uit een eindstandige GlcNAc met daaraan een β1,4-galactose en α1,3-fucose gekoppeld. Deze suikerresiduen worden door de enzymen β1,4-galactosyltransferase en α1,3-fucosyltransferase IXa toegevoegd en deze enzymen komen van nature niet voor in planten. De gelijktijdige productie van deze enzymen met omega-1 resulteerde in een N-glycaan profiel dat vergelijkbaar was met natuurlijk omega-1.

We ontdekten ook dat ~90% van het geproduceerde omega-1 werd uitgescheiden door plantencellen. Hierdoor was de zuivering van dit eiwit relatief eenvoudig. Dit demonstreert dat planten prima in staat zijn eiwitten op grote schaal uit te scheiden, waar het tegenovergestelde in het verleden werd gesuggereerd. Ondanks het feit dat we dezelfde strategie gebruikt hebben om secretie van IgG en omega-1 door plantencellen te bewerkstelligen, werd maximaal 28% IgG uitgescheiden tegenover ~90% omega-1. Welke inherente eiwiteigenschappen uitscheiding beïnvloeden is nog onbekend. Meer onderzoek naar de secretie van eiwitten door plantencellen zou een oplossing kunnen bieden voor de moeizame zuivering van eiwitten uit planten, het huidige knelpunt van veruit de meeste plant-productiesystemen.

In hoofdstuk 6 beschrijven we niet de productie van een specifiek eiwit, maar richten we ons op een algemeen aspect van eiwitproductie; de vertaling van gen naar eiwit. Eiwitten worden gecodeerd door genen die opgebouwd zijn uit vier verschillende bouwstenen; de nucleotiden cytosine (C), guanine (G), adenine (A) en thymine (T). De combinatie van drie nucleotiden wordt een codon genoemd en codeert voor een specifiek aminozuur. Aminozuren worden in een keten geschakeld en kunnen vervolgens met elkaar interacties aangaan om zo een eiwit te vormen. Eukaryote eiwitten kunnen maximaal 20 verschillende aminozuren bevatten en elk aminozuur kan door 1, 2, 4 of 6 codons gecodeerd worden. In andere woorden maximaal 6 codons coderen voor hetzelfde aminozuur. Het gebruik van de verschillende codons is nooit gelijk en welk codon het meest gebruikt wordt is afhankelijk van de soort. Je zou dus kunnen spreken van 'dialecten' tussen verschillende soorten. Hoe minder verwant de soorten zijn, hoe groter het verschil in het dialect. Omdat de soorten die gebruikt worden om menselijke eiwitten te maken een ander dialect hebben dan mensen, vinden deze soorten het soms moeilijk de menselijke genen te 'verstaan'. Als gevolg hiervan zal er minder van het eiwit gemaakt worden dan wellicht mogelijk is. Daarom zijn vele verschillende methodes ontwikkeld om menselijke genen te 'vertalen' naar het dialect van producerende soorten. Echter zijn de resultaten behaald met deze methodes erg variabel. Ook wij hebben een strategie ontwikkeld voor de vertaling van genen die verrassend succesvol bleek. Onze strategie is niet gebaseerd op het dialect van ons productie platform *N. benthamiana*, maar het dialect in een bepaalde set genen uit verschillende ver-verwante plantensoorten; namelijk de genen die erg veel gebruikt worden. Uit voorgaand onderzoek was gebleken dat verschillende plantensoorten, inclusief ver-verwante plantensoorten, allemaal met hetzelfde dialect spreken als het hun veel gebruikte genen betreft. Onze strategie, gebaseerd op dit 'veel gesproken' dialect, resulteerde in aanzienlijk hogere eiwitproductie in planten. We laten dit niet alleen zien voor een menselijk eiwit, maar ook voor een kippen- en kwalleneiwit. Dit stimuleerde het bestuderen van veel gebruikte genen in niet-plantensoorten en we ontdekten dat dit veel gesproken dialect van planten grotendeels gedeeld wordt door ver-verwante soorten uit verschillende biologische rijken. We leveren daarnaast ook indirect bewijs voor een onderliggende selectiedruk.

In dit proefschrift laten wij zien dat naast het aanpassen van het 'dialect' van genen het aanpassen van regulerende elementen waarschijnlijk niet meer zal leiden tot significante productietoename, maar dat veel eiwitten eiwit-specifieke productie-knelpunten hebben. De opbrengst van humaan IL-10 werd beperkt door multimerisatie, sIgA complexvorming wordt mogelijk beperkt door de inefficiënte N-glycosylering van het staartstuk van IgA, zowel IgA als IgG worden inefficiënt uitgescheiden vergeleken met omega-1 en vertoonde IgA een afwijkend N-glycaan profiel ten opzichte van IgG. Daarom is meer kennis vereist over hoe eiwit-inherente eigenschappen de kwaliteit en opbrengst van eiwitten beïnvloeden. Daarbij is productie van eiwitten in *N. benthamiana* een ideaal wetenschappelijk instrument om deze eiwit-inherente eigenschappen en hun effect op eiwit opbrengst te onderzoeken.

Acknowledgements

As most of you will know, I have had the great privilege of not walking the PhD 'pathway' alone :-P. For those of you who may not know, this does not mean I had a great supervisor. No, no, no, this also doesn't mean I didn't have a great supervisor, because I did. Thank you Arjen for your direct manner of communication, your patience with me and my direct manner of communication, your checking, double-checking, triple-checking and quadruple-checking of my writing, your overall positive demeanour and always 'having my back'. Thank you also Jaap, for always taking the time to shine your light on my manuscripts with your helicopter view and for your guidance throughout my PhD.

But no, with not alone I mean I had someone with whom I did experiments together. No, I don't mean our fantastic technician Debbie. But on that note, thank you dear Debbie for all the blood (yes blood as well), sweat and tears you put into not only my thesis, but also in being part of the LMA-team. The best times of my PhD I have enjoyed with you doing 'Arjen's-LMA-style' ATTAs, 'conveyor-belt-style' ELISAs and stressing out over the organisation of the practical of the Plant Biotechnology course.

The person I am referring to even did experiments for me on occasion. A student? No. Although, it is true, I had lots of students doing lots of work for me! I found teaching a 'mixed-emotions' job at times, but can honestly say I have enjoyed the guidance of all 'my' students in their endeavour called a BSc or MSc thesis. Not all their work ended up in this thesis, but all have contributed to this thesis in the sense that they taught me a great deal about myself, forced me to (re)dive into specific subject matter and always created a motivational, exciting atmosphere. Therefore, in chronological order, Grazyna Truszkowska, Tineke Vliek, Jose Ramon, Jannie Tilma, Marieke Beijer, Luuk Belgers, Bart Nijland, Mohamed Abdi Hassan, Kevin van der Eijken, Joyce Quist, Francisco Marques, Kelly Heckman, Ngoc Tram Hong, Sonja Warmendam, Ira Chestakova, Edgar Wils, Koen Varossieau, Diecke de Ronde, Che Chong, Rachel Begheyn, Bob Engelen, Aleksandra Syta, Tim Warbroek, Lauri Reuter, Michelle Yang, Tam Nguyen, Simone Oostindie, Jorik Swier, Kirstin Jansen, Etienne van Dalen, Myrte Huijskes, Kassiani Kytidou, Kim van Noort, Marloes van Wijk, Eva Capuder, Alja van der Schuren, Thomas Roodsant, Levine Zinger, Allard Stellingwerf and Winnie van den Boogaard, thank you for tagging along on my PhD track for a while. I hope you all are well and perhaps are even able to attend my defense.

I could really talk to this one person, about anything related to work. A really great colleague perhaps? Yes, thanks to all my colleagues who made the laboratory of Nematology more than only a place to work! With special thanks to JanR for opening the door to the world of DNA, restriction enzymes and cloning. You literally got me started and I have benefitted from your exhaustive knowledge from the beginning until today. And JanvdV, you were part of the LMA-team for the greatest part of my PhD thesis and I have enjoyed your (and Henk's!) company on many occasions in and outside work. I hope we will stay in touch and I can keep enjoying your unique view and love of life. Pjotr, Mark, Basten, Martijn and Hans, you have been directly evolved in the realisation of chapter 6 and I thank you for all your enthusiasm and effort. Dear Jet, thank you for your smiles, rolling eyes and pats on the back, and of course 'the walk' ;-). Jose, thank you for always being the rebel, your positivity and friendship. Lisette, thank you for always being in control, aware and on top of everything related to the organization of our chair group. Joost, Sven, Paul, Hein, Rikus, CasperS and of course again Debbie thank you for organising the lab so well, and in particular Joost, for always locating the box I couldn't find, Rikus for your 'apparatus-karma' and protein-work assistance and, Sven and Paul for the inappropriate comments and CasperS (and also Cindy!) for being so colourful.

This person is a colleague, but also a friend and family. He calls my dear mom and dad 'schoonouders', whom I thank for always listening to my ramblings over a glass of wine and providing me with the means and mental support to succeed in life. Like me, he calls our friends 'family', whom I thank for their friendship. Thank you dear, Femke for always knowing what to do, Ronald for always bearing with us girls, Annemarie for always analysing the situation perfectly and saying the right thing, Ana for always being there when needed, Kamila and Tomek for all the good times, Tineke for your kindness, Sanne and Andre for knowing what it is like (you'll be next!).

The person I have shared the PhD pathway and my life with these past years is Ruud. Thank you dear dear Ruud, for literally everything. We met in the tiny student room in the Binnenhaven building, where behind your computer we made our very first team-effort on the IL-10 project. We became good friends. After your MSc thesis you worked for Jose for a while, but just before you went abroad to do an internship in Oxford, we became more than friends. Typical… Although you were offered a PhD position there, I very selfishly convinced you to start a PhD in the same research group I was in. Although we have our own projects, we have always been very involved in each other's work, and even swapped projects on occasion. I have never made a better team with anyone, and I guess you feel the same, because we got married two years ago this coming April. We also embarked on a new adventure and became parents to our beautiful son Flip. Finally, work is not the centre of our universe anymore :-P. Nevertheless, I hope we will be able to work and be together for a long time to come. Dear Ruud, again, thank you for your involvement in every single step of my PhD. I didn't do it alone, because I had you, and now you have me.

Halfway, halverwege.

Curriculum vitae

Lotte Bregje Westerhof was born on the 9th of August 1982 in Groningen, The Netherlands. She graduated from the high school Nienoord College in Leek in 2000. Thereafter she joined the BSc program Biotechnology of Noordelijke Hogeschool Leeuwarden and Hogeschool van Hall Larenstein (Leeuwarden, The Netherlands). During her studies she specialised in plant biotechnology and spend half a year at Plant Research International (Wageningen, The Netherlands) working on edible vaccines under the supervision of dr. ir. Dion Florack and half a year at The University of Western Australia (Perth, Australia) working on Wandoo crown decline under the supervision of dr. ir. Pieter Poot. After obtaining her BSc in 2004 she continued her specialisation in plant biotechnology by joining the MSc program Plant Biotechnology at Wageningen University (Wageningen, The Netherlands). She obtained her MSc with the specialisation 'Molecular Farming' and 'Plant Pathology and Pest' control in 2007. During her MSc studies she started working on the plant-based expression of interleukin-10 under the supervision of dr ir. Arjen Schots at the Laboratory of Nematology of Wageningen University. In the same group she continued to do research on plant-production of (potential) biopharmaceuticals as a PhD and will be given the opportunity to defend her PhD thesis on the 3rd of December of 2014. Since September 2013 she holds a Postdoc position at Wageningen University and lives in Wageningen with her husband Ruud Wilbers and their son Flip.

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Lotte Westerhof
Date: 3 December 2014
Group: Nematology, Wageningen University & Research Centre

1) Start-up phase	date
► **First presentation of your project**	
Plant-based IL-10 production	Oct 09, 2006
► **Writing or rewriting a project proposal**	
► Writing a review or book chapter	
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*1.5 credits**

2) Scientific Exposure	date
► **EPS PhD Student Days**	
EPS PhD Student Day, Wageningen University	Sep 13, 2007
EPS PhD Student Day, Leiden University	Feb 26, 2009
► **EPS Theme Symposia**	
EPS Symposium 'Plant Endomembranes', University of Amsterdam	Jul 02, 2010
EPS Theme Symposium 3 'Metabolism and Adaptation', Wageningen University	Feb 10, 2011
► **NWO Lunteren days and other National Platforms**	
► **Seminars (series), workshops and symposia**	
Immunesystem in Darwinian times	May 20, 2009
Synthon Workshop Plant biotechology	Jul 31, 2009
COST meeting 2010	Jan 25-26, 2010
► **Seminar plus**	
► **International symposia and congresses**	
Cytokines in Health and Disease 2007 - San Fransisco, USA	Oct 26-30, 2007
European Mucosal Immunology Group meeting 2008 - Milan, Italy	Oct 08-10, 2008
14th International congress of mucosal immunology 2009 - Boston	Jul 05-09, 2009
► **Presentations**	
Poster presentation - Cytokines in Health and Disease 2007 - San Fransisco	Oct 26-30, 2007
Oral presentation Synthon Workshop Plant biotechnology	Jul 31, 2009
Oral presentation COST meeting 2010 - Wageningen	Jan 25-26, 2010
Oral presentations Synthon work progress	2008-2010
► **IAB interview**	Dec 04, 2009
► **Excursions**	
Subtotal Scientific Exposure	*19.7 credits**

3) In-Depth Studies	date
► **EPS courses or other PhD courses**	
Plant-based vaccinces and antibodies - Verona, Italy	Jun 2009
Production of heterologous proteins in plants - Helsinki, Finland	Oct 17-28, 2011
► **Journal club**	
Particiption in a literature discussion group at Nematology	2007-2010
► **Individual research training**	
Subtotal In-Depth Studies	*8.9 credits**

4) Personal development	date
► **WGS: PhD competence assessment**	Sep 2007
Scientific Writing	Nov 2007
WGS Course: Project and Time Management	Nov-Dec, 2007
Supervising Thesis	Jan 17-18, 2008
Scientific Artwork	Feb 16-17, 2009
Scientific Publishing	Nov 19, 2009
► **Organisation of PhD students day, course or conference**	
EPS Student Day 2007	2007
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*7.2 credits**

TOTAL NUMBER OF CREDIT POINTS*	**37.3**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

This research was conducted at the Laboratory of Nematology, Wageningen University, (Wageningen, The Netherlands) and financially supported in part by Synthon (Nijmegen, The Netherlands) and a grant from the Dutch Ministry of Economic Affairs (PID07124).

Cover picture: Ruud H. P. Wilbers
Cover design: Proefschriftmaken.nl || Uitgeverij BOXPress
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, 's-Hertogenbosch