

RSC Advances



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This *Accepted Manuscript* will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



Journal Name

ARTICLE

Learnings from quantitative structure activity relationship (QSAR) studies with respect to food protein-derived bioactive peptides: A review

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Alice B. Nongonierma^a and Richard J. FitzGerald^{a*}

The generation of bioactive peptides (BAPs) from dietary proteins has been widely studied. One of the main limitations of a broader application of BAPs in functional foods may arise from their low potency. Therefore, the search for more potent structures is crucial. Quantitative structure activity relationship (QSAR) has been widely applied in drug discovery and some examples may also be found in the study of BAPs. The aim of this review was to assess the efficiency of QSAR for the discovery of novel and potent BAPs, derived from food protein sources. A wide range of bioactive properties including antioxidant, antimicrobial, angiotensin converting enzyme (ACE), renin and dipeptidyl peptidase IV (DPP-IV) inhibition as well as bitter peptides have been investigated with QSAR. Some studies have identified structural requirements for specific bioactivities, which generally confirmed findings from earlier studies carried out on those BAPs. However, discrepancies are found across analyses, possibly due to the quality of the peptide datasets as well as the descriptors used to build QSAR models. It appears to date that only a limited number of QSAR studies conducted with BAPs have subsequently carried out confirmatory studies and evaluated promising peptide sequences *in vivo*. This suggests that more research is needed in order to advance knowledge in the area of BAP discovery using QSAR.

1 Introduction

Food proteins contain peptide sequences which have been shown, using different *in vitro* bioassays, to positively affect specific health markers. Such peptides, which are called bioactive peptides (BAPs), may be released from a wide range of dietary proteins.¹ The efficacy and activity of BAPs has been studied *in vivo*, suggesting in certain instances that food protein-derived BAPs may be relevant to human health. However, contradictory results are still reported in the scientific literature in relation to their health benefits.²⁻⁴

An increasing number of peptide sequences identified in various food protein hydrolysates have been reported in the literature over the past few years. The technology used to identify these sequences within food protein hydrolysates or their fractions has greatly evolved. Earlier studies have mainly relied on immunoreactive techniques (e.g., enzyme-linked immunoassay - ELISA) or Edman degradation of the N-terminal amino acids linked to peptide sequencers to identify positive peptide candidates.⁵⁻⁷ However, advances in mass spectrometric (MS) analysis have allowed greater capability in the identification of peptides within complex mixtures and also

identification of a large number of peptide sequences in an automated manner, for reviews, see:⁸⁻¹¹. Nevertheless, challenges such as the reliable identification of short peptide sequences (< 5 amino acid residues) are still an issue for a more comprehensive understanding of dietary BAP structures.¹²⁻¹⁴

Overall, an understanding of the structural requirements for peptides having specific bioactivities has increased. This has been made possible by (1) novel strategies to release and/or isolate BAPs from food protein hydrolysates, (2) the increasing number of BAP sequences which have been identified, (3) access to a higher number of bioinformatic, peptidomic and proteomic tools as well as (4) structure-activity relationship studies.^{15,16} Conventionally, structure-activity relationships for BAPs have been developed using empirical knowledge based on previously known BAP sequences. Recently, the structure-function of peptides has been reported for a wide range of bioactive properties including mineral binding, angiotensin converting enzyme (ACE) and renin inhibition, antithrombic, antidiabetic, antimicrobial, immunomodulatory, antioxidant and opioid activities.¹⁷

Quantitative structure activity relationship (QSAR) modelling is a well-accepted approach for the study of active molecules, which is extensively used in drug discovery.¹⁸ QSAR has been developed to elucidate novel drug molecules displaying higher activity, produced at lower cost or which mediate less side-

^a Department of Life Sciences and Food for Health Ireland (FHI), University of Limerick, Limerick, Ireland. Email: dick.fitzgerald@ul.ie; Fax: + 353 -61331490; Tel: +353-61202598.

effects. Furthermore, several examples of QSAR are found for the study of food protein-derived BAPs, for reviews, see: ^{15,19}. QSAR studies have been applied to bioactive properties such as ACE inhibition, antimicrobial, anticancer and antioxidant activities, for reviews, see: ²⁰⁻²². Additional studies are also found where QSAR approaches have been employed to try to understand the link between peptide structure and bitterness. ²³⁻²⁵

A summary of the structural features of ACE inhibitory peptides, based on the QSAR outcomes, has been compiled in a recent review by Iwaniak, et al. ¹⁵ However, to our knowledge, studies summarizing the structural requirements for other BAPs, as determined by QSAR analysis, are rare. A better understanding of peptide structural requirements for bioactivity may lead to the discovery of novel peptide sequences with enhanced bioactivities. ²⁶ Therefore, the aim of this review was to assess existing QSAR approaches in order to identify common peptide motifs in BAPs which may be used to design novel dietary protein-derived BAP sequences. The search period covered by the review was from 1990 to date. QSAR studies were classified by bioactive properties, i.e., antioxidant, antimicrobial, ACE, renin and dipeptidyl peptidase IV (DPP-IV) inhibitory as well as bitter peptides.

2 QSAR methodology applied to BAPs

QSAR is defined as a relationship linking structural characteristics of molecules to their biological or chemical properties, for reviews, see: ^{20,27,28}. The general work flow to design a QSAR strategy, which is summarised in Fig. 1, will be described in the following subsections.

2.1 Building the BAP library

The first step in all QSAR analyses is to collate the sequences of target peptides to be used to build the QSAR model(s). These sequences generally arise from in-house or publicly available BAP databases. Various publicly available BAP databases have been listed in recent reviews. ^{15,19} BIOPEP, for example, is a database which is commonly employed for the study of dietary peptides. ²⁹

The peptide library used to construct the QSAR model contains quantitative biological outputs, e.g., half maximal inhibitory concentration (IC₅₀) values for enzyme inhibition, minimal inhibitory concentration (MIC) for microbial strains, half maximal effective concentrations (EC₅₀) or scavenging activities for antioxidant peptides, etc. ^{24,26,30,31} For comparative purposes, it is recommended that the bioactivity output should be obtained under similar experimental conditions. ^{18,27,32} The peptide database may be restricted to sequences which originate from a single protein ²⁶ or a group of proteins found within certain species (e.g., *Homo sapiens*, *Bos taurus*, etc.). ³³ In addition to the peptide origin, the dataset can also be made of peptides having a defined amino acid length or incorporate peptides having various amino acid lengths. ²⁴ In other instances, QSAR strategies have focused on peptide analogs of a so-called "lead peptide". The lead peptide may be used to design specific analogs which are then employed to construct minimum analogue peptide sets (MAPS). ^{31,34} The sequence of the peptide candidates to be included in the peptide library and subsequently tested to generate biological activity data may be defined using a full or fractional factorial design. The MAPS is designed in such a way as to incorporate the minimum number of peptides in the dataset while covering a wide range of amino acid physicochemical properties and while simultaneously varying their positions within the peptide sequences. The number of peptide analogs generated depends on the number of amino acid positions varied within the peptide sequence. ³⁴ The inclusion of peptide analogs in datasets has been described, for example, for derivatives of lactoferricin (LFcin), a peptide with antimicrobial properties. ^{35,36}

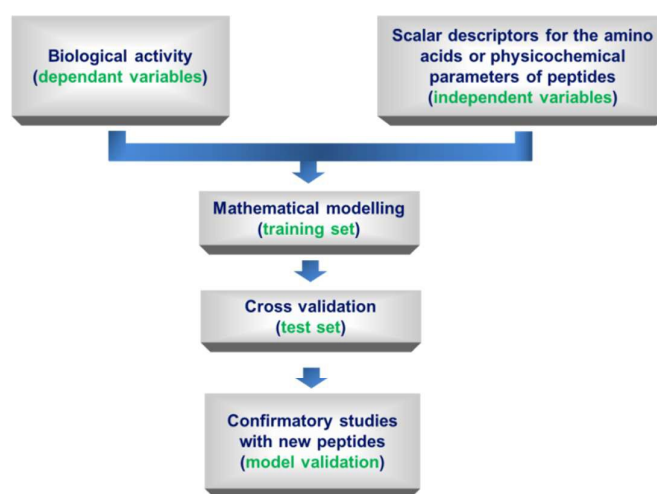


Fig. 1 Schematic of the general approach used in quantitative structure activity relationship (QSAR) methodologies for the study of bioactive peptide (BAP) sequences (adapted from Hellberg et al. ³¹ and Iwaniak et al. ¹⁵).

Journal Name

ARTICLE

It has been stated that selection/design of the peptide set is the most crucial step to conduct successful QSAR studies.³¹ Once the peptide library has been compiled, the peptides are generally classified into a “training set” and a “test set”. This classification generally consists in randomly excluding a certain number of peptides from the peptide library. These excluded peptides will then be part of the test set, which is subsequently used for cross validation of the QSAR model. The peptides within the test set may also be chosen in order to cover a wide range of structural characteristics. For example, this may be achieved using a statistical molecular design approach which is based on a fractional design methodology.²⁷

2.2 Description of the peptides using scalar descriptors of the constituent amino acids

The next step of the QSAR study consists in describing the peptides incorporated in the training set. This may be achieved either using selected physicochemical descriptors or amino acid scalar descriptors. Different scalar descriptors for amino acids have been developed.^{26,34} These may be classified as 1- (molecular formula), 2- (topological), 3- (conformational) or 4D (orientation and time-dependent) descriptors.¹⁸ Examples of 2- and 3D amino acid descriptor scales are outlined in Tables 1 and 2, respectively. One of the popular amino acid descriptor scales used in the early QSAR studies is the 3-z score scales which was developed by Hellberg, et al.³⁴ The 3-z values, which were determined using a principal component analysis (PCA) combining 29 physicochemical characteristics of the 20 conventional amino acids, correspond to the hydrophilicity (z1), size (z2) and charge (z3) of the amino acids.³⁴ Later on, a 5-z scale was developed by Sandberg et al.³⁷, incorporating the 12 previously used physicochemical properties for the development of an extended 3-z scale³⁸ and 14 other physicochemical properties to describe 87 amino acids. The 5-z scale descriptors incorporated lipophilic, steric and electronic properties of amino acid side chains. The v-scale of Lin et al.³⁹ incorporates three structural characteristics (van der Waals volume, net charge index and hydrophobicity) of the 20 conventional amino acids. The v-scale was developed with the view of providing more meaningful physicochemical properties to describe structural parameters of amino acid side chains.

Other amino acid descriptors have been described to develop 3D scales. The T-scale has been developed by Tian et al.⁴⁰ from a PCA of 67 structural and topological variables of 135 amino acids. A PCA was carried out on the hydrogen bonding (5), electronic (23), steric (37) and hydrophobic (54) properties of the 20 conventional amino acids, yielding 10 descriptors termed as divided physicochemical

property scores (DPPS).⁴¹ The 3D scale developed by Collantes and Dunn⁴² combines amino acid side-chains descriptors, i.e., the isotropic surface area (ISA, hydrophobic character of the side chain substituent) and the electronic charge index (ECI, charge concentration of the amino acid).

The main limitation of those scalar descriptors may be the difficulty in relating biological activity to specific physicochemical characteristics of the peptides.²⁷ To overcome this issue, QSAR modelling analyses combining a z-score approach and individual physicochemical characteristics of the peptides have been carried out.³²

2.3 QSAR mathematical modelling

The third step of the QSAR analysis consists in building the QSAR model(s). Each peptide in the training set is described by the descriptors of its constitutive amino acids or its overall physicochemical characteristics. QSAR modelling consists in mathematical equations, which are based on computational methods (e.g., multiple linear regression (MLR), partial least square regression (PLSR), PCA, artificial neural network (ANN), etc.).²¹ The QSAR model is a multivariate data approach linking the structural characteristics (peptide descriptors - independent variables) to the biological activity (dependent variable) of the peptides.

When the peptide descriptors are defined by the characteristics of its constitutive amino acids, a QSAR model may be described by eqn (1):

$$A_i = \alpha_0 + \sum_1^j \sum_1^k \alpha_{j,k} AA_{j,k} + \varepsilon \quad (1)$$

With A_i : the biological activity of the peptide i ; j : the position of the amino acid within the peptide i ; k : the number of the amino acid descriptor; $\alpha_{j,k}$: the coefficients of the model; α_0 : the constant of the model and ε the residual.

Some peptide QSAR approaches have studied peptide sets made of one defined amino acid length (e.g., di- or tripeptides). It is however possible to develop QSAR models using peptide libraries composed of sequences having variable amino acid lengths. This has been addressed by Pripp, et al.³² and Li and Li⁴³ in the development of a QSAR approach for ACE-inhibitory peptides (2-6 amino acid in length) and antioxidant peptides (3-20 amino acid in length). In such QSAR studies, the length (l) of the smallest peptide of the dataset is generally taken into consideration in order to build the model. The QSAR model includes peptide descriptors of the l amino acids located at the N- and C-terminus of each peptide (eqn 2):

$$A_i = \alpha_0 + \sum_1^j \sum_1^k \alpha_{j,k} AA_{j,k} + \sum_1^j \sum_1^k \alpha_{j,k} N_{j,k} + \sum_1^j \sum_1^k \beta_{j,k} C_{j,k} + \varepsilon \quad (2)$$

With A_i : the biological activity of the peptide i ; j : the N- or C-terminal position of the amino acid of the peptide i (varying between 1 and l); k : the number of the amino acid descriptor; $\alpha_{j,k}$, $\beta_{j,k}$ and α_0 : the coefficients constant of the model and ε the residual.

The QSAR model is subsequently statistically cross-validated using the peptides within the test set (originally excluded from the training set) to verify the ability of the QSAR model to be applied to unknown compounds.

2.4 Confirmatory studies with synthetic peptides

In certain instances, the QSAR analysis may allow determination of structural features of peptides which are linked with enhanced bioactive properties. These features may subsequently be used to design peptides which are (1) predicted to be highly bioactive and (2) novel in their sequence. The novelty of the peptide can be assessed after searching the sequence in BAP databases as well as in the relevant patent literature. Confirmatory studies allow external validation of the QSAR model and testing its robustness in terms of its ability to predict the bioactivity of peptides.

To date, the highest number of QSAR studies which are relevant to food protein-derived peptides appear to have been conducted with ACE inhibitory or antimicrobial peptides (AMPs). The following

sections review QSAR studies which have been classified according to the target *in vitro* bioactivity (i.e., antioxidant, antimicrobial, ACE, renin and DPP-IV inhibition) as well as bitterness of the peptides.

3 QSAR studies with specific BAPs

3.1 ACE inhibitory peptides

Several studies employing QSAR approaches have focused on ACE inhibitory peptides, for reviews, see:^{21,44}. In most of these studies, QSAR has been used to better understand the structural features of peptides which correlate with *in vitro* ACE inhibitory activity and in certain cases to predict novel and potent ACE inhibitory peptides. The main outcomes of these QSAR studies are summarised in Table 3. The peptide descriptors used to construct the QSAR models were either based on physicochemical characteristics^{32,45} or scalar amino acid descriptor scales.^{30-32,46} Most of the QSAR studies have successfully proposed the structural requirements linked with the ACE inhibitory properties of peptides. These consisted of specific amino acids located at different positions in the peptide sequence (Table 3). However, depending on the QSAR study, different structural features for ACE inhibitory peptides have been proposed. This may arise from the training set used, the experimental conditions employed to obtain the biological activity of the peptides, the size and diversity of the peptides and the amino acid descriptors used to construct the QSAR model.

Table 1. Examples of amino acid descriptors which have been employed for 2D quantitative structure activity relationship (QSAR) modelling of bioactive peptides (BAPs).

Amino acid	3 z-score ³⁴			v-scale ³⁹			5 z-score ³⁷				
	z1	z2	z3	v1	v2	v3	z1	z2	z3	z4	z5
A	0.07	-1.73	0.09	0.05702	0.007187	0.42	0.24	-2.32	0.60	-0.14	1.30
R	2.88	2.52	-3.44	0.58946	0.043587	-1.37	3.52	2.50	-3.50	1.99	-0.17
N	0.71	-0.97	4.13	0.22972	0.005392	-0.82	3.05	1.62	1.04	-1.15	1.61
D	-1.39	2.32	0.01	0.21051	-0.02382	-1.05	3.98	0.93	1.93	-2.46	0.75
C	0.92	-2.09	-1.4	0.14907	-0.03661	1.34	0.84	-1.67	3.71	0.18	-2.65
Q	3.64	1.13	2.36	0.34861	0.049211	-0.30	1.75	0.50	-1.44	-1.34	0.66
E	3.08	0.39	-0.07	0.32837	0.006802	-0.87	3.11	0.26	-0.11	-3.04	-0.25
G	2.23	-5.36	0.30	0.00279	0.179052	0.00	2.05	-4.06	0.36	-0.82	-0.38
H	2.41	1.74	1.11	0.37694	-0.01069	0.18	2.47	1.95	0.26	3.90	0.09
I	-4.44	-1.68	-1.03	0.37671	0.021631	2.46	-3.89	-1.73	-1.71	-0.84	0.26
L	-4.19	-1.03	-0.98	0.37876	0.051672	2.32	-4.28	-1.30	-1.49	-0.72	0.84
K	2.84	1.41	-3.14	0.45363	0.017708	-1.35	2.29	0.89	-2.49	1.49	0.31
M	-2.49	-0.27	-0.41	0.38872	0.002683	1.68	-2.85	-0.22	0.47	1.94	-0.98
F	-4.92	1.30	0.45	0.55298	0.037552	2.44	-4.22	1.94	1.06	0.54	-0.62
P	-1.22	0.88	2.23	0.2279	0.239531	0.98	-1.66	0.27	1.84	0.70	2.00
S	1.96	-1.63	0.57	0.09204	0.004627	-0.05	2.39	-1.07	1.15	-1.39	0.67
T	3.22	1.45	0.84	0.19341	0.003352	0.35	0.75	-2.18	-1.12	-1.46	-0.40
W	-4.75	3.65	0.85	0.79351	0.037977	3.07	-4.36	3.94	0.59	3.44	-1.59
Y	2.18	0.53	-1.14	0.6115	0.023599	1.31	-2.54	2.44	0.43	0.04	-1.47
V	-2.69	-2.53	-1.29	0.25674	0.057004	1.66	-2.59	-2.64	-1.54	-0.85	-0.02

Journal Name

ARTICLE

Table 2. Examples of amino acid descriptors which have been employed for 3D quantitative structure activity relationship (QSAR) modelling of bioactive peptides (BAPs).

Amino acid	Collantes scale ⁴²		DPPS scale ⁴¹					T scale ⁴⁰									
	ISA	ECI	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	T1	T2	T3	T4	T5
A	62.90	0.05	-1.02	-2.88	-0.56	0.36	-6.15	-1.68	0.04	-2.51	-1.94	-0.01	-9.11	-1.63	0.63	1.04	2.26
R	52.98	1.69	1.99	4.13	-4.41	-1.02	4.78	3.04	-9.06	6.71	4.41	0.07	0.23	3.89	-1.16	-0.39	-0.06
N	17.87	1.31	-2.19	1.86	0.38	-0.13	-2.30	1.41	-5.71	-1.11	1.73	-0.19	-4.62	0.66	1.16	-0.22	0.93
D	18.46	1.25	-6.60	3.32	1.61	0.36	-3.25	1.95	-7.36	0.14	1.24	-0.15	-4.65	0.75	1.39	-0.40	1.05
C	78.51	0.15	0.21	1.12	3.42	-0.68	-2.27	-1.22	3.11	-2.98	-1.70	1.57	-7.35	-0.86	-0.33	0.80	0.98
Q	19.53	1.36	-0.47	1.16	-0.57	0.69	0.39	1.93	-5.46	-0.84	1.93	0.85	-3.00	1.72	0.28	-0.39	0.33
E	30.19	1.31	-5.39	0.65	-0.98	1.39	-0.23	2.51	-6.84	-0.68	1.41	1.28	-3.03	1.82	0.51	-0.58	0.43
G	19.93	0.02	-2.86	-5.00	-2.97	0.53	-11.45	1.89	-2.11	-3.99	-2.16	-0.76	-10.61	-1.21	-0.12	0.75	3.25
H	87.38	0.56	0.73	2.68	-0.66	-1.89	1.60	1.13	-1.94	-0.11	0.44	0.15	-1.01	-1.31	0.01	-1.81	-0.21
I	149.77	0.09	1.91	-3.13	0.01	1.14	2.70	-4.55	8.93	0.18	-1.10	-0.76	-4.25	-0.28	-0.15	1.40	-0.21
L	154.35	0.10	1.64	-2.57	0.00	1.35	2.62	-2.65	7.72	0.05	-1.03	-1.81	-4.38	0.28	-0.49	1.45	0.02
K	102.78	0.53	2.47	1.54	-4.28	-0.86	2.77	2.06	-6.18	2.05	2.19	-1.65	-2.59	2.34	-1.69	0.41	-0.21
M	132.22	0.34	1.93	-0.01	1.21	0.99	2.79	-0.56	5.33	-0.87	-0.99	-1.09	-4.08	0.98	-2.34	1.64	-0.79
F	189.42	0.14	2.68	0.84	2.22	0.71	5.02	-0.30	8.60	1.13	-1.40	-0.28	0.49	-0.94	-0.63	-1.27	-0.44
P	122.35	0.16	0.45	-2.89	1.77	-5.81	-3.79	-0.61	0.70	1.21	-1.67	1.79	-5.11	-3.54	-0.53	-0.36	-0.29
S	19.75	0.56	-1.76	-0.19	1.06	-0.69	-5.72	0.14	-4.14	-2.42	-0.13	0.69	-7.44	-0.65	0.68	-0.17	1.58
T	59.44	0.65	-0.55	-0.66	0.13	-0.31	-2.76	-1.56	-2.46	-2.12	0.17	0.08	-5.97	-0.62	1.11	0.31	0.95
W	179.16	1.08	3.88	1.78	1.68	2.00	9.31	0.89	7.53	4.27	-0.23	-1.42	5.73	-2.67	-0.07	-1.96	-0.54
Y	132.16	0.72	2.10	1.26	1.15	0.91	5.90	0.74	3.71	3.32	0.25	1.33	2.08	-0.47	0.07	-1.67	-0.35
V	120.91	0.07	0.83	-3.02	-0.22	0.97	0.05	-4.55	5.61	-1.41	-1.44	0.30	-5.87	-0.94	0.28	1.10	0.48

^a DPPS: divided physicochemical property scores

While different structural features for potent ACE inhibitory peptides have been reported, the C-terminal sequence of the peptide appear to have a major contribution to ACE inhibition while a minor contribution of the N-terminal sequence has been proposed in several of the QSAR studies conducted to date.^{30,32,33,45,46} In particular, the C₁ (C-terminal) amino acid of peptides is thought to have a major effect on the *in vitro* ACE inhibitory properties of peptides. Overall, most QSAR studies have indicated that the presence of aliphatic hydrophobic and small amino acids (Ala, Trp, Pro, Phe, Gly, Cys, Leu and Ile) at the C₁ position of peptides was a good predictor for potent *in vitro* ACE inhibitory activity. Amino acids located at other positions also appear to play a role in modulating the overall ACE inhibitory properties of peptides. It was reported, for instance, that the nature of C₂ to C₄ amino acid in peptides had an effect on the ACE

inhibitory potential.^{45,46} The importance of the C-terminal sequence in ACE inhibition may come from the specific mode of action of the enzyme. ACE is a dipeptidyl carboxypeptidase, meaning that peptide binding to its active site occurs through the C-terminal dipeptide sequence.⁴⁷ More particularly, hydrophobic amino acids have been described to bind to the hydrophobic S'₂ subsite of the ACE active site.⁴⁵ Zhou, et al.⁴⁸ have shown that there was a relationship between the binding energy of peptides to ACE and their inhibitory potency, with a positive contribution of hydrophobicity to peptide binding. In addition, they demonstrated using truncated versions of Gln-Pro-Leu-Ile-Tyr-Pro that the C-terminal dipeptide (Tyr-Pro) played a preponderant role in peptide binding to ACE and that the C₅ and C₆ amino acids hardly affected binding.⁴⁸

To date, most QSAR studies on ACE inhibition appear to have been applied to relatively short peptides (≤ 8 amino acid residues, Table 3). A recent QSAR study on ACE inhibitory activity of peptides has been applied to the largest peptide ($> 1,400$ peptides) dataset employed to date.⁴⁹ The originality of this study also lied in the fact that it took into account peptides with a broad size range (from 2 to > 12 amino acid residues). The peptides were classified according to their amino acid length as “tiny” (≤ 3 amino acids), “small” (4–6 amino acids), “medium” (7–12 amino acids) and “large” (> 12 amino acids). When analyzing the amino acid composition of ACE inhibitory peptides, Gly ($\geq 13\%$) was found to be the most abundant amino acid in dipeptides while Pro ($\geq 14\%$) was the most abundant in peptides having ≥ 3 amino acids. This result is in agreement with QSAR outcomes from the study of Wang, et al.⁵⁰ also predicting that Gly was the preferred amino acid residue both at the N- and C-terminal position of ACE inhibitory dipeptides.

Despite the very large number of QSAR studies focusing on ACE inhibitory peptides, only a few have subsequently been utilised as predictive tools to design peptides with potent ACE inhibitory properties (Table 3). In a few instances, QSAR studies have led to the identification of novel ACE inhibitory peptides following confirmatory *in vitro* studies with synthetic peptides.^{30,33,51} Novel ACE inhibitory peptides, Tyr-Phe and Leu-Arg-Phe, being 2–10 times more potent than the well-known ACE inhibitory lactotriptide Ile-Pro-Pro, were reported.³⁰ Similarly, Ile-Val-Pro ($IC_{50} = 49.7 \pm 4.2 \mu\text{M}$) and Val-Ile-Pro ($IC_{50} = 26.1 \pm 0.8 \mu\text{M}$) were found to be relatively potent ACE inhibitors.³³ Huang, et al.⁵¹ specifically applied QSAR to C-terminal Pro containing tripeptides and successfully identified two novel ACE inhibitory peptides, Ile-Ile-Pro and Ile-Val-Pro having IC_{50} values of 1.39 and 1.58 μM , respectively.

To our understanding, new ACE inhibitory peptides identified using QSAR approaches have been evaluated in one study for their *in vivo* hypertensive properties.³³ Ile-Val-Pro and Val-Ile-Pro were administered to spontaneously hypertensive rats (SHRs) at a dose of 0.75 mg kg⁻¹. Interestingly, these treatments resulted in a significantly higher reduction (~ 3 times, $p < 0.05$) in systolic blood pressure (SBP) than Ile-Pro-Pro, when evaluated at the same dose.³³

3.2 Renin inhibitory peptides

Besides ACE, the inhibition of renin may also be targeted for a better regulation of blood pressure. To our knowledge, only one QSAR study has been conducted to date on renin inhibitory peptides.⁵² This may be because this area of research is relatively new and only a limited number of renin inhibitory peptide sequences have been identified to date. The QSAR study was developed using the percentage of renin inhibition of 11 dipeptides (evaluated at 3.2 mM) originally identified in a pea protein hydrolysate having renin inhibitory properties. The outcomes of the QSAR analysis showed that peptides with potent renin inhibitory properties were predicted to possess a bulky amino acid (Trp, Tyr or Phe) at the C-terminus while hydrophobic and small amino acids (Val, Leu, Ile and Ala) may be found at the N-terminus of the

dipeptide (Table 3). Peptides with a Trp at the C-terminus were predicted to be the most potent renin inhibitors. Confirmatory studies were conducted with four dipeptides (Leu-Trp, Ile-Trp, Ala-Trp and Val-Trp). Two of the peptides (Ala-Trp and Val-Trp) were inactive towards renin, while the renin inhibitory activity determined experimentally was lower or of the same order as the predicted inhibition for⁵³ Leu-Trp and Ile-Trp, respectively. In addition, Ile-Trp ($IC_{50} = 2.3 \pm 0.07 \text{ mM}$) was a newly identified renin inhibitory peptide also displaying high ACE inhibitory activity ($IC_{50} = 4.74 \pm 0.04 \mu\text{M}$). Interestingly, Ile-Trp had previously been shown to induce a significant reduction in SBP in SHR studies.^{54,55}

3.3 DPP-IV inhibition

Several studies have demonstrated that certain peptide sequences can inhibit DPP-IV *in vitro* and also in small animals, for reviews, see:^{53,56–60} To date, it is still not clear which structural features of peptides are required to induce high DPP-IV inhibition. A number of structural studies based on peptide alignment approaches have shown that the presence of Pro or Ala at position 2 of the peptide as well as certain amino acids located at the N-terminal position of the peptide such as Trp, along with branched chain amino acids (Leu, Ile) correlated with relatively potent *in vitro* DPP-IV inhibition.^{61,62} To date, only one study appears to have applied a QSAR approach to better understand the structural characteristics of milk protein-derived peptides in relation to their DPP-IV inhibitory properties.⁶³ During this study, the issue of incorporating IC_{50} data obtained under different experimental conditions as well as peptides which were not competitive DPP-IV inhibitors into the QSAR model was highlighted. However, the QSAR model failed to allow an accurate prediction of the IC_{50} data for competitive DPP-IV inhibitory peptides. However, the ranking for their DPP-IV inhibitory potency was conserved. Additional QSAR studies incorporating a larger number and more diverse (length and amino acid composition) peptides may help to improve the predictive ability of QSAR models as applied to DPP-IV inhibitory peptides.

The outcomes of the two QSAR models (developed with a 3-z and 3-v scales, Table 3) applied to DPP-IV inhibitory peptides revealed the importance of hydrophobic amino acids (Trp, Ile, Leu and Phe) at the N-terminal position of the peptide.⁶³ In addition, the 3-v scale model also showed the importance of hydrophobic amino acids located at position 2 of the peptide. These findings were in agreement with earlier structural studies showing the importance of hydrophobic amino acids at the N-terminal side of peptides with DPP-IV inhibitory properties.^{61,62} The aim of this QSAR study was not to design peptides with more potent DPP-IV inhibitory properties. However, the QSAR models were used as a tool to predict the DPP-IV inhibitory properties of a large number of peptides which have previously been identified in the gastrointestinal tract of humans following milk ingestion. Confirmatory studies allowed the identification of milk protein derived peptides relevant to humans with relatively high *in vitro* DPP-IV inhibitory potency such as Leu-Pro-Val-Pro-Gln displaying an IC_{50} value of $43.8 \pm 8.8 \mu\text{M}$.

Journal Name

ARTICLE

Table 3. Features of bioactive peptides (BAPs) as determined by quantitative structure activity relationship (QSAR) studies

Bioactivity ^a	Peptide length ^b	Scale used ^c	Favorable amino acid ^d at the peptide position (C ₁ is the C-terminal amino acid)					Confirmatory studies ^d	Reference	
			C ₆	C ₅	C ₄	C ₃	C ₂			C ₁
ACE inhibition	2 (n = 58)	3-z scale					hydrophobic & positively charged Xaa	hydrophobic & bulky Xaa	none	31
	2-8 (n = 36)	8 physicochemical characteristics					small Xaa	hydrophobic & non-positively charged Xaa	none	32
	2-6 (n = 29)	3-z scale					small & non-positively charged Xaa	non-positively charged Xaa	none	
	2 (n = 168)	3-z scale					F/Y/W	F/Y/W/P	FW, WW, YW	30
	3 (n = 140)	3-z scale				V/L/I	K/R	P/F/W	VRF, IKP, LRW, LRF	
	2 (n = 168)	3- and 5-z scales					F/Y/W	F/Y/W	none	46
	3 (n = 140)					V/L/I	R/K	P/F/W	none	
	4 (n = 79)			V/I/V/M	R/H/W/F	F	Y/P/F	none		
	≥ 5 (n = 226)			W	I/L/V/M	H/W/M	Y/C	none		
	2 (n = 58)	3-z, 5-z, 8-v, 38 physicochemical characteristics, ISA and ECI scales					hydrophobic Xaa	hydrophobic, small Xaa	none	23
	3 (n = 55)	17 different scales				bulky, charged Xaa		G	none	39
	2 (n = 58)	10 different scales					G	G	none	50
	3 (n = 55)	3-z, 5-z and 8-G scales				G	G	G	none	
	9 (n = 19)	3-z, 5-z and 8-G scales		P					none	
≥ 5 (n = 245/18)	38 physicochemical characteristics		V/I	W/Y/C	D/N/K	R/V/T	G/L/A/V/I	none	45	
3 (n = 17)	3-z scale				V/L/I	P/C	P/F/W	IVP, INP, IQP, VIP	33	
3 (n = 38)	3-z scale						IIP, IVP	51		
Renin inhibition	2 (n = 11)	3-z and 5-z scales					V/L/I/A	W/Y/F	LW, IW, AW, VW	52
Antioxidant	3 (n = 143/71)	DPPS, HESH, ISA-ECI, MS-WHIM, VHSE and 3-z scales				A/G/V/L/E	R/K/H	-	64	

ARTICLE								Journal Name		
	3 (n = 143/71)	8-v scale			A/V/L	R/K/H/D/E/T /S/N/Q	W/E/L/I/M/V/Y	none	65	
	4 (n = 12)									
	2 (n = 32)	DPPS, ISA-ECl, 3-z and 5-z scales				Y	hydrophobic, small, low hydrogen bond and electronic Xaa		66	
						W	bulky, hydrophobic Xaa			
Bitterness	2 (n = 48)						hydrophobic & bulky Xaa	hydrophobic & bulky Xaa	none	31
	2 (n = 77)	3-z scale					hydrophobic, polar/charged Xaa	hydrophobic or bulky Xaa		25
	3 (n = 52)						bulky Xaa	hydrophobic, bulky Xaa		
	4 (n = 23)				basic, bulky Xaa			basic, bulky, hydrophobic Xaa		
	5 (n = 12)			basic, bulky Xaa				bulky, hydrophobic Xaa		
	6 (n = 20)		basic, bulky, hydrophilic Xaa					bulky, hydrophobic Xaa		
	7 (n = 16)						bulky, hydrophobic Xaa	bulky, basic Xaa		
	2 (n = 48)	3-z, 5-z, 8-v, 38 physicochemical characteristics, ISA and ECI scales					hydrophobic, small Xaa	hydrophobic, small Xaa	none	23
	2 (n = 53)	3-z scale					bulky Xaa	hydrophobic Xaa	none	24
	3 (n = 55)	3-z scale					bulky Xaa (W/R/Y)	hydrophobic Xaa		
	2 (n = 48)	17 different scales					W/Y/F	W/Y/F		39
Favorable amino acid ^d at the peptide position (N ₁ is the N-terminal amino acid)										
			N ₁	N ₂	N ₈	N ₉	N ₁₃			
DPP-IV inhibition	2-5 (n = 21/5)	3-z scale	W/I/F/L	W/I/F/L				VPGEIVE, LPQNIPPLT, LPLPLL, QPLPPT, LPVPQ	63	
		3-v scale	W/I/F/L							
Anti-microbial	15 (n = 19)	12 murine LFcin analogs	W		C	C	V/L/I/M	none	67	

^a ACE: angiotensin converting enzyme; DPP-IV: dipeptidyl peptidase IV. ^b the number of peptides used to build the QSAR model is given into bracket (when two numbers are provided, the first and second ones correspond to the number of peptides in the training and the test set, respectively). ^c DPPS: divided physicochemical property scores; ECI: electronic charge index; HESH: hydrophobic, electronic, steric and hydrogen; ISA: isotropic surface area; MS-WHIM: molecular surfaces-weighted holistic invariant molecular; VHSE: vectors of hydrophobic, steric and electronic properties. ^d The amino acids are coded with their one letter code. Xaa: amino acid.

Journal Name

ARTICLE

3.4 Antimicrobial peptides

AMP sequences are found within several food proteins. They have been proposed as a means, for example, to (1) reduce infection in living organisms, (2) address the issue of antibiotic resistant strains, (3) combat the formation of biofilms or (4) extend the shelf-life of food products.^{68,69} The application of QSAR to AMPs has been reviewed by Jenssen.²⁰ A large number of QSAR studies carried out on AMPs have been published.

Several studies have utilised QSAR approaches to better understand the structural requirements of the antimicrobial lactoferrin (LF)-derived pentadecapeptide, LFCin, for reviews, see:^{18,70} In these studies, LFCin from different species was used as a lead peptide to construct peptide analogs which were then employed to build a QSAR model in order to define structural characteristics of potent AMPs. A QSAR study on bovine LFCin (LF f(14-41)) analogs (8 peptides with 12-19 amino acid residues) was conducted to study antimicrobial activity against *Escherichia coli* and *Staphylococcus aureus*.⁷¹ Large and negatively charged peptides or peptides with a high hydrophobic moment (μ) displayed high antimicrobial activity against *E. coli* and *S. aureus*, respectively. In another study, 19 murine LFCin (LF f(16-30)) analog peptides were used to build a QSAR model in relation to their antimicrobial activity against *E. coli* and *S. aureus*.³⁵ The importance of the N-terminal amino acid was highlighted. It is thought that the N-terminal amino acid may establish electrostatic interactions with the negatively charged phospholipids of bacterial membranes. Thereafter, it was hypothesised that hydrophobic amino acids (Trp, Tyr) may bind to the interface through hydrogen bonding, causing the phospholipid membranes to leak. While this study concluded that good antibacterial activity was obtained by replacement of several amino acids within the peptide sequence of murine LFCin, it was shown that the parameters of importance were the net charge and micelle affinity of the peptide. The most effective murine LFCin analog was found to be LFCin Arg1, 9 Trp8 Tyr13 (with an Arg at position 1 and 9, a Trp at position 8 and a Tyr at position 13).

Most QSAR studies conducted with antimicrobial LFCin peptides appear to have been designed with descriptors consisting of physicochemical parameters. Therefore, they took into account the whole characteristics of the peptide as opposed to the properties of its individual amino acids, making them more complicated to interpret and to employ for subsequent design of novel peptide sequences. Very few QSAR studies have attempted to elucidate the amino acid descriptors which correlated with the antimicrobial activity of LFCin peptide analogs.⁶⁷ The 3-z scale was used for the amino acids located at the 4 varied positions (1, 9, 8 and 13) of 19

murine LFCin analogs. The preferred amino acids at each position were determined following QSAR analysis (Table 3). A larger peptide dataset of human, bovine, caprine and murine LFCin analogs (52 peptides) was employed in another QSAR study using the 3-z scale for peptide descriptors.⁷² The most important properties that governed the antimicrobial activity (*E. coli* and *S. aureus*) was z1 (hydrophilicity) for amino acids located at positions 1, 3, 4 and 14, z2 (size) for position 10 and 14 and z3 (charge) for position 4.

The ability of human LFCin analog peptides to act as cell membrane permeabilising agents of *Pseudomonas aeruginosa*, as a means to subsequently enhance the action of a synthetic antibiotic (novobiocin) and to avoid antibiotic resistance mechanisms, was studied using a QSAR approach.^{73,74} There was no direct relationship between peptide antimicrobial activity and cell permeabilising activity. The QSAR analysis revealed a positive correlation between antimicrobial activity and peptide hydrophobicity, the number of Trp and aromatic residues as well as the percentage of hydrophobic plus basic residues. On the other hand, peptides with cell permeabilising activity generally possessed aromatic and positively charged residues and had an amphiphilic structure.

Antimicrobial peptides other than LFCin have also been studied by QSAR.⁷⁵⁻⁷⁷ For example, the search for non-hemolytic (to reduce the risk of endotoxic shock) cyclic cationic peptides having high antibacterial properties has been aided by QSAR studies.⁷⁷ It was shown that charge and amphipaticity was responsible for increased antibacterial activity. On the other hand, the anti-hemolytic effect was linked to lower peptide lipophilicity, in particular that of the residues involved in the nonpolar face of the peptides, which are likely to form a β -hairpin-like structure.

While some QSAR studies have concluded in structural characteristics relevant to AMPs, clear structural requirements for AMPs still do not appear to be available, as stated earlier.¹⁸ This has been explained by the fact that the structural requirements of AMPs appear to involve a quite complex combination of specific physicochemical properties (hydrophobicity, cationic residues, amphipaticity). In addition, the mechanism of action and specific target of AMPs are still not fully understood.⁷⁴

3.5 Antioxidant peptides

Antioxidant peptides have also been studied using QSAR approaches. The structural characteristics of peptides with antioxidant activity, as determined by QSAR studies, are

summarised in Table 3. QSAR study of the radical scavenging (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and 2,2'-azobis(2-aminopropane) dihydrochloride (AAPH)) ability of dipeptides containing Trp and Tyr at the N-terminus was carried out.⁶⁶ Differences in the preferred C-terminal amino acid residue of dipeptides with ABTS scavenging properties were seen when a Tyr or Trp was found at the N-terminus of the dipeptide (Table 3). For larger peptides (tri- and tetrapeptides), it was shown that the N-terminal (high hydrophobicity and low electronic properties) and next to the C-terminal (hydrophilic) amino acids were the most important in the antioxidant (ferric thiocyanate method, autoxidation system of linoleic acid) activity.^{64,65} It was suggested that radical scavenging peptides could interact with the lipid phase owing to their hydrophobic N-terminal amino acid while the amino acid at position 2 (next to the C-terminus) of the peptide could act through its ability to form hydrogen bonds with free radicals.⁶⁵ 2- and 3D descriptors were used to conduct a QSAR study with antioxidant (superoxide (SO), ABTS and AAPH radical scavenging) peptides possessing 2-20 amino acid residues.⁴³ In this study, different amino acids were shown to be important depending on the radical scavenging activity considered. For the ABTS, AAPH and SO scavenging, the C₂ and N₂ terminal amino acids were the most important in relation to the antioxidant activity, while for both AAPH and SO scavenging the C₃, C₄ and C₁ were also linked with antioxidant activity. Overall, the study of Li and Li⁴³ highlighted that the C-terminal amino acids had a higher impact on the antioxidant (radical scavenging) properties than the N-terminal amino acids. The electronic properties were shown to play the most important role in radical scavenging ability of the peptides. In addition, the presence of a bulky hydrophobic amino acid at the C-terminus was also shown to correlate with the radical scavenging ability of peptides.⁴³

A recent QSAR study has utilised peptide descriptors linked with electron transfer properties, i.e., energy of highest occupied molecular orbital (E_{HOMO}) and bond length of active sites ($L_{(\text{X-H})}$), to study the antioxidant activity ($\cdot\text{OH}$ and $\text{O}_2\cdot^-$ scavenging) of di- to heptapeptides.⁷⁸ The $\text{O}_2\cdot^-$ scavenging ability of peptides correlated with high E_{HOMO} and long $L_{(\text{X-H})}$. However, the predictive ability of the model designed for $\cdot\text{OH}$ scavenging was not very accurate, possibly due to interferences of the assay reagents (Cu^{2+}). Overall, this study has highlighted the role of electron effects on the radical scavenging ability of peptides.

A systematic evaluation of the antioxidant capacity (ferric reducing antioxidant power - FRAP) of all possible tripeptides (172 unique sequences) from β -Lg has been carried out.²⁶ The QSAR analysis showed that antioxidant activity was governed by the electronic and hydrogen-bonding properties of all amino acids within the peptide. For the N- and C-terminal amino acids, it was also shown that the steric properties of the amino acids were important. Cys- and Trp-containing tripeptides were associated with high antioxidant activity. The effect of both of these amino acid residues was explained by their ability to interact with free radicals through

hydrogen (-SH and indole group), electron (S) or proton (aromatic ring of Trp) donation. It was found that 3 β -Lg tripeptides (Leu-Thr-Cys, Cys-Gln-Cys and Gly-Thr-Trp) had a higher antioxidant activity than the well-known physiological antioxidant glutathione (Glu-Cys-Gly).

The antioxidant properties of peptides may be determined using a wide range of assays. These assays are targeted at quite different oxidative species, therefore, leading to different outcomes in terms of the peptides' potential ability to reduce oxidation.⁶⁶ These differences highlight the challenges in attempting to find a consensus in the structural properties for antioxidant peptides. As highlighted in the recent review of Li and Yu,²² no clear consensus between peptide structure and antioxidant peptides has been established using QSAR studies, possibly due to the lack of knowledge of the mechanism of action of such peptides.

4 Bitter peptides

Besides bioactive properties, the sensory attributes of peptides have also been studied using QSAR approaches. The need to develop bioactive protein hydrolysates with low bitterness has been highlighted in several studies due to the fact that certain food protein hydrolysates display bitter attributes, for reviews, see:^{79,80}. The characteristics of bitter peptides as determined by QSAR studies are summarised in Table 3.

In several studies, the R^2 determined with QSAR models developed for bitter peptides was shown to be quite low. This was related to the difficulty of accurately measuring bitterness as its threshold is subject to interindividual variation in humans.⁴⁸ Asao, et al.⁸¹ conducted a QSAR study on short (di- and tri-) peptides and their derivatives and showed a positive link between lower bitterness threshold and the peptides' physicochemical properties (i.e., length of the carbon backbone and octanol/water partition coefficient). In line with structural studies conducted on bitter short (di- and tri-) peptides, other QSAR studies have shown that hydrophobicity and size of the amino acids at the C₁ and C₂ position of peptides, respectively, correlated positively with bitterness.^{23,24}

For datasets incorporating larger peptides (2-14 amino acid residues, $n = 224$), it was shown using the 3-z scale that the bitterness was correlated with hydrophobic amino acids in the C₁ position and bulky, basic and hydrophilic amino acids in N₁ (N-terminal) position of the peptides.²⁵ The molecular mass, hydrophobicity, number of amino acid residues in the peptide together with the amino acid descriptors of the 3-z scale were incorporated in a PLSR. The previous physicochemical parameters had a higher influence on the bitterness than the 3-z scale. This suggests that the bitterness depends on the overall properties of the peptide rather than on the properties of specific amino acids within the peptide sequence.²⁵

A number of QSAR studies have attempted to identify peptides displaying both good bioactivity profiles and low bitterness. It has

been shown that several potent ACE inhibitory or antioxidant short peptides (2 to 3 amino acid residues) were also bitter.^{23,40,48,82} In contrast, in another study,²⁴ no direct correlation was seen between ACE inhibitory properties and bitterness of di- or tripeptides. The number of possible dipeptide combinations ($20^2 = 400$) is lower than that of larger peptides, increasing the structural diversity of larger peptides. Therefore, the likelihood of identifying highly BAP candidates with low bitterness is increased with larger peptides. In addition, the C₁ to C₄ amino acid residues are important in peptide binding to ACE. Hence the interest of identifying peptides with 3–4 amino acid residues having high ACE inhibitory activity and a good sensory profile.^{23,48}

5 Limitations of QSAR for the study of BAPs

The main limitations of QSAR analysis, which relate to peptide descriptors, dataset and the methodology used for model development and validation, have been comprehensively reviewed by Cherkasov, et al.¹⁸ It has been pointed out on a number of occasions that some of the amino acid scales may be challenging to interpret especially when the different components of the scale do not represent a defined physicochemical and/or structural property.⁶⁴ However, it is still possible to identify preferred amino acids in specific locations on peptides which would correlate with high bioactivity. The choice of peptide descriptors to build the QSAR model may have an impact on model accuracy. Over the years, different amino acid scales have been developed, incorporating a higher number of physicochemical properties or structural characteristics which are deemed important for the bioactive properties of peptides (Table 1 and 2). Generally, QSAR models are constructed with several amino acid scales or a wide range of peptide physicochemical properties. Subsequently, peptide descriptors giving the highest correlation with the bioactivity are selected. Several QSAR approaches have suggested that specific scales allowed obtention of QSAR models with higher correlation and cross-validation coefficients (R^2 and Q^2 , respectively), which has generally been the case with the more recently developed amino acid descriptor scales.^{37,39,49,83,84}

Some QSAR studies incorporating relatively large numbers of BAPs in the dataset have been conducted with bioactivity data obtained using different experimental conditions (enzyme:substrate ratio, source of enzymes, substrate, temperature, etc.), which makes the validity of the models developed questionable. Several studies have also not taken into account the mode of action of the peptides with certain biological receptors. This is particularly relevant for enzyme inhibition assays where peptides may act at the active site (competitive enzyme inhibition) or outside the active site (modes of inhibition other than competitive) of the target enzyme(s). Recently, Nongonierma and FitzGerald⁶³ have shown when using QSAR that it was only possible to obtain a statistically significant correlation between DPP-IV inhibition of peptides and their descriptors when applying a series of filters such as IC₅₀ data obtained under the same experimental conditions for inhibitors

with the same mode of DPP-IV inhibition (competitive). While many QSAR studies have yielded models with statistically significant correlations, other studies have not achieved this result, possibly due to the relatively low number of peptides in the training set, structural diversity and/or structural relevance of the peptides therein.^{31,51,66} This again highlights the importance of the quality/heterogeneity of the data which may be included in the QSAR model.¹⁸ In addition, when the mechanism of action is not fully understood, e.g., mode of enzyme inhibition, it can be challenging to select meaningful peptide descriptors to build QSAR models.

In addition to selecting the appropriate peptide sequences to build QSAR models, a limited number of studies have attempted to train the models with so-called “negative datasets” which incorporate peptides which do not display any bioactivity. In their QSAR approach developed with ACE inhibitory peptides, Kumar, et al.⁴⁹ have also included negative datasets to develop classification models in order to assign unknown peptides to a category of active or inactive peptides. These models have been incorporated in a freely available web resource (<http://crdd.osdd.net/raghava/ahtpin>).

6 Utilisation of specific features in the targeted release of BAPs from dietary proteins during enzymatic hydrolysis

Studies which have embarked on translating QSAR findings to the targeted release of BAPs through dietary protein hydrolysis are scarce. The development of enzymatic strategies to release promising BAPs from food proteins is of significant interest to the food industry. Enzymatic hydrolysis is well accepted as a food-grade approach for the large-scale production of BAPs in a relatively cost-efficient manner.^{1,17} In order to release specific BAP sequences from food proteins, knowledge of both the amino acid sequence of the protein as well as the enzyme specificity and kinetics is required.

Using QSAR models applied to ACE inhibitory peptides, an *in silico* approach was developed in order to predict dietary protein substrates which would act as good precursors of ACE inhibitory peptides.⁸⁵ The ACE IC₅₀ value of peptides predicted *in silico* to be released from the major proteins present in 15 food commodities by thermolysin and combinations of thermolysin and pepsin then thermolysin, pepsin and trypsin were determined by QSAR. This analysis allowed the identification of meat (pork, beef and chicken) proteins as rich sources for potent ACE inhibitory peptides (IC₅₀ < 10 μM).

Validation of the *in silico* results with *in vitro* testing of the hydrolysates is needed to confirm their potential bioactivity. The prediction ability of QSAR combined with peptide cutters to release ACE inhibitory peptides from the major egg white proteins was studied by Majumder et al.⁸⁶ They predicted that thermolysin

followed by pepsin digestion of ovotransferrin would allow the release of potent ACE inhibitory peptides (Ile-Arg-Try, Leu-Lys-Pro and Ile-Gln-Try). The three target peptides could not be identified by liquid chromatography (LC)-MS/MS. However, precursors of the target peptides were found within the digest. The inability to release the three target peptides was imputed to the limited access of the enzymes to certain peptide bonds possibly due to the globular structure of ovotransferrin. Based on the knowledge of ACE inhibitory peptide structure, which generally possess Pro residues at their C-terminal region, a post-Pro endoproteinase preparation from *Aspergillus niger* (An-PEP) was employed to hydrolyse a Pro-rich protein substrate, i.e. bovine β -casein.⁸⁷ The hydrolysate generated after 24 h incubation of β -casein with An-PEP at pH 6.0 and an enzyme to substrate (E:S) ratio of 2.5 (w/w) was a particularly potent inhibitor of ACE, having an IC_{50} value of $16.41 \pm 6.06 \mu\text{g mL}^{-1}$. Subsequently, LC-MS characterisation of the peptides within this hydrolysate followed by confirmatory studies with synthetic peptides revealed that the bioactivity was linked to the presence of several C-terminal Pro containing peptides, some of which had ACE IC_{50} values in the μM range.

Other examples illustrating the use of *in silico* predictions to guide hydrolysis generation may be found in the development of DPP-IV inhibitory peptides. Tulipano et al.⁶² predicted, using a peptide cutter approach, that gastrointestinal digestion of β -lactoglobulin (β -Lg) would yield a higher number of DPP-IV inhibitory peptides than that of α -lactalbumin (α -La). Their prediction was subsequently validated by the fact that a β -Lg gastrointestinal digest had a higher DPP-IV inhibitory potency than an α -La digest. Recently, the targeted release of known DPP-IV inhibitory peptides from α -La has been studied.⁸⁸ Approximately 64% of the peptide sequences predicted to be released *in silico* by digestion of α -La with elastase were identified by LC-MS/MS in an α -La elastase digest. The differences between *in silico* predictions and *in vitro* peptide release were possibly due to the presence of disulphide bonds within the α -La sequence. All five DPP-IV inhibitory peptides predicted to be released *in silico* were identified in the α -La elastase digest. Currently, the number of studies translating *in silico* results to BAP release *in vitro* are limited. However, the above studies have demonstrated the benefit of employing *in silico* approaches as a means to select enzyme \times substrate combinations which may result in the release of potent BAPs.

In addition to enzyme specificity and protein sequence knowledge, a wide range of physicochemical parameters (pH, temperature, E:S ratio, protein concentration, etc.) can affect both the enzyme activity and the protein conformation. Therefore, peptide release is highly dependent on the conditions employed during food protein hydrolysis. Understanding of the impact of hydrolysis parameters on hydrolysate bioactivity has been systematically studied using multifactorial experimental design and response surface methodology (RSM) approaches.⁸⁹ In particular when a specific peptide sequence is being targeted for enzymatic release, RSM may be employed to determine the hydrolysis parameters for an

optimum yield of the peptide. For example, RSM has been applied to optimally release an ACE inhibitory peptide, His-Leu-Pro-Leu-Pro (β -casein f(134–138)), from casein using Corolase PP, an intestinal enzyme preparation.⁹⁰ The optimum conditions were found to be 24 h hydrolysis using an E:S ratio of 6% (w/w).

7 Conclusions

QSAR approaches have successfully been used to increase the level of understanding of the structural characteristics of peptides which are linked with specific bioactivities. In agreement with earlier structural studies, different QSAR analyses have generally identified the same structural features in peptides with potent *in vitro* bioactivity. To date, it appears that only a limited number of QSAR studies have allowed the design and discovery of more potent BAPs. Interesting studies have been recently conducted to also take into account the sensory properties of the peptides. Therefore, allowing, for example, the identification of peptides with high *in vitro* bioactivities and low bitterness attributes.

The production of food protein hydrolysates enriched with these potent BAPs may be achieved through the utilisation of design of experiments combined with RSM approaches to predict the optimum hydrolysis parameters which would yield the enzymatic release of BAPs. Combination of QSAR with other *in silico* tools and peptide library approaches will allow the development of systematic methods for the discovery of novel and potent BAPs. For instance, QSAR models have been used to predict the bioactivity of very large sets of peptide sequences with unknown bioactivities.^{26,63} These large sets of peptides may correspond, for example, to all possible amino acid combinations to generate a specific peptide size, peptides which may be present within a specific proteome, novel peptide sequences as well as protein-derived peptides identified within humans.

Overall, many of the QSAR studies appear to have highlighted the importance of hydrophobic amino acids (Pro, Trp, Leu, Ile, Val and Ala) for a wide range of bioactive properties (ACE, DPP-IV and renin inhibition). Interestingly, the presence of some of these residues (i.e., Pro) within peptide sequences have been linked with gastrointestinal or serum stability *in vivo*, which potentially makes them interesting candidates for the development of functional foods targeted at human nutrition.

To date, it appears that certain bioactivities (e.g., renin and DPP-IV inhibition) have not been extensively studied with QSAR approaches. More research in this area will allow the development of more potent BAPs and ultimately the identification of enzymatic hydrolysis strategies to optimally release such peptide sequences. The main challenge in applying QSAR approaches to certain bioactive properties lies in the fact that the target and mode of action of the peptides is not known, making it virtually impossible to develop meaningful models. Therefore, more research is also needed in this area.

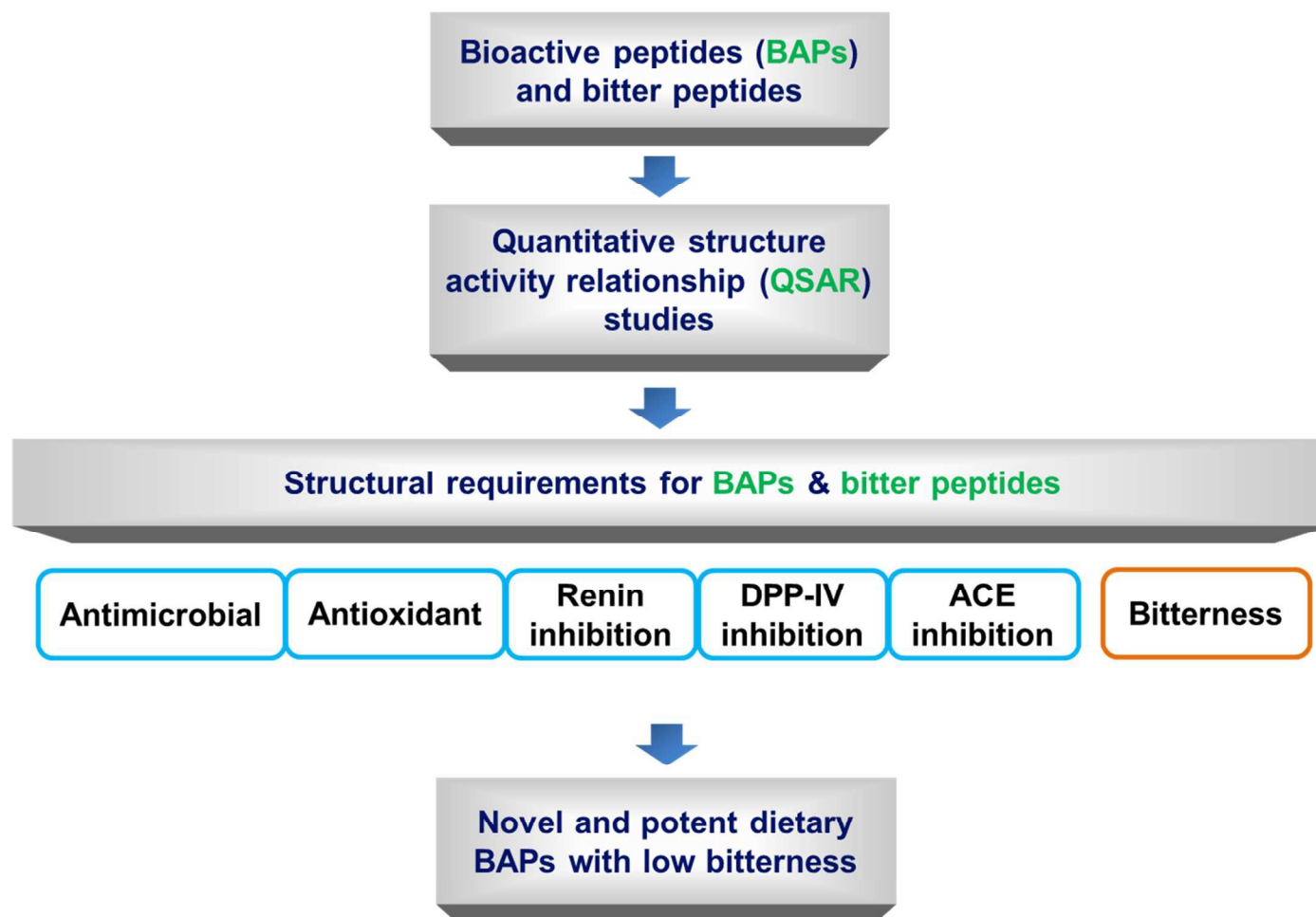
Acknowledgements

The work described herein was supported by Enterprise Ireland under Grant Number TC2013-0001.

Notes and references

- E. C. Li-Chan, *Curr. Opin. Food Sci.*, 2015, **1**, 28-37.
- A. B. Nongonierma and R. J. FitzGerald, *J. Funct. Foods*, 2015, **640**, 640-656.
- A. B. Nongonierma and R. J. FitzGerald, *Caries Res.*, 2012, **46**, 234-267.
- W. M. Miner-Williams, B. R. Stevens and P. J. Moughan, *Nutr. Res. Rev.*, 2014, **27**, 308-329.
- B. Chabance, P. Jollès, C. Izquierdo, E. Mazoyer, C. Francoual, L. Drouet and A.-M. Fiat, *Br. J. Nutr.*, 1995, **73**, 583-590.
- N. Ledoux, S. Mahe, M. Dubarry, M. Bourras, R. Benamouzig and D. Tome, *Food/Nahrung*, 1999, **43**, 196-200.
- H. Meisel, S. Fairweather-Tait, R. J. FitzGerald, R. Hartmann, C. N. Lane, D. McDonagh, B. Teucher and J. M. Wal, *Br. J. Nutr.*, 2003, **89**, 351-358.
- A. Panchaud, M. Affolter and M. Kussmann, *J. Proteomics*, 2012, **75**, 3546-3559.
- L. Sánchez-Rivera, D. Martínez-Maqueda, E. Cruz-Huerta, B. Miralles and I. Recio, *Food Res. Int.*, 2014, **63**, Part B, 170-181.
- M. del Mar Contreras, I. Lopez-Exposito, B. Hernandez-Ledesma, M. Ramos and I. Recio, *J. AOAC Int.*, 2008, **91**, 981-994.
- A. L. Capriotti, C. Cavaliere, S. Piovesana, R. Samperi and A. Laganà, *Anal. Bioanal. Chem.*, 2016, **408**, 2677-2685.
- S. Le Maux, A. B. Nongonierma and R. J. FitzGerald, *Food Chem.*, 2015, **175**, 847-854.
- S. Le Maux, A. B. Nongonierma, B. Murray, P. M. Kelly and R. J. FitzGerald, *Food Res. Int.*, 2015, **77**, 534-539.
- M. B. O'Keeffe and R. J. FitzGerald, *Food Chem.*, 2015, **184**, 140-146.
- A. Iwaniak, P. Minkiewicz, M. Darewicz, M. Protasiewicz and D. Mogut, *J. Funct. Foods*, 2015, **16**, 334-351.
- D. Agyei and M. K. Danquah, *Biotechnol. Adv.*, 2011, **29**, 272-277.
- A. B. Nongonierma, M. B. O'Keeffe and R. J. FitzGerald, in *Advanced Dairy Chemistry*, eds. P. L. H. McSweeney and J. A. O'Mahony, Springer-Verlag, New-York, USA, 4 edn., 2016, pp. 417-482.
- A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin and R. Todeschini, *J. Med. Chem.*, 2014, **57**, 4977-5010.
- A. B. Nongonierma and R. J. FitzGerald, *Trends Food Sci. Technol.*, 2016, **50**, 26-43.
- H. Jenssen, *Expert Opin. Drug Discovery*, 2011, **6**, 171-184.
- R. Jahangiri, S. Soltani and A. Barzegar, *Pharm. Sci.*, 2014, **20**, 122.
- Y. Li and J. Yu, *J. Med. Food*, 2015, **18**, 147-156.
- A. H. Pripp and Y. Ardö, *Food Chem.*, 2007, **102**, 880-888.
- J. Wu and R. E. Aluko, *J. Pept. Sci.*, 2007, **13**, 63-69.
- H.-O. Kim and E. C. Y. Li-Chan, *J. Agric. Food Chem.*, 2006, **54**, 10102-10111.
- M. Tian, B. Fang, L. Jiang, H. Guo, J. Cui and F. Ren, *Dairy Sci. Technol.*, 2015, **95**, 451-463.
- A. H. Pripp, T. Isaksson, L. Stepaniak, T. Sørhaug and Y. Ardö, *Trends Food Sci. Technol.*, 2005, **16**, 484-494.
- Q.-S. Du, R.-B. Huang and K.-C. Chou, *Curr. Protein Pept. Sci.*, 2008, **9**, 248-259.
- P. Minkiewicz, J. Dziuba, A. Iwaniak, M. Dziuba and M. Darewicz, *J. AOAC Int.*, 2008, **91**, 965-980.
- J. Wu, R. E. Aluko and S. Nakai, *J. Agric. Food Chem.*, 2006, **54**, 732-738.
- S. Hellberg, L. Eriksson, J. Jonsson, F. Lindgren, M. Sjöström, B. Skagerberg, S. Wold and P. Andrews, *Int. J. Pept. Protein Res.*, 1991, **37**, 414-424.
- A. H. Pripp, T. Isaksson, L. Stepaniak and T. Sørhaug, *Eur. Food Res. Technol.*, 2004, **219**, 579-583.
- P. Jing, B. Qian, Y. He, X. Zhao, J. Zhang, D. Zhao, Y. Lv and Y. Deng, *Int. Dairy J.*, 2014, **35**, 95-101.
- S. Hellberg, M. Sjöström, B. Skagerberg and S. Wold, *J. Med. Chem.*, 1987, **30**, 1126-1135.
- M. B. Strøm, W. Stensen, J. S. Svendsen and Ø. Rekdal, *J. Pept. Res.*, 2001, **57**, 127-139.
- S. Nakai, J. C. Chan, E. C. Li-Chan, J. Dou and M. Ogawa, *J. Agric. Food Chem.*, 2003, **51**, 1215-1223.
- M. Sandberg, L. Eriksson, J. Jonsson, M. Sjöström and S. Wold, *J. Med. Chem.*, 1998, **41**, 2481-2491.
- J. Jonsson, L. Eriksson, S. Hellberg, M. Sjöström and S. Wold, *Quantitative Structure-Activity Relationships*, 1989, **8**, 204-209.
- Z.-h. Lin, H.-x. Long, Z. Bo, Y.-q. Wang and Y.-z. Wu, *Peptides*, 2008, **29**, 1798-1805.
- F. Tian, P. Zhou and Z. Li, *J. Mol. Struct.*, 2007, **830**, 106-115.
- F. Tian, L. Yang, F. Lv, Q. Yang and P. Zhou, *Amino Acids*, 2009, **36**, 535-554.
- E. R. Collantes and W. J. Dunn, *J. Med. Chem.*, 1995, **38**, 2705-2713.
- Y.-W. Li and B. Li, *J. Theor. Biol.*, 2013, **318**, 29-43.
- R. López-Fandiño, J. Otte and J. van Camp, *Int. Dairy J.*, 2006, **16**, 1277-1293.
- I. Sagardia, R. H. Roa-Ureta and C. Bald, *Food Chem.*, 2013, **136**, 1370-1376.
- J. Wu, R. E. Aluko and S. Nakai, *QSAR Comb. Sci.*, 2006, **25**, 873-880.
- H.-S. Cheung, F.-I. Wang, M. A. Ondetti, E. F. Sabo and D. W. Cushman, *J. Biol. Chem.*, 1980, **255**, 401-407.
- P. Zhou, C. Yang, Y. Ren, C. Wang and F. Tian, *Food Chem.*, 2013, **141**, 2967-2973.
- R. Kumar, K. Chaudhary, J. S. Chauhan, G. Nagpal, R. Kumar, M. Sharma and G. P. Raghava, *Sci. Rep.*, 2015, **5**, 12512.
- X. Wang, J. Wang, Y. Lin, Y. Ding, Y. Wang, X. Cheng and Z. Lin, *J. Mol. Model.*, 2011, **17**, 1599-1606.
- G.-D. Huang, R. Zhang, Y.-P. Luo, W. Zeng and W.-J. Wu, *Chinese J. Struct. Chem.*, 2014, **33**, 1741-1748.
- C. C. Udenigwe, H. Li and R. E. Aluko, *Amino Acids*, 2011, **42**, 1379-1386.
- I. M. Lacroix and E. C. Li-Chan, *Trends Food Sci. Technol.*, 2016, **54**, 1-16.
- H. Fujita, K. Yokoyama and M. Yoshikawa, *J. Food Sci.*, 2000, **65**, 564-569.
- M. Sato, T. Hosokawa, T. Yamaguchi, T. Nakano, K. Muramoto, T. Kahara, K. Funayama, A. Kobayashi and T. Nakano, *J. Agric. Food Chem.*, 2002, **50**, 6245-6252.

56. A. B. Nongonierma and R. J. FitzGerald, *Curr. Opin. Food Sci.*, 2016, **8**, 19-24.
57. C.-L. Jao, C.-C. Hung, Y.-S. Tung, P.-Y. Lin, M.-C. Chen and K.-C. Hsu, *Biomedicine*, 2015, **5**, 9-15.
58. O. Power, A. Nongonierma, P. Jakeman and R. FitzGerald, *Proc. Nutr. Soc.*, 2014, **73**, 34-46.
59. P. Patil, S. Mandal, S. K. Tomar and S. Anand, *Eur J Nutr*, 2015, **54**, 863-880.
60. I. M. E. Lacroix and E. C. Y. Li-Chan, *Mol. Nutr. Food Res.*, 2014, **58**, 61-78.
61. A. B. Nongonierma and R. J. FitzGerald, *Food Chem.*, 2014, **165**, 489-498.
62. G. Tulipano, L. Faggi, A. Nardone, D. Cocchi and A. M. Caroli, *Int. Dairy J.*, 2015, **48**, 66-72.
63. A. B. Nongonierma and R. J. FitzGerald, *Peptides*, 2016, **79**, 1-7.
64. Y.-W. Li, B. Li, J. He and P. Qian, *J. Mol. Struct.*, 2011, **998**, 53-61.
65. Y.-W. Li, B. Li, J. He and P. Qian, *J. Pept. Sci.*, 2011, **17**, 454-462.
66. L. Zheng, Y. Zhao, H. Dong, G. Su and M. Zhao, *J. Funct. Foods*, 2016, **21**, 485-496.
67. T. Lejon, M. B. Strøm and J. S. Svendsen, *J. Pept. Sci.*, 2001, **7**, 74-81.
68. C. D. Fjell, J. A. Hiss, R. E. Hancock and G. Schneider, *Nat. Rev. Drug Discov.*, 2012, **11**, 37-51.
69. N. K. Brogden and K. A. Brogden, *Int. J. Antimicrob. Agents*, 2011, **38**, 217-225.
70. K. Hilpert, C. D. Fjell and A. Cherkasov, *Peptide-Based Drug Design*, 2008, **494**, 127-159.
71. Ø. Rekdal, J. Andersen, L. H. Vorland and J. S. Svendsen, *J. Pept. Sci.*, 1999, **5**, 32-45.
72. T. Lejon, T. Stiberg, M. B. Strøm and J. S. Svendsen, *J. Pept. Sci.*, 2004, **10**, 329-335.
73. S. Sánchez-Gómez, B. Japelj, R. Jerala, I. Moriyón, M. Fernández Alonso, J. Leiva, S. E. Blondelle, J. Andrà, K. Brandenburg, K. Lohner and G. Martínez de Tejada, *Antimicrob. Agents Chemother.*, 2011, **55**, 218-228.
74. G. Maccari, R. Nifosi and M. Di Luca, in *Microbial pathogens and strategies for combating them: Science, technology and education*, ed. A. Méndez-Vilas, Formatex Research Center, Bajadoz, Spain, 2013, pp. 1265-1277.
75. L. Yang, M. Shu, K. Ma, H. Mei, Y. Jiang and Z. Li, *Amino Acids*, 2009, **38**, 805-816.
76. H. Jenssen, C. D. Fjell, A. Cherkasov and R. E. W. Hancock, *J. Pept. Sci.*, 2008, **14**, 110-114.
77. V. Frecer, B. Ho and J. L. Ding, *Antimicrob. Agents Chemother.*, 2004, **48**, 3349-3357.
78. Y. Cheng, F. Luo, Z. Zeng, L. Wen, Z. Xiao, H. Bu, F. Lv, Z. Xu and Q. Lin, *Struct. Chem.*, 2015, **26**, 739-747.
79. R. J. FitzGerald and G. O'Cuinn, *Biotechnol. Adv.*, 2006, **24**, 234-237.
80. B. C. Saha and K. Hayashi, *Biotechnol. Adv.*, 2001, **19**, 355-370.
81. M. Asao, H. Iwamura, M. Akamatsu and T. Fujita, *J. Med. Chem.*, 1987, **30**, 1873-1879.
82. J.-H. Wang, Y.-L. Liu, J.-H. Ning, J. Yu, X.-H. Li and F.-X. Wang, *J. Mol. Struct.*, 2013, **1040**, 164-170.
83. F. Tian, P. Zhou, F. Lv, R. Song and Z. Li, *J. Pept. Sci.*, 2007, **13**, 549-566.
84. J.-B. Tong, J. Chang, S.-L. Liu and M. Bai, *J. Serb. Chem. Soc.*, 2015, **80**, 343-353.
85. Y. Gu, K. Majumder and J. Wu, *Food Res. Int.*, 2011, **44**, 2465-2474.
86. K. Majumder and J. Wu, *Food Res. Int.*, 2010, **43**, 1371-1378.
87. R. Norris, A. Poyarkov, M. B. O'Keefe and R. J. FitzGerald, *Food Chem.*, 2014, **156**, 29-36.
88. A. B. Nongonierma, S. Le Maux, J. Hamayon and R. J. FitzGerald, *Food Funct.*, 2016, **Advance article**, DOI: 10.1039/C1036FO00239K
89. C. van der Ven, H. Gruppen, D. B. A. de Bont and A. G. J. Voragen, *Int. Dairy J.*, 2002, **12**, 813-820.
90. A. Quirós, B. Hernández-Ledesma, M. Ramos, P. J. Martín-Álvarez and I. Recio, *J. Dairy Sci.*, 2012, **95**, 4280-4285.



QSAR studies may help to better understand structural requirements for peptide bioactivity and therefore to develop potent BAPs.