A heuristic, computer-driven and top-down approach to identify novel bioactive peptides: A proof-of-principle on angiotensin I converting enzyme inhibitory peptides

Food Research International

# A heuristic, computer-driven and top-down approach to identify novel bioactive peptides: a proof-of-principle on angiotensin I converting enzyme inhibitory peptides

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | FOODRES-D-21-03008R1 |
| **Article Type:** | Research Paper |
| **Keywords:** | bioactive peptides; top-down approach; angiotensin I converting enzyme; in silico screening; egg proteins |
| **Corresponding Author:** | Luca Dellafiora<br><br>ITALY |
| **First Author:** | Carmen Lammi |
| **Order of Authors:** | Carmen Lammi |
| | Giovanna Boschin |
| | Carlotta Bollati |
| | Anna Arnoldi |
| | Gianni Galaverna |
| | Luca Dellafiora |
| **Abstract:** | Bioactive peptides refer to as short peptides (3–20 amino acid residues in length) endowed of specific biological activities. The identification and characterization of bioactive peptides of food origin are crucial to better understand the physiological consequences of determined foods, as well as to design novel foods, ingredients, supplements, and diets to counteract the onset or worsening of metabolic disorders. For this reason, the identification of bioactive peptides is also relevant from a pharmaceutical standpoint. Nevertheless, the systematic identification of bioactive sequences of food origin are still challenging and relies mainly on the so defined “bottom-up” approaches, which rarely results in the total identification of most active sequence. Conversely, “top-down” approaches pursue the identification of bioactive sequences with certain features and may be more suitable for the precise identification of very potent bioactive peptides. In this context, this work presents a top-down, computer-assisted and hypothesis-driven identification of potent angiotensin I converting enzyme inhibitory tripeptides, as a proof of principle. A virtual library of 6840 tripeptides was screened in silico to identify potential highly potent inhibitory peptides. Then, computational results were confirmed in experimental trials leading to the identification of a very potent novel sequence, LMP, with an $IC_{50}$ of 15.8 and 6.8 μM in cell-free and cell-based assays, respectively. In addition, a bioinformatic approach was used to search potential food sources for LMP and yolk proteins were identified as a possible relevant source worthy of being analyzed further. Overall, the method presented may represent a powerful and versatile framework for a systematic, high-throughput and top-down identification of bioactive peptides. |

- top-down, computer-assisted strategy for ACE inhibitory tripeptides identification
- virtual library of 6840 tripeptides was screened in silico to identify ACE inhibitory peptides
- LMP is a new potent ACE-inhibitory peptide with an $IC_{50}$ of 15.8 μM
- LMP inhibits the cellular ACE activity with $IC_{50}$ 6.8 μM

# A heuristic, computer-driven and top-down approach to identify novel bioactive peptides: a proof-of-principle on angiotensin I converting enzyme inhibitory peptides

Carmen Lammi[1], Giovanna Boschin[1], Carlotta Bollati[1], Anna Arnoldi[1], Gianni Galaverna[2], Luca Dellafiora[2]*

[1] Department of Pharmaceutical Sciences, University of Milan, Via Mangiagalli 25, 23133, Milan, Italy

[2] Department of Food and Drug, University of Parma, Parco Area delle Scienze 27/A, 43124, Parma, Italy

Declarations of interest: none

* Correspondence to:

Luca Dellafiora, Department of Food and Drug, University of Parma, Parco Area delle Scienze 27/A, 43124, Parma, Italy. Phone: +39 0521-906079. Email: luca.dellafiora@unipr.it

**Abstract**

Bioactive peptides ~~refer to~~ are~~as~~ short peptides (3–20 amino acid residues in length) endowed of specific biological activities. The identification and characterization of bioactive peptides of food origin are crucial to better understand the physiological consequences of ~~determined~~ food~~s~~, as well as to design novel foods, ingredients, supplements, and diets to counteract ~~the onset or worsening of~~ mild metabolic disorders. For this reason, the identification of bioactive peptides is also relevant from a pharmaceutical standpoint. Nevertheless, the systematic identification of bioactive sequences of food origin ~~are~~ is still challenging and relies mainly on the so defined "bottom-up" approaches, which rarely results in the total identification of most active sequences. Conversely, "top-down" approaches ~~pursue~~ aim at ~~the~~ identifying ~~ication of~~ bioactive sequences with certain features and may be more suitable for the precise identification of very potent bioactive peptides. In this context, this work presents a top-down, computer-assisted and hypothesis-driven identification of potent angiotensin I converting enzyme inhibitory tripeptides, as a proof of principle. A virtual library of 6840 tripeptides was screened *in silico* to identify potential highly potent inhibitory peptides. Then, computational results were confirmed ~~in~~ experimentally and ~~trials leading to the identification of~~ a very potent novel sequence, LMP was identified. LMP showed ~~, with~~ an $IC_{50}$ of 15.8 and 6.8 µM in cell-free and cell-based assays, respectively. In addition, a bioinformatics approach was used to search potential food sources ~~for~~ of LMP. ~~and y~~Yolk proteins were identified as a possible relevant source to ~~worthy of being~~ analyze~~d~~ in further experiments. Overall, the method presented may represent a powerful and versatile framework for a systematic, high-throughput and top-down identification of bioactive peptides.

*Keyword*: bioactive peptides, top-down approach, angiotensin I converting enzyme, *in silico* screening, egg proteins

## 1. Introduction

Bioactive peptides are amino acid sequences, generally 3–20 ~~amino acid~~ residues in length, endowed of ~~and refer to as~~ some biological activities, which may be ~~ly active peptides that are~~ encrypted into proteins ~~and~~ remaining inactive until they are released upon protein hydrolysis (Mora et al., 2019). In the last decade, bioactive peptides of food origin have gained a growing interest as they potentially ~~determine a measurable~~ modulate~~ion~~ ~~of~~ essential biological functions, including the regulation of blood pressure, cholesterol level, and glucose metabolism, among the others (Moller et al., 2008; Peighambardoust et al., 2021). From a food science standpoint, the identification and characterization of bioactive peptides released and made available upon food digestion is ~~crucial~~ an important piece of information ~~for~~ to understand~~ing~~ the ~~possible health~~ effects of ~~specific~~ food on human health~~s~~, as well as to rationally design specific diets, foods, or food supplements to counteract ~~the onset or the aggravation of~~ mild metabolic disorders (Dellafiora et al., 2015; Li et al., 2018). Nevertheless~~, thanks to the broad array of biological activities~~, the identification of bioactive sequences is also relevant to the ~~from a~~ pharmaceutical field. In this respect, bioactive peptides ~~purpose either~~ can be considered ~~as~~ drug candidates *per se* or leads to derive peptido-mimetic molecules with enhanced pharmacological properties (D'Annessa et al., 2020; Lenci and Trabocchi, 2020; Uhlig et al., 2014). In both cases, the development of reliable and high-throughput methods to identify and characterize bioactive sequences is getting more and more ~~desirable~~advisable.

Nowadays, despite the advances in the analytical methods, the comprehensive identification of bioactive peptides encrypted in food is still challenging and mainly relies on the so defined “bottom-up” approaches (Schrader, 2018). These approaches are often time-consuming and typically ~~based~~ require~~on~~: i) the digestion of food using consolidated *in vitro* digestive models (either static or dynamic) or the extraction of proteins from food ~~sources of interest and their~~followed by their ~~digestion~~hydrolysis using enzymes or other chemical means; ii) the description of peptidomic fractions made ~~either~~ available ~~or accessible~~for the absorption; and iii) the biological testing of a selection of sequences (Capriotti et al., 2016; Lammi et al., 2019). Frequently, the high number of sequences identified in digested samples needs to be rationally hierarchized to select a manageable number of hi~~n~~ts to test experimentally~~in experimental trials~~. In this respect, the use of computational methods may

estimate the bioactivity of identified sequences ~~either qualitatively or quantitatively helping~~ to ~~identify~~ highlight those ~~most worthy of being~~to test ~~analyzed experimentally~~ with high priority (Agyei et al., 2018; Marseglia et al., 2019; Yu et al., 2019). In this context, the choice of digestion protocols~~, along with,~~ the analytical precision to describe ~~the~~ peptidomic profiles~~profile,~~ and the reliability of computational prediction are key factors to ensure the ~~comprehensive~~ successful decryption of potent bioactive sequences included in food ~~proteins~~. However, such analysis~~this type of approach~~ is inherently complex, and it is likely to fail the full mining of bioactive sequences ~~considering~~ due to the possible methodological inaccuracies, instrumental and experimental variability, or readout fluctuations.

~~As opposed to such approaches~~Conversely, "top-down" methods ~~can~~ refer to techniques to create organized bioactive structures, including peptides, either by etching down a bulk material or by ~~manipulating~~ engineering ~~components~~ them into specific locations (Smith et al., 2011). In the context of bioactive peptides of food origin, "top-down" methods are typically independent from food matrices. They target ~~and aim at~~the identification of ~~ying~~ bioactive sequences with specific features~~,~~ that ~~which~~ could be searched afterwards ~~on a second instance~~ among those made available upon digestion of given foods. These methods provide a high-throughput framework of analysis where~~, once define a chemical space to search,~~ all the possible sequences within a given chemical space defined *a priori* may be ~~theoretically~~ analyzed. As a second step of analysis, specific active sequences might be searched within proteins, including but not limited to those of food origin, to identify the ~~diverse types of~~ sources potentially encrypting the bioactive sequences of interest. However, the experimental setup to characterize properly the bioactivity of peptides may be burdensome, limiting in practice the number of sequences undergoing testing. Therefore, similarly to "bottom-up" methods, "top-down" approaches require a careful prioritization of sequences to test and a rational design of experiments to make feasible the analysis in a real-world scale.

In this context, as a proof-of-principle, this work dealt with a top-down, computer-assisted and hypothesis-driven identification of potent angiotensin I converting enzyme (ACE; EC 3.4.15.1) inhibitory tripeptides. ACE is a key player in the blood pressure regulation as it is a carboxy-dipeptidase transforming the inactive peptide angiotensin I into the vasoconstrictor peptide angiotensin II (Vasquez-Villanueva et al., 2019). Therefore, ACE inhibition may result in an appreciable reduction of the blood pressure in hypertensive subjects (Kaur et al.,

2021). In this respect, ~~, and to that end~~ many promising peptides of food origin have been identified over the years (e.g. (Cao et al., 2020; Vasquez-Villanueva et al., 2019)). Specifically, the present work relied on a knowledge-based *in silico* screening of a ~~semi-~~ combinatorial library of tripeptides including all the possible non-repeated permutations using the 20 proteinogenic amino acids (6840 different sequences in total) to identify potent *in vitro* ACE inhibitory sequences never described before. The analysis focused on tripeptides as they can have a relatively high degree of bioavailability and shown an appreciable ab~~d~~sorption by enterocytes (van der Wielen et al., 2017). ~~Once ascertained t~~The screening performance was assessed first. ~~,~~ Then, a selection of top-scored sequences never studied before ~~to the best of our knowledge~~ for ACE inhibition underwent a 3D modeling study to identify ~~entries~~ peptides to test experimentally.~~for the biological activity assessment.~~ LMP was the best sequence identified in the *in silico* studies and it underwent ACE inhibition testing *in vitro* using cell-free and cell-based trials. ~~In particular, the effect of the best scored sequence (LMP) on the *in vitro* and cellular ACE activity was investigated using the porcine kidney recombinant ACE enzyme.~~ The biological characterization of LMP activity confirmed its~~a~~ high inhibitory potential. Finally, the presence of ~~activity of LMP, whose relevance~~ LMP in proteomes of organisms relevant to food production and its possible release upon proteases activity has been investigated through bioinformatic means.~~in food area was estimated through a bioinformatic approach searching its presence in the proteomes of organisms relevant to the food production.~~

## 2. Material and methods

### 2.1 Computational analysis

#### 2.1.1 Buildup of tripeptides library

The virtual library of tripeptides included all the possible ~~non-repeated~~ permutations considering the 20 proteinogenic amino acids, while excluding peptides with repeated amino acids in their sequence (6840 sequences in total ware included). The 3D structures were built in the Trypos .mol2 format using the Biopolymer tool implemented in Sybyl, version 8.1 (https://www.certara.com) taking advantage of an *ad hoc* spl (sybyl programming language) script ~~to ensure the automatic and bulk production of the entire~~

~~library of peptides~~. The C-terminal and N-terminal were set de-protonated and protonated, respectively.

### 2.1.3 Ligand-based virtual screening of tripeptides library

A multiple-model ligand-based virtual screening was applied to mine highly active sequences from the virtual library. ~~To that end, a~~A selection of the most active sequences (i.e. with a reported $IC_{50}$ lower than 1 μM) ~~reported~~ recorded in ~~the~~ two ~~reference~~ benchmark database ~~for~~ of bioactive peptides, i.e. BIOPEP-UWM (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) (Minkiewicz et al., 2019) and AHTPDB (https://webs.iiitd.edu.in/raghava/ahtpdb/index.php) (Kumar et al., 2015), were selected, along with captopril (a strong ACE inhibitory drug), as the reference template molecules to perform 5 independent ligand-based virtual screening using the LiSiCA (Ligand Similarity using Clique Algorithm) algorithm (Legnik et al., 2015). This algorithm provides a fast ligand-based virtual screening platform to ~~search~~ quantify ~~for~~ chemical similarities between a reference template (strong ACE inhibitors in this case) and a database of target compounds. LiSiCA~~, and it~~ expresses similarities using the Tanimoto coefficient, a gold standard to quantify chemical analogies. LiSiCA's default parameters were used, with the exception of considering the 3D structures of ligands and setting ~~with the exception of considering the 3D structure of ligands with~~ the maximum allowed atom spatial distance for 3D product graph ~~set~~ at 2. ~~Specifically~~

In more detail, LRW (reported $IC_{50}$=0.2 μM), IVY (reported $IC_{50}$=0.5 μM), GEP (reported $IC_{50}$=0.3 μM), and IVR (reported $IC_{50}$=0.8 μM) were arbitrarily selected to represent the chemical heterogeneity of the 14 sequences with an $IC_{50}$ lower than 1 μM reported in the database at the time of analysis (i.e. LRW, IKW, GEP, LKP, MKP, IVY, MRW, IRY, IRW, LGP, LIY, MAP, IVR , LRP and VHW; for further details see section 3.1.1 and Table 1S, Supporting material). The structure of captopril was retrieved form PubChem (https://pubchem.ncbi.nlm.nih.gov; compound CID: 44093) and set deprotonated for the analysis. The sequences included in the library then underwent 5 independent virtual screening (i.e. one for each reference template) and the best score out of the five independent screening applied was used to estimate the likeliness of sequences to act as strong ACE inhibitor.

### 2.1.4 Docking studies

Docking studies were performed on a selection of sequences to study the interaction with ACE from a structural standpoint. Specifically, docking simulations were performed using the GOLD software (Genetic Optimization for Ligand Docking, version 2020) as it already showed reliability to compute protein-ligand interactions (e.g. ref. (Maldonado-Rojas and Olivero-Verbel, 2011; Rollinger et al., 2006)). Docking protocol was set in agreement with previous studies that already succeeded to estimate the ACE inhibitory activity or peptides ~~on ACE~~ (Dellafiora et al., 2015; Dellafiora et al., 2020). ~~Briefly,~~ The models for both C- and N-domains of ACE were derived from the Protein Data Bank (http://www.rcsb.org) structures having PDB codes 4APH and 4BZS, respectively (Kramer et al., 2014; Masuyer et al., 2012) as previously described (Dellafiora et al., 2015). The docking software GOLD implements a genetic algorithm that may cause fluctuations of scores. Therefore, each docking simulation was performed in triplicate and scores are expressed as means ± standard deviations.

### 2.1.5 Pharmacophoric analysis

The pockets of ACE were defined using GetCleft (Gaudreault et al., 2015), while the respective pharmacophoric ~~imagines~~ fingerprints were derived using the IsoMIF (Chartier and Najmanovich, 2015). Default parameters were used. As exception, the maximum distance value between the grid and residues atoms was set at 3, and a grid resolution of 1 Å was used.

### 2.1.6 Molecular dynamic simulations

Molecular dynamic simulations were performed to investigate the geometrical stability of peptide-ACE complexes over the time, in agreement with a previous study (Dellafiora et al., 2020). Briefly, simulations were performed using GROMACS (version 5.1.4) (Abraham et al., 2015) with CHARMM27 all-atom force field parameters support (Best et al., 2012). Input structures were solvated with SPCE waters in a cubic periodic boundary condition, and counter ions ($Na^+$ and $Cl^-$) were added to neutralize the system. Prior to simulation, each system was energetically minimized to avoid steric clashes and to correct improper geometries using the steepest descent algorithm with a maximum of 5,000 steps. Afterwards, all the systems underwent isothermal (300 K, coupling time 2psec) and isobaric

(1 bar, coupling time 2 psec) 100 psec simulations before running 50 nsec simulations (300 K with a coupling time of 0.1 psec and 1 bar with a coupling time of 2.0 psec).

### 2.1.7 Statistical analysis of docking results

The docking simulations were run in triplicates and the statistical analysis of docking results was performed using IBM SPSS Statistics for Linux, version 25 (IBM Corp., Armonk, NY). The data was analyzed by one-way ANOVA ($\alpha = 0.05$), followed by post hoc Fisher's LSD test ($\alpha = 0.05$). Of note, the score assignment was found satisfyingly stable for the purpose of this work in all measurements (with a coefficient of variation lower than 3.5 %) and further replicates were considered not needed.

### 2.1.8 Bioinformatic search of LMP in possible food sources

The relevance of certain foods as ~~a~~ possible source of LMP was estimated using a straightforward bioinformatic approach through a peptide search into the proteome of chickens (*Gallus gallus*), garden peas (*Pisum sativum*) and Spirulina platensis (*Arthrospira platensis*) stored in the UniProt Proteomes repository (https://www.uniprot.org/proteomes; taxa ID 9031, 3888 and 118562, respectively). Only the sequences annotated as "reviewed" (i.e. records with information extracted from literature and curator-evaluated computational analysis) were considered for the analysis (i.e. 2,297; 398 and 2 sequences for chickens, garden peas and Spirulina platensis, respectively).

The possible release of peptides from proteins has been calculated using the PeptideCutter tool (Gasteiger et al., 2005), which has been developed by the Swiss-Prot group and supported by the SIB Swiss Institute of Bioinformatics (https://www.sib.swiss/). The protein sequences in the FASTA format were used as input selecting the all set of proteases available.

## 2.2 Experimental analysis

### 2.2.1 Chemicals and sampling

All chemicals (reagents and solvents) were from Sigma-Aldrich (St. Louis, MO, USA). Caco-2 cells were obtained from INSERM (Paris, France; Dulbecco's modified Eagle's medium (DMEM), stable L-glutamine, foetal bovine serum (FBS), phosphate buffered saline (PBS),

penicillin/streptomycin and 96-well plates were purchased from Euroclone (Milan, Italy). The ACE1 Activity Assay Kit was from Biovision (Milpitias, CA, USA). The peptides LKP and LMP were synthetized by GenScript (Piscataway, NJ, USA) at >95% purity.

### 2.2.2 *In vitro* evaluation of ACE inhibitory activity

Peptides were tested as already described (Boschin et al., 2014a, b) evaluating hippuric acid (HA) formation from hippuryl-histidyl-leucine (HHL), as mimic substrate for angiotensin I.

### 2.2.3 Cell line culture

Caco-2 cells were routinely sub-cultured at 50% density and maintained at 37°C in a 90% air/10% $CO_2$ atmosphere in DMEM containing 25 mM of glucose, 3.7 g/L of $NaHCO_3$, 4 mM of stable L-glutamine, 1% non-essential amino acids, 100 U/L of penicillin and 100 µg/L of streptomycin (complete medium), supplemented with 10% heat-inactivated foetal bovine serum (FBS; Hyclone Laboratories, Logan, UT, USA).

### 2.2.4 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) Assay

A total of 3 x $10^4$ Caco-2 cells/well were seeded in 96-well plates and treated with 0.1 – 100 µM of LKP and LMP, or vehicle ($H_2O$) in complete growth media for 48 h at 37 °C under 5% $CO_2$ atmosphere. Subsequently, the treatment solvent was aspirated and 100 µL/well of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) filtered solution added. After 2 h of incubation at 37 °C under 5% $CO_2$ atmosphere, 0.5 mg/mL solution was aspirated and 100 µL/well of the lysis buffer (8 mM HCl + 0.5% NP-40 in DMSO) added. After 10 min of slow shaking, the absorbance at 575 nm was read on the Synergy H1 fluorescence plate reader (Biotek, Bad Friedrichshall, Germany).

### 2.3.5 Cell-Based ACE Activity Assay

For the experiments, cells were seeded on 96-well plates at a density of 5 × $10^4$ cells/well for 24 h. The following day, cells were treated with 100 µL of LKP and LMP peptides (from 1.0 to 50.0 µM) or vehicle in growth medium for 24 h at 37 °C. The next day, cells were collected and lysed ~~scraped~~ in 30 µL of ice-cold ACE1 lysis buffer and transferred in an ice-cold microcentrifuge tube. After centrifugation at 13,300 *g* for 15 min at 4 °C, the supernatant

was recovered and transferred into a new ice-cold tube. Total proteins were quantified by Bradford method, and 1.5 µg of total proteins (the equivalent of 1.5 µL) were added to 18 µL of ACE1 lysis buffer in each well in a black 96-well plate with clear bottoms. For the background control, 20 µL of ACE1 lysis buffer were added to 20 µL of ACE1 assay buffer. Then, 20 µL of 4% of ACE1 substrate (in assay buffer) was added in each well except the background one and the fluorescence (Ex/Em 330/430 nm) was measured in a kinetic mode for 10 min at 37°C.

#### 2.3.6 Statistical analysis of biological experiments

Statistical analyses of *in vitro* and cellular ACE activity data set were carried out by Student's t-test using Graph-pad Prism 9 (SanDiego, CA, USA). Values were expressed as means ± sd; *p*-values < 0.05 were significant.

## 3. Results

### 3.1.1 Benchmarking and assessment of screening performances

The computational analysis was benchmarked against the data reported so far in specific reference databases to develop the screening strategy and to check computational performances. In particular, the two reference databases used in this study were BIOPEP-UWM (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) (Minkiewicz et al., 2019) and AHTPDB (https://webs.iiitd.edu.in/raghava/ahtpdb/index.php) (Kumar et al., 2015). The extraction of ~~all~~ the all set of sequences matching the chemical criteria used to build the library under analysis (i.e. tripeptides with not repeated amino acids) was performed first. At the time of analysis (last database access 20$^{th}$ of May 2021), BIOPEP-UWM included 1051 sequences annotated as ACE inhibitors. Among them, 241 tripeptides were listed in total, with 171 sequences ~~having~~ showing non-repeated amino acid residues and having~~explicit molar~~ $IC_{50}$ values reported. Concerning AHTPDB, 1463 ACE inhibitory sequences ~~annotated as ACE inhibitors~~ were found. Among them, 398 tripeptides were listed, with 156 entries ~~having a~~showing non-repeated sequences and having ~~explicit molar~~ $IC_{50}$ values reported. After removing redundancies, 213 tripeptides with non-repeated sequence with $IC_{50}$ values ranging from 0.02 to 2700 µM were listed in total (Supporting material, Table 1S).

The final list of peptides was then analyzed to extract a selection of theose most active sequences (with $IC_{50}$ values arbitrarily set as ≤ 1 μM) to be used as template molecules in the ligand-based virtual screening ~~of the library under analysis~~. Specifically, the ligand-based screening used in the present work ~~refers to as a~~consisted in the computational hierarchization ~~of library's~~of ~~-~~tripeptides included in the virtual library according to their chemical similarity to the strong tripeptides ~~already characterized and~~ recorded in the ~~reference~~ benchmark databases. ~~In this respect~~In principle, the higher the chemical similarity ~~of uncharacterized sequences~~ to verythe ~~most~~ potent ~~ones~~ peptides already characterized, the higher the chance to identify uncharacterized peptide with a high inhibitory activity. However, in principle this ~~more likely the possibility to detect novel potent inhibitory sequences is~~may reduce the chemical space under analysis likely leading to identify sequences with a marked chemical similar to those previously described. Considering the range of activity ($IC_{50}$) reported in the reference databases (i.e. 0.02 – 2700.00 μM), the threshold for the most potent sequences was arbitrarily set at 1 μM. The list of 213 peptides extracted from BIOPEP-UWM and AHTPDB included 15 sequences with $IC_{50}$ ≤ 1 μM (Supporting material, Table S1) and, among them, 4 sequences were selected after a visual inspection as reference to describe the chemical heterogeneity observed: LRW (reported $IC_{50}$=0.2 μM), IVY (reported $IC_{50}$=0.5 μM), GEP (reported $IC_{50}$=0.3 μM) and IVR (reported $IC_{50}$=0.8 μM). The strong ACE inhibitory drug captopril ($IC_{50}$=0.007 μM) (Li et al., 2019) was included in the list of reference templates too.

Of note, the library under analysis was likely to include, in addition to LRW, IVY, GEP and IVR, other sequences previously characterized and ~~a certain number of sequences previously characterized and listed already~~listed in BIOPEP-UWM or AHTPDB. Therefore, ~~This piece of information~~this information ~~-~~was~~as~~ used to assess the computational performance in terms of extraction efficiency of~~to extract~~ highly active sequences (with $IC_{50}$ values arbitrarily set ~~as~~ < 3 μM). Specifically, 33 sequences with an $IC_{50}$ < 3 μM ~~(excluding IVY, IVR, GEP and LRW)~~ included in the virtual library were already annotated in BIOPEP-UWM or AHTPDB (excluding IVY, IVR, GEP and LRW). T~~o that end, t~~he 5 reference templates mentioned above (i.e. LRW, IVY, GEP, IVR and captopril) were used as a template to run 5 independent screenings of the whole virtual library. ~~Specifically, 33 sequences with an $IC_{50}$ < 3 μM (excluding IVY, IVR, GEP and LRW) included in the library were already annotated in BIOPEP-UWM or AHTPDB.~~The ranking analysis of those 33 sequences in respect to the

whole library and according to the 5 hierarchies obtained in the five independent screenings gave the enrichment capacity of the method used. The strongest sequences (i.e. the 9 sequences with an $IC_{50}$ < 1 μM) were in the output's top-ranked 7% in at least one of the 5 independent screenings run, while the 94% of those with $IC_{50}$ < 3 μM (i.e. 31 out of 33 sequences) were in the top-ranked 10% in at least one of the 5 independent screenings. This outcome confirmed the efficacy of the multiple-model ~~scoring~~ screening method ~~applied~~ to identify ~~enrich the top-ranked hierarchy with~~ highly active tripeptides (Figure 1).

### 3.1.2 Heuristic definition of the chemical space to analyze via molecular modeling

The next step of computational analysis was the 3D molecular modeling of the most promising sequences highlighted by virtual screenings to better estimate their capability to interact and inhibit ACE.

~~Once ascertained the capability of the procedure to enrich the top-ranked positions of virtual screening outcomes with potent ACE inhibitory peptides, a~~A knowledge-based approach to ~~resize~~ reduce the number of sequences to analyze *via* ~~the chemical space to search with~~ molecular modeling was applied~~. This step was necessary~~ due to the higher computational ~~cost~~ demand of 3D modeling compared to the ligand-based screening that prevented ~~that made unfeasible~~ its application on the whole virtual ~~entire~~ library. The analysis of the most active sequences reported in the reference database provided the rational basis to identify a small set of ~~resize the number of~~ sequences to ~~a set suitable for being analyzed via~~analyze using molecular modeling. As shown in Table 1, the 13 most active tripeptides ($IC_{50}$ ≤ 1 μM) had in the position 1 and 3 hydrophobic residues, with the exception of only one sequence with an arginine in position 3 (IVR). Specifically, I and L were prevalent in the first position while P was prevalent in the third position. Conversely, it was not possible to define a specific ~~characteristic~~ feature for the second position as both hydrophobic and hydrophilic residues were observed~~found there~~. Therefore, it was inferred that branched hydrophobic amino acids in the first position with a proline in the third position are ~~suitable~~ important ~~characteristic~~ chemical features ~~for~~ of strong inhibitory peptides, in agreement with previous studies (Wu et al., 2006).

On this basis, all the sequences in the library under analysis with a leucine and proline in the first and third position, respectively, were considered for the 3D modeling (19 tripeptides in total; Table 2). Among those, nine sequences were found in the list of tripeptides already

characterized and listed in BIOPEP-UWM or AHTPDB and therefore they were not ~~ex~~included ~~from~~ in the analysis. The remaining nine sequences, never tested before to the best of our knowledge for ACE inhibitory activity, underwent the virtual screening procedure to identifying the most promising sequences to analyze with 3D molecular modeling. As shown in Table 3, the three best ranked sequences ~~(i.e. LIP, LTP and LMP)~~ according to the average rank ~~out~~ of the 5 independent virtual screenings (i.e. LIP, LTP and LMP) underwent docking analysis. In agreement to previous studies, docking simulations may ~~to better~~ evaluate~~e~~ the~~ir~~ capability of tripetides to interact with both the C- and N-terminal domain of ACE~~, in agreement with previous studies~~ (Dellafiora et al., 2015; Dellafiora et al., 2020). The structurally related tripetide LKP, which was already characterized and annotated in the BIOPEP-UWM among the strongest sequences ($IC_{50}$ of 0.3 μM), was used as ~~also included in the analysis as a~~ reference peptide. ~~in the light of its structural analogy.~~

Keeping in mind that ~~the~~ docking score~~s~~ may ~~be proportional~~correlate with ~~to~~ the ~~theoretical~~ strength~~en~~ of ligand-pocket interaction, LMP was deemed the most promising as it recorded scores significantly higher than LIP and LTP ($p < 0.05$) in both ACE's domain. Comparing LMP and LKP, the respective scores were not significantly different (p=0.57 and p=0.09 in N- and C-domain, respectively) and for them a similar interaction could be hypothesized. The calculated binding poses were then analyzed in the light of the pharmacophoric fingerprint of ACE's pockets. Although the two pockets differ for some residues, they have a very similar pharmacophoric fingerprint, as previously described (Dellafiora et al., 2015). The small differences were found not relevant for the sake of this study and therefore only results concerning the C-terminal domain are presented for simplicity. As shown in Figure 2A, all the four tripeptides analyzed had the amino and carboxy termini engaged in polar contacts with the ACE binding site. ~~Of note, t~~The diverse capacity of the side chain in position #2 to satisfy the pharmacophoric requirements of the pocket could explain the diverse scores collected for LKP, LMP, LIP and LTP. Specifically, the space of the pocket receiving the side chain of residues in position #2 is mainly hydrophobic with a polar upper ~~terminus~~ part able to receive polar groups like hydrogen bond donors or positively charged bases. In this respect, the methionine of LMP was found better embedded into the hydrophobic space of the pocket compared to the isoleucine and threonine of LIP and LTP, respectively. This could determine a higher hydrophobic contribution to the binding event. LTP recorded the theoretically worst interaction as it

arranged into such a hydrophobic region the hydrophilic moiety of its threonine residue (i.e., the hydroxyl group of the 3-hydroxybutanoic side chain). Concerning LKP, the lysine side chain was found able to satisfy the hydrophobic region, but also to use the amino group to form additional polar interactions with the upper hydrophilic portion mentioned above. This additional contact, which is missing in the other 3 tripeptide analyzed, could suggest a potentially stronger interaction compared to LIP, LTP or LMP that mightpossibly resulting in a stronger inhibitory activity.

On this basis, LMP was deemed the most promising among the set under investigation and it was further analyzed, in comparison with LKP, in molecular dynamic studies to check the geometrical stability of ACE-peptide complexes over the time, in agreement with a previous study work (Dellafiora et al., 2020). As shown in Figure 2B, according to the root-mean squared deviation, the geometrical stability of ACE-LKP complex was found constant over the time and similar to that of ACE-LKP complex. Therefore, molecular dynamics confirmed the capability of LMP to stably interact with ACE possibly determining a certain degree of inhibition., and likely inhibit ACE.

### 3.2. Experimental assessment of in vitro and cellular ACE inhibitory potential of LMP

The *in vitro* biological assessmentACE inhibitory potential of LMP on the *in vitro* ACE activity wasere evaluated using the porcine kidney recombinant form of the enzyme, using in comparison with LKP, as reference peptide. Both peptides efficiently inhibited the ACE activity by 97.4± 0.15 % and 89.8 ± 0.12 %, respectively, at 250 µM. In addition, LKP displays an $IC_{50}$ value equal to 9.23±0.6 µM, whereas LMP equal to 15.8±0.2 µM (p<0.001, (Table 4). Before cellular evaluation, however, it was necessary to perform MTT experiments to exclude any potential cytotoxic effect in the test system used (i.e., human intestinal Caco-2 cells). Results suggested that in the range of concentration 0.1– 100 µM, no cytotoxic effects were observed for both peptides (Figure S1; Supporting material). Therefore, to evaluate their effects on the ACE activity expressed at a cellular level, human intestinal Caco-2 cells were treated with LKP and LPM in the range of concentration 0.1– 100 µM for 24 h. After cell lysis, the ACE activity was measured in the presence of a fluorescent substrate. In this assay, LKP and LMP inhibited the cellular ACE activity with a dose-response trend and $IC_{50}$ values equals to 3.8±0.23 and 6.8±0.34 µM, respectively, without significant difference (Figure 3).

### 3.3 Bioinformatic search of possible food-related sources of LMP

As a proof of principle, ~~Once ascertained *in vitro* the activity of LMP as potent ACE inhibitory tripeptide, t~~the existence of possible food-related sources of LMP was ~~search~~ investigated ~~through~~ using a bioinformatic~~s~~ approach searching the presence of LMP sequence~~. In particular, LMP was searched~~ into the proteome of chickens, peas and *Spirulina platensis*~~, as a proof of principle~~.

As reported in Table 2S and 3S, 65, 12 and 1 LMP-containing proteins were identified in chicken, peas and *S. platensis*, respectively. The relevance of LMP-containing proteins of peas and *S. platensis* as a possible source of LMP could not be easily inferred due to the shortage of data on their actual abundance in food. Conversely, among the 62 proteins of chicken, 2 were deemed relevant as a possible source of LMP from a real world perspective (i.e. vitellogenin-1 and vitellogenin-2; UniProt ID P02845 and P87498) based on ~~due to~~ their abundance in ~~importance in~~ eggs ~~being precursors of the major yolk proteins lipovitellins and phosvitin~~ (Wang et al., 2020). In addition, based on the computational prediction of peptidase-mediated hydrolysis of those proteins, LMP was found possibly released from vitellogenin-2 upon cleavage by chymotrypsin and thermolysin. This evidence further supported the possible importance of this protein as a source of LMP.

## 4. Discussion

Bioactive peptides have gained a growing interest in the past years thanks to their potential to counteract mild metabolic disorders. From a food science standpoint, the identification of bioactive peptides from food sources still poses a major challenge thought their systematic identification is a key piece of information to rationalize ~~the~~ some biological effects ~~outcome~~ of ~~certain~~ food~~s or to design specific diet~~. Nowadays, “bottom-up” approaches are primarily used ~~in the research of~~to bioactive sequences of food origin. Although they may ensure the identification of potent bioactive sequences made available upon digestion, they can’~~no~~t guarantee ~~neither~~ the identification of most active sequences ~~nor the totality of bioactive peptides~~ encrypted in a given ~~source~~matrix. Conversely, ~~T~~”top-down” approaches ~~refers instead to as searching methods that~~ are typically matrix-independent and they may

provide~~. These methods may complement the canonical bottom-up analysis providing~~ a high-throughput platform to mine very potent bioactive sequences integrating the canonical bottom-up analysis~~, regardless the source they may be included in~~. In this context, as a proof of principle, the present work, ~~as a proof of principle,~~ dealt with a matrix-independent computer-assisted study where a ~~semi~~ combinatorial~~y~~ virtual library of 6840 tripeptides was screened to identify novel very potent *in vitro* ACE inhibitory sequences. The focus on tripeptides was based on the evidence that short sequences may have a higher epithelial permeability and bioavailability compared to longer sequences (van der Wielen et al., 2017). Computer-supported methods are nowadays well consolidated means to study the bioactivity of peptides, as previously described also for ACE inhibitory peptides. However, wide libraries of peptides are rarely investigated although their screening already identified ~~proved to be an effective mean to identify~~ novel bioactive peptides (e.g. (Chen et al., 2021)). Of note, ACE inhibitory peptides do not cause the possible adverse effects of ACE inhibitory drugs, although they typically a have less potent activity. As an example, captopril shows $IC_{50}$ values in the nM range (Li et al., 2019)) while inhibitory peptides are typically active in the µM range. Keeping in mind that the most potent ACE inhibitory peptides identified so far showed activity in the low µM range, as per LKP ($IC_{50}$ 0.4 µM according to BIOPEP-UWM), VPP and IPP ($IC_{50}$ 9 and 5 µM, respectively) (Li et al., 2019), the workflow ~~succe~~successfully ~~eded to~~ identifie~~d~~ a novel and very potent ACE inhibitory sequence. Indeed, LMP was described for the first time to the best of our knowledge as a potent ACE inhibitory peptide with an observed $IC_{50}$ in cell-free and cell-based assays of 15.8 and 6.8 µM, respectively. LKP was included in this study as a reference compound, since its hypotensive effect has been previously characterized (Majumder et al., 2015; Majumder and Wu, 2010). Notably, LKP~~, which~~ can ~~comes~~ be released ~~from~~ by the enzymatic digestion of egg white protein ovotransferrin and it proved to reduce ~~, is a tripeptide that through the inhibition of ACE activity led to a significant reduction of~~ the blood pressure (~30 mmHg) ~~after orally administration~~ in spontaneously hypertensive rats (SHRs) after oral administration via ACE inhibition. Interestingly, the change in blood pressure was accompanied by the preservation of nitric oxide (NO) dependent vasorelaxation and lowering of plasma angiotensin (Ang) II levels (Majumder et al., 2015). On the basis of the strong structural analogies between LKP and LMP, a certain ~~di~~ degree of activity *in vivo* can

be reasonably expected for LMP ~~as well that is worth of further investigations with priority~~ deserving further dedicated investigations.

~~Of note, i~~In the present study, LKP gave $IC_{50}$ values slightly higher than that~~at~~ reported in BIOPEP-UWM (i.e. 9.2 and 3.7 µM ~~obtained at~~ in cell-free and cell-based trials~~*in vitro* and cellular level~~, respectively, against 0.4 µM as per BIOPEP-UWM). This discrepancy could be due to the diverse experimental ~~asset~~ setup used, in agreement with a previous study~~, which~~ suggesting that *in vitro* assays carried out with different methods may give slightly different ~~numerical~~ results~~, even though they are in the same range and therefore, comparable~~ (Hernandez-Ledesma et al., 2003). In facts, the $IC_{50}$ ~~reported~~ recorded ~~by~~ in BIOPEP-UWM was ~~calculated~~ obtained ~~using the~~ *in vitro* using ~~assay in which~~ the recombinant enzyme from rabbit lung ~~was employed~~ (Majumder and Wu, 2010), ~~on the contrary,~~while ~~in the present study,~~ the recombinant porcine kidney enzyme was used in the present study. This difference could partially explain the small difference observed. ~~.~~ ~~Notwithstanding the different experimental protocol~~Nevertheless, LKP was confirmed as a very potent ACE inhibitory peptide with an observed $IC_{50}$ in the low µM range. In addition, its ACE-inhibitory mechanism of action was also investigated in a more realistic way using an assay based on human intestinal Caco-2 cells, which was recently optimized and successfully applied to study other food derived peptides (Lammi et al., 2020; Li et al., 2021).

The significantly higher potency of LKP compared to LMP observed in cell-free assay ~~and~~ ($p < 0.001$), ~~which was not however significant~~but not in cell-based ~~trials~~ experiments, ~~*in vitro* and in ce~~, ~~lls~~($p < \ldots$ and $\ldots$, respectively) could be partially explained in the light of the pharmacophoric analysis of docking poses. In this respect, the importance of both hydrophobic and polar interactions at the second amino acid position was ~~defined on the basis of distribution of hydrophobic and hydrophilic characteristics of the space able to receive the amino acid side chain~~described. Specifically, the lysine of LKP was found able to satisfy both ~~those~~ these key interactions. ~~using the polar side chain terminal to engage with polar contact the ACE binding site.~~ Conversely, LMP's methionine~~, which has no polar group in the side chain,~~ could interact only via hydrophobic-hydrophobic interactions. This line of interpretation could also provide a mechanistic explanation to the high variability of residues observed in position 2. Indeed, both hydrophobic and hydrophilic side chains can contribute to the binding event whenever they can provide a proper geometry of interaction~~–~~ avoiding hydrophobic-hydrophilic interference~~matching the pharmacophoric~~

~~properties of that pocket region and avoiding polar-hydrophobic interferences~~. In this respect, LTP gave the worst computational records as it arranged the hydrophilic ~~threonine~~ hydroxyl group of threonine into the hydrophobic region of the pocket. In line with this interpretation, all the tripeptides listed in BIOPEP-UWM and AHTPDB with a threonine in position 2 had ~~relatively worse~~ $IC_{50}$ values worse ~~compared to~~than LKP.

On this basis, this manuscript provided a compelling line of evidence ~~pointing~~ suggesting the relevance of ~~to~~ the ~~effectiveness of the~~ workflow presented ~~for the sake of~~to identify~~ing~~ potent bioactive peptides. Moreover, the 3D modeling provided a mechanistic interpretation for the data collected to further understand the structural requirements of ACE inhibitory peptides.

Concerning the identification of possible LMP sources, chickens and related products were described as potential candidates, although the relevance of peas and *S. platensis* could not be excluded completely. Indeed,~~,~~ LMP was found in chicken vitellogenin-1 and vitellogenin-2~~, among the others~~. These are precursors of the main yolk proteins of chicken eggs lipovitellins and phosvitin. Therefore, chicken egg yolk has been identified as a possible source of LMP ~~worthy of being~~deserving further dedicated studied ~~further~~. ~~In this respect~~Specifically, based on computational predictions, LMP was found potentially released from vitollogenin-2 upon cleavage by chymotrypsin and thermolysin. This evidence suggested that the release of LMP from vitellogenins and derived proteins might happen either during the gastrointestinal digestion (e.g. due to chymotrypsin action) or upon a certain food processing as thermolysin is a bacterial peptidase with a potential multi-purpose use in food technologies (Ke et al., 2013; Tavano et al., 2018). Notably, the release of ACE-inhibitory peptides upon cleavage by thermolysine of food matrices has been previously reported (Tavano et al., 2018). In this respect, it must be noted that the hydrolysis yield of yolk proteins, including those containing LMP, could be purposely set acting on processing and treatment conditions to maximize the release of LMP.

Moreover, yolk oligopeptides have been described to have an anti-hypertensive action (Grootaert et al., 2019). It is interesting to underline the potential synergistic activity of both LKP and LMP after egg protein consumption for lowering blood pressure. ~~Specifically, the release of LMP among the fraction made disposable to living organisms might have a role to determine the anti-hypertensive effects and it deserves further specific studies to better understand the biological impact of egg proteins.~~ In addition, although the epithelia

abdsorption and the actual internal disposability need to be assessed in further studies, the data collected in experimental trials support the possible relevance of LMP to the *in vivo* situation as it showed a potent inhibitory activity ($IC_{50}$ of 6.7 µM) against the human enzyme expressed by Caco-2 cells.

## 5. Conclusion

~~Generally speaking, t~~This study described an effective knowledge-based method that took advantage of the bulk of bioactivity data reported in reference databases to support the design of analytical strategies to mine highly active sequences against a specific biological endpoint. ~~Specifically,~~ LMP was mined from a virtual library of 6840 sequences and described for the first time as a potent ~~and novel~~ ACE inhibitor with $IC_{50}$ values in the low µM range. The procedure provided a three-tier approach where experimental confirmations followed fast ligand-based screenings and slower molecular modeling studies~~, which~~ providinged mechanistic information to better understand the *in vitro* data from a molecular point of view. The inhibition of ACE was used as a proof of principle, but the reference database used here list some additional biological activity including antibacterial, hypocholesterolemic and antidiabetic activity. Therefore, the workflow presented could be easily moved to other types of activity and extended to longer peptide sequences. In addition, the search of possible protein sources, which have been showcased here for few proteomes as a proof of principle, can be implemented in systematic, recursive and high throughput searching methods to mine the sequences of interest from difference data sources. ~~–~~Food proteins, proteins from food waste or alternative protein sources (i.e. proteins not related to food production) are some possible examples. Therefore, the method presented may represent a powerful and versatile framework for a systematic, high-throughput and top-down identification of bioactive peptides. In this respect, search strategies are advised to be performed on well-characterized proteomes and for species with a deep understanding of protein expression level in the various tissues and organs to maximize the search success.

## Funding

This work was supported by the project “Mime4Health” founded by Emilia-Romagna Region, Italy (986/2018 PORFESR_2014_2020).

**Acknowledgement**

The authors would like to acknowledge Dario Manfredi for programming support. This research also benefits from the HPC (high performance computing) facility of the University of Parma, Italy.

**References**

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers SoftwareX 1-2, 19-25.

Agyei, D., Tsopmo, A., Udenigwe, C.C., 2018. Bioinformatics and peptidomics approaches to the discovery and analysis of food-derived bioactive peptides. Analytical and Bioanalytical Chemistry 410, 3463-3472.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M., Mackerell, A.D.J., 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. J. Chem. Theory Comput. 8, 3257-3273.

Boschin, G., Scigliuolo, G.M., Resta, D., Arnoldi, A., 2014a. ACE-inhibitory activity of enzymatic protein hydrolysates from lupin and other legumes. Food Chemistry 145, 34-40.

Boschin, G., Scigliuolo, G.M., Resta, D., Arnoldi, A., 2014b. Optimization of the Enzymatic Hydrolysis of Lupin (Lupinus) Proteins for Producing ACE-Inhibitory Peptides. Journal of Agricultural and Food Chemistry 62, 1846-1851.

Cao, S.M., Wang, Y., Hao, Y.J., Zhang, W.G., Zhou, G.H., 2020. Antihypertensive Effects in Vitro and in Vivo of Novel Angiotensin-Converting Enzyme Inhibitory Peptides from Bovine Bone Gelatin Hydrolysate. J. Agr. Food Chem. 68, 759-768.

Capriotti, A.L., Cavaliere, C., Piovesana, S., Samperi, R., Lagana, A., 2016. Recent trends in the analysis of bioactive peptides in milk and dairy products. Analytical and Bioanalytical Chemistry 408, 2677-2685.

Chartier, M., Najmanovich, R., 2015. Detection of Binding Site Molecular Interaction Field Similarities. J. Chem. Inf. Model. 55, 1600-1615.

Chen, R.Y., Miao, Y.L., Hao, X., Gao, B., Ma, M.Z., Zhang, J.Z.H., Wang, R., Li, S., He, X., Zhang, L.J., 2021. Investigation on the characteristics and mechanisms of ACE inhibitory peptides by a thorough analysis of all 8000 tripeptides via binding free energy calculation. Food Science & Nutrition 9, 2943-2953.

D'Annessa, I., Di Leva, F.S., La Teana, A., Novellino, E., Limongelli, V., Di Marino, D., 2020. Bioinformatics and Biosimulations as Toolbox for Peptides and Peptidomimetics Design: Where Are We? Frontiers in Molecular Biosciences 7.

Dellafiora, L., Paoella, S., Dall'Asta, C., Dossena, A., Cozzini, P., Galaverna, G., 2015. Hybrid in Silico/in Vitro Approach for the Identification of Angiotensin I Converting Enzyme Inhibitory Peptides from Parma Dry-Cured Ham. J. Agr. Food Chem. 63, 6366-6375.

Dellafiora, L., Pugliese, R., Bollati, C., Gelain, F., Galaverna, G., Arnoldi, A., Lammi, C., 2020. "Bottom-Up" Strategy for the Identification of Novel Soybean Peptides with Angiotensin-Converting Enzyme Inhibitory Activity. J. Agr. Food Chem. 68, 2082-2090.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: J.M. Walker (Ed.), The Proteomics Protocols Handbook. Humana Press.

Gaudreault, F., Morency, L.P., Najmanovich, R.J., 2015. NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID. Bioinformatics 31, 3856-3858.

Grootaert, C., Voorspoels, S., Jacobs, G., Matthijs, B., Possemiers, S., Van der Saag, H., Van Camp, J., Lucey, A., 2019. Clinical aspects of egg bioactive peptide research: a review. International Journal of Food Science and Technology 54, 1967-1975.

Hernandez-Ledesma, B., Martin-Alvarez, P.J., Pueyo, E., 2003. Assessment of the spectrophotometric method for determination of angiotensin-converting-enzyme activity: Influence of the inhibition type. J. Agr. Food Chem. 51, 4175-4179.

Kaur, A., Kehinde, B.A., Sharma, P., Sharma, D., Kaur, S., 2021. Recently isolated food-derived antihypertensive hydrolysates and peptides: A review. Food Chem. 346.

Ke, Q.D., Chen, A., Minoda, M., Yoshida, H., 2013. Safety evaluation of a thermolysin enzyme produced from Geobacillus stearothermophilus. Food Chem. Toxicol. 59, 541-548.

Kramer, G.J., Mohd, A., Schwager, S.L.U., Masuyer, G., Acharya, K.R., Sturrock, E.D., Bachmann, B.O., 2014. Interkingdom Pharmacology of Angiotensin-I Converting Enzyme Inhibitor Phosphonates Produced by Actinomycetes. Acs Medicinal Chemistry Letters 5, 346-351.

Kumar, R., Chaudhary, K., Sharma, M., Nagpal, G., Chauhan, J.S., Singh, S., Gautam, A., Raghava, G.P.S., 2015. AHTPDB: a comprehensive platform for analysis and presentation of antihypertensive peptides. Nucleic Acid Res. 43, D956-D962.

Lammi, C., Aiello, G., Boschin, G., Arnoldi, A., 2019. Multifunctional peptides for the prevention of cardiovascular disease: A new concept in the area of bioactive food-derived peptides. J. Funct. Foods 55, 135-145.

Lammi, C., Aiello, G., Dellafiora, L., Bollati, C., Boschin, G., Ranaldi, G., Ferruzza, S., Sambuy, Y., Galaverna, G., Arnoldi, A., 2020. Assessment of the Multifunctional Behavior of Lupin Peptide P7 and Its Metabolite Using an Integrated Strategy. J. Agr. Food Chem. 68, 13179-13188.

Legnik, S., Stular, T., Brus, B., Knez, D., Gobec, S., Janezic, D., Konc, J., 2015. LiSiCA: A Software for Ligand-Based Virtual Screening and Its Application for the Discovery of Butyrylcholinesterase Inhibitors. J. Chem. Inf. Model. 55, 1521-1528.

Lenci, E., Trabocchi, A., 2020. Peptidomimetic toolbox for drug discovery. Chemical Society Reviews 49, 3262-3277.

Li, S.S., Bu, T.T., Zheng, J.X., Liu, L., He, G.Q., Wu, J.P., 2019. Preparation, Bioavailability, and Mechanism of Emerging Activities of Ile-Pro-Pro and Val-Pro-Pro. Comprehensive Reviews in Food Science and Food Safety 18, 1097-1110.

Li, S.S., Liu, L., He, G.Q., Wu, J.P., 2018. Molecular targets and mechanisms of bioactive peptides against metabolic syndromes. Food & Function 9, 42-52.

Li, Y., Aiello, G., Fassi, E.M.A., Boschin, G., Bartolomei, M., Bollati, C., Roda, G., Arnoldi, A., Grazioso, G., Lammi, C., 2021. Investigation of Chlorella pyrenoidosa Protein as a Source of Novel Angiotensin I-Converting Enzyme (ACE) and Dipeptidyl Peptidase-IV (DPP-IV) Inhibitory Peptides. Nutrients 13, 1624.

Majumder, K., Chakrabarti, S., Morton, J.S., Panahi, S., Kaufman, S., Davidge, S.T., Wu, J.P., 2015. Egg-derived ACE-inhibitory peptides IQW and LKP reduce blood pressure in spontaneously hypertensive rats. J. Funct. Foods 13, 50-60.

Majumder, K., Wu, J.P., 2010. A new approach for identification of novel antihypertensive peptides from egg proteins by QSAR and bioinformatics. Food Res. Int. 43, 1371-1378.

Maldonado-Rojas, W., Olivero-Verbel, J., 2011. Potential interaction of natural dietary bioactive compounds with COX-2. J. Mol. Graph. Model. 30, 157-166.

Marseglia, A., Dellafiora, L., Prandi, B., Lolli, V., Sforza, S., Cozzini, P., Tedeschi, T., Galaverna, G., Caligiani, A., 2019. Simulated Gastrointestinal Digestion of Cocoa: Detection of Resistant Peptides and In Silico/In Vitro Prediction of Their Ace Inhibitory Activity. Nutrients 11.

Masuyer, G., Schwager, S.L.U., Sturrock, E.D., Isaac, R.E., Acharya, K.R., 2012. Molecular recognition and regulation of human angiotensin-I converting enzyme (ACE) activity by natural inhibitory peptides. Sci. Rep. 2.

Minkiewicz, P., Iwaniak, A., Darewicz, M., 2019. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. Int. J. Mol. Sci. 20.

Moller, N.P., Scholz-Ahrens, K.E., Roos, N., Schrezenmeir, J., 2008. Bioactive peptides and proteins from foods: indication for health effects. Eur. J. Nutr. 47, 171-182.

Mora, L., Aristoy, M.C., Toldrá, F., 2019. Bioactive Peptides, in: L. Melton, F. Shahidi, P. Varelis (Eds.), Encyclopedia of Food Chemistry. Elsevier Inc., pp. 381-389.

Peighambardoust, S.H., Karami, Z., Pateiro, M., Lorenzo, J.M., 2021. A Review on Health-Promoting, Biological, and Functional Aspects of Bioactive Peptides in Food Applications. Biomolecules 11.

Rollinger, J.M., Schuster, D., Baier, E., Ellmerer, E.P., Langer, T., Stuppner, H., 2006. Taspine: bioactivity-guided isolation and molecular ligand-target insight of a potent acetylcholinesterase inhibitor from Magnolia x soulangiana. J. Nat Prod. 69, 1341-1346.

Schrader, M.M.M.B.-d.-.-.-.-P., 2018. Origins, Technological Development, and Applications of Peptidomics, in: M. Schrader, L. Fricker (Eds.), Peptidomics. Methods in Molecular Biology. Humana Press, New York, NY, pp. 3-39.

Smith, K.H., Tejeda-Montes, E., Poch, M., Mata, A., 2011. Integrating top-down and self-assembly in the fabrication of peptide and protein-based biomedical materials. Chemical Society Reviews 40, 4563-4577.

Tavano, O.L., Berenguer-Murcia, A., Secundo, F., Fernandez-Lafuente, R., 2018. Biotechnological Applications of Proteases in Food Technology. Comprehensive Reviews in Food Science and Food Safety 17, 412-436.

Uhlig, T., Kyprianou, T., Martinelli, F.G., Oppici, C.A., Heiligers, D., Hills, D., Calvo, X.R., Verhaert, P., 2014. The emergence of peptides in the pharmaceutical business: From exploration to exploitation. EuPA Open Proteomics 4, 58-69.

van der Wielen, N., Moughan, P.J., Mensink, M., 2017. Amino Acid Absorption in the Large Intestine of Humans and Porcine Models. Journal of Nutrition 147, 1493-1498.

Vasquez-Villanueva, R., Orellana, J.M., Marina, M.L., Garcia, M.C., 2019. Isolation and Characterization of Angiotensin Converting Enzyme Inhibitory Peptides from Peach Seed Hydrolysates: In Vivo Assessment of Antihypertensive Activity. J. Agr. Food Chem. 67, 10313-10320.

Wang, H., Qiu, N., Mine, Y., Sun, H.H., Meng, Y.Q., Bin, L., Keast, R., 2020. Quantitative Comparative Integrated Proteomic and Phosphoproteomic Analysis of Chicken Egg Yolk Proteins under Diverse Storage Temperatures. J. Agr. Food Chem. 68, 1157-1167.

Wu, J.P., Aluko, R.E., Nakai, S., 2006. Structural requirements of angiotensin I-converting enzyme inhibitory peptides: Quantitative structure-activity relationship study of di- and tripeptides. J. Agr. Food Chem. 54, 732-738.

Yu, Z.P., Chen, Y., Zhao, W.Z., Zheng, F.P., Ding, L., Liu, J.B., 2019. Novel ACE inhibitory tripeptides from ovotransferrin using bioinformatics and peptidomics approaches. Sci. Rep. 9.

**Table 1.** Heuristic selection ~~of Visual inspection~~ of most active sequences for visual inspection ~~in~~ from those reported in the reference databases *[1]

| Sequence | $IC_{50}$ (µM)[2] | 1st Position | 2nd Position | 3rd Position |
|---|---|---|---|---|
| LRW | 0.2 | L | R | W |
| IKW | 0.2 | I | K | W |
| GEP | 0.3 | G | E | P |
| LKP | 0.4 | L | K | P |
| MKP | 0.4 | M | K | P |
| IVY | 0.5 | I | V | Y |
| MRW | 0.6 | M | R | W |
| IRY | 0.6 | I | R | Y |
| IRW | 0.6 | I | R | W |
| LGP | 0.7 | L | G | P |
| LIY | 0.8 | L | I | Y |
| MAP | 0.8 | M | A | P |
| IVR | 0.8 | I | V | R |
| VHW | 0.9 | V | H | W |
| LRP | 1.0 | L | R | P |
| | **Polarity** | All hydrophobic | Mixed | Mainly hydrophobic |
| | **Residues** | V/L/I/G/M | R/K/E/V/G/I/H | W/P/Y + R |
| | **Prevalence** | I/L | --- | P |

*[1] BIOPEP-UWM (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) and~~;~~ AHTPDB (https://webs.iiitd.edu.in/raghava/ahtpdb/index.php)

[2] $IC_{50}$ stands for the half maximal inhibitory concentration

**Table 2.** List of ~~library's~~ tripeptides from the in-house virtual library with leucine and proline in first and third position, respectively

| Sequence | ~~Already~~ Previously reported inhibitory activity ~~characterized*~~ [1] | Experimental activity [2] ~~$IC_{50}$ (μM)*~~ |
|---|---|---|
| LVP | Yes | 9.9 |
| LIP | No | Not reported~~nr~~ |
| LTP | No | Not reported~~nr~~ |
| LEP | No | Not reported~~nr~~ |
| LAP | Yes | 3.5 |
| LQP | Yes | 1.9 |
| LKP | Yes | 0.3 |
| LRP | Yes | 1.0 |
| LDP | No | Not reported~~nr~~ |
| LSP | Yes | 1.7 |
| LMP | No | Not reported~~nr~~ |
| LGP | Yes | 0.7 |
| LNP | Yes | 43.0 |
| LCP | No | Not reported~~nr~~ |
| LHP | No | Not reported~~nr~~ |
| LFP | No | Not reported~~nr~~ |
| LYP | Yes | 6.6 |
| LWP | No | Not reported~~nr~~ |

[1]* According to the data reported in the publicly available database BIOPEP-UWM (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) or AHTPDB (https://webs.iiitd.edu.in/raghava/ahtpdb/index.php); ~~nr → not reported~~

[2] Experimental activity, expressed as $IC_{50}$ (half maximal inhibitory concentration), reported in the publicly available database BIOPEP-UWM (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) or AHTPDB (https://webs.iiitd.edu.in/raghava/ahtpdb/index.php)

700

**Table 3.** Multiple-model virtual screening (VS) results of tripeptides developed using IVR, LRW, IVY and GEP tripeptides or captopril as template molecule

| Sequence | VS with IVR | | VS with captopril | | VS with LRW | | VS with IVY | | VS with GEP | | Average rank |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | VS score | Rank | VS score | Rank | VS score | Rank | VS score | Rank | VS score | Rank | |
| LIP* | 0.76 | 1 | 0.46 | 3 | 0.57 | 1 | 0.68 | 1 | 0.67 | 1 | 1 |
| LTP* | 0.72 | 2 | 0.48 | 2 | 0.54 | 2 | 0.65 | 2 | 0.69 | 2 | 2 |
| LMP* | 0.66 | 4 | 0.48 | 1 | 0.54 | 3 | 0.59 | 3 | 0.62 | 3 | 3 |
| LEP | 0.68 | 3 | 0.44 | 5 | 0.51 | 4 | 0.61 | 4 | 0.84 | 4 | 4 |
| LCP | 0.65 | 6 | 0.46 | 4 | 0.49 | 6 | 0.58 | 6 | 0.61 | 6 | 6 |
| LHP | 0.61 | 7 | 0.38 | 7 | 0.50 | 5 | 0.54 | 5 | 0.62 | 5 | 6 |
| LDP | 0.65 | 5 | 0.41 | 6 | 0.49 | 7 | 0.58 | 7 | 0.80 | 7 | 6 |
| LWP | 0.54 | 9 | 0.33 | 9 | 0.45 | 8 | 0.49 | 8 | 0.55 | 8 | 8 |
| LFP | 0.59 | 8 | 0.37 | 8 | 0.45 | 9 | 0.53 | 9 | 0.55 | 9 | 9 |

* indicates the three best sequences according to the average ranking out of the five independent screening applied

701

702

703

**Table 4.** Docking ~~results~~ scores and half maximal inhibitory concentration ($IC_{50}$) of the *in vitro* ACE-inhibitory activity

| Sequence | N-domain ~~Run 1Run 2Run~~ Docking scores ~~Mean ± SD~~[1] | C-domain ~~Run 1Run 2Run 3~~Docking scores ~~Mean ± SD~~[1] | Experimental assessment $IC_{50}$ (μM) [2] |
|---|---|---|---|
| LMP | ~~76.577.076.5~~76.6 ± 0.3 | ~~91.490.389.0~~90.2 ± 1.2 | 15.8±0.2 |
| LTP | ~~72.972.673.9~~73.1 ± 0.6 | ~~81.383.982.7~~82.3 ± 1.4 | ~~nd*~~Not determined |
| LIP | ~~70.573.071.9~~71.8 ± 1.3 | ~~75.667.076.6~~736.1 ± 0.5~~.3~~ | Not determined~~nd*~~ |
| LKP | ~~75.777.375.5~~76.2 ± 1.0 | ~~94.592.398.8~~95.2 ± 3.3 | 9.23±0.6 [#] |

[1] Docking scores are expressed as mean values ± standard deviation of three independent docking simulations

[2] $IC_{50}$ stands for the half maximal inhibitory concentration and it is expressed as a mean value of four independent experiments in triplicate ± standard deviation

[#] Statistical differences between the two treatments at each time-point were calculated by Student's t-test (***p < ~~...~~0.001)~~* nd stands for not determined in the present study~~

704

705

**Figures**

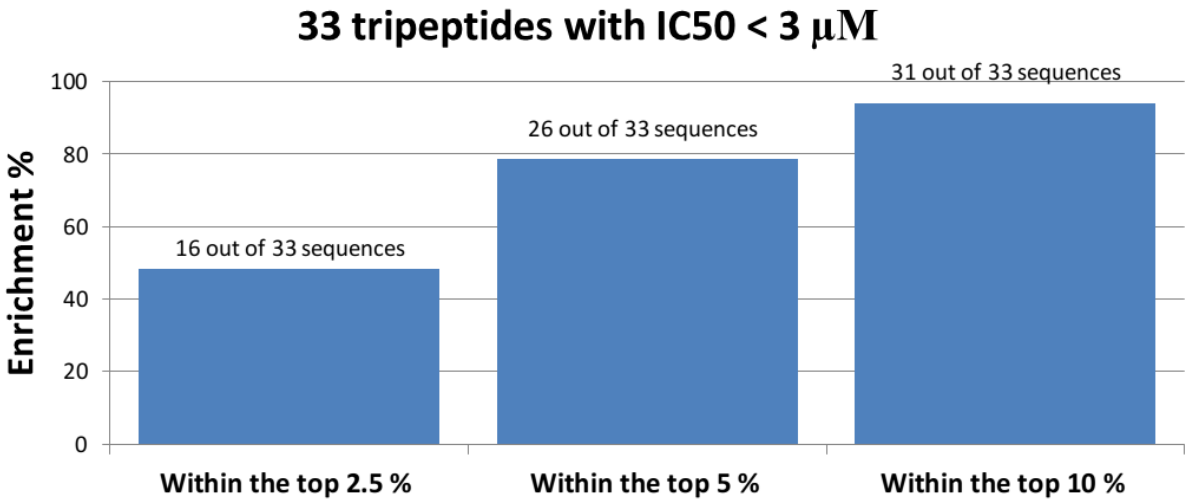


**Figure 1.** Enrichment plot of already characterized potent ACE inhibitory tripeptides (i.e. $IC_{50}$ < 3 µM) in the top-ranked region of library hierarchization according to at least one of the five independent screening.

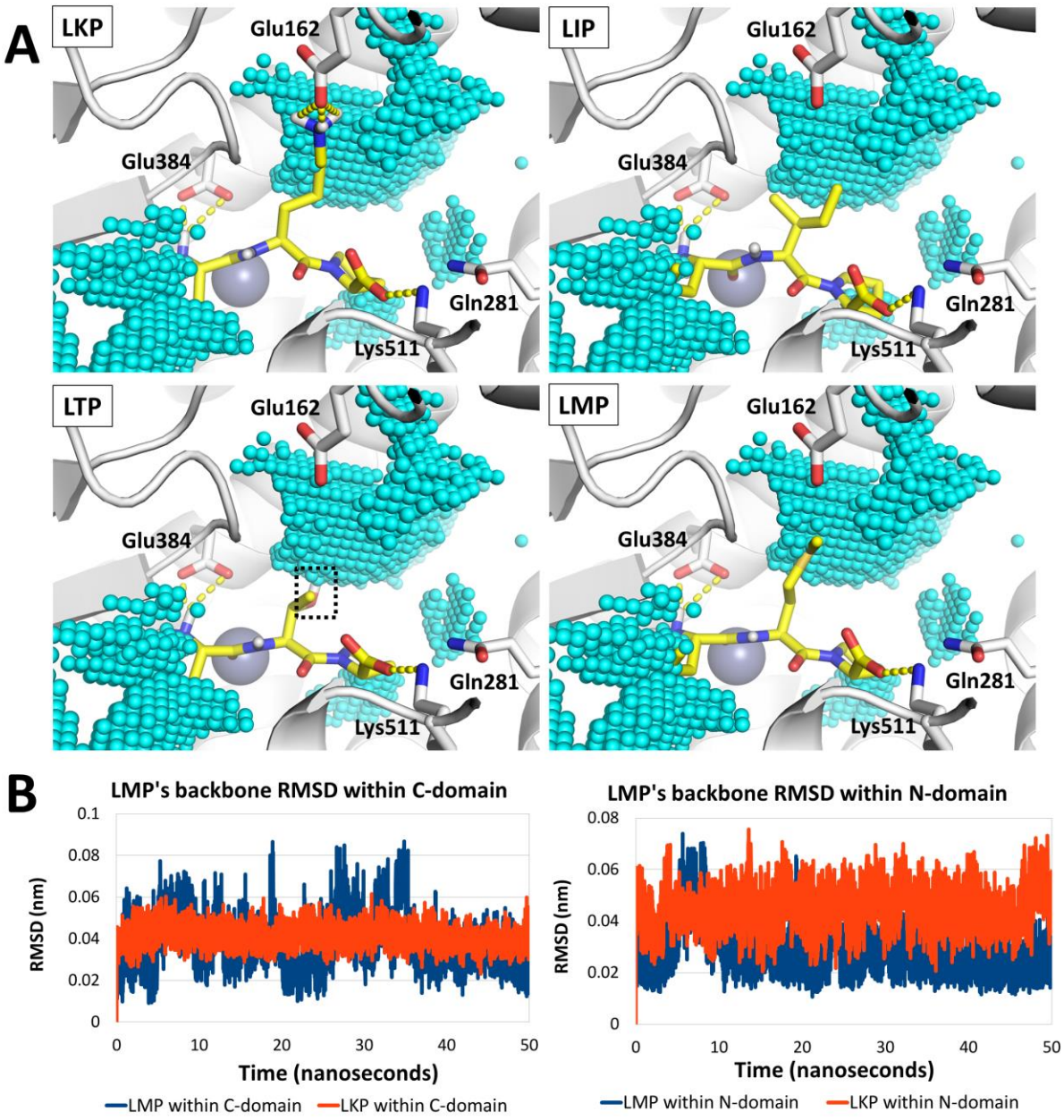


**Figure 2.** Results of ~~M~~molecular modeling ~~results~~**. A.** Binding poses of LKP, LMP, LIP and LTP at the binding site of C-terminal domain of ACE. Polar contacts are represented by yellow dotes lines, while the black dashed box indicates the improper arrangement of threonine's hydroxyl group into a space energetically suitable to receive hydrophobic groups (represented by cyan spheres). The protein is represented in white cartoon, peptides are represented in sticks and

the catalytic Zn ion is represented by the grey sphere. **B.** Molecular dynamic results of LMP and LKP. The RMSD plot of LMP's backbone in complex with the N- and C-terminal is shown.

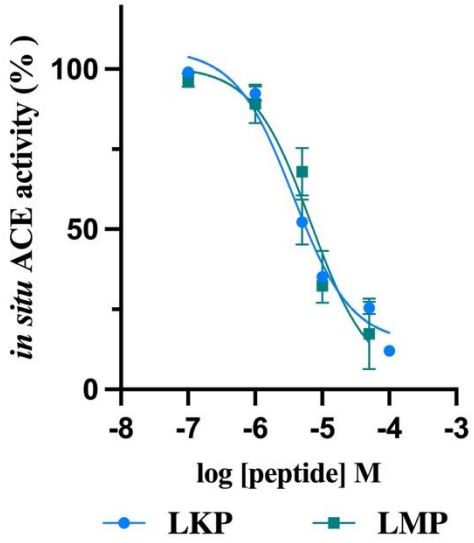


**Figure 3.** Evaluation of the inhibitory effects of LKP (blue line) and LMP (green line) peptides on ACE expressed by Caco-2 cell membranes. Each point represents the mean ± sd of four independent experiments in triplicate.

Graphical Abstract