Check for updates

Received: 6 October 2020 | Revised: 30 November 2020 | Accepted: 25 December 2020

DOI: 10.1111/cts.12985

ARTICLE

Estimating peptide half-life in serum from tunable, sequence-related physicochemical properties

Marco Cavaco[1,2] | **Javier Valle**[2] | **Isabel Flores**[3] | **David Andreu**[2*] | **Miguel A. R. B. Castanho**[1*]

[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

[2]Proteomics and Protein Chemistry Unit, Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain

[3]Instituto das Políticas Públicas e Sociais do ISCTE, Lisboa, Portugal

Correspondence
Miguel A. R. B. Castanho, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa. Av. Prof. Egas Moniz, 1649–028 Lisboa, Portugal.
Email: macastanho@medicina.ulisboa.pt

David Andreu, Proteomics and Protein Chemistry Unit, Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona Biomedical Research Park, Dr. Aiguader 88, 08003 Barcelona, Spain.
Email: david.andreu@upf.edu

Funding information
This research was supported by the Portuguese Fundação para a Ciência e a Tecnologia (FCT; grants PD/BD/128281/2017, PTDC/BBB-NAN/1578/2014, PTDC/BIA-VIR/29495/2017, UID/Multi/04349/2019, and PTDC/QUI-NUC/30147/2017), the Spanish Ministry of Economy and Innovation (MINECO, grants AGL2014-52395-C2-2-R and AGL2017-84097-C2-2-R, and Maria de Maeztu Program for Centers of Excellence); the European Union H2020-MSCA-RISE-2014 program (grant no. 828774), and the "La Caixa" Banking Foundation (grant HR17-00409).

Abstract

Proteolytic instability is a critical limitation for peptide-based products. Although significant efforts are devoted to stabilize sequences against proteases/peptidases in plasma/serum, such approaches tend to be rather empirical, unspecific, time-consuming, and frequently not cost-effective. A more rational and potentially rewarding alternative is to identify the chemical grounds of susceptibility to enzymatic degradation of peptides so that proteolytic resistance can be tuned by manipulation of key chemical properties. In this regard, we conducted a meta-analysis of literature published over the last decade reporting experimental data on the lifetimes of peptides exposed to proteolytic conditions. Our initial database contained 579 entries and was curated with regard to amino acid sequence, chemical modification, terminal half-life ($t_{1/2}$) or other stability readouts, type of stability assay, and biological application of the study. Although the majority of entries in the database corresponded to (slightly or substantially) modified peptides, we chose to focus on unmodified ones, as we aimed to decipher intrinsic characteristics of peptide proteolytic susceptibility. Specifically, we developed a multivariable regression model to unravel those peptide properties with most impact on proteolytic stability and thus potential $t_{1/2}$ predicting ability. Model validation was done by two different approaches. First, a library of peptides spanning a large interval of properties that modulate stability was synthesized and their $t_{1/2}$ in human serum were experimentally determined. Second, the $t_{1/2}$ of 21 selected peptides approved for clinical use or in clinical trials were recorded and matched with the model-estimated values. With both approaches, good correlation between experimental and predicted $t_{1/2}$ data was observed.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Stability is a hot topic in the peptide field, as a frequently serious drawback in peptide drug application. Unfortunately, knowledge on the physicochemical factors affecting

*These authors contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of the American Society for Clinical Pharmacology and Therapeutics.

peptide stability is scarce. During the development of new therapeutic peptides, researchers tend to rely solely on empirical experience to improve peptide sequence lifetimes.

WHAT QUESTION DID THIS STUDY ADDRESS?

Which are the physicochemical properties affecting the stability of peptides?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study reveals major predictors of peptide stability. In addition, and because the determination of experimental half-life is demanding, we have used the knowledge from the study to develop a multivariable regression model that predicts peptide stability in a reliable way.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

Our study will help researchers improve their therapeutic peptide candidates more rapidly, thus contribute to decrease the number of peptide withdrawals in preclinical and clinical studies.

INTRODUCTION

Peptide drugs are steadily making inroads into the clinics. However, poor pharmacokinetics (PKs), mainly due to degradation by body proteases, often critically compromises the successful development of a bioactive peptide into a drug lead.[1,2] Degradation occurs mainly in plasma but also in other locations, such as the gastrointestinal system, the liver, and in immune cells, the outcome often being lifetimes so short that oral delivery is unfeasible, with high doses often require even for parenteral administration.[3–5] On the face of these limitations, developing protease-resistant versions is essential if peptide drugs are to fulfill their promise. Moreover, development of such versions should ideally involve minimal alterations in structure, hence preferably rely on natural, unmodified amino acids, so that activity and cost-effectiveness are not compromised.

Although practically all peptides are by definition protease-susceptible, some sequences are sporadically found to be long-lived in serum, plasma, or other biological fluids. The determinants (e.g., composition and structure) for such exceptional behavior are often enigmatic, and efforts toward unraveling them are neither plentiful nor intense.[6,7]

Typically, for peptides with short bloodstream terminal half-lives ($t_{1/2}$), peptide chemists are successfully developing strategies to increase lifespan by resorting to chemical modifications, such as end-group capping, *D*- or non-coding amino acid substitution, or cyclization, as well as conjugation to macromolecules, encapsulation, or other approaches.[1,8–10] In general, most of these modifications rely on a mixture of empiric and chemical intuition criteria. Nonetheless, knowledge-based compilation and systematization of pertinent data would seem a preferable course of action for integrating proteolytic stability into PK improvement strategies for peptide drug lead optimization.

In an effort to shed light onto this problem, herein we report results of a meta-analysis conducted on publications over the last decade in which experimentally determined peptide stabilities toward proteases are reported. Specifically, we have worked on a database built with 579 entries where each peptide is annotated by designation, sequence, modifications, stability data, stability assay, and biological application.

Whereas most entries in the database represent slightly or substantively modified peptides, studying the stability of unmodified ones is fundamental for understanding biological significance and ensuring biotechnological cost-effectiveness. For these reasons, we focused on a subset of unmodified peptides and developed a multivariable regression model to identify key chemical parameters that determine $t_{1/2}$ and may be used to predict peptide stability on the basis of a priori known physicochemical properties. The model has been validated by two distinct approaches. First, a library of 16 de novo designed peptides with sequences spanning a broad range of key parameters affecting peptide stability determined within this study was synthesized and the respective $t_{1/2}$ in human serum were experimentally determined. Second, we selected 21 peptides already approved for clinical use or in clinical trials for which $t_{1/2}$ were available from the literature. The model-predicted stabilities for these 21 sequences were next determined and compared with the experimental values. Our model was shown capable of predicting the stability of the peptides with more than reasonable accuracy.

METHODS

Database design and data collection

Peptide-related research is published over a wide variety of journals. We focused our search over 33 journals providing a broad coverage of peptide science areas, from academic

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

scientific research to health applications (Table S1). A curated database of peptides with experimentally determined $t_{1/2}$ was created by following the multistep strategy presented in Figure 1. Briefly, to detect relevant papers, we used the search engine of each journal webpage with "peptides" and/or "characterization," "stability," "proteolysis," "degradation," and "cleavage" as keywords over the 2010 to 2019 decade. To ensure relevance, the actual literature reference of the identified papers was also searched manually to verify that the required data were indeed included. Of 4150 papers, 3571 were excluded from the study for lack of accurate quantitative or semiquantitative data on stability, identification of the stability assay applied, explicit peptide sequence, or specified sequence modifications. The comparison of these data is not straightforward. Most frequently, only semiquantitative data are reported (e.g., percentage of peptide that remains intact after a certain period), whereas the most informative stability parameter (i.e., $t_{1/2}$) is not given.[11] This lack of harmonization in published stability data is one of the most limiting factors hampering proper data comparison. Hence, a categorization of these data was needed (Tables S2–S3) to properly visualize the stability distribution within our database (Figure 2). To allow the building of a multivariable regression model, the remaining 579 entries in the database had to be further refined, with 528 being excluded for not providing precise, experimentally determined $t_{1/2}$ values. Thus, the regression model took into consideration the $t_{1/2}$ values and chemical properties associated to the amino acid sequences of the remaining 51 selected papers (129 peptides).

Quantitative data analysis

As mentioned above, only unmodified natural peptides with experimentally determined $t_{1/2}$ were considered (Figure 1) to identify intrinsic determinants of proteolytic resistance. For each entry in the refined database, the following properties were calculated using either online tools[12] or information available in the literature: (i) molecular weight; (ii) isoelectric point (pI); (iii) UV-Vis absorption extinction coefficient (M^{-1} cm^{-1}) at 280 nm; (iv) net charge at pH 7.0; (v) hydrophobicity (H); (vi) hydrophobic moment (μH); (vii) presence of nonpolar residues (%); and (viii) secondary structure.

As molecular weight, pI, net charge, hydrophobicity, hydrophobic moment, and presence of nonpolar residues are continuous variables, descriptive statistics was performed to evaluate these entries (Figure S1). They were found to be evenly distributed, thus requiring no further action. A scatterplot was then used to evaluate the correlation between each of the variables and peptide $t_{1/2s}$ (Figures S2–S5).

Because pI was not evenly distributed through the entries of the database, we categorized this variable into 4 ranges: pI < 7; 7 < pI < 10; 10 < pI < 12; and pI > 12. Distribution among those groups was homogeneous (Table S4). The net charge was again unevenly distributed. In this case, we classified peptides as anionic (ch <0) and cationic (ch >0). As the second group was significantly larger than the first, we further categorized the variable into: 0 < ch < 3; 3 < ch < 7; and ch > 7, which made the groups more homogenous (Table S5).

The UV-Vis extinction coefficient reports mainly the abundance of tryptophan (Trp; 5690 M^{-1} cm^{-1}) and tyrosine

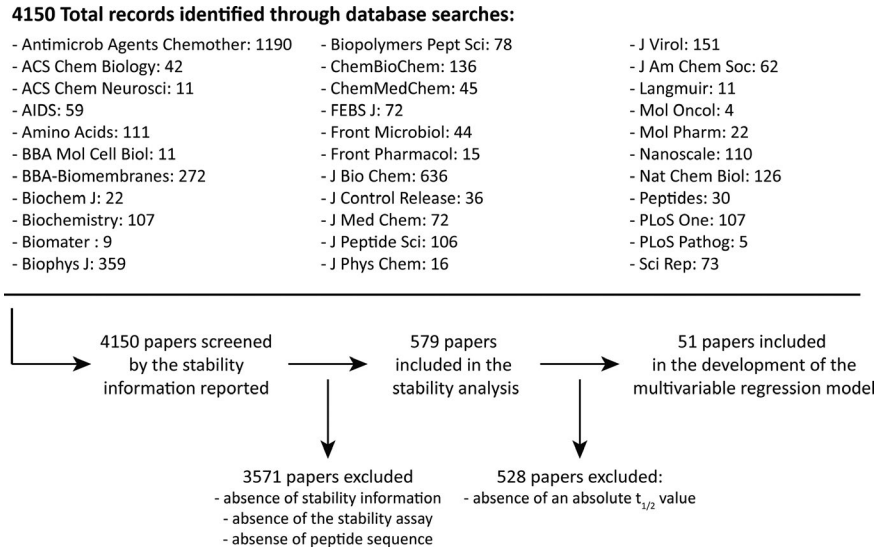


FIGURE 1 Multistep strategy in data acquisition. The meta-analysis followed three different steps: (1) apply search engines of the selected journals with the keywords: "peptides" and/or "characterization," "stability," "proteolysis," "degradation," and "cleavage" to identify relevant papers (output = 4150 papers); (2) the papers not reporting stability information, type of stability assay performed, or specified peptide sequence were excluded (output = 579 papers); (3) to develop the multivariable regression model, precise terminal half-life ($t_{1/2}$) values were required; thus papers without this information were excluded (output = 51 papers)

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

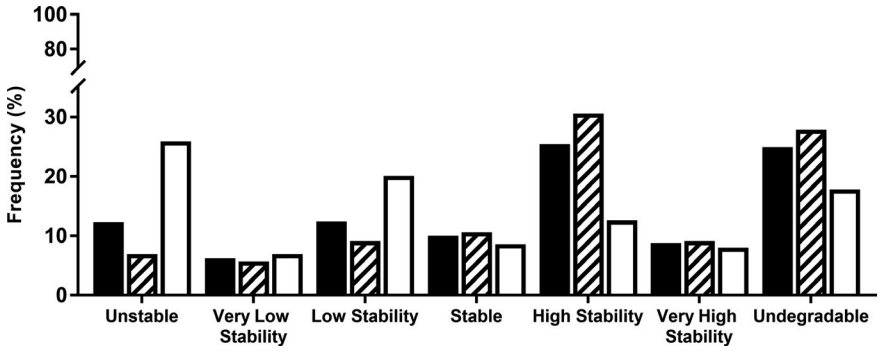


FIGURE 2 Stability distribution of peptides within the database. The distribution of all peptides (black columns) after categorization shows that most peptides presented a high stability, very high stability or were considered undegradable. The existence of over 70% of modified peptides (dashed columns) clearly dominated the stabilities reported. Nevertheless, among the unmodified peptides (white columns), although unstable, very low, and low stability populations predominated, significant populations of high stability, very high stability, and undegradable peptides could also be found

(Tyr; 1280 M^{-1} cm^{-1}) residues in the sequence, with smaller contributions from phenylalanine (200 M^{-1} cm^{-1}) and cysteine (120 M^{-1} cm^{-1}). This is not a continuous variable and was also not evenly distributed. Given the different contributions of Trp and Tyr, we categorized the entries for the presence/absence of these residues (Tables S6 and S7). The new groups thus defined were more homogenous.

Secondary structure is a parameter more difficult to assess and the sample size within our database was too small to allow a proper statistical assessment (Table S8). In addition, the frequency of each possible secondary structure was low. Consequently, we excluded this variable from further analysis.

This initial statistical assessment allowed us to categorize variables not homogenously distributed and to identify continuous variables that might correlate with $t_{1/2}$. Then, for each variable, we performed a one-way analysis of variance (ANOVA) with a Bonferroni's multiple comparison test. Based on these results, and to allow the development of the multivariable regression model, we categorized noncontinuous variables as dummies (binary variables). These dummies were compared using an independent *t*-test to verify statistically significant differences ($\alpha < 0.05$). Table S17 shows the new descriptive statistics. Next, a Spearmen correlation test was performed to identify the correlation between the variables studied and the $t_{1/2}$. Finally, the variables showing a statistically significant correlation ($\alpha < 0.05$) were included into the multivariable regression model.

All computational statistical analyses were performed on the IBM SPSS statistics version 25. All related plots were obtained on GraphPad Prim version 7 (GraphPad Software, San Diego, CA).

Peptide synthesis and purification

For experimental validation of the multivariable regression model, 16 peptides (Table 1), were synthesized in C-terminal carboxamide form in a Prelude automated synthesizer (Gyros Protein Technologies, Tucson, AZ) running Fmoc protocols at 0.1 mmol scale on an Fmoc-Rink-amide ChemMatrix resin. Side chain functionalities were protected with *tert*-butyl (Tyr, Ser), N^G-2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Arg), and tert-butoxycarbonyl (Trp) groups. Eight-fold excess of Fmoc-*L*-amino acids and HBTU, in the presence of a double molar amount of DIEA, were used for the coupling steps, with DMF as solvent. After chain assembly, full deprotection and cleavage was carried out with TFA/H_2O/TIS (95:2.5:2.5 v/v, 90 min, r.t.). The peptides were then precipitated from the cleavage solution by addition of cold diethyl ether, redissolved in H_2O and lyophilized. They were checked for purity by analytical reverse-phase high-performance liquid chromatography (RP-HPLC) and purified by preparative RP-HPLC as described below. Fractions of greater than 90% purity and correct mass by liquid chromatography-mass spectrometry (LC-MS) were pooled and lyophilized. Peptide stock solutions were prepared in sterile deionized water and stored at −20°C.

The $t_{1/2}$ of the peptides was determined experimentally using the method described below and matched against that obtained by the multivariable regression model for the respective amino acid sequence.

RP-HPLC and LC-MS analysis

Analytical RP-HPLC was performed on a LC-20AD instrument (Shimadzu, Japan) equipped with a Luna C18 column (4.6 × 50 mm, 3 μm; Phenomenex, USA) using 5–60% linear gradients of solvent B (0.036% TFA in MeCN) into A (0.045% TFA in H_2O) at a flow rate of 1 ml/min and UV detection at 220 nm. For more polar peptides 2 (GSSQSSGSGSSQSSG) and 6 (GSSRSSGSGSSRSSG) (Table 1), a linear 0–40% B gradient was used. Preparative RP-HPLC was performed

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Peptides synthesized to validate the multivariable regression model

Peptide	Amino acid sequence	Molecular weight, Da, calculated (found)	HPLC t_R (min)[a]	Purity, %
1	GAAQAAGSGAAQAAG	1157.2 (1158.1)	3.187	93.4
2	GSSQSSGSGSSQSSG	1285.2 (1286.1)	3.207	94.7
3	GAARAAGSGAARAAG	1213.3 (1214.2)	5.055	95.4
4	GAAQAYGSGYAQAAG	1342.2 (1341.4)	5.068	95.0
5	GAAQAAGWGAAQAAG	1257.1 (1256.3)	4.281	97.3
6	GSSRSSGSGSSRSSG	1342.1 (1341.3)	3.175	98.7
7	GSSQSYGSGYSQSSG	1437.8 (1437.4)	3.903	98.4
8	GAARAYGSGYARAAG	1397.6 (1397.5)	5.148	97.0
9	GSSQSSGWGSSQSSG	1385.2 (1384.3)	3.289	98.9
10	GAARAAGWGAARAAG	1313.3 (1312.4)	4.764	90.0
11	GAAQAYGWGYAQAAG	1441.1 (1440.5)	6.939	96.1
12	GSSRSYGSGYSRSSG	1494.2 (1493.5)	3.901	94.3
13	GSSRSAGWGASRSSG	1409.3 (1408.4)	4.369	94.4
14	GSSQYAGWGAYQSSG	1505.1 (1504.5)	5.262	96.9
15	GAARYAGWGAYRAAG	1497.4 (1496.6)	5.212	98.8
16	GSSRYSGWGSYRSSG	1593.3 (1592.69	5.210	94.2

Abbreviation: HPLC, high-performance liquid chromatography.

[a]See experimental part for details.

on an LC-8 instrument (Shimadzu) fitted with a Luna C18 column (21.2 × 250 mm, 10 μm; Phenomenex) using linear gradients of solvent B (0.1% TFA in MeCN) into A (0.1% TFA in H_2O) with a flow rate of 25 ml/min and UV detection at 220 nm. LC-MS was performed in an LC-MS 2010EV instrument (Shimadzu) fitted with an XBridge C18 column (4.6 × 150 mm, 3.5 μm; Waters, Spain), eluting with linear gradients of 0.08% formic acid in MeCN into 0.1% formic acid in H_2O over 15 min at 1 ml/min. Electrospray ionization was performed with a detector voltage of 1.5 kV, in the positive mode, with a nebulizing gas flow of 1.5 L/min, a 1 sec event time and a scan speed of 2000, in the 100–2000 m/z mass range.

Serum stability of peptides

1 mM peptide solutions in H_2O were mixed 1:1 (v/v) with human serum (Sigma-Aldrich, USA), and incubated at 37°C with gentle shaking. Then, 120-µL aliquots were taken at different timepoints, and protease activity was stopped with 20 µL of 10% (v/v, in H_2O) trichloroacetic acid. After 30 min at 4°C, samples were centrifuged at 13,000 *g* for 10 min to remove serum proteins, and the supernatants were analyzed by analytical RP-HPLC and LC-MS, as described above. Percent of intact peptide was calculated by peak integration, expressed as percent of the amount at t_0, and data were fitted to a one-phase exponential decay model using GraphPad Prism version 7 to estimate the $t_{1/2}$.

Software tool development

A user-friendly platform freely usable to predict the $t_{1/2}$ of any given peptide sequence was developed. In addition to the $t_{1/2}$, other sequence-related characteristics are also available (Table S22). For the sake of clarity, we organized the software tool outputs in three different sections, namely "Basic information," "pH and Isoelectric Point," and "Peptide Stability." In the first section, the user can visualize the number of residues, chemical formula, molecular weight, and extinction coefficient. In the pH and isoelectric point section, basic information related to charge and pI are presented. Finally, the $t_{1/2}$ and the respective categorization of peptides with regard to stability are presented in the last section.

The platform was programmed on JavaScript version 8, update 221. The respective integration and framework development was performed on the http://electronjs.org website.

Results

Multivariable regression model

We used the refined database of 51 papers on 129 unmodified peptides with experimentally determined $t_{1/2}$ to identify determinants responsible for proteolytic resistance. As mentioned above, the physicochemical properties investigated were: (i) molecular weight; (ii) isoelectric point (pI); (iii) UV-Vis extinction coefficient (M^{-1} cm^{-1}) at 280 nm; (iv) net

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

charge at pH 7.0; (v) hydrophobicity (H); (vi) hydrophobic moment (μH); (vii) presence of nonpolar residues (%); and (viii) secondary structure. The selection of these variables is based on their highly informative nature, easy computation, and practical reasoning, as directly derived from amino acid sequence.

The molecular weight, pI, net charge, hydrophobicity, hydrophobic moment, and presence of nonpolar residues were treated as continuous variables. Among these, molecular weight, hydrophobicity, hydrophobic moment, and presence of nonpolar residues presented a homogenous distribution (Figure S1). A scatterplot for each variable was used to verify its correlation with peptide $t_{1/2}$ ($\ln[t_{1/2}]$) (Figures S2–S5). Molecular weight (Figure S2) and hydrophobic moment (Figure S4) showed a random distribution (i.e., no correlation was observed). In contrast, hydrophobicity (Figure S3) and presence of nonpolar residues (Figure S5) showed a positive correlation.

The pI and net charge present a nonhomogenous distribution. Consequently, we categorized them into groups, in search of homogeneity. Concerning pI, categorizing this variable into four groups (pI <7; 7 < pI <10; 10 < pI <12; and pI >12) provided a homogeneous distribution (Table S4), but the sample size within each group was too small. To overcome this limitation, we performed a one-way ANOVA and a box-plot to assess differences, which allowed a proper categorization into two homogenous groups. Figure S6 and Table S9 show a tendency of peptides with pI greater than 10 to have a lower $t_{1/2}$ (i.e., basic residues [high pI] decreased stability). On this basis, the pI variable was categorized into two groups, namely pI less than 10 (flagged "pI = 0") and pI greater than or equal to 10 (flagged "pI = 1"), which were equally distributed. Next, a *t*-test was performed to assess the existence of significant differences between groups. The results demonstrated a statistically significant difference ($\alpha = 0.023$; Figure S7 and Table S10). Concerning net charge, we categorized the variable into four groups (Table S5). However, owing to the high heterogeneity still observed, the variable was excluded from the study.

The UV-Vis extinction coefficient reports mainly the abundance of tryptophan (Trp; 5690 M^{-1} cm^{-1}) and tyrosine (Tyr; 1280 M^{-1} cm^{-1}) residues in the sequence, with smaller contributions from phenylalanine (200 M^{-1} cm^{-1}) and cysteine (120 M^{-1} cm^{-1}). This semicontinuous variable was not evenly distributed. Given the different contributions of Trp and Tyr, we categorized the entries for the presence/absence of these amino acid residues (Tables S6 and S7). The new groups were more homogenous.

Concerning Trp, we divided the entries into three categories depending on the absence, the presence of one Trp, or of more than one Trp. We then performed one-way ANOVA with a Bonferroni's multiple comparison test. Variance analysis among these three groups demonstrated a statistically significant difference ($\alpha = 0.038$; Figure S8 and Table S11). However, the multiple comparison test did not report statistically significant differences (Table S12). Nevertheless, based on the tendency observed on the boxplot and α values in the Bonferroni's test, the absence of Trp seemed to increase peptide stability. Thus, for the correlation analysis, the variable was further categorized as either absence (flagged "W = 0") or presence (flagged "W = 1) of Trp. A *t*-test analysis performed after this categorization showed a statistically significant difference between the new groups ($\alpha = 0.013$; Figure S9 and Table S13).

The same procedure was applied for analyzing the effect of Tyr. Similar to Trp, this variable was categorized in terms of absence, presence of one Tyr, and presence of more than one Tyr. Variance analysis showed a statistically significant difference ($\alpha = 0.004$) among all groups (Figure S10 and Table S14) and multiple comparisons confirmed the initial assessment (Table S15). In this case, the presence of more than one Tyr residue increased peptide stability. The variable was consequently categorized as absence or presence of one Tyr (flagged "Y = 0") or of more than one Tyr (flagged "Y = 1"). Group analysis revealed a statistically significant difference ($\alpha = 0.001$; Figure S11 and Table S16).

Secondary structure is a parameter more difficult to assess and the sample size within our database was too small to allow a proper statistical analysis (Table S8). In addition, the frequency of each possible secondary structure was too low. Consequently, we excluded this variable from further analysis.

Table S17 shows the results of the descriptive statistics of the variables after categorization. Finally, a nonparametric correlation analysis (Spearmen correlation test) of all variables was performed (Table 2); the results demonstrated that hydrophobicity ($\rho = 0.259$; $\alpha = 0.05$), % of nonpolar residues ($\rho = 0.414$; $\alpha = 0.001$), pI ($\rho = -0.254$; $\alpha = 0.029$), presence/absence of Trp ($\rho = -0.279$; $\alpha = 0.016$), and presence/absence of Tyr ($\rho = 0.370$; $\alpha = 0.001$) correlated with peptide $t_{1/2\ s}$. Next, we assessed correlation among these variables to establish their independence from each other. A strong correlation ($\rho = 0.821$; $\alpha = 0.000$) was observed between percentage of nonpolar residues and hydrophobicity. Considering that both variables are related, the strong, very significant correlation found was plausible and expected. To avoid redundancy, only percentage of nonpolar residues was used, in preference to hydrophobicity because it displays a higher correlation coefficient. In sum, percentage of nonpolar residues, and of Tyr, had positive influence on proteolytic stability, whereas pI and presence of Trp negatively affected stability.

With the variables demonstrating correlation with peptide stability, multiple regression analysis was performed. The analysis gave an R^2 of 0.392 (Table S18) and variance showed a statistically significant difference ($\alpha = 0.001$) with a *Z*-test of 8.715 (Table S19). An equation, reflecting the

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Nonparametric correlation analysis of all the variables

			$\ln[t_{1/2}]$	Molecular weight, Da	Hydrophobicity (H)	Hydrophobic moment (μH)	Nonpolar residues (%)	Dummy tryptophan	Dummy tyrosine	Dummy pI
Spearman rô	$\ln[t_{1/2}]$	Correlation coefficient	1.00	0.205	0.259*	0.058	0.414**	−0.279*	0.370**	−0.254*
		Sig (2 extremities)		0.080	0.050	0.664	0.001	0.01	0.001	0.029
		N	109	74	58	58	59	74	74	74
	Molecular Weight (Da)	Correlation coefficient	0.205	1.000	−0.150	−0.005	−0.091	−0.045	0.452**	−0.257**
		Sig (2 extremities)	0.080		0.131	0.964	0.360	0.611	0.000	0.003
		N	74	129	103	103	104	129	129	129
	Hydrophobicity (H)	Correlation coefficient	0.259*	−0.150	1.000	0.157	0.821**	0.331**	0.063	−0.211*
		Sig (2 extremities)	0.050	0.131		0.113	0.000	0.001	0.525	0.033
		N	58	103	103	103	103	103	103	103
	Hydrophobic moment (μH)	Correlation coefficient	0.058	−0.005	0.157	1.000	0.060	0.107	−0.191	0.110
		Sig (2 extremities)	0.664	0.964	0.113		0.548	0.282	0.054	0.267
		N	58	103	103	103	103	103	103	103
	Presence Nonpolar residues	Correlation coefficient	0.414**	−0.091	0.821**	0.060	1.000	0.167	0.263**	−0.167
		Sig (2 extremities)	0.001	0.360	0.000	0.548		0.091	0.007	0.090
		N	59	104	103	103	104	104	104	104
	Dummy tryptophan	Correlation coefficient	−0.279*	−0.045	0.331**	0.107	0.167	1.000	−0.060	−0.042
		Sig (2 extremities)	0.016	0.611	0.001	0.282	0.091		0.499	0.638
		N	74	129	103	103	104	129	129	129
	Dummy tyrosine	Correlation coefficient	0.370**	0.452**	0.063	−0.191	0.263**	−0.060	1.000	−0.209*
		Sig (2 extremities)	0.001	0.000	0.525	0.054	0.007	0.499		0.018
		N	74	129	103	103	104	129	129	129
	Dummy pI	Correlation coefficient	−0.254*	−0.257*	−0.211*	0.110	−0.167	−0.042	−0.209*	1.000
		Sig (2 extremities)	0.0029	0.003	0.033	0.267	0.090	0.638	0.018	
		N	74	129	103	103	104	129	129	129

Abbreviations: pI, isoelectric point; $t_{1/2}$, terminal half-life.

*Significant correlation in 0.05 (2 extremities).

**Significant correlation in 0.01 (2 extremities).

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

statistical validation of the model, estimated peptide $t_{1/2}$ (in minutes) as a function of presence of nonpolar residues, presence/absence of Trp (W), presence/absence of Tyr (Y), and pI, as follows (Figure S12 and Table S20):

$$\ln\left[t_{1/2}\right] = 2.226 + 0.053 \times \mathrm{NP}_{\mathrm{residues}}\,[\%] - 1.515 \times W[0,1] + 1.290 \times Y[0,1] - 1.052 \times pI[0,1] \quad (1)$$

where $\mathrm{NP}_{\mathrm{residues}}[\%]$ is presence of nonpolar residues; W[0,1] is the absence (0) or presence of at least one Trp (1); Y[0,1] is the absence or presence of just one Tyr (0) or of at least two Tyr (1); and pI[0,1] is pI less than 10 (0) or pI greater than 10 (1).

Interestingly, according to Equation 1, the most proteolytically stable peptides are those combining pI greater than or equal to 10 with a high percentage of nonpolar residues, two or more Tyr, and not more than one Trp residue.

Experimental validation of the model

In order to validate the predictive power of the model (Equation 1), we followed two different but complementary approaches. The first relied on a library of 16 peptides especially designed to cover a wide range of $\mathrm{NP}_{\mathrm{residues}}[\%]$, W[0,1], Y[0,1], and pI[0,1] combinations (Table 1 and Table S21). The second approach made use of literature data for peptides of pharmacological significance (Table 3). Although the first approach is more robust from the statistical and conceptual point of view, the second constitutes a good gauge of the potential applicability of the equation to "real life" situations.

Validation using a library of designed peptides

The peptide library to validate the model used as starting template the 15-residue peptide GAAQAAGSGAAQAAG (Table 1, entry 1), a palindromic sequence containing 3 types of "low profile"—small size, non-charged—residues such as Gly (neutral), Ala (hydrophobic) and Ser (mildly polar), plus a medium-size, relatively polar, and equally noncharged residue, Gln. This initial template was modified in the subsequent entries on the basis of the criteria developed from the correlation analysis by, for example, modifying the hydrophobic content (Ala Ser conversion, entry 2), increasing the cationic character hence the pI (Arg replacements, entries 3, 6, etc.), exploring the impact of Tyr (entry 4) or Trp (entry 5) replacements, and judicious combinations thereof. The physicochemical combinations performed are shown in Table S21.

The $\ln[t_{1/2}]$ of the 16 peptides in human serum ranged from 2 to 6 (i.e., 7.4 to 403.4 min), covering a wide interval as anticipated (Figure 3A and Figure S13). Peptides 8 and 11 were the most resistant ($t_{1/2} > 100$ min), followed by 3, 4,

TABLE 3 Peptides approved or in clinical trials used for the validation of the model

Peptide	Approval status	Amino acid sequence
Aviptadil	Approved (2000)	HSDAVFTDNYTRLRKQMAVKKYLNSILN
Bivalirudin	Approved (1999)	FPRPGGGGNGDFEEIPEEYL
Corticotropin	Approved (1952)	SYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEF
Enfuvirtide	Approved (2003)	YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF
Exenatide	Approved (2005)	HGEGTFTSDLSKQMEEEAVRLFIEWLKNGGPSSGAPPPS
Glucagon	Approved (1989)	AQDFVQWLMNT
P−15	Approved (1999)	GTPGPQGIAGQRGVV
Pramlintide	Approved (2005)	KCNTATCATQRLANFLVHSSNNFGPILPPTNVGSNTY
Teriparatide	Approved (2002)	SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF
Tesamorelin	Approved (2010)	YADAIFTNSYRKVLGQLSARKLLQDIMSRQQGESNQERGARARL
Tetracosactide	Approved (1980)	SYSMEHFRWGKPVGKKRRPVKVYP
Thymopentin	Approved (1985)	RKDVY
EA−230	Phase II	LQGV
Ghrelin	Phase II	GSSFLSPEHQRVQQRKESKKPPAKLQPR
Hlf1-11	Phase I/II	GRRRRSVQWCA
Dusquetide	Preclinical	RIVPA
R7	Phase II	KLAKLAK
TAT	Phase II	GRKKRRQRRRPQ
PTD4	Phase II	YARAAARQARA
MTS	Phase II	AAVALLPAVLLALLAP
MBI−226	Phase II	ILRWPWWPWRRK

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

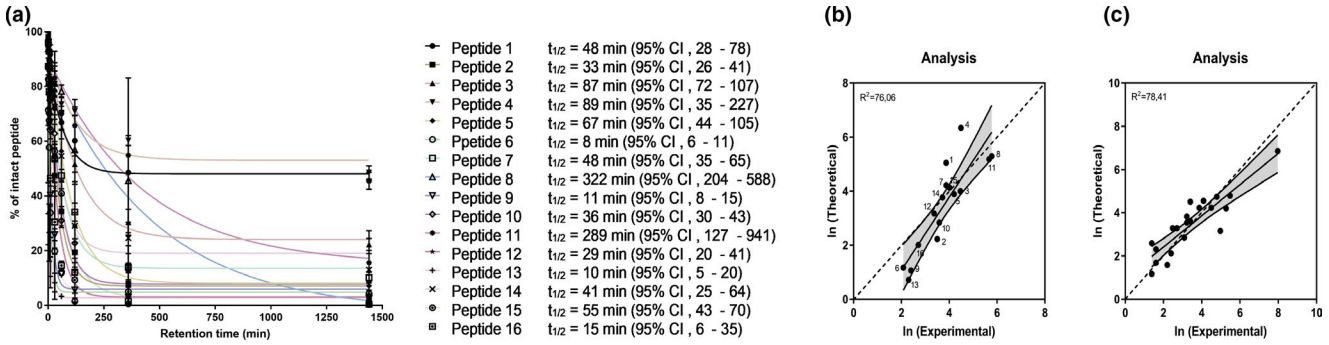


FIGURE 3 Stability of peptides 1–16 in 50% (v/v) human serum. (a) Time course plots of peptides 1–16 in the presence of human serum obtained from chromatogram peak integration. The terminal half-life ($t_{1/2}$ s) estimated by fitting experimental data to an exponential decay model and the corresponding 95% confidence interval (CI) are shown at right and are the mean of three experiments for each peptide in the library. (b) Linear regression plot of experimentally obtained $t_{1/2}$ s of the peptide library versus values derived from the statistical prediction model (Equation 1). (c) Linear regression plot of literature-reported experimental $t_{1/2}$ s from representative therapeutic peptides compared with values obtained from the statistical prediction model

and 5 (60 min < $t_{1/2}$ < 100 min). Most peptides in the library (e.g., entries 1, 2, 7, 10, 14, and 15) had modest lifetimes in the 30 min less than $t_{1/2}$ less than 60 min range, whereas entries 6, 9, 12, 13, and 16 were more ephemeral ($t_{1/2}$ < 30 min; Figure 3A and Figure S13).

The ln[$t_{1/2}$] estimated from Equation 1, compared with the experimental values (Figure 3B), could be fitted to a linear regression plot from which an R^2 of 0.76 was retrieved, demonstrating a rather suitable correlation of the model predictions with experimental data.

Validation using pharmaceutically relevant peptides

The $t_{1/2}$ predictive potential of the model was also tested on "real life" unmodified peptides currently marketed by pharmaceutical companies or at various stages of clinical trials in the industry pipelines. The corresponding experimental $t_{1/2}$ values were obtained from the literature and, as in the previous approach, the ln[$t_{1/2}$] values calculated from Equation 1 were matched against the experimental ones in a linear correlation plot (Figure 3C) from which an R^2 of 0.78 was retrieved, again demonstrating reasonable correlation.

DISCUSSION

Researchers report the stability of peptides in wholly different ways. [13] The most informative is peptide $t_{1/2}$ but most frequently only semiquantitative data are reported (e.g., percentage of peptide that remains intact after a certain period). To complicate matters further, among the 579 papers initially selected for our meta-analysis database, 70% had modifications. As expected in medicinal chemistry approaches, such modifications are very heterogeneous in nature, which makes data parametrization totally impossible. [10,14] In this work, we did not address post-translational modifications because for artificial peptides, which account for most pharmacological and biotechnological applications, these modifications are restricted to a few niche studies. On the other hand, 30% of the initially selected peptides in our database were unmodified and reasonably resistant to proteolysis, the main cause for the low $t_{1/2}$ of peptides. Nevertheless, it is necessary to bear in mind that there are other causes for peptide instability, such as hydrolysis or oxidation. [11] In this study, we focused on the intrinsic properties conferring such peptides unusually long $t_{1/2s}$, with the development of a multivariable regression model able to predict peptide $t_{1/2}$ from amino acid sequence as the main goal of our work.

From a set of sequence-dependent variables potentially bearing on peptide serum lifetime, four were found to impact significantly on $t_{1/2}$, namely, the presence of nonpolar residues, the presence/absence of Trp and/or Tyr, and electric charge as gauged by the isoelectric point, pI. High contents of nonpolar residues and the presence of two or more Tyr residues increased stability, whereas, on the other hand, the presence of at least one Trp residue and an elevated pI increased the vulnerability of the sequence. Equation 1 is the result of applying a multivariable regression model using the properties mentioned above as independent variables. This equation can be used for virtually any peptide sequence to estimate $t_{1/2}$ in an easy-to-use way.

To validate Equation 1 and ascertain its suitability for use in peptide therapeutics pipelines, we followed two complementary strategies. First, using a library of tailor-designed peptides covering a wide range of structural feature combinations affecting the independent variables, experimental and estimated $t_{1/2}$ were matched for correlation. The $t_{1/2}$ in serum was chosen as exclusive matching parameter, as it is a consensual gold standard in measuring peptide resistance

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to proteolysis.[5,15] Human serum is a cocktail of proteins, enzymes, hormones, electrolytes, and other blood components, thus a good model for blood itself. Nevertheless, caution on the use of commercial human serum is advisable given limitations, such as lack of homogeneity between batches giving rise to different activities.[11] Use of specific proteases, such as trypsin, chymotrypsin, or proteinase K, is operationally simpler but constitutes a more simplistic approach. On the other hand, peptide clearance in vivo, although undeniably valuable, does reflect a complex combination of effects, such as enzymatic digestion, peptide biodistribution into different organs, and physiological elimination, all of which make specific $t_{1/2}$ calculation in vivo essentially unfeasible.

Second, a similar approach was followed for a set of peptides in clinical use or under assay for future clinical use. As in the peptide library-based instances, estimated $t_{1/2}$ matched experimentally determined values with reasonable accuracy. Equation 1 is therefore not only a valid explanatory tool, but also shows proven predictive power for use in peptide drug development strategies.

ACKNOWLEDGMENT

The authors are very grateful to Alex Cavaco for the help in the development of the user-friendly software.

CONFLICT OF INTEREST

All authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

M.C., J.V., I.F., D.A., and M.A.R.B.C. wrote the manuscript. M.C., D.A., and M.A.R.B.C. designed the research. M.C. and J.V. performed the research. M.C. I.F., D.A., and M.A.R.B.C. analyzed the data. I.F. contributed new analytical tools.

REFERENCES

1. Cavaco M, Castanho MARB, Neves V. Peptibodies: an elegant solution for a long-standing problem. *Pept Sci*. 2018;110:e23095.
2. Otvos L, Wade JD. Current challenges in peptide-based drug discovery. *Front Chem*. 2014;2(62):1-4.
3. Erak M, Bellmann-Sickert K, Els-Heindl S, Beck-Sickinger AG. Peptide chemistry toolbox – Transforming natural peptides into peptide therapeutics. *Bioorg Med Chem*. 2018;26:2759-2765.
4. Yin N, Brimble MA, Harris PWR, Wen J. Enhancing the oral bioavailability of peptide drugs by using chemical modification and other approaches. *Med Chem*. 2014;4:463-769.
5. Böttger R, Hoffmann R, Knappe D. Differential stability of therapeutic peptides with different proteolytic cleavage sites in blood, plasma and serum. *PLoS One*. 2017;12:e0178943.
6. Werner Halina M, Cabalteja Chino C, Horne WS. Peptide backbone composition and protease susceptibility: impact of modification type, position, and tandem substitution. *ChemBioChem*. 2015;17:712-718.
7. Pérez-Peinado C, Dias SA, Mendonça DA, Castanho MARB, Veiga AS, Andreu D. Structural determinants conferring unusual long life in human serum to rattlesnake-derived antimicrobial peptide Ctn[15-34]. *J Pept Sci*. 2019;25:e3195.
8. Weinstock MT, Francis JN, Redman JS, Kay MS. Protease-resistant peptide design—empowering nature's fragile warriors against HIV. *Pept Sci*. 2012;98:431-442.
9. Pollaro L, Heinis C. Strategies to prolong the plasma residence time of peptide drugs. *MedChemComm*. 2010;1:319-324.
10. Buckley ST, Hubálek F, Rahbek UL. Chemically modified peptides and proteins - critical considerations for oral delivery. *Tissue Barriers*. 2016;4:e1156805.
11. Cavaco M, Andreu D, Castanho MARB. The challenge of peptide proteolytic stability studies: scarce data, difficult readability, and the need for harmonization. *Angew Chem Int Ed*. 2021;60(4):1686–1688.
12. bioSYNTHESIS. Peptide Property Calculator. Vol. ver 3.1 https://www.biosyn.com/peptidepropertycalculator/PeptidePropertyCalculator.aspx (2013).
13. Mathur D, Prakash S, Anand P, et al. PEPlife: a repository of the half-life of peptides. *Sci Rep*. 2016;6: 36617, 1-7.
14. Luca G, De Marco R, Lucia C. Chemical modifications designed to improve peptide stability: incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Curr Pharm Des*. 2010;16:3185-3203.
15. Jenssen H, Aspmo SI. Serum stability of peptides. In: *Peptide-Based Drug Design* (ed. Otvos L). Totowa, NJ: Humana Press; 2008:177–186.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Cavaco M, Valle J, Flores I, Andreu D, A. R. B. Castanho M. Estimating peptide half-life in serum from tunable, sequence-related physicochemical properties. *Clin Transl Sci*. 2021;14:1349–1358. https://doi.org/10.1111/cts.12985

17528062, 2021, 4, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.12985 by Cochrane Portugal, Wiley Online Library on [06/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License