Tuber storage proteins as potential precursors of bioactive peptides: an *in silico* analysis

Mohammed Auwal Ibrahim[a,c], Megan J. Bester[b], Albert W. Neitz[a] and Anabella R.M. Gaspar[a]

[a]*Department of Biochemistry, University of Pretoria, Pretoria 0002, South Africa*
[b]*Department of Anatomy, University of Pretoria, Pretoria 002, South Africa*
[c]*Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria*

***Correspondence to:** Dr. Mohammed Auwal Ibrahim, Department of Biochemistry, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0002, South Africa E mail: mauwalibrahim@gmail.com or maibrahim@abu.edu.ng

Abstract

In silico analysis was used to evaluate storage proteins from plant tubers as potential precursors of bioactive peptides after simulated gastrointestinal digestion. Proteins derived from potato (patatins), sweet potato (sporamins), yam (dioscorins) and taro (tarin) were subjected to *in silico* gastrointestinal digestion with a combination of pepsin, chymotrypsin and trypsin in the BIOPEP database which led to the release of 387 peptide fragments which were predicted to have bioactivities such as dipeptidyl peptidase IV (DPP-IV), angiotensin converting enzyme (ACE), antioxidative and antithrombotic activities. Prediction of antimicrobial activity of the released peptides using the Collection of Antimicrobial Peptides (CAMP) database indicated 28 peptides as potential antimicrobial peptides (AMPs) with varied percentage similarity with known AMPs in Antimicrobial Peptide Database (APD). Furthermore, 32 peptides with potential anticancer activity were predicted using the AntiCP database while 9 peptides were predicted to be bioactive according to peptideRanker but the precise bioactivity was not identified. The patatins seem to be the richest source of DPP-IV inhibitory and antimicrobial peptides while dioscorins yielded the highest amount of antihypertensive and anticancer peptides. The data suggests that the tuber storage proteins could release an array of non-toxic and species-specific bioactive peptides with health promoting effects after gastrointestinal digestion.

Key words: Anticancer; Antidiabetic; Antihypertensive; Bioactive peptides; *in silico* study; Tuber storage proteins

List of abbreviations:

ACE: Angiotensin converting enzyme

AMPs: Antimicrobial peptides

ANN: Artificial neural network

APD: Antimicrobial peptide database

CAMP: Collection of antimicrobial peptides

CaMPDE: Calmodulin-dependent phosphodiesterase

DA: Discriminant analysis

DPP-IV: Dipeptidyl peptidase IV

GIP: Glucose-dependent insulinotropic polypeptide

GLP-1: Glucagon-like peptide-1

RF: Random forest

SVM: Support vector machines

1. Introduction

Bioactive peptides are specific fragments of proteins that modulate physiological functions through interactions with specific receptors and consequently induce a physiological response leading to beneficial health effects (Fitzgerald and Murray, 2006). These peptides are usually inactive sequences within the parent protein molecule but are released via enzyme-catalyzed protein hydrolysis especially during gastrointestinal digestion of dietary proteins or fermentation (Moller *et al.,* 2008). Thus, dietary proteins are presently considered as an important source of health promoting agents in terms of bioactive peptides in addition to their role as sources of amino acids to the body.

Numerous bioactive peptides have been described from different food materials of animal or plant origin (Malaguti *et al.,* 2014; Park and Nam, 2015) and are shown to possess antimicrobial (Wang *et al.,* 2016), antioxidant (Byun *et al.,* 2009), antihypertensive (Lafarga *et al.,* 2014), antidiabetic (Mojica and de Majia, 2016; Zhang *et al.,* 2016), anti-inflammatory (Korhonen and Pihlanto, 2007), anticancer (Yin *et al.,* 2013), antithrombotic (Park and Nam, 2015) activities among others. Indeed, some bioactive peptides have already been approved by the FDA for commercialization as therapeutic peptides and these are documented in the therapeutic peptides database (http://crdd.osdd.net/raghava/thpdb/). Consequently, a considerable amount of research efforts has focused on bioactive peptides with a view to investigate such peptides as functional food ingredients targeted at health promotion in humans. Additionally, these bioactive peptides have the potential to be formulated as nutritional supplements to provide their beneficial health effects to the general populations (Nongonierma and FitzGerald, 2016). This might subsequently promote the cultivation and consumption of the specific food types.

Plant tubers are major staple foods that form an important source of dietary proteins and carbohydrates, especially in African countries where the consumption of protein-rich animal products is low (Shewry, 2003). The major types of tubers that account for >99 % of the world tuber consumption are potato (*Solanum tuberosum*), cassava (*Manihot esculenta*), sweet potato (*Ipomea batatas*), yam (*Dioscorea spp*) and taro (*Colocasia esculenta*). Interestingly, in addition to their roles as sources of dietary carbohydrates, these tubers are also endowed with various proportions of storage proteins which serve as the main source of dietary proteins upon their consumption. The main tuber storage proteins content of potato, sweet potato, yam and taro are

patatins, sporamins, dioscorins and tarin respectively (Shewry, 2003), while the existence of true storage proteins in cassava is not yet fully characterized and established. Hence, the relevance of these tuber storage proteins, especially to African populations, cannot be overemphasized considering the role of tubers as major staples in the continent.

The use of *in silico* analysis is a recent and useful technique for predicting the release of bioactive peptides from known protein sequences as well as in the selection of precursor proteins not previously investigated as sources of bioactive peptides (Lafarga *et al.*, 2014). The use of *in vitro* methodologies is considered expensive and time consuming, and the *in silico* analysis provides rapid and important information on the peptide sequence prior to laboratory-based evaluation and further clinical trials. Moreover, there is generally a very good agreement between the *in silico* prediction and the *in vitro* bioactive effects of the peptides (Nongonierma and FitzGerald, 2016). Consequently, a number of authors have utilized *in silico* analysis to predict the release of bioactive peptides from food substances, especially of animal origin, such as meat (Minkiewicz *et al.,* 2011), milk (Nongonierma *et al.,* 2014) and fish (Huang *et al.,* 2015) with very limited data on plant-derived dietary proteins. Therefore, considering the crucial position of plant-derived dietary proteins to African populations, there is an urgent need to investigate the profiles of potential bioactive peptides of the proteins with a view to decipher their possible health beneficial effects.

In this article, *in silico* analysis was used to predict the profiles and potential bioactive peptide sequences released after enzymatic digestion of tuber storage proteins in the gastrointestinal tract.

2. Materials and methods

2.1 Materials

Patatins, sporamins and dioscorins are categorized into A and B subtypes each while the major storage proteins in taro are categorized into G1 (tarin) and G2 (a and b) globulins. The amino acid sequence and molecular mass of each of these proteins were retrieved from the Uniprot Knowledgebase available at http://www.uniprot.org/. Other relevant details of the analyzed proteins are summarized in Table 1.

Table 1 Background information on the selected tuber storage proteins used to conduct the *in silico* studies

Protein	**Source**	**Uniprot Entry identifier**	**Uniprot Entry name**	**Number of amino acids residues**	**Molecular mass (Da)**
Patatin-01 (A)	Potato	Q2MY50	PAT01_SOLTU	387	42,608
Patatin-B1 (B)	Potato	P15476	PATB1_SOLTU	386	42,609
Sporamin A	Sweet potato	P14715	SPOR1_IPOBA	219	24028
Sporamin B	Sweet potato	P14716	SPOR2_IPOBA	216	24032
Dioscorin A	Yam	Q9M4Z0	Q9M4Z0_DIOAL	273	30982
Dioscorin B	Yam	Q9M501	Q9M501_DIOAL	273	31351
Tarin	Taro	Q43418	Q43418_COLES	253	27694
*Globulin G2a	Taro	Q9S8T7	Q9S8T7_COLES	29	3153
*Globulin G2b	Taro	Q9S8T8	Q9S8T8_COLES	25	2861

* Partial sequences as the complete sequence is not yet available

2.1 In silico simulated gastrointestinal digestion of the tuber storage proteins

The amino acid sequences of each of the retrieved proteins were subjected to proteolysis to simulate gastrointestinal digestion in the BIOPEP database (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) (Minkiewicz *et al.*, 2008). The "Enzyme(s) action" application in the BIOPEP database was used to simultaneously digest each of the proteins by using digestive enzymes in the gastrointestinal tract; pepsin (pH 1.3), chymotrypsin and trypsin to mimick the actual *in vivo* digestion.

2.3 Prediction of the bioactivities of the generated peptides using the BIOPEP database

After the simultaneous digestion by the combination of chymotrypsin, trypsin and pepsin, the released peptide fragments were further subjected to "search for active fragments" tool of the database where exact peptides sequences with known bioactivities as documented in the BIOPEP database were identified.

2.4 Prediction of antimicrobial and anticancer activities of the released peptides

The potential of the released peptides to serve as AMP was predicted using the predictive tool in the Collection of Antimicrobial Peptides (CAMP) Database (http://www.camp.bicnirrh.res.in/). Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Network (ANN) and Discriminant Analysis (DA) were the 4 multivariate statistical models used for the analysis in the CAMP database which was described in detail by Waghu *et al*. (2014). The results for the prediction were presented with the relevant probability scores, except for ANN. The peptides were categorized as AMPs or non-AMPs. The criteria used for the classification of a peptide as AMP was when the resulting score was ≥ 0.5 and positive recognition as AMP was obtained for at least two statistical models. All the peptides that met the above inclusion criteria were further subjected to the Antimicrobial Peptide Database (APD) (http://aps.unmc.edu/AP/main.php) to determine the net charge, hydrophobicity and percentage similarity to AMPs deposited in APD (Wang *et al.*, 2009).

The anticancer potential of the released peptides was predicted using the anticancer predictive tool in the AntiCP database (http://crdd.osdd.net/raghava/anticp/) which uses the SVM based

prediction method (Tyagi *et al.*, 2013). In this study, all the released peptides were submitted in to the database and the SVM threshold value of 1.0 was selected for the anticancer prediction.

2.5 Peptide ranking for bioactivity

In order to determine the possible existence of other potential bioactive peptides whose bioactivity was not predicted in the above studies, the released peptides from each of the proteins were subjected to peptideRanker (http://bioware.ucd.ie/~compass/biowareweb/Server_pages/peptideranker.php) and their peptide scores were calculated (Mooney *et al.*, 2012). All peptides with scores > 0.7 were considered to be bioactive and using this criterion, the potential bioactive peptides from each protein were manually searched to determine whether, or not, they were predicted to possess any of the bioactivity as earlier analyzed by the BIOPEP, antimicrobial and anticancer databases. All the potential bioactive peptides without a prior identified bioactivity were selected.

2.6 Peptide toxicity prediction

The released peptides were analyzed for potential toxicity using the ToxinPred tool available at http://crdd.osdd.net/raghava/toxinpred/ (Gupta *et al.*, 2013). In this study, the SVM based prediction method with a threshold value of 1.0 was selected for toxicity prediction.

3. Results

Complete sequences of seven tuber storage proteins with a range of 253-387 amino acid residues were analyzed alongside incomplete sequences of globulins G2a and G2b from taro (Table 1). Subsequently, each of the proteins was digested using the "Enzyme action" tool of the BIOPEP database with a combination of pepsin, chymotrypsin and trypsin. The full fragmentation pattern of each of the proteins in the presence of the three enzymes is presented in the supplementary file with 387 generated peptides from all the proteins. All the analyzed proteins released bioactive peptides of peptides with 2-18 amino acid residues ranging from 38 for sporamin A to 74 for patatin B1 (Table 2). According to the BIOPEP database, all the proteins (except G2 globulins) released known bioactive peptide sequences. For instance, patatin 01, sporamin A, dioscorin A and tarin generated 21, 11, 20 and 16 known bioactive peptide sequences respectively. The profiles of the potential biological activities of the peptides following the *in silico* analysis of every protein

Table 2. BIOPEP analysis of the bioactive peptides predicted to be released from tuber storage proteins after digestion with pepsin, chymotrypsin and trypsin

Protein	Total number of released peptides	Bioactivity of the peptides					
		DPP IV inhibition	ACE inhibition	Antioxidative	Renin inhibition	Glucose uptake stimulation	CaMPDE inhibition
Patatin- 01	71	SL (2); IL; ML (2); MY; NL; NY; PF; QW; SF; TK	MY; QK; NY; SF; ASL	MY	EF (2)	IL	EF (2)
Patatin-B1	74	SL (3); IL; ML; MY; NL; NY; PF; SF; TK; TY	MY; GPL; NY; SF;	MY; TY;	EF (2)	IL	EF (2)
Sporamin A	38	AL (4); SL; MK; TL; TY; VF	VF;	TY	-	-	-
Sporamin B	39	AL (2); SL; VR; AF; MK; TR; VF; VK; VL	VF; VK; AF; VR; CF	SDF	-	VL	-
Dioscorin A	50	HL; PL; IR; ML; NY; QL (2); SK; SY;	IR; VAY; PL; HL; GK; NY; SY;	HL; IR	IR	-	IR
Dioscorin B	51	HL; PL; ML; QF; QL (2) SY	PR; PL; HL; GR; GK; QK; SY	HL	-	-	-
Tarin	54	AL; GL; AF; GW; NR; TL; VK; VL; VVF	VK; GW; AF; GL	TDY; NHK	-	VL	-
*Globulin G2a	5	0	0	0	0	0	0
*Globulin G2b	5	0	0	0	0	0	0

Other bioactive peptides were GPR and DY as antithrombotic and ion flow regulating peptides respectively, released by patatin B1. The number in parenthesis denotes the total number of peptide fragments with the respective bioactivity. * Partial sequences as the complete sequence is not yet available

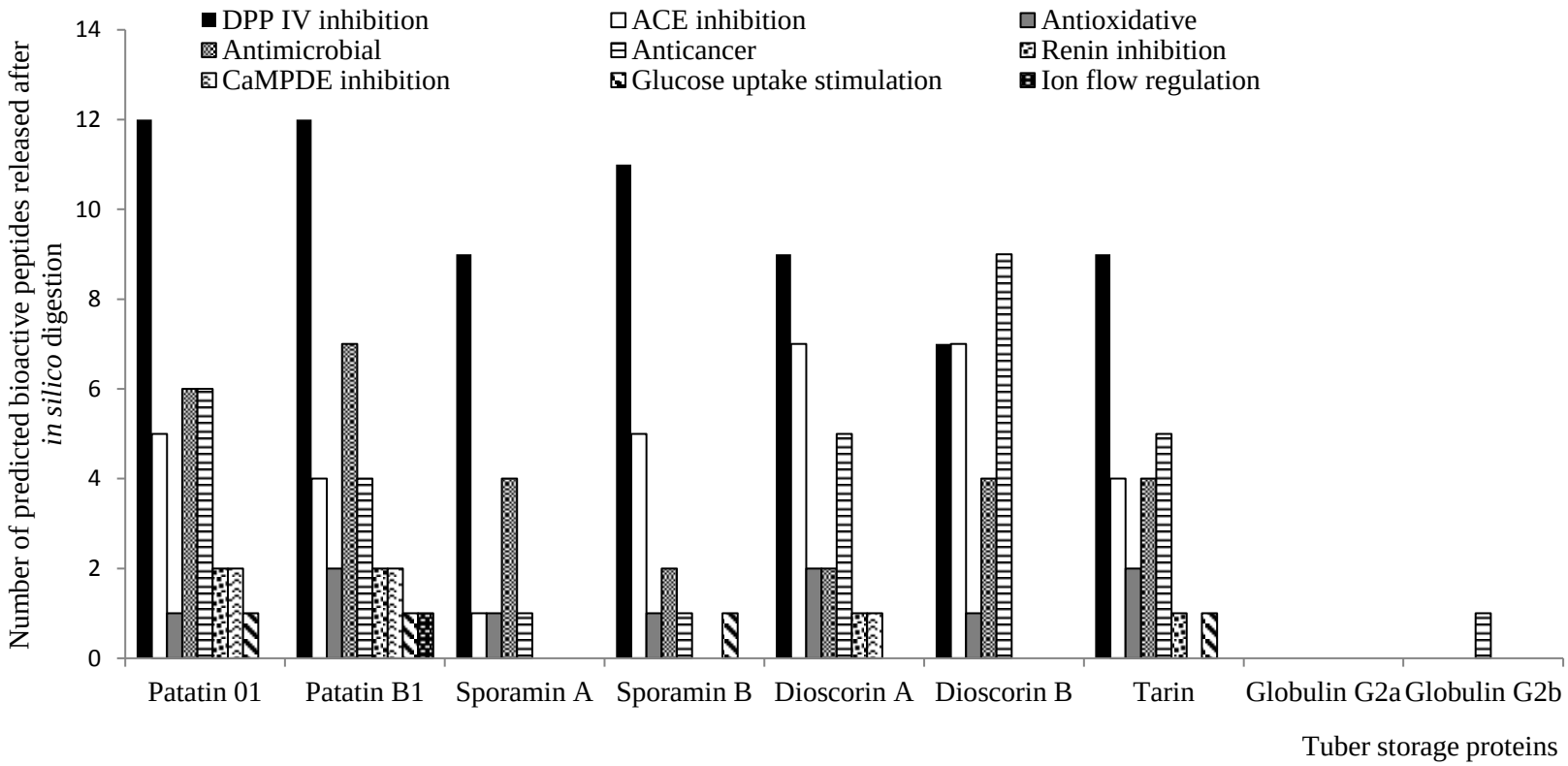

Figure 1: Profiles of predicted bioactive peptides released from the tuber storage proteins after *in silico* simulated gastrointestinal digestion

revealed that dipeptidyl peptidase IV (DPP-IV) inhibitory dipeptides were the predominant group of bioactive peptides in all the proteins. Indeed, all the seven tuber storage proteins with complete sequences released an array of potential DPP-IV inhibitory dipeptides (Table 2). However, the highest number (11) of the DPP-IV inhibitory dipeptides was released by patatins (01 and B1) (Figure 1). It is also noteworthy that the dipeptides SL, AL and QL were the predominantly released DPP-IV inhibitory dipeptides from patatins, sporamins and dioscorins respectively.

According to the BIOPEP database, all the proteins (except G2 globulins) generated known bioactive peptides with angiotensin converting enzyme (ACE) inhibitory and antioxidative activities. However, dioscorins (A and B) generated the highest number (7 di- and tripeptides) of ACE inhibitory peptides, while one antioxidative peptide was released by each of the proteins with the exception of patatin B1 and tarin with two antioxidative peptides each (Table 2). In fact, two bioactive tripeptides were released by patatin B1 (GPL, GPR) and tarin (TDY, NHK) unlike a single or no bioactive tripeptide released by the other proteins. The bioactive tripeptides released by tarin are predicted to be antioxidative peptides. Other peptides with predicted bioactivity are renin and calmodulin-dependent phosphodiesterase (CaMPDE) inhibitors, stimulators of glucose uptake, antithrombotic activity and ion flow regulators were released in varying numbers by the proteins. It was interesting that the bioactive peptides with renin and CaMPDE inhibitory activities were identical to EF (2) and IR released by the patatins and dioscorin A respectively (Table 2).

Prediction of the generated peptides to serve as AMPs revealed that all the proteins could release potential AMPs upon gastrointestinal digestion (Table 3). Based on the selection criteria of positive results in at least two algorithms, twenty eight peptides were predicted as potential AMPs from all the analyzed proteins while patatins had the highest number of predicted AMPs ($\geq$ 5), followed by dioscorin B and tarin with 4 predicted AMPs each (Table 3). Among all the predicted AMPs, DIVPF was the only one with positive prediction in three algorithms (SVM, RF and ANN) but no closely similar sequence was found in the APD database as it was only similar to a known AMP (WAIVLL) by 33.33 % in the APD database. It is also noteworthy that the potential AMP released from dioscorin B (DGDDDF) had positive AMP prediction in two algorithms (SVM and RF) and a high similarity of 66.66% with a known AMP (DEDDD) in the APD database (Table 3). Generally, most of the predicted AMPs were anionic with only five

Table 3. Potential antimicrobial peptides released during the *in silico* digestion as predicted by the algorithms available at the CAMP database (http://www.camp.bicnirrh.res.in/) and similar sequences found in the APD (http://aps.unmc.edu/AP/main.php) database

Protein	Released peptide	SVM	RF	ANN	DA	Net Charge	Hydrophobicity (%)	*Similar AMP in APD	*Percentage similarity (%)
Patatin 01	EEMVTVL	0.865 (AMP)	0.3825 (NAMP)	AMP	0.003 (NAMP)	-2	57	ELLVDLL	42.85
	GIIPGTIL	0 (NAMP)	0.4215 (NAMP)	AMP	0.845 (AMP)	0	50	FLPLIGKILGTIL	46.15
	SPEL	1 (AMP)	0.5445 (AMP)	NAMP	0.027 (NAMP)	-1	25	EL	50
	DNPETY	0.999 (AMP)	0.6685 (AMP)	NAMP	0.133 (NAMP)	-2	0	NGVQPKY	33.33
	EEAL	1 (AMP)	0.441 (AMP)	NAMP	0.008 (NAMP)	-2	50	ELLL	50
Patatin B1	EEMVTVL	0.865 (AMP)	0.3825 (NAMP)	AMP	0.003 (NAMP)	-2	57	ELLVDLL	42.85
	DIVPF	0.950 (AMP)	0.503 (AMP)	AMP	0.011 (NAMP)	-1	60	WAIVLL	33.33
	SPEL	1 (AMP)	0.5445 (AMP)	NAMP	0.027 (NAMP)	-1	25	EL	50
	PPHHF	1 (AMP)	0.452 (NAMP)	NAMP	0.844 (AMP)	+2	20	FFHLHFHY	37.5
	SSIK	1 (AMP)	0.459 (NAMP)	AMP	0.043 (NAMP)	+1	25	SLLSLIRKLIT	36.36
	DSPETY	1 (AMP)	0.6805 (AMP)	NAMP	0.004 (NAMP)	-2	0	KTCENLADTY	30
	EEAL	1 (AMP)	0.441 (AMP)	NAMP	0.008 (NAMP)	-2	50	ELLL	50
Sporamin A	AGGNY	0.892 (AMP)	0.304 (NAMP)	AMP	0.005 (NAMP)	0	20	QGGQANQ	42.85
	DMMSK	1 (AMP)	0.567 (AMP)	AMP	0.002 (NAMP)	0	40	AMVSS	40
	HDHML	0.999 (AMP)	0.759 (AMP)	NAMP	0.018 (NAMP)	+1	40	HLGHHALDHLLK	33.33

	PTDV	1 (AMP)	0.503 (AMP)	NAMP	0.001 (NAMP)	-1	25	TVVTQA	28.57
Sporamin B	HDSESGQY	1 (AMP)	0.647 (AMP)	NAMP	0.026 (NAMP)	-1	0	GSEIQPR	33.33
	IEVVNDNL	0.985 (AMP)	0.5145 (AMP)	NAMP	0.045 (NAMP)	-2	50	ELLVDLL	44.44
Dioscorin A	TVIK	0 (NAMP)	0.531 (AMP)	AMP	0.047 (NAMP)	+1	50	WLLVNK	33.33
	QPTNF	1 (AMP)	0.48 (NAMP)	NAMP	0.808 (AMP)	0	20	TYVTNA	33.33
Doscorin B	DGDDDF	1 (AMP)	0.518 (AMP)	NAMP	0 (NAMP)	-4	16	DEDDD	66.66
	SDPF	0.030 (NAMP)	0.5505 (AMP)	NAMP	0.530 (AMP)	-1	25	RPPGFSPFR	30
	DDPAY	1 (AMP)	0.711 (AMP)	NAMP	0.045 (NAMP)	-2	20	RYPAVGYT	37.5
	TVIK	0 (NAMP)	0.531 (AMP)	AMP	0.047 (NAMP)	+1	50	WLLVNK	33.33
Tarin	HPDGR	0.717 (AMP)	0.6025 (AMP)	NAMP	0 (NAMP)	+1	0	PLGG	40
	GPSVF	0.85 (AMP)	0.345 (NAMP)	AMP	0.011 (NAMP)	0	40	RPPGFSPFR	40
	NDPW	0.001 (NAMP)	0.513 (AMP)	NAMP	0.873 (AMP)	-1	25	DEKGPKWKR	30
	PAIW	0.003 (NAMP)	0.355 (NAMP)	AMP	0.827 (AMP)	0	75	WAIVLL	33.33

Only those sequences with positive AMP prediction in at least two of the models are presented. SVM, RF, ANN and DA refer to support vector machine, rain forest, artificial neural network and discriminant analysis respectively whilst CAMP and APD means collection of antimicrobial peptides and antimicrobial peptides database respectively. NAMP means non-antimicrobial peptide

cationic AMPs (PPHHF and SSIK (patatin B); HDHML (sporamin A); TVIK (dioscorins A and B); HPDGR (tarin)).

Thirty two bioactive peptides with anticancer potential were also predicted to be generated after the simulated gastrointestinal digestion of the analyzed proteins (Table 4) using the selected criteria of SVM threshold value of 1.0. In this instance, dioscorin B had the highest number of predicted anticancer peptides (9) followed by patatin 01 with six predicted peptides. Indeed, patatin 01 generated a peptide THTAEETAK with the highest likelihood to serve as potential anticancer peptide (SVM score = 1.43). However, among the peptides released by dioscorin B, TAPPCTEDITW had the highest prediction of anticancer activity (SVM score = 1.25). Another interesting observation in the anticancer prediction was that globulin G2b, without a prior bioactive prediction of any released peptide, also generated a peptide ANPVL with anticancer potential. Overall, the predicted anticancer peptides were also anionic with the exception of cationic ISQISR and QPGPIR from dioscorin B and tarin respectively, (Table 4). The predicted anticancer peptides also had greatly varied hydrophobicity ranging from 14.29% for NSSTGQF to 80% for ANPVL.

Upon the identification of bioactive peptides in the analyzed proteins, it was observed that not all the released peptides had predicted bioactivity. Hence, peptideRanker was used to rank the generated peptides from every protein according to their potential to be bioactive using a peptideRanker score of > 0.70 as a cutoff point. A total of 9 peptides were predicted to be bioactive based on the selected criteria but no specific bioactivity could be identified for them (Table 5). Each of the patatins as well as tarin had two of such peptide sequences, although PVIF was common among the patatins. The peptide GPADPF from dioscorin A with a peptideRanker score of 0.93 had the highest likelihood to be bioactive among all the peptides with unpredicted bioactivity (Table 5).

Prediction of the toxicity profiles of the released bioactive peptides using the ToxinPred tool indicated that none of the peptides was potentially toxic within the selection criteria.

Figure 1 presents the summary of the potential bioactive peptides released from each of the analyzed proteins *vis* a *vis* the identified bioactivity. The patatins generated the highest number of potential DPP-IV inhibitors and AMPs while dioscorins had the highest number of potential ACE inhibitors and anticancer peptides.

Table 4. Potential anticancer peptides released from the *in silico* digestion of tuber storage proteins with pepsin, chymotrypsin and trypsin as predicted by the anticancer peptides (AntiCP) database (http://aps.unmc.edu/AP/main.php)

Protein	Released peptide	SVM score	Net charge	Hydrophobicity (%)
Patatin 01	EGQL	1.11	-1	25
	NSSTGQF	1.00	0	14.29
	TEVAISSF	1.01	-1	50
	THTAEETAK	1.43	-1	22.22
	MTDY	1.00	-1	25
	EEAL	1.17	-2	50
Patatin B1	EGQL	1.11	-1	25
	MTDY	1.00	-1	25
	ISTVF	1.15	0	60
	EEAL	1.17	-2	50
Sporamin A	PTDV	1.16	-1	50
Sporamin B	PTDM	1.12	-1	50
Dioscorin A	PQQAEDEF	1.14	-3	37.5
	TCGNGMEQSPIQL	1.05	-1	30.77
	ATDAR	1.26	0	40
	MGSY	1.01	0	25
	TAPPCTEDITW	1.25	-2	45.45
Doscorin B	MSSTL	1.00	0	40
	TPIL	1.04	0	75
	INQVEY	1.06	-1	33.33
	EVQMVHESQDQR	1.10	-2	25
	SDPF	1.06	-1	50
	ISQISR	1.02	+1	33.33
	DDPAY	1.04	-2	40
	MGSY	1.01	0	25
	TAPPCTEDITW	1.25	-2	45.45
Tarin	DGSTVW	1.10	-1	33.33
	NDPW	1.05	-1	50
	VPGL	1.13	0	75
	QPGPIR	1.02	+1	50
	SAPL	1.10	0	75
Globulin G2b	ANPVL	1.06	0	80

Only those sequences with positive anticancer prediction (SVM $\geq$ 1.0) in the AntiCP database are presented. The net charge of the peptides was computed peptide property calculator available at http://www.biosyn.com/peptidepropertycalculator/peptidepropertycalculator.aspx . Hydrophobicity was calculated from peptide hydrophobicity/hydrophilicity analysis program at (http://peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php)

Table 5. Peptides released after the *in silico* digestion of tuber storage proteins without identified bioactivity

Protein	Released peptide	PeptideRanker score	Net charge	Hydrophobicity (%)
Patatin 01	PPHY	0.77	+1	50
	PVIF	0.71	0	100
Patatin B1	SGSIF	0.79	0	40
	PVIF	0.71	0	100
Sporamin A	SNSPF	0.81	0	40
Sporamin B	SNIPF	0.86	0	60
Dioscorin A	GPADPF	0.93	-1	66.67
Doscorin B	AVIAIMF	0.76	0	100
Tarin	NGDF	0.81	-1	25
	NGNW	0.80	0	25

The potential for bioactivity was predicted by a high peptideranker ((http://bioware.ucd.ie/~compass/biowareweb/Server_pages/peptideranker.php) score (> 0.7). The net charge of the peptides was computed peptide property calculator available at http://www.biosyn.com/peptidepropertycalculator/peptidepropertycalculator.aspx. Hydrophobicity was calculated from peptide hydrophobicity/hydrophilicity analysis program at (http://peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php)

4. Discussion

The African continent is plaqued by a myriad of human health challenges including communicable and non-communicable diseases with very limited treatment and management options. Thus, understanding the potential of biopeptides derived from African major staples might serve as a strategy for the utilization of those staples as functional foods. Interestingly, these tubers are also available in other parts of the world. The present study demonstrated that tubers as major staples in Africa contain proteins that could release an array of bioactive peptides with diverse bioactivities after gastrointestinal digestion.

A total of 387 peptides were released from the tuber proteins following the *in silico* simulation of gastrointestinal digestion. Bioactivities predicted for the released peptides include DPP-IV, ACE and renin inhibitions, antioxidative and antithrombotic activities as well as glucose uptake stimulation. Among all the proteins, patatins derived from potato generated the highest amount of DPP-IV inhibitors. DPP-IV rapidly inactivates incretin hormones such as glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) that cause glucose-dependent insulin secretion, increase β- cells mass and decrease glucagon secretion (Ahren *et al.,* 2013). Therefore, DPP-IV inhibitors increase the half-life of the incretins and consequently maintain glucose-induced insulin secretion and reduce hyperglycemia in diabetes. A number of food-derived peptides have been predicted to possess DPP-IV inhibitory activity which also was confirm using *in vitro* studies (Lafarga *et al.,* 2014; Minkiewicz *et al.,* 2011). These observations might suggest that the 11 DPP-IV inhibitory bioactive peptides released by patatins may help in maintaining glucose-induced insulin secretion. However, unlike the general thought that most food-derived DPP-IV inhibitory peptides contain a proline residue (Lafarga *et al.,* 2014; Minkiewicz *et al.,* 2011), this was not observed in this study because only two dipeptides, PF and PL from patatins and dioscorins respectively, were found with the above-mentioned feature.

Antimicrobial peptides were also predicted to be released by all the analysed proteins using the CAMP database with results accuracy of 87 – 93 % for each of the statistical model (Thomas *et al.,* 2010). In this instance, the patatins also generated 9 different predicted AMPs which was highest among all the analysed proteins. Notably, the peptides SPEL and EEAL generated by the two patatins had 50 % similarity with the previously known AMPs, EL and ELLL, respectively. EL and ELLL are AMPs derived from marine bacterium *Bacillus subtilis* and active against both

bacteria and fungi (Thomas *et al.,* 2010). AMPs are generally classified into cationic and anionic AMPs, while the anionic account for approximately 30 % of food-derived predicted AMPs (Tareq *et al.,* 2014). However, the predicted AMPs from the tuber storage proteins, including the patatins, were mostly anionic (net charge -1 to -4) peptides which were also reported to form amphiphilic structures, such as α-helix or β-sheet, that are crucial for their activity and exert various effects on sensitive microorganisms (Dziuba and Dziuba, 2011). Considering the above observations, it is plausible to suggest that patatins derived from potatoes may serve as source of bioactive peptides with DPP-IV inhibitory and antimicrobial activity.

Hypertension along with the associated cardiovascular diseases is a major global public health issue and ACE inhibitors are the first line of therapy (Pfeffer and Frohlich, 2006). Thus, ACE inhibitory peptides are among the most widely studied bioactive peptides and all the proteins analysed in this study released an array of ACE inhibitors while dioscorins were the most promising in this regard. Five out of the ten ACE inhibitory peptides released by dioscorins (IR, GK, PR, GR and QK) have positively charged amino acid residue (R or K) at the C- terminal position while the remaining five peptides (PL, VAY, HL, NY and SY) have hydrophobic (aromatic or branched side chains) amino acid residues at the C-terminal position. Interestingly, these two features are known attributes of potent ACE inhibitory peptides (Bhat *et al.,* 2015). In fact, as an example, IR, PL and VAY are food-derived ACE inhibitory peptides with IC_{50} values of 659, 337.32 and 16 µM respectively (Miyoshi *et al.,* 1999; FitzGerald and Meisel, 1999; Byun and Kim, 2002).

The observed generation of potential anticancer peptides by the proteins under study was interesting because bioactive peptides in cancer therapy are receiving considerable attention and a number of peptide-based anticancer therapies are at various phases of preclinical and clinical trials (Gregorc *et al.,* 2011). However, the definition surrounding the structure-activity relationship of anticancer peptides is still under debate (Gaspar *et al.,* 2013) but current information on anticancer peptides revealed that the presence of C, G, I, K and W dominated at various position is a crucial structural feature (Tyagi *et al.,* 2013). All the predicted biopeptides had at least one of these amino acids. In this context, dioscorins generated the highest number of predicted anticancer peptides including the peptide TAPPCTEDITW with a C and W residues in addition to a high SVM score of 1.25. Furthermore, this peptide has a high hydrophobicity of 45.45 % which is also a feature

that was severally reported to enhance the anticancer properties of peptides (Gaspar *et al.*, 2013). However, it is noteworthy to state that anticancer drugs target a wide range of cellular pathways, which include cell cycle, DNA replication among others, but the AntiCP database used in this prediction was not designed to identify the precise mechanism of action of the anticancer peptides. Based on the foregoing along with the limited understanding of the structure-activity relationship of anticancer peptides, it may be worthwhile to further probe, using cell based assays, these predicted anticancer peptides from tuber proteins, especially TAPPCTEDITW generated by dioscorins A and B. On a general note, the propensity of ACE inhibitory and potential anticancer peptides released by dioscorins compared to other tuber proteins should stimulate research interest in the investigation of yam as a potential source of functional food relevant in the management of hypertension and cancer.

With the exception of the partially sequenced G2 globulins, sporamins generated the least amount of peptides possibly because of its shorter chain length (Table 1). In spite of this observation, it was noted that sporamins produced different kinds of DPP-IV and ACE inhibitory peptides such as VR and VF respectively, which were not released by any of the tuber proteins. Additionally, the sporamins released a rarely encountered glucose uptake stimulating peptide (VL). It seems therefore that the sporamins are generally rich in valine containing bioactive dipeptides (Table 2).

Using BIOPEP, the observed release of lesser number of peptides with antioxidative, antithrombotic, renin (hypotensive) and CaMPDE inhibitory properties as well as glucose uptake stimulating and ion flow regulating activities is consistent with previous reports (Minkiewicz *et* 2011; Keska and Stadnik, 2016). In this context, patatin B1 was the only analysed protein that generated a number of peptides with all the above mentioned bioactivities. In fact, the release of GPR and DY with antithrombotic and ion flow regulating activities may suggest the potential of patatins in the development of functional food for thrombosis management and ion flow regulations. Furthermore, the release of known CaMPDE inhibitors, EF and IR by the patatins and dioscorin A respectively, suggest that these proteins may provide a therapeutic effect against neurodevelopmental conditions and degenerative disorders such as fetal alcohol spectrum disorder and Alzheimer's disease respectively (Medina, 2011). This is because CaMPDE inhibitors are known to increase the levels of second messengers cAMP/cGMP leading to the expression of

neurotrophic factors, neuronal plasticity-related genes and other neuroprotective molecules (Medina, 2011).

The peptideRanker scores obtained for the peptides in Table 5 is a good prediction of a potential bioactivity for those peptides because the peptideRanker score ranges from 0 (no bioactivity) to 1 (definitive bioactivity) (Mooney *et al.*, 2012). The inability to predict any bioactivity within our defined criteria for those peptides may suggest that, perhaps, those peptides are potentially novel bioactive peptides. More *in silico* and/or *in vitro* analyses may be needed in this regard.

An important parameter for a functionally relevant bioactive peptide is non-toxicity. Therefore, the released peptides were analyzed using the ToxinPred programme (http://crdd.osdd.net/raghava/toxinpred/). It was interesting to note that none of the released peptides from all the proteins was predicted to be toxic which suggested that the peptides could potentially be used as ingredients with health promoting properties for human consumption.

In conclusion, *in silico* simulated gastrointestinal digestion of the analyzed tuber storage proteins indicated that the proteins released an array of non-toxic bioactive peptides with health promoting properties. Among all the analyzed tuber storage proteins, potato derived bioactive peptides seemed to have better antidiabetic and antimicrobial potentials, while yam derived bioactive peptides showed better potential as antihypertensive and anticancer agents. Furthermore, the potato derived bioactive peptides had the widest range of predicted bioactivities. Thus, this study has identified an important knowledge-gap in the search for novel bioactive peptides, especially from African foods, and the information may be use to promote the consumption of this tubers as functional foods and/or could be exploited for the development of novel neutraceuticals from the African tubers. Future studies should focus on the synthesis of the peptides and *in vitro* validation of the predicted bioactivities especially the antimicrobial and anticancer effects because prediction of other bioactivities by BIOPEP was based on previously reported *in vitro* analysis of the peptide sequences.

Acknowledgement

We acknowledge the National Research Foundation of South Africa and the University of Pretoria for the financial supports. The first author also acknowledges the University of Pretoria for the

award of postdoctoral fellowship position in Biochemistry and Ahmadu Bello University, Zaria, Nigeria for the award of a study fellowship.

Conflict of interest

The authors declare that they have no conflicts of interest

Reference

Ahren, B., Schweizer, A., Dejager, S., Villhauer, E.B., Dunning, B.E., & Foley, J.E. (2013). Mechanisms of action of the dipeptidyl peptidase-4 inhibitor vildagliptin in humans. *Diabetes, Obesity and Metabolism, 13*, 775-783.

Bhat, Z.F., Kumar, S., & Bhat, H.F. (2015). Bioactive peptides of animal origin: a review. *Journal of Food Science and Technology, 52,* 5377 -5392.

Byun, H.G., & Kim, S.K. (2002). Structure and activity of angiotensin I-converting enzyme inhibitory peptides derived from Alaskan pollack skin. *Biochemistry and Molecular Biology, 35,* 239-243.

Byun, H.G., Lee, J.K., Park, H.G., Jeon, J.K., & Kim, S.K. (2009). Antioxidant peptides isolated from the marine rotifer, *Brachionus rotundiformis*. *Process Biochemistry, 44*, 842-846.

Clare, D.A., Catignani, G.L., & Swaisgood, H.E. (2003). Biodefense properties of milk: The role of antimicrobial proteins and peptides. *Current Pharmaceutical Design 9,* 1239-1255.

Dziuba, B., & Dziuba, M. (2011). New milk protein-derived peptides with potential antimicrobial activity: an approach based on bioinformatic studies *International Journal of Molecular Sciences, 15,* 14531-14545.

FitzGerald, R.J., & Murray, B.A. (2006). Bioactive peptides and lactic fermentations *International Journal of Dairy Technology*, *59,* 118.

FitzGerald, R.J., & Meisel, H. (1999). Lactokinins: Whey protein-derived ACE inhibitory peptides. *Nahrung, 43*, 165-167.

Gaspar, D., Veiga, A.S., & Castanho, M.A. (2013). From antimicrobial to anticancer peptides. A review. *Frontiers in Microbiology, 4*, 294.

Gregorc, V., De Braud, F.G., De Pas, T.M., Scalamogna, R., Citterio, G., Milani, A., Boselli, S., Catania, C., Donadoni, G., Rossoni, G., Ghio, D., Spitaleri, G., Ammannti, C., Colombi, S., Caligaris-Cappio, F., Lambiase, A., & Bordignon, C. (2011). Phase I study of NGR-hTNF, a selective vascular targeting agent, in combination with cisplatin in refractory solid tumors. *Clinical Cancer Research, 17*, 1964–1972.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G.P.S. (2013). *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One, 8,* e73957.

Huang, B., Lin, H., & Chang, Y. (2015). Analysis of proteins and potential bioactive peptides from tilapia (*Oreochromis* spp) processing co-products using proteomic techniques coupled with BIOPEP database. *Journal of Functional Foods, 19*, 629–640.

Korhonen, H., & Pihlanto, A. (2007). Food-derived bioactive peptides-opportunities for designing future foods. *Current Pharmaceutical Design, 9,* 1297-1308.

Keska, P., & Stadnik, J. (2016). Porcine myofibrillar proteins as potential precursors of bioactive peptides – an *in silico* study. *Food and Functions, 7,* 2878–2885.

Lafarga, T., Connor, P.O., & Hayes, M. (2014). Identification of novel dipeptidyl peptidase-IV and angiotensin-I-converting enzyme inhibitory peptides from meat proteins using *in silico* analysis. *Peptides, 59,* 53-62.

Malaguti, M., Dinelli, G., Leoncini, E., Bregola, V., Bosi, S., Cicero, A.F., & Hrelia, S. (2014). Bioactive peptides in cereals and legumes: agronomical, biochemical and clinical aspects *International Journal of Molecular Sciences, 15,* 21120-21135.

Medina, A.E. (2011). Therapeutic utility of phosphodiesterase type I inhibitors in neurological conditions *Frontiers in Neuroscience 5,* 21.

Minkiewicz, P., Dziuba, J., & Michalska, J. (2011). Bovine meat proteins as potential precursors of biologically active peptides: a computational study based on the BIOPEP database. *Food Science and Technology International, 17,* 39–54.

Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., & Darewicz, M. (2008). BIOPEP database and other programs for processing bioactive peptide sequences. *Journal of AOAC International 91,* 965.

Mojica, L., & de Majia, M.G. (2016). Optimization of enzymatic production of anti-diabetic peptides from black bean (*Phaseolus vulgaris* L.) proteins, their characterization and biological potential. *Food and Functions, 7,* 713-727.

Möller, N.P., Scholz-Ahrens, K.E., Roos, N., & Schrezenmeir, J. (2008). Bioactive peptides and proteins from foods: indication for health effects. *European Journal of Nutrition, 47,* 171.

Mooney, C., Haslam, N.J., Pollastri, G., Shields, D.C. (2012). Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity. *PLoS One 7,* e45012.

Nongonierma, A.B., & FitzGerald, R.J. (2016). Strategies for the discovery, identification and validation of milk protein-derived bioactive peptides. *Trends in Food Science and Technology,* 50, 26-43.

Nongonierma, A.B., Mooney, C., Shields, D.C., & FitzGerald, R.J. (2014). *In silico* approaches to predict the potential of milk protein-derived peptides as dipeptidyl peptidase IV (DPP-IV) inhibitors. *Peptides, 57,* 43-51.

Park, Y.W., & Nam, M.S. (2015). Bioactive peptides in milk and dairy products: a review. *Korean Journal of Food Science and Animal Resources, 35,* 831-840.

Pfeffer, M.A., & Frohlich, M.D. (2006). Improvements in clinical outcomes with the use of angiotensin-converting enzyme inhibitors: Cross-fertilization between clinical and basic investigation. *American Journal of Physiology: Heart and Circulation Physiology 291,* 2021–2025.

Shewry, P.R. (2003). Tuber storage proteins. *Annals of Botany, 91,* 755 -769.

Tareq, F.S., Lee, M.A., Lee, H.S., Lee, Y.J., Lee, J.S., Hasan, C.M., Islam, M.S., & Shin, H.J. (2014). Gageotetrins A-C, non-cytotoxic antimicrobial linear lipopeptides from a marine bacterium *Bacillus subtilis. Organic Letters, 16,* 928-931.

Thomas, S., Karnik, S., Barai, R.S., Jayaraman, V.K., & Idicula-Thomas, S. (2010). CAMP: A useful resource for research on antimicrobial peptides. *Nucleic Acids Research 38,* D774–D780.

Tyagi, A., Kapoor, P., Kumar, R., Chaudhary, K., Gautam, A., & Raghava, G.P.S. (2013) *In silico* models for designing and discovering novel anticancer peptides. *Scientific Reports 3,* 2984.

Waghu, F.H., Gopi, L., Barai, R.S., Ramteke, P., Nizami, B., & Idicula-Thomas, S. (2014). CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Research, 42,* D1154–D1158.

Wang, G.S., Li, X., & Wang, Z. (2009). APD2: The updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Research, 37*, D933–D937.

Wang, S., Zeng, X., Yang, Q., & Qiao, S. (2016). Antimicrobial peptides as potential alternatives to antibiotics in food and animal industry. *International Journal of Molecular Sciences, 17*, 603.

Yin, C.M., Wong, J.H., Xia, J., & Ng, T.B. (2013). Studies on anticancer activities of lactoferrin and lactoferricin. *Current Protein and Peptide Science 14,* 492–503.

Zhang, Y., Wang, N., Wang, W., Wang, J., Zhu, Z., & Li, X. (2016). Molecular mechanisms of novel peptides from silkworm pupae that inhibit α-glucosidase. *Peptides, 76,* 45-50.