Expression of four *S. pneumoniae* type 2 polysaccharide biosynthetic enzymes utilising the endogenous Kex2 protease activity in tobacco

Kafeel Ahmad[1]*, David Twell[1], Kevin Christopher Gough[2] and Ben Charles Maddison[3]
[1]Department of Biology, Adrian Building, University of Leicester, University Rd, Leicester, UK LE1 7RH
[2]Department of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Loughborough, Leicestershire, UK
[3]ADAS Biotechnology Group, Department of Biology, University of Leicester, University Rd, Leicester, UK LE1 7RH

Abstract: In order to express multisubunit proteins, or to manipulate metabolic pathways in plants it is essential to be able to efficiently express multiple proteins within the same plant cell. To increase the efficiency of multi-protein expression, we demonstrate the use of the Golgi localized Kex2 protease activity in tobacco to process a large polyprotein precursor consisting of four individual protein domains into its individual protein constituents. Four genes encoding enzymes involved in the biosynthesis of *S. pneumoniae* type 2 polysaccharide were assembled into a single expression cassette as a large polyprotein driven by a single cauliflower mosaic virus (CaMV) 35S promoter. Each of the individual protein domains were separated by three sequential Kex2 protease digestion sites. At the N-terminus a Pr1b signal peptide was incorporated for efficient targeting of the polyprotein to the apoplast. Each individual protein domain was tagged with its own immuno-tag. The construct was used for the transformation of *Nicotiana tabacum* and stable lines were selected. All four processed proteins could be immunologically detected in protein extracts using Western blotting indicating correct expression and Kex2 processing. Utilisation of the Kex2 protease system represents an efficient way of expressing multiple proteins in the same plant. This method simplifies the transformation procedures, and presents a method for expression of multiple proteins within the same plant.

Keywords: Kex2 protease, *S. pneumoniae*, polyprotein precursor, transgenic

INTRODUCTION

Historically, plants have been transformed with single genes to investigate plant growth, improve product characteristics, study plant gene expression, and to express foreign proteins of industrial or pharmaceutical interest. So-called first generation genetically modified crops, have involved the expression of single genes to confer traits such as herbicide tolerance or insect resistance. More complex expression systems where plants need to be engineered to express multiple genes include the expression of multimeric proteins such as the co-expression of antibody fragments (Hiatt *et al.*, 1989; Saxena *et al.*, 2016), the expression of multiple biosynthetic enzymes to produce useful compounds in planta (Nawrath *et al.*, 1994; Luo *et al.*, 2015), or the expression of enzymes involved in biosynthetic pathways whose product can increase the nutritive value of foods (Ye *et al.*, 2000). There is therefore considerable pharmaceutical/commercial interest in being able to exploit methods that simplify plant manipulation and enable efficient expression of multiple proteins for the synthesis of a variety of valuable products. Multi gene expression has historically been achieved in different ways; independent transgenic lines expressing single genes can be produced and these transgenic lines can be crossed to combine all the genes responsible for the traits under study in a single plant (Ma *et al.,* 2005). This is a time consuming strategy and is limited by the number of genes that can be introduced. A plant could be sequentially transformed with the transgenes of interest one by one or in units consisting of more than one gene. This again is time consuming and requires use of a different selection marker for each transformation event, and the number of available markers is limited. Multiple genes could also be expressed in the form of multiple expression cassettes linked together, each cassette with its own promoter and terminator (Slater *et al.,* 1999; Belmont, 2016). However multiple copies of the same promoter used for transgenes of interest can lead to gene silencing (Meyer and Saedler, 1996). Co-transformation is another method for multiple gene expression that involves the simultaneous transformation of plants with the transgenes of interest (Zhang and Fauquet, 1998; Ninh *et al.*, 2015). Internal ribosome entry sites (IRES) have been used primarily for the expression of two genes in the form of a bicistronic message (Gouiaa *et al.,* 2012). IRES regulated gene expression is generally lower than the cap-dependant gene upstream of the IRES sequence (Houdebine and Attal, 1999). Multiple genes can also be expressed in the chloroplast genome (Daniell and Dhingra, 2002), however, this method is unsuitable for proteins that need to be targeted to cell compartments other than the chloroplast. A further strategy that has previously been described, but which has had little uptake by plant biotechnologists involves linking multiple coding

---

**Corresponding author:* e-mail: Kafeelpbg@gmail.com

sequences or open reading frames (ORFs) by the Kex2 protease cleavage sites (Jiang and Rogers, 1999).

The *Saccharomyces cerevisiae* Kex2 protein is a type 1 integral membrane protein that resides in the *trans*-Golgi network and functions in the eukaryotic secretory pathway to process polypeptide prohormone at pairs of basic residues (Julius *et al.,* 1984; Fuller and Thorner, 1989; Roongsawang *et al*., 2016). The Kex2 gene codes for a Ca^{2+}-dependant serine protease (Fuller and Thorner, 1989) and homologues of the Kex2 protease have been found in a diverse group of organisms (Thomas *et al.,* 1991; Hayflick *et al.,* 1992). A Kex2 protease-like pathway has previously been demonstrated in tobacco (Jiang and Rogers, 1999; Kinal *et al.*, 1995), however, the enzymes involved in such processing have to date not been characterised. Kinal *et al* (1995) expressed the KP6 killer preprotoxin of *Ustilago maydis* in transgenic tobacco to study its processing in tobacco. The preprotoxin is not active until it is processed into α and β subunits, and this processing is performed by Kex2 protease (Tao *et al.,* 1990). The protoxin was correctly processed into α and β subunits in transgenic tobacco and the mature toxin produced killer activity similar to the native toxin. Jiang and Rogers (1999) demonstrated the substrate specificity and Golgi-localization of the plant Kex2 like protease in tobacco using a reporter protein by fusing a mutated proaleurain to a domain of the BP-80 protein. Efficient Kex2 processing was observed in constructs containing three Kex2 cleavage sites between the mutated proaleurain and BP-80, with a large portion of mutated proaleurain secreted into the culture medium.

This study utilised the plant Kex2 like protease system for the co-expression of four *Streptococcus pneumoniae* type 2 capsular polysaccharide biosynthetic genes in tobacco plants for evaluation of the suitability of this system for multiple protein production. *S. pneumoniae* can cause a variety of pneumococcal infections such as meningitis, septicaemia and pneumonia (world health organization). Each year 1.6 million people including more than 800,000 children die due to pneumococcal related diseases (world health organization). A major virulence factor of *S. pneumoniae* is a polysaccharide capsule, which protects the bacterial cell against host phagocytes (Moxon and Kroll, 1990, De Vos *et al*., 2015). About 90 different polysaccharide serotypes (Jedrzejas, 2004) have been identified. The role of the pneumococcal capsule in infection has been demonstrated by the fact that anti-capsular antibodies have been proved effective against the infection (Snippe *et al.,* 1983; Hausdorff *et al*., 2015). Among the different polysaccharide types, one of particular importance is the type 2 capsular polysaccharide consisting of singly branched hexasaccharide repeating units, composed of one D-glucuronic acid, two D-glucose, and three L-rhamnose residues (Jansson *et al.,* 1975). Pneumococcal polysaccharides are constituents of effective polyvalent vaccines against this organism in humans.

In the D39 strain of *Streptococcus pneumoniae* (gene bank accession no. AF029368) 17 genes are responsible for the synthesis of type 2 polysaccharide arranged in a single transcriptional unit (Iannelli *et al.,* 1999). Ten genes *cps2*A, *cps2*B, *cps2*C, *cps2*D, *cps2*K, *cps2*P, *cps2*L, *cps2*M, *cps2*N and *cps2*O are involved in the biosynthesis of simple sugar monomers (Garcia and Lopez, 1997; Guidolin *et al.,* 1994; Kolkman *et al.,* 1997; Morona *et al.,* 1997; Munoz *et al.,* 1997; Nassau *et al.,* 1996). Seven central genes *cps2*E, *cps2*T, *cps2*F, *cps2*G, *cps2*H, *cps2*I and *cps2*J encode five putative glycosyltransferases, a polysaccharide polymerase, and a repeat unit transporter (Iannelli *et al.,* 1999).

In this study four genes (*cps2*E, *cps2*T, *cps2*F and *cps2*G) were expressed in a single tobacco plant using a multigene expression cassette where three repeated Kex2 cleavage sites were introduced at the junctions between genes for the Golgi localized Kex2- like protease processing of the expressed proteins. The PR1b endoplasmic reticulum targeting signal was used for ER targeting of the polyprotein which is followed by transport to the Golgi for processing and then secretion to the apoplast by the default pathway. Each protein product was tagged with its own unique immuno-tag to enable the monitoring of expression of each protein by immunoblotting. Correct processing of the *S. pneumoniae* 'polyprotein' by the Kex-2 protease system is a step towards the expression of a complement of proteins involved in the synthesis of *S. pneumoniae* polysaccharide within plants which could be used as cost effective commercial scale factories for the production of vaccine against this important human pathogen.

MATERIALS AND METHODS

PCR amplification of pneumococcal type 2 polysaccharide biosynthetic genes

The pneumococcal type 2 polysaccharide biosynthesis gene sequences were obtained from NCBI gene bank (accession number: AF026471). DNA oligonucleotides were synthesised containing sequences for peptide tags for their expression at the C terminus of each protein. 3 tandem repeated Kex2 cleavage sequences were included C-terminal to the immuno-tags. Tag coding sequences were optimized for expression in tobacco utilizing the GeneDesign software (Sarah *et al.,* 2006).

Four type 2 polysaccharide genes (*cps2*E, *cps2*T, *cps2*F and *cps2*G) were PCR amplified from *S. pneumoniae* strain D39 genomic DNA. DNA from Pneumococcal strain D39 was kindly provided by Professor Peter W Andrew (Department of Infection Immunity and Inflammation, University of Leicester).

Table 1: List of DNA oligonucleotides used in this study

Primer name	Primer sequence (5′-3′) and description
KA014	GCGACTAGTATGAAGAAGTCAGTTTATATC Description: Forward PCR primer for cps2T with *Spe*1 site
KA017	CGCCCATGGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATAGCATAATCTGGAACATCATATGGATACTTCGCTCCATCTCTCATAAATAC Description: Reverse primer for *cps2*E with HA tag-3×*Kex*2-*Nco*1 site
KA018	CTCACTTTTTCCCCCTTCAAAC Description: Reverse primer for *cps2*T
KA019	CGCCTGCAGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATCTTATCATCATCATCCTTATAATCCTCACTTTTTCCCCCTTCAAAC Description: Reverse PCR primer for *cps2*T with Flag tag-3×Kex2- *Pst*1 site
KA020	GCGCTGCAGATGAATGGAAAAACAGTAAAG Description: Forward primer for *cps2*E with *Pst*1 site
KA021	ACTTCGCTCCATCTCTCATAAATAC Description: Reverse primer for *cps2*E
KA024	GCTCAAAGATACAGGCTTC Description: Forward primer for *cps2*T
KA025	GCGCCATGGATGGCAACTTATAATGGAGC Description: Forward primer for *cps2*F with *nco*1 site
KA026	CGCGCATGCTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATATGATGATGATGATGAATAAACATTAACTCACCGATTAC Description: Reverse primer for *cps2*F with His tag-3×*Kex*2-*sph*1 site
KA028	GCGGCATGCATGACAGGTGGAATAAGAG Description: Forward primer for *cps2*G with *Sph*1 site
KA 029	CGCGGGCCCCTGCAGTTATCTATCTCTAAAATCAGAAGAAGTAGAAGATTTACCGTTTTCAATATATACC Description: Reverse primer for *cps2*G with E2 tag-stop codon-*Pst*1-*Apa*1 site

PCR products of the pneumococcal polysaccharide gene sequences were amplified using Phusion DNA polymerase (NEB). Details of primers used for gene amplification are detailed in table 1: PCR was carried out in a 50µl reaction volume of 1x phusion reaction buffer containing 200 µM dNTP mix, 2µM each primer, 50ng of template DNA and 1 U phusion DNA polymerase, 1.5mM $MgCl_2$. Amplifications were carried out for 30 cycles consisting of a denaturation at 95°C for 30 seconds, annealing at 52°C for 30 seconds and extension at 72°C for 90 seconds.

Overview of cloning the four type 2 polysaccharide biosynthesis genes (cps2T, cps2E, cps2F and cps2G)

The *cps2*T gene was amplified from *S. Pneumoniae* genomic DNA using the primers KA014 and KA019, the amplified products were cut with SpeI and PstI and then cloned directly into the *Spe*1/*Pst*1 site of the plasmid construct pKAM1. pKAM1 is a pCAMBIA2301 (Centre for the Application of Molecular Biology to the International Agriculture of Canberra, Australia) derived construct. This construct, driven by CaMV 35S promoter contains a PR1b signal sequence (Cutt *et al.,* 1988) immediately upstream of a Spe1 restriction site, and any insert is terminated at a 3′ rbcs terminator sequence which is flanked by a unique pst-1 restriction site (fig. 1). The cps2T insert was cloned in frame with the PR1b signal sequence and this was confirmed by DNA sequencing. The resulting construct is referred to as pKAM2. The remaining three genes (*cps2*E, *cps2*F and *cps2*G) were PCR amplified from genomic DNA, restriction digested with appropriate enzymes and sequentially cloned into the pGEM DNA cloning vector (Promega) in the following order; *Pst*1-*cps2*E-HA tag-3×Kex2-*Nco*1-*cps2*F-His tag-3×Kex2-*Sph*1-*cps2*G-E2 tag-stop-*Pst*1-*Apa*1 to generate construct pGEMKA3 (fig. 1).

The *cps2E* gene was amplified from *S. pneumoniae* strain D39 genomic DNA using phusion DNA polymerase with forward primer KA020 (with a *Pst*1 site at 5′ end) reverse primer KA017 containing an Nco1 site. The PCR product was digested with *Pst*1/*Nco*1 and then cloned into the *Pst*1/*Nco*1 sites of pGEM to produce the construct PGEMKA1.

The *cps2F* gene was amplified from *S. pneumoniae* strain D39 genomic DNA with forward primer KA025 (containing a *Nco*1 site at 5′ end) and the reverse primer KA026. This amplification attached a His tag, 3×Kex2 cleavage sites and a *sph*1 site to the 3′ end of the gene. The *cps2*F fusion gene was digested with *Nco*1/*Sph*1 and cloned into the *Nco*1/*Sph*1 sites of pGEMKA1 to generate pGEMKA2, keeping the cps2F gene sequence in frame with cps2E.

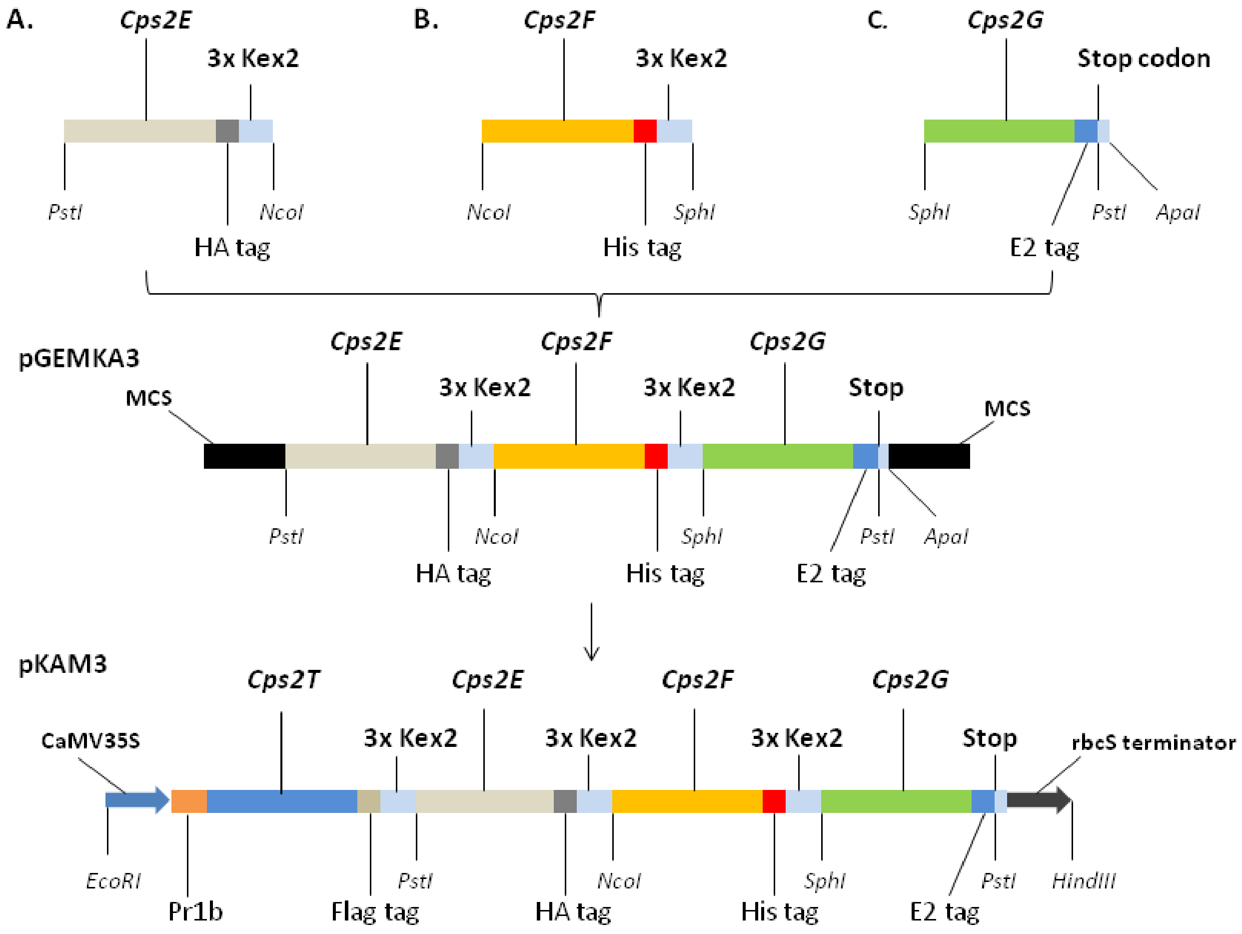

Fig. 1: Cloning of *cps2*E, *cps2*F and *cps2*G: Type 2 polysaccharide biosynthesis genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) were amplified by PCR from *S. pneumoniae* genomic DNA. Additional sequences were incorporated into the PCR products during primer synthesis: A. *cps2*E with HA tag, 3×*Kex2* sites and *Nco*1 site at 3′ end and *Pst*1site at 5′ end B. *cps2*F with His tag, 3×Kex*2* sites and *Sph*1 site at 3′ end, and *Nco*1 site at 5′ end; C. *cps2*G with E2 tag, 3×Kex*2* sites, stop codon, *Pst*1 and *Apa*1 sites at 3′ end and *Sph*1 site at 5′ end. The cps2T gene was amplified and then cloned directly into the construct pKAM1 in frame with a Pr1b signal sequence at the spe1 and pst1 restriction sites. The cps2T gene was tagged with in-frame coding sequence for a Flag tag peptide. The remaining three polysaccharide genes (*cps2*E, *cps2*F and *cps2*G) were sequentially cloned into pGEM resulting in the final construct pGEMKA3. Briefly; the *cps2*E PCR product was cloned into the *Pst*1/*Nco*1 sites of pGEM to generate pGEMKA1 (not shown). The *cps2*F gene PCR product was cloned into the *Nco*1/*Sph*1 sites of pGEMKA1 in frame with *cps2*E to generate pGEMKA2 (not shown). The *cps2*G PCR product was cloned into the *Sph*1/*Apa*1 sites of pGEMKA2 in frame with *cps2*F to generate pGEMKA3. The DNA segment consisting of the three genes was then excised from pGEMKA3 at the two flanking *Pst*1 sites and cloned into the *Pst*1 site of pKAM2. Clones in the correct orientation were fully sequenced to verify that cps2E, 2F and 2G were in frame with *cps2*T, generating the construct pKAM3. In pKAM3, the expression cassette for the four type 2 polysaccharide genes is controlled by CaMV35S promoter. The N-terminal PR1b signal targets the polyprotein to the endoplasmic reticulum.

Amplification of *cps2G* from *S. pneumoniae* genomic DNA was carried out with forward primer KA028 (with *Sph*1 site at 5′ end) and reverse primer KA029 (carrying an E2 tag, stop codon, a *Pst*1 and an *Apa*1 site). The *cps2*G PCR product and construct pGEMKA2 were digested with *Sph*1/*Apa*1 restriction enzymes. The digested gene product was cloned into the *Sph*1/*Apa*1 sites of pGEMKA2 to generate pGEMKA3. In pGEMKA3 the 3483 bp DNA segment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) is flanked by two *Pst*1 sites.

The DNA segment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) was subcloned by excision from pGEMKA3 using *PstI* and ligated into the *PstI* site of pKAM2 in frame with *cps2*T to generate the final binary vector pKAM3 (fig. 1). The Cps2T-2E-2F-2G insert within the pKAM3 construct was fully sequenced by

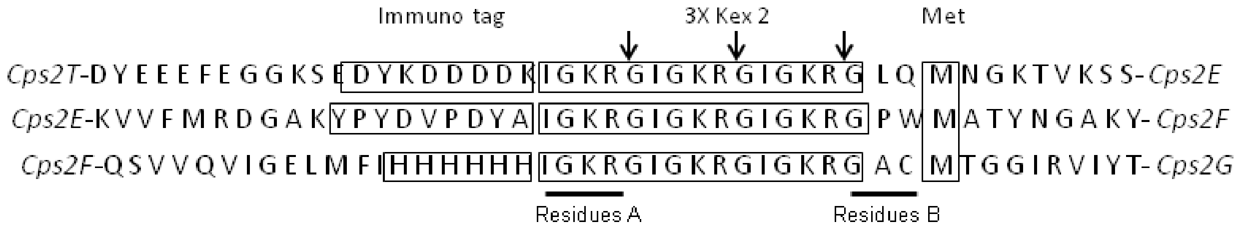

Fig. 2: Transgenic protein N- and C- termini: The junction regions between each transgenic protein are depicted for Cps2T, Cps2E, Cps2F and Cps2G. Depicted are the C terminal immuno tag, the 3 Kex-2 protease digestion peptides, with the proposed cleavage sites shown by an arrow. The first methionine of the authentic bacterially expressed protein is indicated (Met). Residues A indicate the 4 amino acids which will remain at the C terminus of the proteins Cps2T, Cps2E and Cps2F after Kex2 cleavage, residues B highlight the 3 amino acids which will be left at the N- terminus of Cps2E, 2F and 2G after Kex2 cleavage.

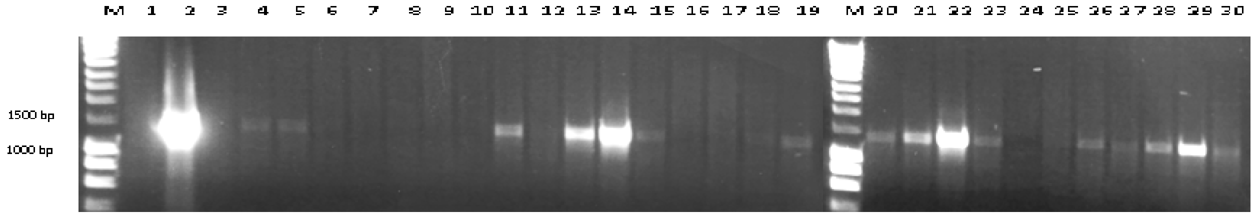

Fig. 3A: PCR analysis of transgene integration in the plant genome (*cps2*E): PCR was carried out with *cps2*E primers KA020 and KA021. lane 1, negative control, untransformed plant; lane 2, positive control pKAM3 plasmid DNA. lane 3-lane 30 contained 28 transgenic lines, (expected PCR product size 1371bp). Hyper ladder 1 (Bioline) (lane M) was used for size comparison.

automated dideoxy Sanger sequencing. All gene sequences were compared and verified as correct against genbank accession AF026471, and additional tag and Kex2 sequences were compared with expected sequence. The border regions of each protein Cps 2T, Cps2E, Cps2F and Cps2G are detailed in fig. 2.

Plant transformations

Tobacco (*Nicotiana tabacumcv* SAMSUN) leaf disks were transformed with the *Agrobacterium* strain GV3101 (carrying the construct and pKAM3) by method of Draper *et al.,* (1988). Leaf discs were placed upside down on MS agar incubated at room temperature for 2 days in the dark then transferred to selective MS agar shoot medium and incubated at 24-26°C in continuous light. Leaf discs were moved to new plates after ten days. Shoot regeneration started 1-2 weeks post infection. Regenerated shoots excised from leaf discs were transferred to MS agar rooting medium containing 3% (w/v) sucrose, 250µg/ml cefotaxime and 100µg/ml kanamycin. After root development, the plantlets were transferred to soil. Plants were allowed to grow until seed production. To avoid cross-fertilization, flower buds were covered with paper bags. For continuous supply of plant leaf material, transgenic plants were clonally propagated from axillary buds in nodal cuttings.

Plant nucleic acid isolation and analysis by PCR

Plant genomic DNA was isolated based on the method of Edwards *et al.,* (1991). Briefly; 100-150mg fresh leaf was frozen in liquid nitrogen, the sample was then ground with an eppendorf grinder and 400µl of PCR extraction buffer (200mM Tris-HCl, pH 7.5; 250mM NaCl; 25mM EDTA; 0.5% SDS) was added. The sample was vortexed, centrifuged at 16100g for 10 min and the supernatant precipitated by addition of 400µl of isopropanol. DNA was recovered by centrifugation at 16100g for 10min. DNA was resuspended in 50µl H_2O and stzred at -20°C. Total RNA was isolated using RNeasy Mini extraction kits (Qiagen), according to the manufacturer's instructions. Nucleic acids were quantified by UV absorbance at 260nm.

PCR and RT-PCR

To confirm transgene integration, genomic DNA from transgenic plant DNA extracts was amplified by PCR comprised of 250µM dNTP mix, 3mM $MgCl_2$, 2.5U Taq polymerase (Bioline), 1µM each primer KA020 and KA021 (within Cps2E) (and 5ng DNA template in a 50µl reaction 1X PCR buffer. Amplification was carried out for 30 cycles consisting of 95°C for 30 seconds, 52°C for 30 seconds and extension at 72°C for 30 seconds.

For the analysis of RNA and the synthesis of first-strand cDNA, 2µg of total plant RNA was mixed with 1µg of gene specific reverse primer (KA018 within the Cps2T gene) in a total volume of 11µl. The mixture was incubated at 70°C for 5 min, chilled on ice for 5 min, and the 3µl of AMV Reverse Transcriptase (Promega), 5µl of AMV Reverse Transcriptase 5 x buffer, 2.5µl of 10mM dNTP mix, 1µl of RNasin Ribonuclease Inhibitor (Promega) and 2.5µl of nuclease-free water was added to the annealed RNA/primer mix in a 25µl reaction volume. The reaction was incubated at 42°C for 60min. Conditions for the amplification of first strand cDNA were the same as that used for the Cps2E amplicon except that primers KA018 and KA024 were used, amplifying a 391bp fragment of Cps2E using 2µl of each cDNA reaction.

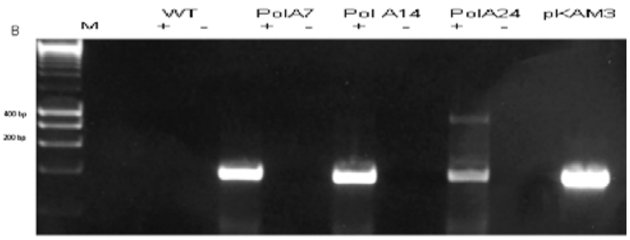

Fig. 3B: RT-PCR to detect transgene expression in different lines: First strand cps2T cDNA was amplified using KA024 and KA018 primers. A 391bp fragment was observed in transgenic line samples (PolA7, PolA14 and PolA24) that underwent reverse transcription reaction prior to PCR and in the positive control pKAM3. There was no amplification from the wild type RNA sample that underwent RT prior to PCR or from any transgenic lines and the wild type RNA sample that had not been reverse transcribed. M; DNA ladder (Bioline) was used for size comparison.

Total plant protein extraction

Fresh leaf material (~0.15g) was ground in liquid nitrogen. Ground tissue was homogenized in 200µl of extraction buffer (100mM Tris, pH 6.8; 10% glycerol; 0.5% SDS; 0.1% Triton X-100; 5mM EDTA, 10mM DTT, 1X protease inhibitor cocktail (Roche) then centrifuged at 16100g at 4°C for 10min. Crude supernatant was transferred to a fresh tube and either used immediately or stored at -20°C.

Western blotting

Proteins were resolved by 12% Tris-Glycine SDS-PAGE and were transferred to either PVDF or nitrocellulose membranes by electro blotting. Membranes were washed in PBS and then blocked in PBS-Tween 0.05%, containing 4% skimmed milk. Membranes were incubated with a dilution of primary antibody, anti Flag (Sigma-Aldrich) at 0.2µg/ml, anti HA (Sigma-Aldrich) at 1/5000 dilution, anti His (Qiagen) at 1/5000 or anti E2 (Abcam) at 1/500 diluted in PBST- 0.5% w/v skimmed milk for 1 h. The membranes were then washed 3 times in PBST-0.5% w/v skimmed milk for 10min before being probed with a 1/2000 dilution of polyclonal goat anti-mouse horse radish peroxidise (HRP) conjugate (Dako Cytomation) in PBST- 0.5% w/v skimmed milk for 75 min. The membranes were then washed 3 times in PBST-0.5% skimmed milk for 10min each. Proteins were detected using the EZ-ECL detection reagent (Geneflow). Proteins on the membrane were detected by exposing the membrane to an ECL Hyperfilm (Amersham) from 30 s to 15min.

RESULTS

Four genes (*cps2*E, *cps2*T, *cps2*F and *cps2*G) were assembled in a single expression cassette driven by the constitutive CaMV 35S promoter. Peptide tags were attached to the C-terminal end of each protein for immuno-detection. Three repeated Kex2 sequences, IGKRGK$_3$, were introduced at the junctions between any two genes, next to the tag sequence of the preceding gene and before the start of the following gene in the expression cassette. The tobacco PR1b endoplasmic reticulum targeting signal sequence (Cutt *et al.,* 1988) was displayed at the 5′ end of the polygene in the expression cassette. The four type 2 polysaccharide synthesis proteins in the form of a single polyprotein were targeted to the endoplasmic reticulum by the PR1b endoplasmic reticulum targeting signal. From the endoplasmic reticulum, the polyprotein would be transported to the Golgi by the default pathway where the Kex2 protease processing would result in the generation of the four type 2 polysaccharide proteins which would then be secreted to the apoplast. The cleavage of the Kex2 tags is expected to leave a four amino acid extension at the C-terminus of proteins CpsT, CpsE and CpsF and three amino acid residues at the N-terminus of CpsE, CpsF and CpsG due to the cleavage of the Kex2 tag and an addition of two residues resulting from the DNA cloning (fig. 2).

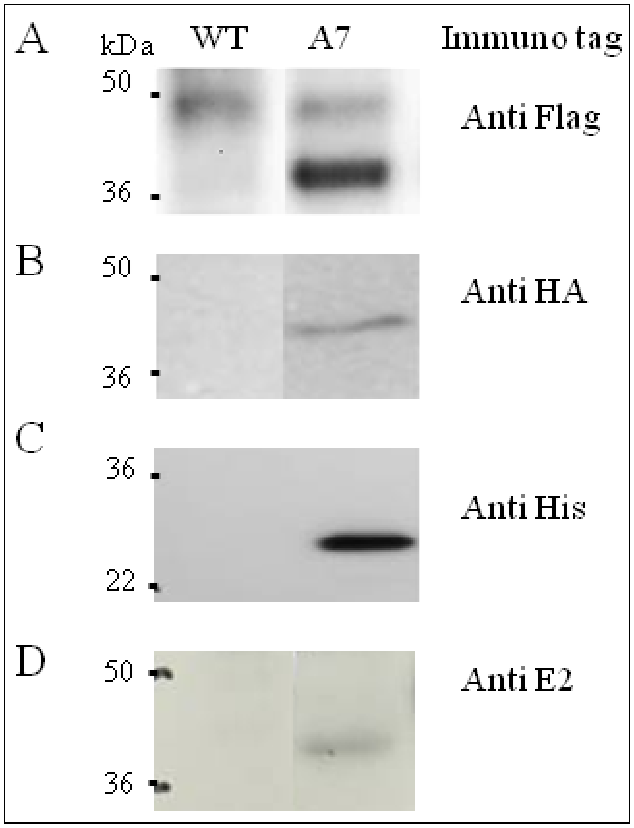

Fig. 4: Western blot detection of transgenic protein product in crude plant extract: Total protein from a transgenic line (PolA7) and one wild type plant were

separated through Tris-Glycine SDS-PAGE and analysed by western blotting. A) A specific band (~41kDa) indicative of the *cps2*T gene product was visualised after probing with anti- flag monoclonal antibody. This band was absent in the wild type plant sample. B) A specific band (~40kDa) indicative of the *cps2*E gene product was visualised after probing with anti- HA monoclonal antibody. This band was absent in the wild type plant sample. C) A specific band (~30kDa) indicative of the *cps2*F gene product was visualised after probing with anti- His monoclonal antibody. This band was absent in the wild type plant sample. D) A specific band with apparent molecular weight of ~40kDa indicative of the *cps2*G gene product was visualised after probing with anti- E2 monoclonal antibody. This band was absent in the wild type plant sample. Molecular mass markers are indicated.

Nicotiana tabacum transformation

Transgenic tobacco lines were produced by *Agrobacterium* mediated transformation of tobacco (*Nicotiana tabacum cv* SAMSUN) leaf disks. Regeneration of shoots from the wounded areas of infected leaf disks was evident 1-2 weeks post infection. After 6-8 weeks, regenerated shoots were excised from leaf discs and transferred to MS agar rooting medium. Rooted transgenic plantlets were transferred to soil and were allowed to grow to maturity. After maturation, self seeds from transgenic lines were harvested. The integration of the four type 2 polysaccharide genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) in the plant genome was confirmed by PCR of genomic DNA from 28 transgenic lines. PCR was carried out with *cps2*E forward primer KA020 and reverse primer KA021 and at least 16 transgenic lines amplified a product of the correct size (fig. 3A). Expression at RNA level was analysed by reverse transcriptase-PCR. An example of RNA level expression is shown in fig. 3B which confirmed the expression of the polycistronic mRNA.

Western blot detection of the four type 2 polysaccharide gene (cps2T, cps2E, cps2F and cps2G) products

Western blot detection of cps2T gene product

Protein was extracted from transgenic lines and a wild type plant control. Detection of the *cps2*T gene product in crude plant extracts was carried out with mouse Anti-Flag M2 monoclonal antibody. A specific band (~41kDa) for *cps2*T was observed in transgenic lines, which was absent in the wild type plant extract. A non-specific band with an apparent molecular weight of about 48 kDa was observed in all samples. Fig. 4A shows an example of this Western blot with the line PolA7.

Western blot detection of cps2E gene product

Protein was extracted from transgenic lines and a wild type plant. Detection of *cps2*E gene product was carried out using an Anti-HA antibody. A specific band (~40 kDa) for the *cps2*E gene product was observed in transgenic lines, which was absent in the wild type plant sample. Fig. 4B shows an example of this Western blot with line PolA7.

Western blot detection of cps2F gene product

Protein was extracted from transgenic lines and a wild type plant and analysed by western blotting detecting *cps2*F gene product using an anti Tetra-His antibody. A specific band (~30kDa) corresponding to *cps2*E gene product was detected in transgenic lines samples, which was absent in the untransformed plant extract control. Fig. 4C shows an example of this Western blot with line PolA7.

Western blot detection of cps2G gene product

Protein was extracted from the transgenic lines and a wild type plant and was analysed by western blotting. Detection of the tagged *cps2*G gene product was carried with an E2 tag antibody. A specific band of around 40 kDa corresponding to *cps2*G protein was detected in the transgenic lines but not in the negative control. Fig. 4D shows an example of this Western blot with line PolA7.

All four of the genes present within the polygene construct were translated and processed into proteins that were specifically immunodetected within the transgenic lines only. Apparent molecular weights were similar to those predicted for the processed gene products.

DISCUSSION

The objective of cloning four pneumococcal type 2 polysaccharide biosynthetic genes was to investigate the potential use of the plant Kex*2* like protease system for the expression of multiple genes from the product of a single polycistronic message. Four type 2 polysaccharide biosynthetic genes from *Streptococcus pneumoniae* were arranged in the form of a single polygene within an expression cassette. The genes were linked end to end separated by three tandemly repeated Kex2 protease digestion sequences for Golgi localized processing by endogenous tobacco Kex2 like activity. The tobacco PR1b endoplasmic reticulum targeting signal sequence was attached to the 5′ end of the polygene. Transgenic lines carrying all four genes were produced and analysed for the production of type 2 polysaccharide biosynthetic enzymes. Expression of these genes in transgenic tobacco was confirmed by RT-PCR and western blotting, the latter specifically detecting all four proteins in crude plant extracts. Detection of all the four type 2 polysaccharide biosynthetic genes protein products suggests that there was correct processing by the Golgi localized Kex2 protease system (Jiang and Rogers, 1999; Haldar, 2016). Using total protein extracts, bands of higher molecular weight of around 150kDa indicative full length unprocessed polyprotein, and intermediate molecular weight species, indicative of processing intermediates, were not seen in Western blots suggesting that the Kex2 processing of the polyprotein is complete.

In the current example, the Kex2 protease site from *Ustilago Maydis* KP6 preprotoxin was displayed as a tripartite linker ($IGKRG_3$) between four pneumococcal type 2 polysaccharide biosynthetic genes in the one expression cassette. The display of the Kex2 protease sites in this format has been shown to improve the access of the Kex2 protease to its substrate within the fusion protein (Jiang and Rogers, 1999). All four type 2 polysaccharide biosynthetic proteins were shown to be processed. This assumes that Kex2 processing is efficient and that processing occurs at all 3 Kex2 recognition sites. The expressed proteins possess the activities, Rhamnosyl transferase, Undecaprenyl-phosphate glucose-1-phosphate transferase, galactosyl transferase and Undecaprenly-phosphate galactose phosphotransferase. These activities are not straightforward to assay biochemically and there are also no immunoreagents readily available for specific detection of each enzyme to demonstrate that each have been correctly expressed and processed. Immunotags were therefore incorporated into each expressed protein in order to detect the processed proteins.

The demonstration of four proteins expressed as a single polyprotein and processed via a Kex 2 like activity extends the utility of the Kex2 system for simultaneous expression of four proteins and its use for multiple heterologous protein production in plants. The Kex2 protease system could be useful for producing plants with multigenic traits, improving or altering metabolic pathways, expressing multimeric foreign proteins and expressing multiple enzymes involved in the synthesis of various compounds that have a pharmaceutical/ commercial interest.

CONCLUSION

In comparison to the existing multigene transformation procedures like IRESs, multi-transformation, sequential transformation, and co-transformation, the use of the Kex2 protease system for multigene engineering could represent a cost effective, less labour intensive method for plant engineering. All of the genes of interest could be integrated at one locus simplifying post-transformation selection and evaluation procedures to get stable transgenic lines expressing all the genes of interest. This strategy could be useful in the production of pharmaceutically important multimeric proteins. The system could also be used for simultaneously improving more than one characteristic of a plant such as resistance to multiple diseases achieved in a single transformation event. Plant metabolic pathways that are of interest for genetic manipulation include carbohydrate metabolism, amino acid and polyamine metabolism, lipid metabolism, and secondary metabolic pathways such as the production of alkaloids, terpenoids, flavonoids, lignins and quinines (Capell and Christou, 2004). There are diverse arrays of opportunities where the Kex2 protease system could be utilized for the expression of multiple genes to improve or alter plant characteristics and to develop multigenic therapeutics.

ACKNOWLEDGEMENTS

We acknowledge the contribution of the late Professor Garry Whitelam, Department of Biology, The University of Leicester to the design and planning of this project.

REFERENCES

Belmont AS (2016). Recombinant gene expression. *U.S Patent*, **9**: 273-324.

Capell T and Christou P (2004). Progress in plant metabolic engineering. *Curr. Opin. Biotech.*, **15**: 148-154.

Cutt JR, Dixon DC, Carr JP and Klessig DF (1988). Isolation and nucleotide sequence of cDNA clones for the pathogenesis-related proteins PR1a, PR1b and PR1c of *Nicotiana tabacum cv. Xanthi NC* induced by TMV infection. *Nucleic Acids Res.*, **16**: 9861-9861.

Daniell H, Dhingra A (2002). Multigene engineering: Dawn of an exciting new era in biotechnology. *Curr. Opin. Biotech.*, **13**: 136-141.

De Vos AF, Dessing MC, Lammers AJ, De Porto AP, Florquin S, De Boer OJ, De Beer R, Terpstra S, Bootsma HJ, Hermans PW and Vant Veer C (2016). The polysaccharide capsule of *Streptococcus pneumonia* partially impedes MyD88-mediated immunity during pneumonia in mice. *PloS. one*., p.e10: 0118181

Draper J, Scott R and Hamil J (1988). *In*: Plant genetic transformation and gene expression (A laboratory manual). Edited by Draper J, Scott R, Armitage P and Walden R. Blackwell Scientific Publications; pp.69-160.

Edwards K, Johnstone C and Thompson C (1991). simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.*, **19**: 1349.

Fuller RS, Brake A and Thorner J (1989). Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2+-dependent serine protease. *Proc. Nat. Acad. Sci.*, **86**: 1434-1438.

Garcia E and Lopez R (1997). Molecular biology of the capsular genes of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.*, **149**: 1-10.

Gouiaa S, Khoudi H, Leidi EO, Pardo JM and Masmoudi K (2012). Expression of wheat Na+/H+ antiporter TNHXS1 and H+-pyrophosphatase TVP1 genes in tobacco from a bicistronic transcriptional unit improves salt tolerance. *Plant Mol. Biol.*, **79**: 137-155.

Guidolin A, Morona JK, Morona R, Hansman D and Paton JC (1994). Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. *Infect. Immun.*, **62**: 5384-5396.

Haldar K (2016). Protein trafficking in apicomplexan parasites: Crossing the vacuolar Rubicon. *Curr. Opin. Microbiol.* **32**: 38-45.

Hausdorff WP, Hoet B and Adegbola RA (2015). Predicting the impact of new pneumococcal conjugate vaccines: serotype composition is not enough. *Expert. Rev. Vaccines*. **14**: 413-428.

Hayflick JS, Wolfgang WJ, Forte MA and Thomas G (1992). A unique Kex2-like endoprotease from Drosophila melanogaster is expressed in the central nervous system during early embryogenesis. *J. Neurosci.,* **12**: 705-17.

Hiatt A, Cafferkey R and Bowdish K (1989). Production of antibodies in transgenic plants. *Nature,* **342**: 76-78.

Houdebine LM and Attal J (1999). Internal ribosome entry sites (IRESs): reality and use. *Transgenic Res.,* **8**: 157-177.

Iannelli F, Pearce BJ and Pozzi G (1999). The Type 2 Capsule Locus of Streptococcus pneumoniae. *J. Bacteriol.,* **181**: 2652-2654

Jansson PE, Lindberg B, Andersson M, Lindquist U and Henrichsen J (1975). The structure of capsular polysaccharide of the pneumococcus type II. *Carbohydr. Res.,* **40**: 69-75.

Jedrzejas MJ (2004). Extra cellular virulence factors of *Streptococcus pneumoniae*. *Front. Biosci.,* **9**: 891-914.

Jiang L and Rogers JC (1999). Functional analysis of a Golgi-localized Kex2p-like protease in tobacco suspension culture cells. *Plant J.,* **18**: 23-32

Julius D, Brake A, Blair L, Kunisawa R and Thorner J (1984). Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-α-factor. *Cell.,* **37**: 1075-1089.

Kinal H, Park CM, Berry JO, Koltin Y and Bruenn JA (1995). Processing and secretion of a virally encoded antifungal toxin in transgenic tobacco plants: Evidence for a Kex2p pathway in plants. *Plant Cell,* **7**: 677-688.

Kolkman MAB, Van Der Zeijst BAM and Nuijten PJM (1997). Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14. *J. Biol. Chem.,* **272**: 19502-19508.

Luo Y, Li BZ, Liu D, Zhang L, Chen Y, Jia B, Zeng BX, Zhao H and Yuan YJ (2015). Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev*., **44**: 5265-5290.

Ma JKC, Drake PMW, Chargelegue D, Obregon P and Prada A (2005). Antibody processing and engineering in plants, and new strategies for vaccine production. *Vaccine,* **23**: 1814-1818.

Meyer P and Saedler H (1996). Homology-dependent gene silencing in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.,* **47**: 23-48.

Morona JK, Morona R and Paton JC (1997). Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. *Mol. Microbiol.,* **23**: 751-763.

Moxon E and Kroll J (1990). The role of bacterial polysaccharide capsules as virulence factors. *Curr. Top. Microbiol. Immunol.,* **150**: 65-85.

Munoz R, Mollerach M, Lopez R and Garcia E (1997). Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: Formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. *Mol. Microbiol.,* **25**: 79-92.

Nassau PM, Martin SL, Brown RE, Weston A, Monsey D, McNeil MR and Duncan K (1996). Galactofuranose biosynthesis in Escherichia coli K-12: Identification and cloning of UDP-galactopyranose mutase. *J. Bacteriol.,* **178**: 1047-1052.

Nawrath C, Poirier Y and Somerville C (1994). Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of arabidopsis thaliana results in high levels of polymer accumulation. *Proc. Nat. Sci.,* **91**: 12760-12764.

Ninh PH, Honda K, Sakai T, Okano K and Ohtake H (2015). Assembly and multiple gene expression of thermophilic enzymes in *Escherichia coli* for *in vitro* metabolic engineering. *Biotechnol. Bioeng.,* **112**: 189-196.

Roongsawang N, Puseenam A, Kitikhun S, Sae-Tang K, Harnpicharnchai P, Ohashi T, Fujiyama K, Tirasophon W and Tanapongpipat S (2016). A novel potential signal peptide sequence and over expression of ER-resident chaperones enhance heterologous protein secretion in thermotolerant methylotrophic yeast *Ogataea thermomethanolica*. *Appl. Biochem. Biotechnol.,* **178**: 710-724.

Sarah M, Richardson SJ, Wheelan RM, Yarrington and Boeke JD (2006). Gene Design: Rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**: 550-556.

Saxena P, Thuenemann EC, Sainsbury F and Lomonossoff GP (2016). Virus derived vectors for the expression of multiple proteins in plants. Recombinant proteins from Plants: Methods and protocols, pp.39-54.

Slater S, Mitsky TA, Houmiel KL, Hao M, Reiser SE, Taylor NB, Tran M, Valentin HE, Rodriguez DJ and Stone DA (1999). Metabolic engineering of Arabidopsis and Brassica for poly (3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. *Nat. Biotechnol.,* **17**: 1011-1016.

Snippe H, Van Houte AJ, Van Dam JE, De Reuver MJ, Jansze M and Willers JM (1983). Immunogenic properties in mice of hexasaccharide from the capsular polysaccharide of *Streptococcus pneumoniae* type 3. *Infect. Immun.,* **40**: 856-861.

Tao J, Ginsberg I, Banerjee N, Held W, Koltin Y and Bruenn JA (1990). Ustilago maydis KP6 killer toxin: structure, expression in Saccharomyces cerevisiae and relationship to other cellular toxins. *Mol. Cell Biol.,* **10**: 1373-1381.

Thomas L, Leduc R, Thorne BA, Smeekens SP, Steiner DF and Thomas G (1991). Kex2-like endoproteases PC2 and PC3 accurately cleave a model prohormone in mammalian cells: Evidence for a common core of neuroendocrine processing enzymes. *Proc. Nat. Acad. Sc.,* **88**: 5297-5301.

World Health Organisation http://www.who.int/nuvi/pneumococcus/en/

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P and Potrykus I (2000). Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science,* **287**: 303-305.

Zhang S and Fauquet M (1998). Expression and inheritance of multiple transgenes in rice plants. *Nat. Biotechnol.,* **16**: 1060-1064.