Author/s (year)	[redacted]
Title (version Nº)	**Protein Stability of *Pichia pastoris* ω3-/Δ15-Desaturase**
Owner	**Nuseed Pty Ltd**
Date	**December 15, 2016**
Project	**Omega 3 Canola**
Report Number	**2016-012**
Testing Facility	**CSIRO Agriculture & Food** **Queensland Biosciences Precinct** **St Lucia, Brisbane, Queensland 4067** **Australia**
Test method	**LC-MRM-MS**
GLP	**No GLP**
Confidentiality	**None**

TITLE:
PROTEIN STABILITY OF *PICHIA PASTORIS* ω3-/Δ15-DESATURASE

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

6500 QTRAP	AB SCIEX 6500 QTRAP LC-MS/MS system
ALA	α-Linoleic acid, 18:3$^{\Delta 9,12,15}$ (ω3)
CE	Collision energy
DHA	Docosahexaenoic acid, 22:6$^{\Delta 4,7,10,13,16,19}$ (ω3)
DHA canola	Genetically modified canola, event NS-B50027-4
DPA	Docosapentaenoic acid, 22:5$^{\Delta 7,10,13,16,19}$ (ω3)
DTT	Dithiothreitol
EPA	Eicosapentaenoic acid, 20:5$^{\Delta 5,8,11,14,17}$ (ω3)
ETA	Eicosatetraenoic acid, 20:4$^{\Delta 8,11,14,17}$ (ω3)
FA	Formic acid
FASP	Filter-assisted sample preparation
FDR	False discovery rate
GFP	Green fluorescent protein
HPLC	High performance liquid chromatography
IAM	Iodoacetamide
kDa	Kilo dalton
LA	Linoleic acid, 18:2$^{\Delta 9,12}$ (ω6)
Lackl-Δ12D	*Lachancea kluyveri* Δ12-desaturase
LC-MS	Liquid chromatography-mass spectrometry
Micpu-Δ6D	*Micromonas pusilla* Δ6-desaturase
MMT	Million metric tons
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
m/z	Mass-to-charge ratio
PBS	Phosphate buffer saline
OA	Oleic acid, 18:1$^{\Delta 9}$
OD_{600}	Optical density at 600 nm wavelength
ω3 LC-PUFA	Omega-3 long-chain (≥C20) polyunsaturated fatty acids
Pavsa-Δ4D	*Pavlova salina* Δ4-desaturase
Pavsa-Δ5D	*Pavlova salina* Δ5-desaturase
Picpa-ω3D	*Pichia pastoris* Δ15-/ω3-desaturase
PMSF	Phenylmethylsulfonyl fluoride
Pyrco-Δ5E	*Pyramimonas cordata* Δ5-elongase
Pyrco-Δ6E	*Pyramimonas cordata* Δ6-elongase
Q1	Quadrupole 1 (referring to the analysis of the precursor ion)

Q3	Quadrupole 3 (referring to the analysis of the fragment ion)
RT	Retention time (min)
SD	Standard deviation
SDA	Stearidonic acid, $18{:}4^{\Delta 6,9,12,15}$ (ω3)
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SGF	Simulated gastric fluid
UA buffer	8 M urea, 0.1 M Tris-HCl, pH 8.5

EXECUTIVE SUMMARY

The purpose of this report was to assess the *in vitro* stability of the *Pichia pastoris* ω3-/Δ15-desaturase (Picpa-ω3D) protein in simulated gastric fluid (SGF) comprising the proteolytic enzyme, pepsin, and in combination with a novel pepsin-trypsin assay employing state-of-the-art mass spectrometric approaches to monitor the precise degradation products. The extent of protein digestion was evaluated by the appearance of peptic products and the disappearance of tryptic peptide products (as a proxy for intact protein). The allergenic potential of a protein is determined by a weight of evidence approach since no single method can predict the allergenicity of a protein. Protein digestibility is one aspect of the overall allergenicity assessment that is conducted for newly introduced proteins into genetically modified crops.

The results of the study demonstrated that greater than 80% was digested within 5 min and greater than 97% of the full-length Picpa-ω3D protein was digested within 60 min of incubation in pepsin when analysed by LC-MS/MS.

The results of this study show that the integral membrane protein Picpa-ω3D was readily digestible in pepsin and/or trypsin. Rapid digestion of the full-length protein is one of many factors that indicate protein safety.

I. INTRODUCTION

The omega-3 long-chain (≥C20) polyunsaturated fatty acids (ω3 LC-PUFA) eicosapentaenoic acid (EPA, 20:5ω3), docosapentaenoic acid (DPA, 22:5ω3) and docosahexaenoic acid (DHA, 22:6ω3) are widely recognised for their beneficial roles in human health, particularly those related to cardiovascular and inflammatory health. EPA, DPA and DHA are primarily sourced from wild-caught fish oils and algal oils, with algae being the primary producer in the marine food web. These sources are under pressure due to increasing demand for ω3 LC-PUFA by aquaculture, nutraceutical and pharmaceutical applications. Additional sources of these fatty acids can be produced by engineering land-based oilseed crops to convert native fatty acids to marine-type ω3 LC-PUFA which are then accumulated in seed oil. Canola is a commonly grown oilseed with 67 million metric tons (MMT) of rapeseed produced globally in 2015/16[1].

In collaboration with the Commonwealth Scientific and Industrial Research Organization (CSIRO), Nuseed Pty Ltd has developed genetically modified canola event NS-B50027-4 (DHA canola), which accumulates significant amounts of DHA in the seed oil.

[1] http://www.ers.usda.gov/data-products/oil-crops-yearbook/oil-crops-yearbook/#World Supply and Use of Oilseeds and Oilseed Products

In this DHA canola, seven fatty acid desaturases and elongases were introduced to convert oleic acid (OA) to DHA in a single pathway expression vector (Figure 1). The pathway consisted of the *Lachancea kluyveri* Δ12-desaturase (Lackl-Δ12D, Watanabe et al. 2004), *Pichia pastoris* Δ15-/ω3-desaturase (Picpa-ω3D, Zhang et al. 2008), *Micromonas pusilla* Δ6-desaturase (Micpu-Δ6D, Petrie et al. 2010b), *Pyramimonas cordata* Δ6-elongase (Pyrco-Δ6E, Petrie et al. 2010a), *Pavlova salina* Δ5-desaturase (Pavsa-Δ5D, Zhou et al. 2007), *P. cordata* Δ5-elongase (Pyrco-Δ5E, Petrie et al. 2010a) and *P. salina* Δ4-desaturase (Pavsa-Δ4D, Zhou et al. 2007). The functionalities and activities of these enzymes have been demonstrated in different heterologous expression systems (see Report Nºs 2016-005, 2016-006, 2016-007, 2016-008, 2016-009, 2016-010, 2010-011) and in transgenic Arabidopsis or Camelina seeds (Petrie et al. 2012, 2014). Based on the sequence similarity and functionality, these seven proteins can be classified into three groups, (1) yeast acyl-CoA type fatty acid desaturases including Lackl-Δ12D and Picpa-ω3D (Figure 1, blue) that introduces a double bond at the Δ12 and Δ15 positions, respectively, (2) algae fatty acid elongases including Pyrco-Δ6E and Pyrco-Δ5E (Figure 1, purple) that add a carbon to the carboxyl end of fatty acids, and (3) algae front-end fatty acid desaturases that introduce a double bond between an existing double bond and the carboxyl end of fatty acids (Zhou et al. 2007) including Micpu-Δ6D, Pavsa-Δ5D and Pavsa-Δ4D (Figure 1, green). One representative from each of these three groups was analysed for protein stability.

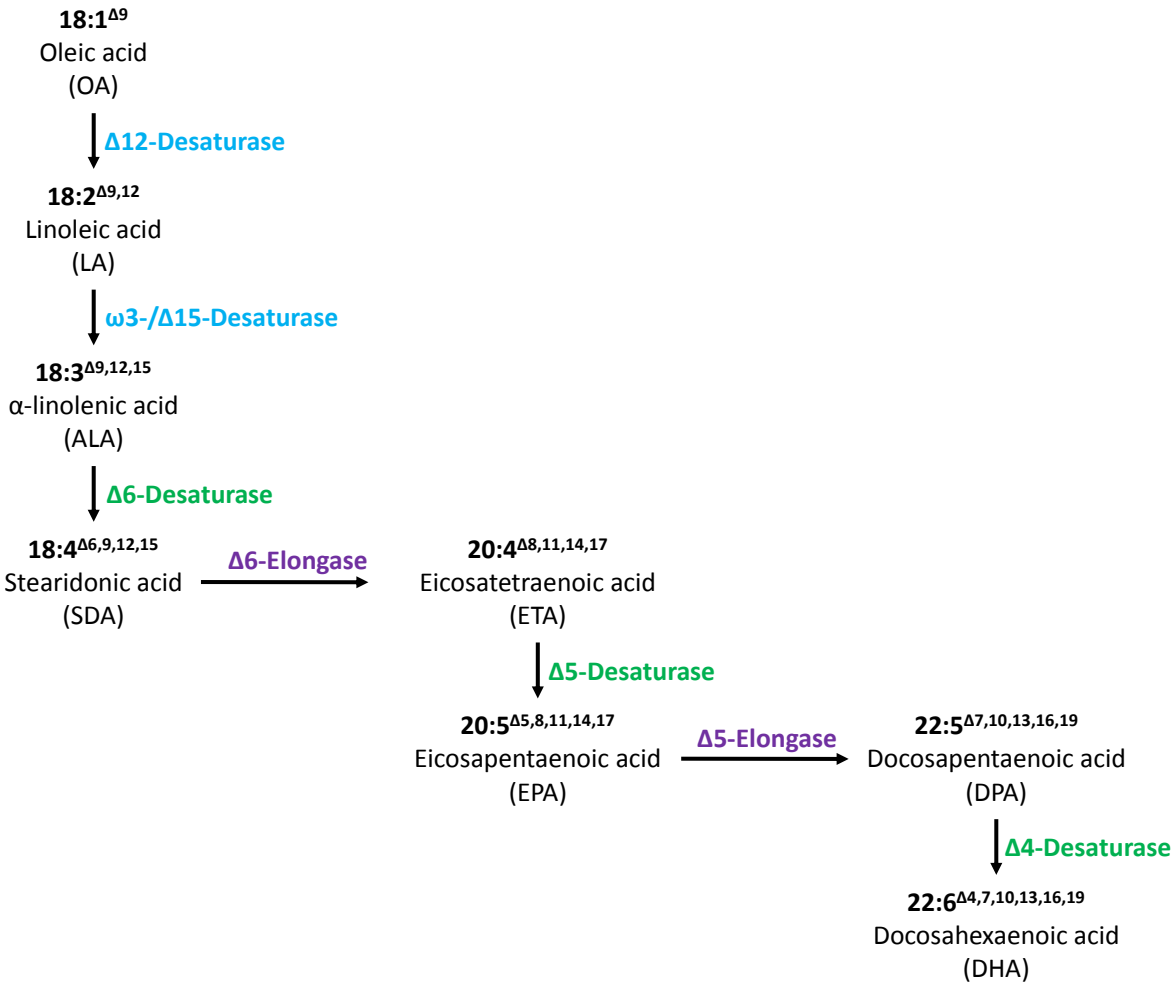

Figure 1. DHA biosynthesis pathway engineered into DHA canola event NS-B50027-4. Seven enzymes introduced in canola to convert oleic acid to the final product docosahexaenoic acid were grouped into three classes: two fatty acid desaturases from yeast in blue, two elongases from microalgae in purple, and three front-end desaturases from microalgae in green (see text for detail).

II. PURPOSE

The likelihood of allergic oral sensitization to a protein is first affected by the stability of the protein to gastrointestinal digestion (Astwood et al. 1996). The purpose of this study was to assess the *in vitro* digestibility of the fatty acid biosynthesis enzymes introduced into DHA canola by digesting with pepsin. In the absence of functioning antibodies against these integral membrane proteins, as typically used for traditional Western blot analysis, a sensitive LC-MS analytical method was developed.

This particular report focuses on the yeast acyl-CoA type fatty acid desaturase representative *P. pastoria* ω3-/Δ15-desaturase (Picpa-ω3D) protein, which was used in the engineering of DHA canola NS-B50027-4 to catalyse the desaturation of linoleic acid (LA) into α-linoleic acid (ALA; $18{:}2^{\Delta 9,12} \rightarrow 18{:}3^{\Delta 9,12,15}$).

III. MATERIALS

A. TARGET PROTEIN

The ω3-/Δ15-desaturase gene used in DHA canola was previously cloned from alga *P. pastoria* (see Report Nº 2016-006 for details). The Picpa-ω3D protein was expressed in *E. coli* C41 strain as a fusion protein with green fluorescent protein (GFP) followed by 8 histidine residues (8 x His) at the N-terminus of the protein (His_8::GFP::Picpa-ω3D) and then purified.

B. OTHER MATERIALS

Sequencing grade porcine trypsin and a highly purified form of pepsin (Catalogue number V195A, 2,500 units/mg) were purchased from Promega (Madison, USA). Mouse anti-His antibody (Catalogue number A7058) was purchased from Sigma-Aldrich (Sydney, Australia).

IV. METHODS

A. PROTEIN EXTRACTION

The Picpa-ω3D protein was expressed in *E. coli* C41 strain as a fusion protein with 8 histidine residues (8 x His) followed by GFP at the N-terminus of the protein (His_8::GFP:: Picpa-ω3D). Cells were grown to an OD_{600} of 0.8 and protein expression was induced with 0.5 mM IPTG at 37°C for 4 h. Cells were then spun down and resuspended in lysis buffer (150 mL per 60 g cell paste) containing 20 mM Hepes pH 7.6, 150 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, 3 Ultra complete protease inhibitor tablets per 150 mL (Roche), 1 mM PMSF, 1 mM DTT and 1200 units of Benzonase (Merck Millipore). Cells were lysed using EmulsiFlexC5 cell homogeniser (Avestin) by three passes at 15,000 psi. After lysis cellular debris was removed by centrifugation, the supernatant was centrifuged at 200,000 x *g* for 90 min at 4°C to isolate the membrane fraction. The membrane pellet was resuspended in 50 mL of HNG buffer (20 mM Hepes, 150 mM NaCl, 10% glycerol, pH 7.6). To solubilise His_8::GFP::Picpa-ω3D from the membrane fraction 1% (w/v) FosCholine-16 (Glycon Biochemicals GmbH) was added and the mixture incubated for 3 h at 4°C. The mixture was then centrifuged for 45 min at 200,000 x *g* at 4°C and the supernatant loaded on a 5 mL HisTrap FF column (GE Healthcare, Australia) in the presence of 10 mM imidazole and 1 mM DTT. The protein was eluted with an imidazole gradient. Fractions were analysed by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with Western blotting. Fractions containing His_8::GFP::Picpa-ω3D fusion protein were pooled and concentrated to 2.5 mL using 100 kDa MWCO concentrators (Millipore). Concentrated sample was injected onto a Superdex 200 16/60 pg gel filtration column (GE Healthcare) equilibrated in HNG buffer in the presence of 0.01% FosCholine-16 and 1 mM DTT. Fractions containing purified His_8::GFP::Picpa-ω3D protein were pooled, concentrated to 1.3 mg/mL, flash frozen in liquid nitrogen and stored at -80°C. Concentrated protein was analysed by SDS-PAGE and Western blotting using an anti-His HRP conjugated antibody (A7058, Sigma-Aldrich) (see Figure 4). The estimated purity was ~90%.

B. METHOD DEVELOPMENT: LC-MS CHARACTERIZATION OF THE PROTEIN AFTER TRYPSIN DIGESTION

The His_8::GFP::Picpa-ω3D protein was diluted in UA buffer (8 M urea, 0.1 M Tris-HCl, pH 8.5) to ~0.125 µg/µL. An aliquot of the protein extract (equivalent to ~25 µg) was subjected to filter-assisted sample preparation (FASP, Wisniewski et al. 2009). The protein extract was applied to a 10 kDa molecular weight cut-off (MWCO) filter (Millipore, Australia), washed with two 200 µL volumes of UA buffer with centrifugation (20,800 x *g*, 15 min). To reduce the protein on the filter, dithiothreitol (50 mM, 200 µL) was added and the solution was incubated at room temperature for 50 min with shaking. The filter was washed with two 200 µL volumes of UA buffer with centrifugation (20,800 x *g*, 15 min). To alkylate the cysteine residues,

iodoacetamide (50 mM, 100 uL) was applied to the protein on the filter with incubation for 40 min at room temperature in the dark. The filter was washed with two 200 µL volumes of UA buffer with centrifugation (20,800 x *g*, 15 min). The buffer was exchanged using 50 mM ammonium bicarbonate (pH 8.0) by two consecutive wash/centrifugation steps. Sequencing grade porcine trypsin at a concentration of 0.01 µg/µL (2 µg in 200 µL of 50 mM ammonium bicarbonate with 1 mM $CaCl_2$) was added to the protein on the 10 kDa filters and incubated for 16 h at 37ºC in a wet chamber. The filter was transferred to a fresh centrifuge tubes and the filtrate (digested peptides) was collected following centrifugation (20,800 x *g*, 10 min). The filters were washed with 200 µL of 100 mM ammonium bicarbonate and the filtrates were combined and lyophilised. The tryptic peptides were resuspended in 50 µL of 1% formic acid (FA) and 10 µL was injected on the LC-MS/MS system.

C. METHOD DEVELOPMENT: LC-MS CHARACTERIZATION OF THE PROTEIN AFTER PEPSIN DIGESTION

The His_8::GFP::Picpa-ω3D protein was diluted in UA buffer to ~0.125 µg/µL. An aliquot of the protein extract (equivalent to ~25 µg) was subjected to FASP. The protein extract was applied to a 10 kDa MWCO filter (Millipore), washed with two 200 µL volumes of UA buffer with centrifugation (20,800 x g, 15 min). The buffer was exchanged using 50 mM ammonium bicarbonate (pH 8.0) by two consecutive wash/centrifugation steps. The pH was adjusted by further washing with acidified 50 mM ammonium bicarbonate (pH 1.2) by two consecutive wash/centrifugation steps. The 10 kDa filter was transferred to a fresh centrifuge tube and 75 µg pepsin (150 µL, 0.5 µg/µL in 50 mM ammonium bicarbonate, pH 1.2) was added to obtain an enzyme to protein ratio of 3:1. The filter was incubated at 37°C. After 120 min the filter was transferred to a clean tube. The filtrate (containing the digested peptides) was collected following centrifugation (20,800 x g, 10 min). The filters were washed with 200 µL of 100 mM ammonium bicarbonate and the filtrates were combined and lyophilised and stored at -20°C until analysis. The resultant peptides were reconstituted in 50 µL of 1% FA of which 10 µL was analysed by LC-MS/MS.

D. METHOD DEVELOPMENT: LC-MS ANALYSIS

Proteolytically digested (either pepsin or trypsin) protein were analysed as described previously (Colgrave et al. 2014) with chromatographic separation (2%/min linear gradient from 2-40% acetonitrile) using a nano HPLC system (Shimadzu Scientific, Rydalmere, Australia) directly coupled to a TripleTOF 5600 MS (AB SCIEX, Foster City CA, USA). ProteinPilot™ 4.0 software (AB SCIEX) with the Paragon Algorithm (Shilov et al. 2007) was used for protein identification. Tandem mass spectrometry data was searched against *in silico* tryptic digests of a custom-built database. The database (57,652 sequences) comprised the *E.*

coli proteins of the Uniprot-KB database (version 2016/02) appended with the transgenic proteins and additionally with a database of contaminant proteins (known as the common repository of adventitious proteins). The search parameters were defined as: (1) no modification to cysteine and pepsin as the digestion enzyme; or (2) iodoacetamide modified for cysteine alkylation and trypsin as the digestion enzyme. Additional modifications and cleavages were defined previously (Colgrave et al. 2014). The database search results were manually curated to yield the protein identifications using a 1% global false discovery rate (FDR) determined by the in-built FDR tool within ProteinPilot software (Tang et al. 2008).

E. IDENTIFICATION OF PROTOTYPIC PEPTIDES FOR PROTEIN DIGESTIBILITY

For the tryptic data, peptide summaries generated by ProteinPilot were used to select peptides that yielded intense peaks and were fully tryptic, *i.e.* no unusual or missed cleavages. For the pepsin data, peptide summaries generated by ProteinPilot were used to select peptides that yielded intense peaks after 120 min incubation with pepsin. As pepsin is non-specific, many of these peptide products were overlapping or contained missed cleavages. MRM transitions (Tables 1-2) were determined for each peptide where the precursor ion (Q1) *m/z* and the fragment ion (Q3) *m/z* values were determined from the data collected in the discovery experiments. Three transitions were used per peptide (with 11 peptic and 8 tryptic peptides from His_8::GFP::Picpa-ω3D), wherein the peak area of the three MRM transitions were summed.

F. DIGESTIBILITY ASSAY

Two test systems, pepsin digestion (representing simulated gastric fluid, SGF) and a combined pepsin-trypsin digestion, were utilized independently to test the stability of the His_8::GFP::Picpa-ω3D protein. SGF contained the proteolytic enzyme pepsin in a buffer adjusted to an acidic pH 1.2, using a highly purified form of pepsin. The SGF was formulated so that an enzyme:protein ratio of 3:1 would be present in the digestion reactions. The digestion of the Picpa-ω3D protein was monitored by LC-MS/MS (as described below).

G. JUSTIFICATION FOR SELECTION OF THE PEPSIN TEST SYSTEM

In vitro digestion models are used widely to assess the nutritional value of ingested proteins based on their amino acid bioavailability. The correlation between protein allergenicity and protein stability in an *in vitro* pepsin digestion assay has been previously established (Astwood et al. 1996). When proteins are found to be highly digestible, the potential for systemic exposure is reduced. The current safety assessment strategy (Codex 2003) is based on a weight-of-evidence approach that recognizes that no single endpoint can predict human allergenicity potential. Based on this strategy, a number of factors are evaluated: the gene

source, determining the similarity of amino acid sequence of the newly expressed protein to known allergens, the abundance of the protein in the crop, and the digestibility of the protein to *in vitro* digestion. The pepsin digestibility assay protocol used in this study was based on the protocol standardized by the International Life Sciences Institute (ILSI) in a multi-laboratory test. The results demonstrated that the *in vitro* pepsin digestion assay is reproducible when a common protocol is followed (Thomas et al. 2004).

H. JUSTIFICATION FOR DESIGN AND APPLICATION OF THE COMBINED PEPSIN-TRYPSIN TEST SYSTEM

The complete digestion of a protein by a single enzyme is difficult to judge, especially when employing a non-specific enzyme such as pepsin. While it is possible to judge the disappearance of the intact protein on a gel or by Western blotting techniques, the protein may be hydrolysed once (cleaved at a single site) or multiple times often resulting in small and overlapping fragments. Allergenic reactions require that a protein or protein fragment simultaneously bind to two IgE molecules in order to induce mast cell degranulation (Goodman 2008). This IgE binding places theoretical limits on the peptide size of between 1500 and 3500 Da. Gel analysis with various staining or antibody techniques is typically able to detect peptides down to approximately 3,000 Da. When employing gel analysis solely, to judge the completeness of digestion, a high level of purity is required. When employing antibodies, the hydrolysis of a protein by a proteolytic enzyme may result in cleavage of the antigenic site (epitope) thus rendering antibody-based detection methods unsuitable. Likewise, cleavage of a protein at a single site may yield two protein fragments, in which one may contain the epitope (recognised by a monoclonal antibody) while the other does not. In this instance, large protein fragments may evade detection.

By using LC-MS/MS analysis, the peptide products resulting from both pepsin and trypsin digestions could first be determined qualitatively and then subsequently a quantitative LC-MS/MS for the detection of these peptide fragments was developed. LC-MS analysis is capable of simultaneously monitoring peptides spanning the entire protein sequence that are generated by proteolytic digestion. The approach to analyse digestibility in this study mimics the typical mammalian digestive system that exposes food proteins to both pepsin (stomach) and trypsin (intestine) enzymes in transit through the gut.

I. PEPSIN DIGESTION

Twenty-five µg of protein (58.4 µL, n=30 comprising 5 replicate digestions and 6 time points) were applied to a 10 kDa molecular weight cut-off filter (Millipore, Australia), washed twice with 200 µL volumes of UA buffer with centrifugation (20,800 x *g*, 15 min). The buffer was

exchanged using 50 mM ammonium bicarbonate (pH 8.0) by two consecutive wash/centrifugation steps. The pH was adjusted by further washing with acidified 50 mM ammonium bicarbonate (pH 1.2) by two consecutive wash/centrifugation steps. The 10 kDa filters were transferred to fresh centrifuge tubes and 84 µg pepsin (150 µL, 0.562 µg/mL in 50 mM ammonium bicarbonate (pH 1.2) was added to obtain an enzyme to protein ratio of 3:1. The replicate tubes were incubated at 37°C for 5 time points (5, 10, 15, 30 and 60 mins). Pepsin was not applied to the 0 time point, which served as an experimental control for acid hydrolysis. The digestion was stopped by the addition of 200 µL of 50 mM ammonium bicarbonate (pH 8.0), which served to irreversibly inactivate the enzyme. The 10 kDa filters were immediately centrifuged (20,800 x *g*, 15 min) and the filtrate containing digested peptides were collected. The filters were washed twice with 200 µL of 50 mM ammonium bicarbonate (pH 8.0) and the filtrates were combined, lyophilised and stored in a -80°C freezer until analyzed. The peptic peptides were resuspended in 50 µL of 1% FA and a 3 µL aliquot run on the 6500 QTRAP LC-MS system and quantified.

J. TRYPSIN DIGESTION

The 10 kDa filters were transferred to fresh centrifuge tubes and the residual protein reduced with 200 µL of 50 mM DTT, 50 mM ammonium bicarbonate (pH 8.5) on mixer at 600 rpm for 45 min prior to centrifugation (20,800 x *g*, 15 min). The protein was alkylated with 200 µL of 50 mM IAM, 50 mM ammonium bicarbonate (pH 8.5) in the dark for 20 min prior to centrifugation (20,800 x *g*, 15 min). The 10 kDa filters were transferred to fresh centrifuge tubes and 2 µg trypsin (200 µL, 0.01 µg/mL in 50 mM ammonium bicarbonate, pH 8.5, and 1 mM $CaCl_2$) was added to obtain an enzyme to protein ratio of 1:15. The replicate tubes were incubated at 37°C for 16 h. The filters were centrifuged (20,800 x *g*, 15 min) and the filtrates containing digested peptides were collected. The filters were washed twice with 200 µL of 50 mM ammonium bicarbonate (pH 8.5) and the filtrates were combined, lyophilised and stored in a -80°C freezer until analyzed. The tryptic peptides were resuspended in 50 µL of 1% FA and a 3 µL aliquot run on the 6500 QTRAP LC-MS and quantified.

K. LC-MRM-MS QUANTIFICATION OF DIGESTION PRODUCTS

Either 3 µL of native peptic peptides (Table 1) or reduced and alkylated tryptic peptides (Table 2) were chromatographically separated on a Shimadzu Nexera UHPLC and analyzed on a 6500 QTRAP mass spectrometer (AB SCIEX, Foster City, USA) as described previously (Colgrave et al. 2014). Quantification was achieved using scheduled MRM scanning experiments using a 60 s detection window for each MRM transition and a 0.3 s cycle time. Peaks were integrated using MultiQuant v3.0 (AB SCIEX) wherein all three transitions were required to co-elute at the same retention time (RT, min) with a signal-to-noise (S/N)>3 for detection and a S/N>5 for

quantification. The graphs showing digestibility of the Picpa-ω3D protein were generated in Graphpad Prism v6.

V. EXPERIMENTAL DESIGN

A. SPECIFICITY OF PROTEOLYTIC ENZYMES USED IN THIS STUDY

For digestibility assay, two enzymes trypsin and pepsin were used. Trypsin is a serine protease that is found in the digestive system. Trypsin cleaves polypeptide chains at the carboxyl side of the basic amino acids lysine (K) or arginine (R), but the cleavage is hindered by the presence of proline as the preceding amino acid (P1' position, Figure 2A). Pepsin is a protease produced in the stomach and is efficient at cleaving the peptide bonds adjacent to aromatic and hydrophobic amino acids phenylalanine (F), tyrosine (Y), tryptophan (W) and leucine (L) (Figure 2B). Histidine (H), lysine (K) and arginine (R) at the P3 position act to hinder proteolysis, while proline (P) at P3 or P4 positions promotes proteolysis.

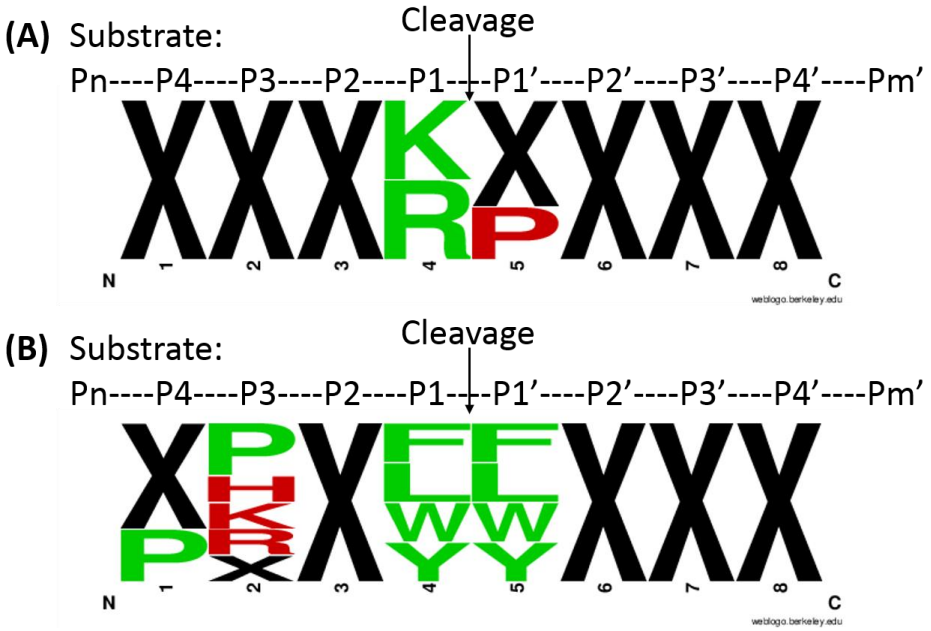

Figure 2. Specificity of proteolytic enzymes used in this study. (A) Trypsin cleavage site. Proline (P) as the preceding amino acid (P1' position) hinders proteolysis is shown in red. (B) Pepsin cleavage sites at both sides of aromatic and hydrophobic amino acids. Amino acids that act to hinder proteolysis at the P3 position are shown in red, whereas those that promote proteolysis (P at P3 or P4) are shown in green. The images were created using WebLogo (Crooks et al. 2004).

B. THEORETICAL PREDICTION OF DIGESTION CURVES

Upon digestion with pepsin alone, there are a number of scenarios that may occur (Figure 3A). The simplest one is when the protein is rapidly digested to produce fully peptic fragments wherein the response rapidly increases reaching a maximum and creating a plateau (filled circle). The second one involves the slow digestion that does not reach a plateau within the experimental duration (filled triangles). This scenario is difficult to judge for completeness as LC-MS monitors the peptide response (peptide peak intensity or area). The third one involves a rapid, but incomplete digestion that may appear to be complete as judged by the plateau in peptide response (empty circles). Lastly, slow and incomplete digestion may be observed (empty triangles).

By employing trypsin post-pepsin (Figure 3B), it is possible to judge the completeness of the digestion by comparison to an experimental control (time 0, no pepsin added) wherein the tryptic peptides liberated appear at the maximum value (in this instance as the MRM peak area). If the protein is not digested, then no decrease in peptide response will be observed (circles, dashed line). If the protein is partially digested, a partial decrease in the peptide response will be observed (squares, dotted line). If the protein is completely digested, the peptide response will drop to zero within the experiment duration (triangles, solid line).

Thus by examining the pepsin proteolytic fragments, the breakdown of a protein could be monitored, but it is noted that determining whether degradation had reached completion is a difficult task. To overcome this deficiency, the tryptic peptide products were used as a proxy for intact protein, wherein in the absence of pepsin, the amount of tryptic peptide was equivalent to 100% of protein being present. In the presence of pepsin (at varying time points during digestion), the level of tryptic peptides would be expected to decrease for peptides that contained a pepsin cleavage site. In this way the complete degradation of the protein could be monitored.

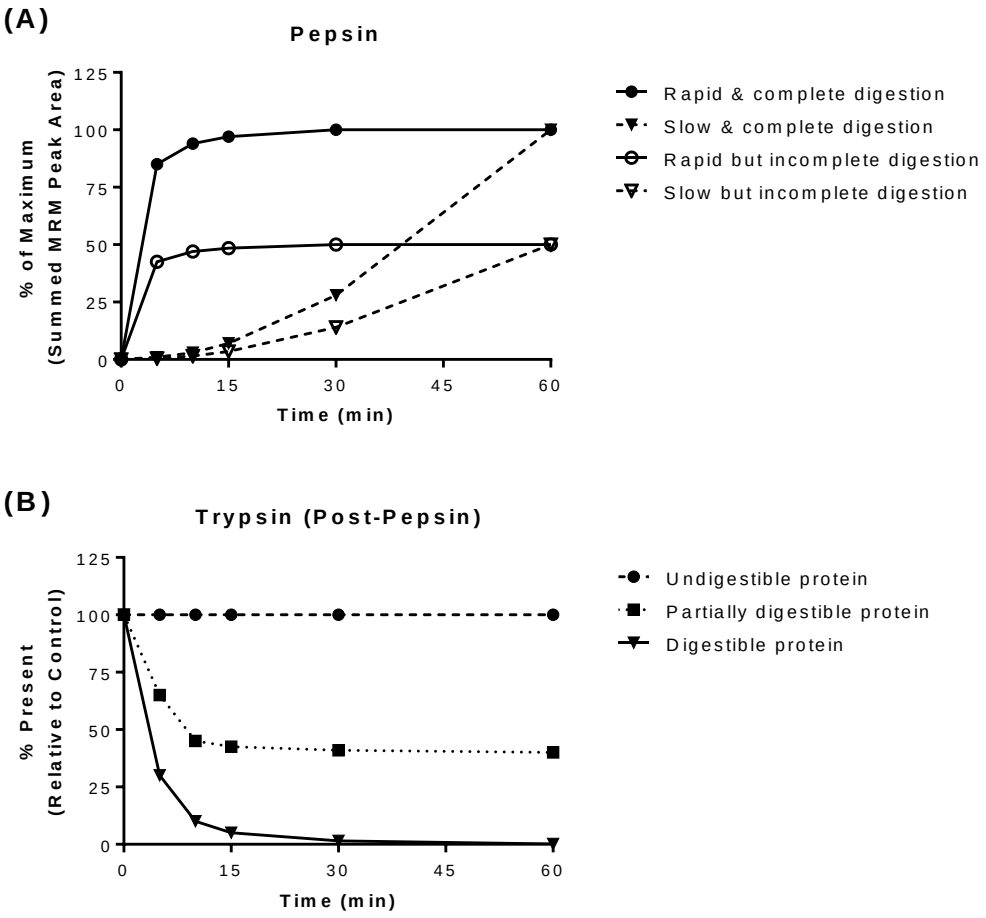

Figure 3. Theoretical digestion curves that could be generated using LC-MS and the proposed digestibility assay. Theoretical digestion curves for pepsin (A) and trypsin post-pepsin (B).

C. SGF DIGESTION

Simulated gastric fluid (SGF) was represented by the proteolytic enzyme pepsin in a buffer adjusted to an acidic pH 1.2. The digestion was performed for 5, 10, 15, 30 and 60 min, with 0 min (no pepsin added) as the control, and each with five replicates. Due to the practical difficulty that was involved in filtering and washing after pepsin digestion with five replicates, the earliest practical time point was 5 min from the addition of pepsin. The increased abundance of targeted peptic peptides was used as indicator of the protein digestibility.

The SGF digestion was further extended by pepsin digestion at the same time point as above, followed by 16 h digestion with trypsin, designated as combined pepsin-trypsin digestion. The

relative abundance of tryptic peptides compared to the abundance of the same peptides in no pepsin digestion (0 min) followed by trypsin digestion was used as an indicator of the protein digestibility.

VI. RESULTS

A. PROTEIN EXTRACTION

The His_8::GFP::Picpa-ω3D fusion protein ran as a doublet with apparent molecular weights of 50 and 60 kDa as determined by SDS-PAGE and Western blotting. The predicted molecular weight of the His_8::GFP::Picpa-ω3D fusion construct is 76.4 kDa, however, a lower than predicted apparent molecular weight on SDS-PAGE is a common and well-documented phenomenon for membrane proteins and is caused by the presence and binding of detergent to the hydrophobic regions (Rath et al. 2009). In addition, the presence of two separate bands (Figure 4, His_8::GFP::Picpa-ω3D) may be due to a population where the GFP in the fusion remains (partially) folded, causing it to migrate faster (lower band) than the population where the GFP is completely denatured (upper band) (Geertsma et al. 2008). Partial proteolysis of the C-terminus of the His_8::GFP::Picpa-ω3D could provide another explanation for the observed doublet, however it is clear from the Coomassie-stained gel that if this is the case, this lower potentially cleaved band is only a minor contaminant. The total protein extracted was estimated to be ~1.3 mg/mL.

The protein was also identified/characterised by LC-MS/MS analysis after proteolytic digestion using both trypsin and pepsin. Because of the difficulty of expressing membrane proteins in general in prokaryotic or eukaryotic systems, the strategy was to express the Picpa-ω3D as a GFP fusion in *E. coli* with a His-tag added to aid in purification. GFP is widely used as a fusion partner for soluble expression and allowing tracking of protein expression by monitoring its fluorescence. The presence of the fusion partner ($His_{8::}$GFP) is unlikely to affect the proteolysis and LC-MS characterisation of the His_8::GFP::Picpa-ω3D protein, nor is the presence of contaminating proteins in the mixture.

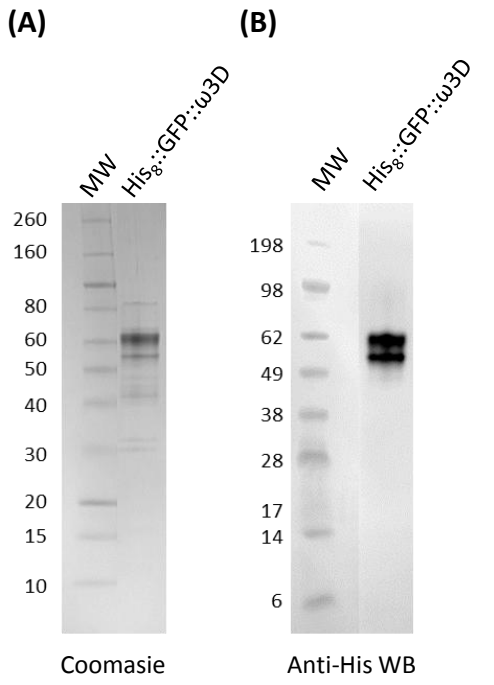

Figure 4. Characterisation of His_8::GFP::Picpa-ω3D protein expressed in E. coli C41. (A) SDS-PAGE and (B) Western blot (WB) analysis of His_8::GFP::Picpa-ω3D protein developed with anti His-tag antibody (1:1000 dilution). Molecular weight markers and 5 µL of His_8::GFP::Picpa-ω3D were run on NuPAGE 4-12% Bis-Tris gel (Invitrogen). MW, protein markers with molecular weight indicated aside.

B. CHARACTERISATION OF THE PICPA-ω3D PROTEIN USING PEPSIN

As depicted in Figure 2, pepsin is a relatively non-specific enzyme that cleavages at Phe (F), Tyr (Y), Trp (W) and Leu (L) resulting in hundreds of possible peptide fragments wherein missed cleavages are commonly observed. *In silico* analysis of the native Picpa-ω3D protein with pepsin digestion suggested the theoretical pepsin cleavage map below.

MSKVTVSGSEILEGSTKTVRRSGNVASFKQQKTAIDTFGNVFKVPDYTIKDILDAIPKH
CYERSLVKSMSYVVRDIVAISAIAYVGLTYIPLLPNEFLRFAAWSAYVFSISCFGFGIW
ILGHECGHSAFSNYGWVNDTVGWVLHSLVMVPYFSWKFSHAKHHKATGHMTRDMVFVPY
TAEEFKEKHQVTSLHDIAEETPIYSVFALLFQQLGGLSLYLATNATGQPYPGVSKFFKS
HYWPSSPVFDKKDYWYIVLSDLGILATLTSVYTAYKVFGFWPTFITWFCPWILVNHWLV
FVTFLQHTDSSMPHYDAQEWTFAKGAAATIDREFGILGIIFHDIIETHVLHHYVSRIPF
YHAREATECIKKVMGEHYRHTDENMWVSLWKTWRSCQFVENHDGVYMFRNCNNVGVKPK
DT

Figure 5. Theoretical pepsin cleavage map. The potential pepsin cleavage sites are indicated in bold red font. Pepsin cleaves at both the amino and carboxyl sides of the highlighted residues. This sequence includes only the native Picpa-ω3D protein.

In this study, the peptide fragments of Picpa-ω3D persisting after pepsin digestion for 120 min were characterised by untargeted LC-MS/MS.

HHHHHHHHSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGK
LPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHN
VEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGI
THGMDELYKENLYFQGGSSKVTVSGSEILEGSTKTVRRSGNVASFKQQKTAIDTFGNVF
KVPDYTIKDILDAIPKHCYERSLVKSMSYVVRDIVAISAIAYVGLTYIPLLPNEFLRFA
AWSAYVFSISCFGFGIWILGHECGHSAFSNYGWVNDTVGWVLHSLVMVPYFSWKFSHAK
HHKATGHMTRDMVFVPYTAEEFKEKHQVTSLHDIAEETPIYSVFALLFQQLGGLSLYLA
TNATGQPYPGVSKFFKSHYWPSSPVFDKKDYWYIVLSDLGILATLTSVYTAYKVFGFWP
TFITWFCPWILVNHWLVFVTFLQHTDSSMPHYDAQEWTFAKGAAATIDREFGILGIIFH
DIIETHVLHHYVSRIPFYHAREATECIKKVMGEHYRHTDENMWVSLWKTWRSCQFVENH
DGVYMFRNCNNVGVKPKDT

Figure 6. Protein sequence coverage obtained after pepsin digestion. Green = peptides identified with >95% confidence; yellow = peptides identified with 50-95% confidence; red = peptides identified with <50% confidence; grey = not detected. Wave underlined is the N-terminal His_8::GFP region and is followed by the second amino acid residue of native Picpa-ω3D (i.e., no methionine) in the fusion protein.

C. CHARACTERISATION OF THE PICPA-ω3D PROTEIN USING TRYPSIN

Trypsin is a relatively specific enzyme cleavages at Lys (K) and Arg (R) resulting in 36 possible peptide fragments, of which 20 were in the mass range suited to LC-MS/MS analysis (Report Nº 2016-015).

MS**K**VTVSGSEILEGST**K**TV**RR**SGNVASF**K**QQ**K**TAIDTFGNVF**K**VPDYTI**K**DILDAIP**K**H
CYE**R**SLV**K**SMSYVV**R**DIVAISAIAYVGLTYIPLLPNEFL**R**FAAWSAYVFSISCFGFGIW
ILGHECGHSAFSNYGWVNDTVGWVLHSLVMVPYFSW**K**FSHA**K**HH**K**ATGHMT**R**DMVFVPY
TAEEF**K**E**K**HQVTSLHDIAEETPIYSVFALLFQQLGGLSLYLATNATGQPYPGVS**K**FF**K**S
HYWPSSPVFD**KK**DYWYIVLSDLGILATLTSVYTAY**K**VFGFWPTFITWFCPWILVNHWLV
FVTFLQHTDSSMPHYDAQEWTFA**K**GAAATID**R**EFGILGIIFHDIIETHVLHHYVS**R**IPF
YHA**R**EATECI**KK**VMGEHY**R**HTDENMWVSLW**K**TW**R**SCQFVENHDGVYMF**R**NCNNVGV**K**P**K**
DT

Figure 7. Theoretical trypsin cleavage map. The potential trypsin cleavage sites for the native Picpa-ω3D protein are indicated in bold blue font. Trypsin cleaves at the carboxyl side of the highlighted residues.

In this study, the peptide fragments present after trypsin digestion for 16 h were characterised by untargeted LC-MS/MS as shown in Figure 8.

HHHHHHHHSK**GEELFTGVVPILVELDGDVNGHK**FSVR**GEGEGDATNGKLTLKFICTTGK**LP
VPWPTLVTTLTYGVQCFSRYPDHMKRHDFFK**SAMPEGYVQERTISFKDDGTYK**TR**AEVKFE**
GDTLVNRIELK**GIDFKEDGNILGHKLEYNFNSHNVYITADK**QKNGIKANFKIR**HNVEDGSV**
QLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEF**VTAAGITHGMDELY**
KENLYFQGGSSKVTVSGSEILEGSTKTVRRSGNVASFKQQKTAIDTFGNVFKVPDYTIKDI
LDAIPKHCYERSLVK**SMSYVVR**DIVAISAIAYVGLTYIPLLPNEFLRFAAWSAYVFSISCF
GFGIWILGHECGHSAFSNYGWVNDTVGWVLHSLVMVPYFSWK**FSHAKHHKATGHMTRDMVF**
VPYTAEEFKEKHQVTSLHDIAEETPIYSVFALLFQQLGGLSLY**LATNATGQPYPGVSK**FFK
SHYWPSSPVFDKKDYWYIVLSDLGILATLTSVYTAYKVFGFWPTFITWFCPWILVNHWLVF
VTFLQHTDS**SMPHYDAQEWTFAKGAAATIDREFGILGII**FHDIIETHVLHHYVSR**IPFYHA**
REATECIKKVMGEHYR**HTDENMWVSLWK**TWR**SCQFVENHDGVYMFR**NCNNVGVKPKDT

Figure 8. Protein sequence coverage obtained after trypsin digestion. Green = peptides identified with >95% confidence; yellow = peptides identified with 50-95% confidence; grey = not detected. Wave underlined is the N-terminal His_8::GFP region followed by the second amino acid of native Picpa-ω3D (i.e., no methionine) in the fusion protein.

D. Development of a quantitative LC-MRM-MS method to assess the protein digestibility

To assess the digestibility of the Picpa-ω3D protein, a targeted LC-MS/MS method based on the use of multiple reaction monitoring (MRM) (Lange et al. 2008) mass spectrometry (MS) was developed. The appearance and the increase of the peptic peptides during the time course of pepsin digestion were used as the evidence of the protein digestibility. Moreover, the rapid

decline of the tryptic peptides subsequent to pepsin digestion served as confirmation of the protein digestibility.

In order to select peptides to quantify in this method, the digestion products resulting from both pepsin and trypsin digestion were first characterised as described previously. Peptides that were identified with 95% confidence and that yielded intense signals in the MS were selected for relative quantification. Eleven peptides were selected from the digestion of the His_8::GFP::Picpa-ω3D protein with pepsin and eight peptides for trypsin digestion are summarized in Tables 1-2. The selected peptides spanned the length of the protein.

Table 1. Peptide MRM transitions for Picpa-ω3D pepsin products.

Peptide	RT (min)[a]	Q1 m/z [a]	z [a]	Q3 m/z [a]	Fragment	CE [a]
FKVPDYTIKDIL	6.66	484.610	3+	601.392	y5+	21.3
				702.440	y6+	21.3
				587.319	b5+	21.3
FKVPDY	4.42	384.702	2+	394.161	y3+	17.9
				621.324	y5+	17.9
				587.319	b5+	17.9
TIKDIL	4.60	351.723	2+	488.308	y4+	16.2
				601.392	y5+	16.2
				571.345	b5+	16.2
TYIPLLPNEF	8.07	603.824	3+	619.309	y5+	27.0
				829.445	y7+	27.0
				701.423	b6+	27.0
TYIPLLPNE	6.75	530.290	3+	682.377	y6+	23.5
				588.339	b5+	23.5
				701.423	b6+	23.5
TYIPLL	7.34	360.221	2+	378.202	b3+	16.7
				475.255	b4+	16.7
				588.339	b5+	16.7
IPLLPNEF	7.36	471.768	2+	437.312	b4+	22.1
				648.408	b6+	22.1
				777.451	b7+	22.1
IPLLPNE	5.74	398.234	2+	324.228	b3+	18.5
				437.312	b4+	18.5
				648.408	b6+	18.5
NATGQPYPGVSKF	4.32	683.344	3+	894.472	y8+	30.8
				1079.552	y9+	30.8
				1180.600	y10+	30.8
FKSHYWPSSPVF	5.69	494.579	3+	560.764	b9++	21.7
				609.291	b10++	21.7
				658.825	b11++	21.7
TAYKVFGF	5.89	466.748	2+	597.340	y5+	21.9
				760.403	y6+	21.9
				831.440	y7+	21.9

Picpa-ω3D sequence:[b]

MSKVTVSGSEILEGSTKTVRRSGNVASFKQQKTAIDTFGNV**FKVPDYTIKDIL**DAIPKHCYERSLVKSMSYVV
RDIVAISAIAYVGL**TYIPLLPNEF**LRFAAWSAYVFSISCFGFGIWILGHECGHSAFSNYGWVNDTVGWVLHSL
VMVPYFSWKFSHAKHHKATGHMTRDMVFVPYTAEEFKEKHQVTSLHDIAEETPIYSVFALLFQQLGGLSLYLA
T**NATGQPYPGVSKFFKSHYWPSSPVF**DKKDYWYIVLSDLGILATLTSVY**TAYKVF**GFWPTFITWFCPWILVNH
WLVFVTFLQHTDSSMPHYDAQEWTFAKGAAATIDREFGILGIIFHDIIETHVLHHYVSRIPFYHAREATECIK
KVMGEHYRHTDENMWVSLWKTWRSCQFVENHDGVYMFRNCNNVGVKPKDT

a. RT, retention time (min); Q1 m/z, precursor ion mass-to-charge ratio (m/z); z, charge state; Q3 m/z, fragment ion m/z; CE, collision energy in V.

b. Picpa-ω3D sequence with mapped peptic peptides (bold, underlined). For pepsin, different cleavage variants were observed owing to the incomplete digestion and these peptides have been differentiated by single, double or waved underline.

Table 2. Peptide MRM transitions for Picpa-ω3D trypsin products.

Peptide	RT (min)[a]	Q1 m/z [a]	z [a]	Q3 m/z [a]	Fragment	CE [a]
VTVSGSEILEGSTK	4.40	703.870	2+	1020.250	y10+	33.5
				1107.550	y11+	33.5
				1206.620	y12+	33.5
SGNVASFK	2.41	405.210	2+	452.250	y4+	18.9
				665.360	y6+	18.9
				551.320	y5+	18.9
TAIDTFGNVFK	5.90	606.820	2+	927.460	y8+	28.7
				1040.540	y9+	28.7
				812.430	y7+	28.7
VPDYTIK	3.42	418.230	2+	736.390	y6+	19.5
				368.690	y6++	19.5
				639.330	y5+	19.5
DILDAIPK	3.42	442.760	2+	656.400	y6+	20.7
				543.310	y5+	20.7
				244.170	y2+	20.7
EATEC[CAM]IK	1.43	425.700	2+	549.240	y4+	19.9
				650.290	y5+	19.9
				325.650	y5++	19.9
HTDENMWVSLWK	6.45	515.910	3+	533.300	y4+	24.3
				632.380	y5+	24.3
				818.450	y6+	24.3
SC[CAM]QFVENHDGVYMFR	5.07	630.260	3+	772.380	y6+	29.9
				887.410	y7+	29.9
				1024.470	y8+	29.9

Picpa-ω3D sequence:[b]

MSK**VTVSGSEILEGSTK**TVRR**SGNVASFK**QQK**TAIDTFGNVFKVPDYTIKDILDAIPK**HCYERSLVKS
MSYVVRDIVAISAIAYVGLTYIPLLPNEFLRFAAWSAYVFSISCFGFGIWILGHECGHSAFSNYGWVN
DTVGWVLHSLVMVPYFSWKFSHAKHHKATGHMTRDMVFVPYTAEEFKEKHQVTSLHDIAEETPIYSVF
ALLFQQLGGLSLYLATNATGQPYPGVSKFFKSHYWPSSPVFDKKDYWYIVLSDLGILATLTSVYTAYK
VFGFWPTFITWFCPWILVNHWLVFVTFLQHTDSSMPHYDAQEWTFAKGAAATIDREFGILGIIFHDII
ETHVLHHYVSRIPFYHAR**EATECIK**KVMGEHYR**HTDENMWVSLWK**TWR**SCQFVENHDGVYMFR**NCNNV
GVKPKDT

a. RT, retention time (min); Q1 m/z, precursor ion mass-to-charge ratio (m/z); z, charge state; Q3 m/z, fragment ion m/z; CE, collision energy in V.
b. Picpa-ω3D sequence with mapped tryptic peptides (bold, underlined). For trypsin, all peptides selected were fully tryptic, *i.e.* contained no missed cleavages. As some of the peptides were adjacent in the sequence, these have been differentiated by single or double underline.

E. DIGESTIBILITY OF PICPA-ω3D PROTEIN

Digestibility of the His_8::GFP::Picpa-ω3D in SGF was assessed by LC-MRM-MS method as described above. Characterisation and quantification of the targeted peptic peptides showed the rapid degradation of the His_8::GFP::Picpa-ω3D protein. The pepsin digestion data has been presented in Figure 9 as the mean of five replicate digests relative percentage of the maximum detected MRM peak area (sum of three transitions) per peptide across the time points (0, 5, 10, 15, 30, 60 min).

Five of the peptides characterised and quantified after pepsin digestion were cleavage variants (Fig. 9A-C, E-F). The black arrows in Figure 9 indicate that the peptide in the upper left panel is cleaved further by pepsin to yield the peptide in the upper right and panel immediately beneath. All peptic peptides monitored were produced rapidly (<15 min). The peptide NATGQPYPGVSKF (Fig. 9D) reached an equilibrium over this time frame and notably was the only peptide that was fully peptic, i.e. could not undergo further hydrolysis by pepsin. The majority of the peptic peptides monitored did not represent the fully cleaved final product as pepsin is relatively non-specific. In some cases, a decrease in peptide level is noted over time, for example, TYIPLLPNEF (Fig. 9E) decreases slowly from ~5 min and its product TYIPLLPNE (Fig. 6F) increases in concentration from 10-60 min. Several other examples of pepsin proteolysis products containing missed cleavages (indicated by red font in peptide sequence), that are therefore susceptible to further degradation, were monitored (Figure 9A-B, E-H). In fact only NATGQPYPGVSKF (Fig. 9D) contains no predicted cleavage sites. The appearance of these peptides in the digest is taken as evidence of the degradation and therefore digestibility of the Picpa-ω3D protein. Three of the eight peptides monitored reached a peak at 5 min. To further illustrate the low specificity of pepsin and the generation of multiple related cleavage products, five cleavage variants of TYIPLLPNEF were monitored (Figure 10). The larger fragments were noted to plateau, whereas the smaller peptide fragments were still increasing in concentration at the conclusion of the 60 min digestion period.

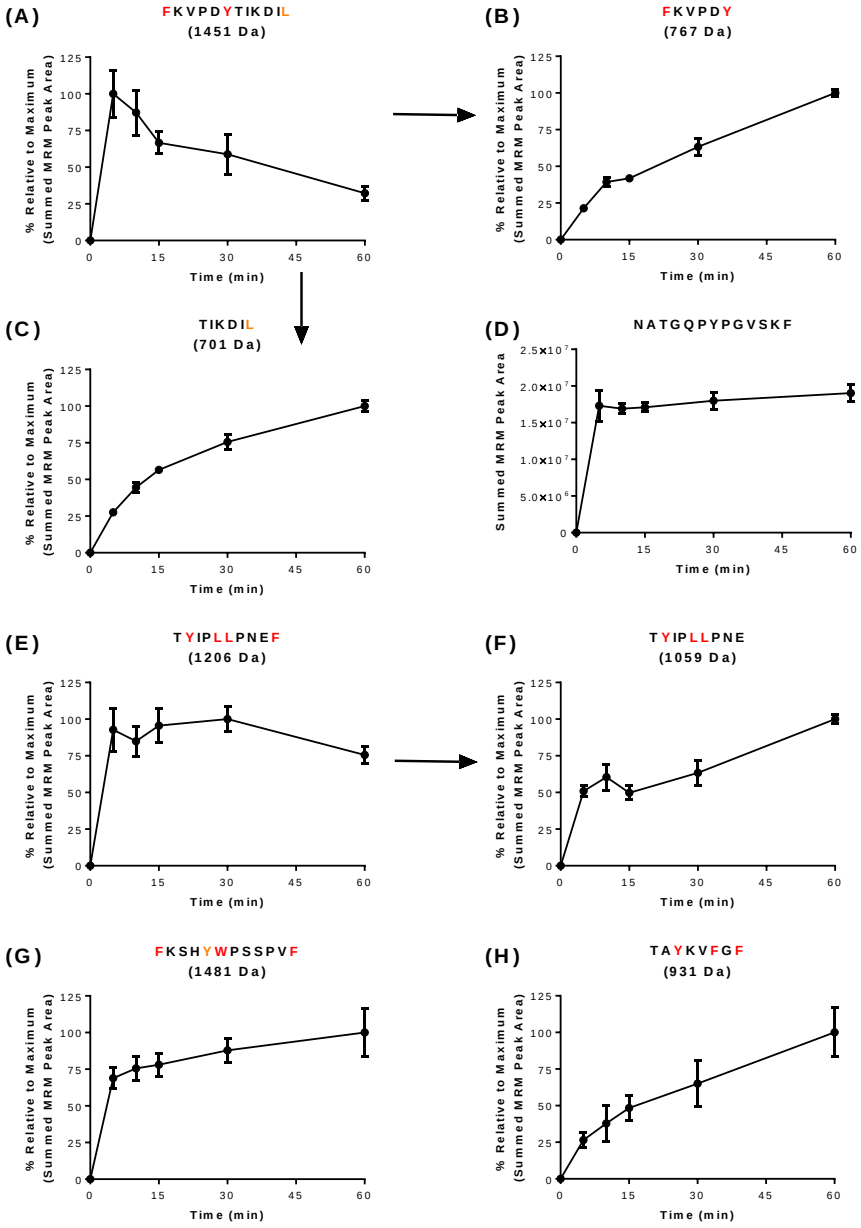

Figure 9. Quantification of the peptic peptides of His_8::GFP::Picpa-ω3D after pepsin digestion. LC-MRM-MS analysis of pepsin proteolytic fragments. The response in the LC-MS system (measured as peak area) was converted to a percentage relative to the maximum peak area observed during pepsin digestion. The experimental control was time 0 with no pepsin addition. The peptides are graphed in order from protein N- to C-terminus. The peptide sequence (and calculated molecular weight) are denoted above each graph. Arrows indicate a subsequent cleavage to yield a secondary cleavage variant. The potential sites for secondary pepsin cleavage are indicated in red font within the sequence. The error bars denote SD.

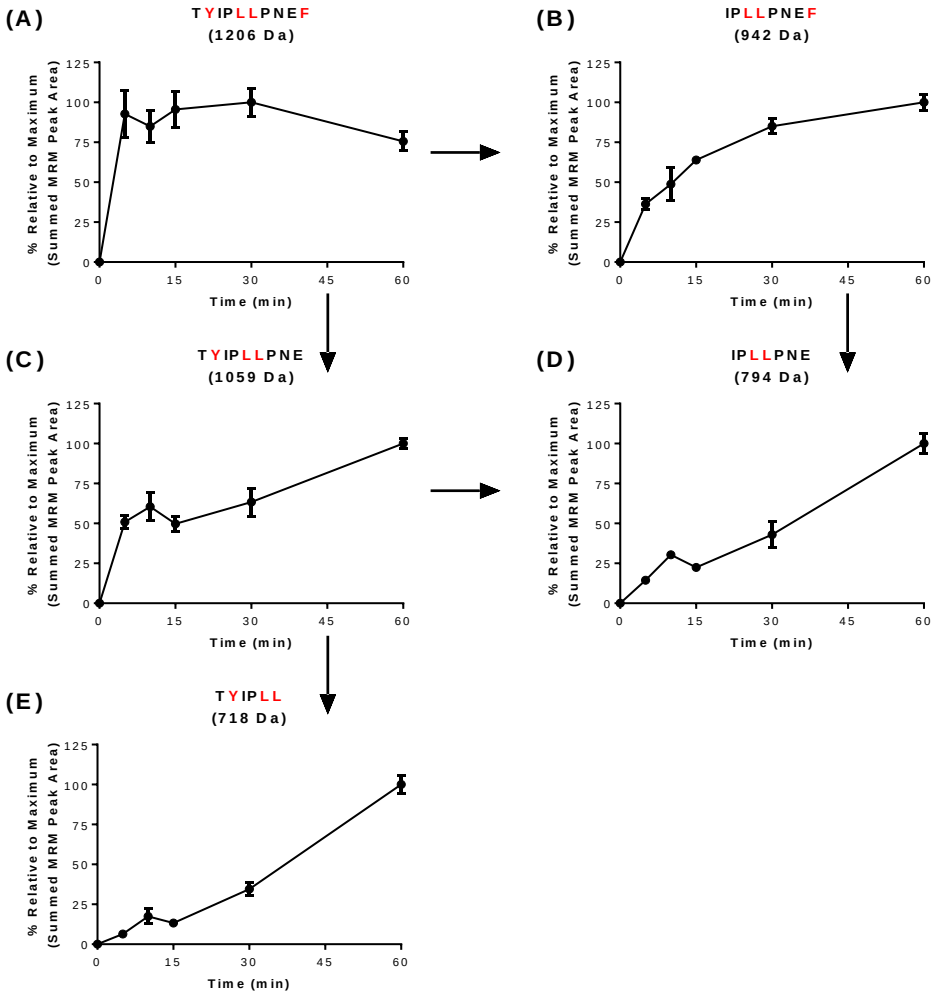

Figure 10. His_8::GFP::Picpa-ω3D cleavage variants produced after pepsin digestion. LC-MRM-MS analysis of pepsin proteolytic fragments. The response in the LC-MS system (measured as peak area) was converted to a percentage relative to the maximum peak area observed during pepsin digestion. The experimental control was time 0 with no pepsin addition. The peptides are graphed in order from protein N- to C-terminus. The peptide sequence (and calculated molecular weight) are denoted above each graph. Arrows indicate a subsequent cleavage to yield a secondary cleavage variant. The potential sites for secondary pepsin cleavage are indicated in red font within the sequence.

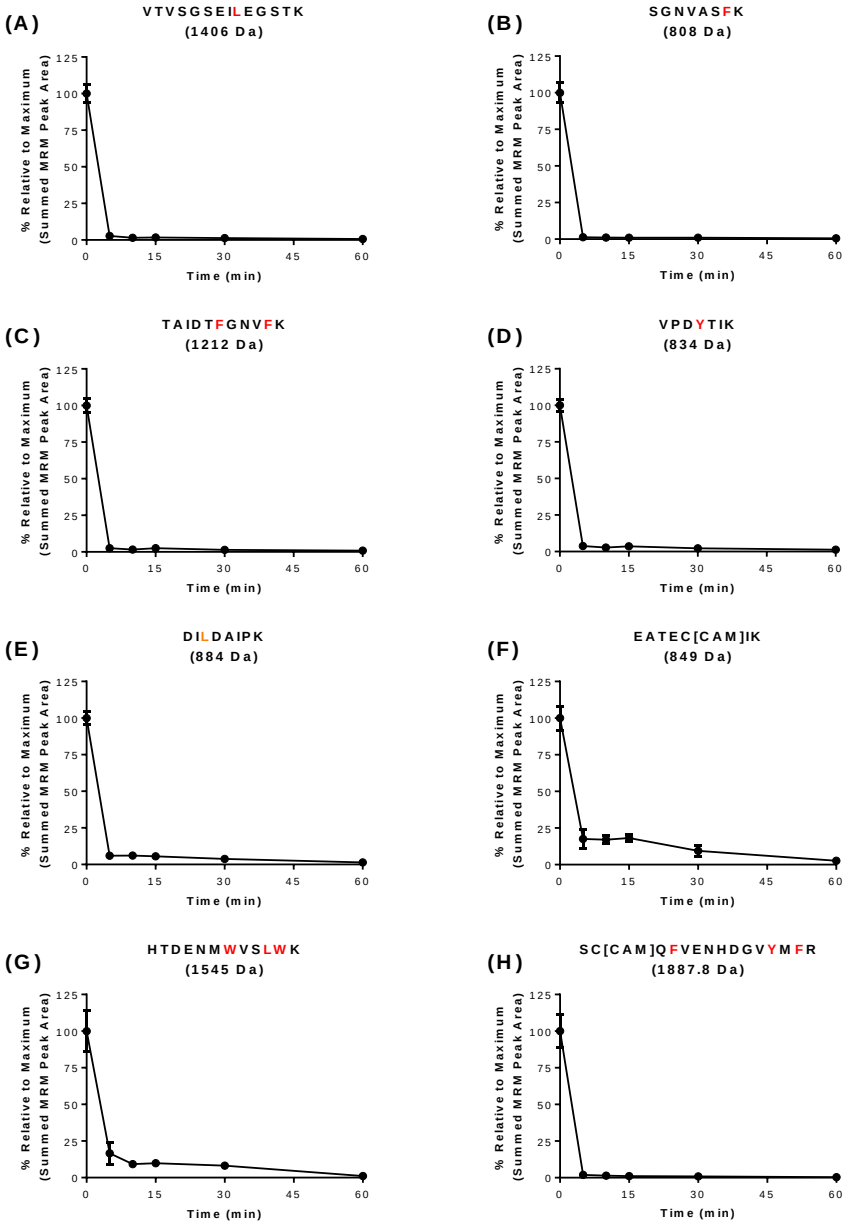

Figure 11. Quantification of the tryptic peptides of His_8::GFP::Picpa-ω3D after combined pepsin-trypsin digestion. The trypsin data has been presented as the mean percentage (n=5 replicate digests) reduction relative to the experimental control at 0 min (no pepsin addition, measured as MRM peak area, sum of three transitions) per peptide across the time points (0, 5, 10, 15, 30, 60 min). The peptides are graphed in order from protein N- to C-terminus. The peptide sequence (and calculated molecular weight) are denoted above each graph. The potential sites for pepsin cleavage of these peptide sequences are indicated in red (expected cleavage) or orange (potentially hindered) font. The error bars denote SD.

The tryptic peptides monitored after the pepsin digest show a rapid decline in the first 5 min and then a further decline over the remainder of the experiment (60 min duration). It is estimated that >97% of the protein is cleaved after 60 min on the basis of the disappearance of these eight tryptic peptides. The peptides containing multiple pepsin cleavage sites: TAIDTF$^{\downarrow}$GNVF$^{\downarrow}$K, HTDENMW$^{\downarrow}$VSL$^{\downarrow}$W$^{\downarrow}$K, SC[CAM]QF$^{\downarrow}$VENHDGVY$^{\downarrow}$MF$^{\downarrow}$R (where X$^{\downarrow}$X represent pepsin cleavage site) are reduced to 0.9, 1.1 and 0.3% of the undigested control (no pepsin digest) respectively. This is supported by analysis of the digested peptides on the TripleTOF 5600 LC-MS/MS, which shows that these peptides are more frequently fragmented to yield smaller fragments after 30-60 min. The tryptic peptides containing fewer sites: VTVSGSEIL$^{\downarrow}$EGSTK, SGNVASF$^{\downarrow}$K and VPDY$^{\downarrow}$TIK (with a single site) or DIL$^{\downarrow}$DAIPK (where the lysine in position P3 is known to hinder pepsin cleavage) were reduced to 0.7, 0.5, 1.3 and 1.4% respectively. The higher percentage of EATECIK observed (17.5% at 5 min and 2.7% at 60 min) can be explained by the absence of peptic digestion sites within this peptide sequence. Overall, it was observed that the peptides from the termini (both N- and extreme C-termini) of the protein were liberated rapidly with <6% remaining after 5 min (Table 3).

Table 3. Percentage of each tryptic peptide remaining during pepsin time course.

Peptide Sequence	Incubation time (min)				
	5	**10**	**15**	**30**	**60**
VTVSGSEILEGSTK	2.8	1.5	1.8	1.3	0.7
SGNVASFK	1.3	1.1	1.0	1.0	0.5
TAIDTFGNVFK	2.5	1.7	2.6	1.4	0.9
VPDYTIK	3.8	2.7	3.6	2.3	1.3
DILDAIPK	6.0	6.1	5.5	3.7	1.4
EATEC[CAM]IK	17.5	17.1	18.1	9.4	2.7
HTDENMWVSLWK	16.5	9.2	9.9	8.2	1.1
SC[CAM]QFVENHDGVYMFR	2.0	1.3	1.1	0.9	0.3

VII. DISCUSSION

Picpa-ω3D is an integral membrane protein. Currently, there is no functional antibody for Western blot analysis available to quantify the transgenic protein content in DHA canola, event NS-B50027-4, or detect the stability of Picpa-ω3D as a native protein. The commercially raised polyclonal and monoclonal antibodies by GenScript (Piscataway, NJ, USA) failed to generate a specific signal towards Picpa-ω3D. The antibodies were raised against the synthetic peptides predicted by GenScript as potential epitopes for antigens (Figure 12).

MSKVTVSGSEILEGSTKTVRRSGNVASFKQQKTAIDTFGNVFKVPDYTIKDILDAIPKHCYERSL
VKSMSYVVRDIVAISAIAYVGLTYIPLLPNEFLRFAAWSAYVFSISCFGFGIWILGHECGHSAFS
NYGWVNDTVGWVLHSLVMVPYFSWKFSHAKHHKATGHMTRDMVFVPYTAEEFKEKHQVTSLHDIA
EETPIYSVFALLFQQLGGLSLYLATNATGQPYPGVSKFFKSHYWPSSPVFDKKDYWYIVLSDLGI
LATLTSVYTAYKVFGFWPTFITWFCPWILVNHWLVFVTFLQHTDSSMPHYDAQEWTFAKGAAATI
DREFGILGIIFHDIIETHVLHHYVSRIPFYHAREATECIKKVMGEHYRHTDENMWVSLWKTWRSC
QFVENHDGVYMFRNCNNVGVKPKDT

Figure 12. Peptides selected for antibody production by GenScript. The peptides for polyclonal antibodies are highlighted in yellow, while the peptide for both polyclonal and monoclonal antibodies is highlighted in cyan.

Although Picpa-ω3D expressed as the His-GFP fusion protein could be analysed by Western blot against the anti-His-tag antibody, such a Western blot analysis could only monitor the fusion region rather than whole protein, when the His-tag is cleaved off, for example after SGF digestion. In addition, the anti-His-tag antibody is not suitable for quantification of the native Picpa-ω3D (unfused) protein in DHA canola. Thus an alternative approach using LC-MRM-MS analysis was developed here, which can be applied both for the quantification of protein expressed in canola and for the stability assays. The results shown here clearly demonstrated that the LC-MS approach is suitable for such an application. This method is as sensitive as traditional Western blot, which can normally detect down to a ng to µg protein scale. The LC-MRM-MS approach can detect as low as 0.6 femtomoles which equates to ~29 pg on a protein scale (Report Nº 2016-015). In addition, Western blot using antibodies might only detect a limited number of epitopes (one or two) from the protein. Here we targeted eleven (peptic) and eight (tryptic) peptides, along with the intact protein, which provides an understanding of the kinetics of digestion and the susceptibility of specific regions of the protein to proteolysis. Due to the technical difficulty that was involved in filtering and washing steps after pepsin digestion with five replicates, the earliest practical time point was 5 min. Nevertheless, the results have shown the successful application of LC-MRM-MS for protein digestibility analysis.

The thermal stability of the trans-membrane enzyme proteins, such as desaturases and elongases, cannot be as easily characterized *in vitro* as it can for non-membrane bound enzymes. Traditionally the isolated protein is heated to various temperatures and its enzymatic activity is assayed to ascertain the level of functionality. A denatured or unfolded protein will lose its enzymatic activity. Desaturase activity has been assayed in crude extracts when the required substrates are added (Jackson et al. 1998) but with DHA canola it is far more difficult because there are multiple desaturases and elongases expressed in the canola seed and the levels of the transgenic proteins in seed were extremely low (see Report Nº 2016-015). Solubilisation using detergents to replace the lipid of the membrane and purification can increase the levels of protein but once away from the membranes, the desaturase or elongase is not assayable most likely due to the requirement of other proteins co-localized in the membrane as well as cofactors, some yet unknown.

Membrane proteins, and especially trans-membrane proteins, are not very thermally stable and are difficult to refold once they are partially or fully denatured (Bowie, 2001). When cells are heated, the lipid membrane becomes more fluid and exposes the hydrophobic regions of the proteins within the membrane. When exposed, the hydrophobic regions tend to lead to protein aggregation and loss of function. This is seen when preparing gel samples of membrane proteins isolated with detergents (Gennis 1989).

SDS-PAGE analysis is not useful for determining thermal denaturation of membrane proteins. It may be used to determine the amount of protein aggregating at a given temperature, with the protein sample being heated, and the aggregated protein removed by centrifugation before the remaining protein in the solution is subjected to SDS-PAGE analysis. However in the absence of detergent, all membrane proteins will be in the precipitate, and in the presence of detergent micelles, unfolded proteins could be artificially solubilised and remain in the soluble fraction, hence providing no useful information about thermal stability or activity.

In the processing of seeds to produce oil, the seed material reaches temperatures ranging from 80ºC to 115ºC. It is improbable that any trans-membrane proteins will be remain in the native folded state at this temperature and even after cooling, it will be unlikely to refold correctly. Therefore, it is expected that Picpa-ω3D will not be an active folded protein after processing. Based on these reasons, the thermal stability of Picpa-ω3D was not included in this study.

The Picpa-ω3D protein belongs to the subfamily of yeast acyl-CoA type fatty acid desaturases that introduce a double bond between the Δ15-position from the carboxyl end of C18 fatty acids, or ω3-position from methyl end. The yeast acyl-CoA type fatty acid desaturases include Δ12- and ω3-/Δ15- desaturases. Some of these yeast acyl-CoA type fatty acid desaturases are also common in food, animal feeds or in food production (see Report Nº 2016-006). The Picpa-ω3D protein was used as the representative of the two yeast acyl-CoA type fatty acid

desaturases (Picpa-ω3D and Lackl-Δ12D) engineered into DHA canola, for stability analysis in this report.

VIII. CONCLUSIONS

The results of this study demonstrated that greater than 80% of the full-length Picpa-ω3D protein was digested within 5 min and greater than 97% of it was digested within 60 min of incubation in pepsin. The combined pepsin-trypsin assay showed a rapid decline in the tryptic peptides that were used as a proxy for the presence of intact protein.

The results of this study show that the integral membrane protein Picpa-ω3D was readily digestible in pepsin and/or trypsin. Rapid digestion of the full-length protein is one of many factors that indicate protein safety.

IX. REFERENCES

Astwood, JD, Leach, JN, and Fuchs, RL. 1996. Allergenicity of foods derived from transgenic plants. *Nature Biotech* 14:1269-1273.

Bowie, JU. 2001. Stabilizing membrane proteins. *Curr Opin Struct Biol* 11:397-402.

Codex Alimentarius Commission. 2003. Alinorm 03/34. Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome Italy 30 June-5 July, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants and Appendix IV, Annex on the assessment of possible allergenicity, pp 47-60.

Colgrave, ML, Goswami, H, Blundell, M, Howitt, CA, Tanner, GJ. 2014. Using mass spectrometry to detect hydrolysed gluten in beer that is responsible for false negatives by ELISA. *J Chromat A* 1370:105-114.

Crooks, GE, Hon, G, Chandonia, JM, Brenner, SE. 2004. WebLogo: A sequence logo generator. *Genome Res* 14:1188-1190.

Geertsma, ER, Groeneveld, M, Slotboom, DJ, Poolman, B. 2008. Quality control of overexpressed membrane proteins. *PNAS USA*, 105:5722-5727.

Gennis, RB. 1989. Biomembranes: Molecular Structure and Function. Springer Advanced Texts in Chemistry. Charles R. Cantor (Editor). Springer Science + Business Media, LLC. Chapter 3.31 pp91-93.

Goodman, R. 2008. Performing IgE serum testing due to bioinformatics matches in the allergenicity assessment of GM crops. *Food Chem Tox*. 46:524-534.

Jackson, FM, Fraser, TC, Smith, MA, Lazarus, C, Stobart, AK, Griffiths, G. 1998. Biosynthesis of C18 polyunsaturated fatty acids in microsomal membrane preparations from the filamentous fungus *Mucor circinelloides*. *Eur J Biochem* 252:513-9.

Lange, V, Picotti, P, Domon, B, Aebersold, R. 2008. Selected Reaction Monitoring for Quantitative Proteomics: a Tutorial. *Mol Syst Biol* 4:222.

Petrie, JR, Liu, Q, Mackenzie, AM, Shrestha, P, Mansour, MP, Robert, SS, Frampton, DF, Blackburn, SI, Nichols, PD, Singh, SP. 2010a. Isolation and characterisation of a high-efficiency desaturase and elongases from microalgae for transgenic LC-PUFA production. *Mar Biotechnol* 12:430-438.

Petrie, JR, Shrestha, P, Belide, S, Kennedy, Y, Lester, G, Liu, Q, Divi, UK, Mulder, RJ, Mansour, MP, Nichols, PD.Singh, SP. 2014. Metabolic engineering *Camelina sativa* with fish oil-like levels of DHA. *PLoS One* 9: e85061.

Petrie, JR, Shrestha, P, Mansour, MP, Nichols, PD, Liu, Q, Singh, SP. 2010b. Metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA Delta 6-desaturase with omega 3-preference from the marine microalga *Micromonas pusilla*. *Metab Eng* 12:233-240.

Petrie, JR, Shrestha, P, Zhou, X-R, Mansour, MP, Liu, Q, Belide, S, Nichols, PD, Singh, SP. 2012. Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLoS One* 7: e49165.

Rath, A, Glibowicka, M, Nadeau, VG, Chen, G, Deber, CM. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS USA*, 106:1760-1765.

Shilov, IV, Seymour, SL, Patel, AA, Loboda, A, Tang, WH, Keating, SP, Hunter, CL, Nuwaysir, LM, Schaeffer, DA. 2007. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* 6:1638-1655.

Tang, WH, Shilov, IV, Seymour, SL. 2008. Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J Proteome Res* 7:3661-3667.

Thomas, K, Aalbers, M, Bannon, GA, Bartels, M, Dearman, RJ, Esdaile, DJ, Fu, TJ, Glatt, CM, Hadeld, N, Hatzos, C, Hee, SL, Heylings, JR, Goodman, RE, Henry, B, Herouet, C, Holsapple, M, Ladics, GS, Landry, TD, MacIntosh, SC, Rice, EA, Privalle, LS, Steiner, HY, Teshima, R, van Ree, R, Woolhiser, M, and Zawodny, J. 2004. A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Reg Tox Pharm* 39:87–98.

Watanabe, K, Oura, T, Sakai, H.Kajiwara, S. 2004. Yeast Δ12 fatty acid desaturase: Gene cloning, expression, and function. *Biosci, Biotechnol Biochem* 68: 721-727.

Wisniewski, JR, Zougman, A, Nagaraj, N, Mann, M. 2009. Universal sample preparation method for proteome analysis. *Nature Meth* 6: 359-363.

Zhang, X, Li, M, Wei, D, Xing, L. 2008. Identification and characterization of a novel yeast ω3-fatty acid desaturase acting on long-chain n-6 fatty acid substrates from *Pichia pastoris*. *Yeast* 25: 21-27.

Zhou, X-R, Robert, SS, Petrie, JR, Frampton, DMF, Mansour, MP, Blackburn, SI, Nichols, PD, Green, AG, Singh, SP. 2007. Isolation and characterization of genes from the marine microalga *Pavlova salina* encoding three front-end desaturases involved in docosahexaenoic acid biosynthesis. *Phytochem* 68: 785-796.

X. UNPUBLISHED REFERENCES

Report Nº 2016-005. [redacted]. Nuseed Pty Ltd.

Report Nº 2016-006. [redacted] Pty Ltd.

Report Nº 2016-007. [redacted] Nuseed Pty Ltd.

Report Nº 2016-008. [redacted] Nuseed Pty Ltd.

Report Nº 2016-009. [redacted] Pty Ltd.

Report Nº 2016-010. [redacted] Nuseed Pty Ltd.

Report Nº 2016-011. [redacted] Nuseed Pty Ltd.

Report Nº 2016-015. [redacted] Nuseed Pty Ltd.